IBM Systems and Technology Group

IBM System Storage-Kompendium

Die IBM Speichergeschichte von 1952 bis 2008

ibm.com/systems/de/storage

Sehr geehrte Kunden, sehr geehrte Geschäftspartner, liebe Leser,

begeben Sie sich auf eine spannende Zeitreise durch die IBM Speichergeschichte. Bereits vor 56 Jahren erfand die IBM die erste externe Speichereinheit in Form des Rollenbandes der IBM 726 mit einer Speicherkapazität von 1.4 MB. Lassen Sie sich faszinieren, was in diesem halben Jahrhundert an Erfindungen und Produkten von IBM im Speicherumfeld entwickelt und auf den Markt gebracht wurde. Vielleicht möchten Sie sich durch dieses Buch auch an Ihre Anfangszeiten in der IT zurück erinnern und gedanklich Ihren beruflichen Werdegang Revue passieren lassen. Aber auch für alle jungen Menschen, die diese Zeitspanne nicht direkt erlebt haben, ist es ein faszinierendes Nachschlagewerk.

Die beschriebenen Zeitepochen reflektieren sehr klar die IBM Entwicklungsstrategie: Rechner und Speichersysteme gehören als Einheit zusammen! Speichertechnologien und Speicherlösungen wurden bis heute immer im Einklang mit den Servertechnologien vorangetrieben, um ausgewogene, leistungsstarke und aufeinander abgestimmte Gesamtlösungen zur Verfügung zu stellen. Diese Strategie brachte viele neue Erfindungen hervor und machte IBM zum weltweiten Patentführer. Und das seit 15 Jahren in Folge.

Im Jahr 2007 wurde der Nobelpreis für Physik an Peter Grünberg vom Forschungszentrum Jülich und Albert Fert von der Universität Paris-Süd vergeben, die 1989 den GMR-Effekt (Giant Magnetoresistance) entdeckt hatten. Wir alle wissen, dass ohne diese Entdeckung die heutigen Festplattenkapazitäten nie möglich geworden wären. Viele wissen allerdings nicht, dass es IBM zu verdanken ist, dass aus der Entdeckung ein Produkt wurde.

Bereits 1990 begann der IBM Forscher Stuart Parkin im IBM Labor Almaden in Kalifornien mit der Umsetzung und erforschte mehr als 30.000 Materialkombinationen, um 1997 den ersten einsetzbaren GMR-Lesekopf für Plattenlaufwerke zur Verfügung zu stellen. Alle drei Forscher wurden 1997 gemeinsam für die GMR-Entwicklung auch mit dem Europhysik-Preis ausgezeichnet.

Vielleicht nicht alles Wichtige, aber alles fundamental Wichtige wurde von IBM im Storage-Bereich entdeckt und entwickelt. Ich möchte Ihnen neben GMR ein paar weitere Eckdaten nennen: 1991 führte IBM drei zukunftsweisende Technologien ein: Dünnfilmbeschichtungen auf den Platten, die ersten MR-Köpfe (Magneto resistive Aufzeichnung) und RAID5-basierende Subsystemarchitekturen. 1999 entdeckte IBM den paramagnetischen Effekt und in demselben Jahr brachte IBM den ersten Micro-Drive auf den Markt, ein Laufwerk in der Größe einer Zwei-Euro-Münze. Im Jahr 2000 erhielt IBM als erste Firma die 'National Medal of Technology for Innovations in Storage' von der amerikanischen Regierung, eine der höchsten technologischen Ehrungen, die bis dahin ausschließlich an Einzelpersonen vergeben wurde. Alle diese faszinierenden Erfindungen und viele mehr finden Sie im Technologie-Anhang dieses Storage-Kompendiums. Die IBM leistete auf dem Gebiet der Speicherentwicklungen Einmaliges und ich bin überzeugt, dass dies auch in der Zukunft der Fall sein wird.

Widmen Sie Ihre Aufmerksamkeit auch besonders der jetzigen Zeitepoche, die vor allem durch Virtualisierung in allen Bereichen der Datenspeicherung geprägt ist. Meinem Mitarbeiter Kurt Gerecke, dem Autor des IBM System Storage-Kompendiums, ist es gelungen, alle im Jahr 2008 aktuellen Speicherprodukte ausführlich und verständlich darzustellen. Das Kompendium ist also nicht nur ein historisches Nachschlagewerk, sondern informiert Sie vor allem über alle aktuellen IBM Speicherentwicklungen.

Ich wünsche Ihnen viel Freude und viel Spass beim Lesen und Studieren des IBM System Storage-Kompendiums und hoffe, dass Sie viele Gelegenheiten haben, diese Informationen für Ihre Zwecke erfolgreich einzusetzen.

Mit ganz herzlichen Grüssen

Ihre

Martina Koederitz

Vice President IBM Deutschland Systems and Technology Group

Die Überarbeitung des IBM System Storage-Kompendiums als Update 2008 erforderte doch wieder viel freie Zeit und viele Wochenenden. Deshalb möchte ich mich zuallererst wieder bei meiner Ehefrau Gabriele für das aufgebrachte Verständnis bedanken. Ihre Geduld war wieder einmal bewundernswert.

Ein grosser Dank gilt auch dem IBM Storage-Produktexpertenteam für seine wertvollen Informationen, Sven Oehme aus Mainz, der so nett war, den Teil GPFS und SOFS Korrektur zu lesen, Stefan Neff für seine De-Dup Bauklötze-Bilder, Josef Weingand für seinen Input zu TS7520 und Barbara Lagemann für ihren Input 'Backup-Konzepte'. Auch bei Herrn Hans W. Spengler vom IBM Museum in Sindelfingen möchte ich mich ganz herzlich für seine tabellarischen Zusammenstellungen bedanken. Allen anderen Kollegen, die mir dabei geholfen haben und jetzt nicht namentlich erwähnt sind, gilt ein ebenso großes Dankeschön.

Mein Dank gilt auch vor allem meinem Management, Herrn Marcellus Scheefer und Herrn Andreas Walter, sowie dem IBM System Storage-Marketing, vor allem Frau Brigitta Haak, Herrn Stephan von Rohden und Herrn Ivano Rodella, welche die Realisation und den Druck dieses Kompendium-Updates ermöglichten.

Besonders möchte ich mich für ihre Unterstützung bei Frau Martina Koederitz, Vice President IBM Deutschland Systems and Technology Group, bedanken, die so nett war, dieses Kompendium mit persönlichen einleitenden Worten zu gestalten.

Die neue Abbildung für den Einband entdeckte ich in einer Bilddatenbank aus dem IBM Labor Almaden. Die Abbildung stammt aus dem Jahre 2000 und zeigt die damals vorgenommenen Versuche in der Holografischen Speicherung. Mit Laser-Strahlen, die eine unterschiedliche Lichtbeugung aufweisen, werden Interferenz-Felder, sogenannte Hologramme, in einem organischen Polymer erzeugt.

Das IBM System Storage-Kompendium beschreibt die IBM Speichergeschichte von 1952 bis einschließlich des 1. Quartals 2008. Alle neuen Produkte, die vor kurzem von IBM angekündigt wurden, sind reflektiert und ausführlich beschrieben. Der Technologie-Anhang wurde maßgeblich erweitert, um noch mehr Verständnis für die Speicherbasistechnologien zu schaffen. Deshalb liegt es mir am Herzen, dass das Kompendium nicht nur als Informationsmaterial oder als Geschenk für Kunden und Geschäftspartner dient, sondern auch gezielt für Schulungszwecke eingesetzt wird. Das gesamte aktuelle IBM Speicherportfolio ist darin ausführlich dokumentiert.

Ein bisschen Reklame möchte ich auch machen! Sollten Sie sich im Raum Stuttgart aufhalten, vergessen Sie nicht das IBM Museum, das Haus der Geschichte der IBM Datenverarbeitung, Bahnhofstrasse 43 in 71063 Sindelfingen, zu besuchen. Dort können Sie 'live' die Geschichte der Informationstechnologie von 1890 bis ins Jahr 2000 erleben. Auch das erste Plattenprodukt der IBM, die RAMAC 350 als Teil des damaligen IBM RAMAC 305 Rechnersystems, können Sie dort in Betrieb sehen. Melden Sie sich rechtzeitig wegen einer Terminabsprache bei mir, Tel. 0170-2207066, oder bei Herrn Hans W. Spengler vom IBM Museum, Tel. 07031-271378.

In diesem Sinne möchte ich mich nochmals bei allen herzlichst bedanken. Ich wünsche Ihnen allen viel Freude und Spass mit dem neuen Update des IBM System Storage-Kompendiums und verbleibe

mit den herzlichsten Grüssen

Ihr

Kurt Gerecke

Allgemeines:

Die Geschichte der IBM Speichertechnologie und der IBM Speichersysteme ist mit ziemlicher Sicherheit eines der faszinierendsten Ereignisse unserer Zeitgeschichte der letzten 56 Jahre. Spannend vom Anfang bis in die heutige Zeit soll diese Broschüre einen Überblick geben und vor allem auch das Verständnis entwickeln, wie erfinderisch die Menschheit sein kann, wenn es darum geht, technologische Grenzen zu überschreiten, um neue Möglichkeiten für die Speicherung von Daten zu eröffnen. Die Vielzahl von Produkten und Erfindungen in diesem Bereich macht es notwendig, die Zeitgeschichte ein bisschen zu ordnen und in zeitgeschichtliche Epochen zu unterteilen. Da dieses Unterfangen bisher nie gemacht wurde, erlaubt sich der Autor, das Speicherzeitgeschehen von 1952 bis in die heutige Zeit in technologische Epochen zu unterteilen, um dann im Detail die einzelnen Epochen von der Produkt- und Produktentwicklungsseite zu beleuchten. Im Anhang dieser Broschüre befindet sich ein Technologie-Teil, der die Entwicklung der Speicherbasistechnologien und Aufzeichnungsverfahren einschließlich anstehender Technologien näher beschreibt.

Entwicklung der Speichersysteme von 1952 bis 2006

Die einzelnen Epochen beschreiben immer eine Zeitphase, die durch besondere Produkte geprägt war, die in dieser Zeitphase auf den Markt kamen. In vielen Fällen wurden die Produkte über die definierte Epoche hinaus aktiv vertrieben und wurden meistens erst in der Folgeepoche oder später vom Marketing zurückgezogen. Die einzelnen Produkte sind in der Epoche, in der die Markteinführung erfolgte, so beschrieben, dass sie den gesamten Lebenszyklus darstellen.

Das Storage-Kompendium erhebt keinen Anspruch auf Vollständigkeit, beschreibt aber die wichtigsten Produkte in der jeweiligen Epoche.

Preisentwicklung Magnetplattenspeicher

Die Schnelligkeit der technologischen Entwicklung bei Magnetplattenspeichern von 1956 bis ins Jahr 2006 lässt sich im besonderen Maße aus dem Preisverfall für ein gespeichertes Megabyte ablesen. Der Verlauf war in etwa folgendermaßen:

Jahr	1956	1964	1975	1987	1991	1997	2000	... 2006
Euro/MB	12000	8000	70	12	9	1	0.25	0.015

Rechnet man den verminderten Platz- und Energiebedarf sowie den geringeren Wartungsaufwand pro Megabyte hinzu, so ist der Trend bei der Kostenentwicklung eher noch günstiger.

Kapazitätsentwicklung Magnetplattenspeicher

Während der Preisverfall für das gespeicherte Megabyte kontinuierlich anhielt, war in den letzten 10 Jahren ein überdimensional hoher Preisverfall zu verzeichnen, der speziell in den letzten Jahren eine Spanne von 40 bis 60 Prozent pro Jahr erreichte. Parallel dazu wurde die Aufzeichnungsdichte durch technologische Innovationen in einem Maße gesteigert, dass sie dem Preisverfall für das gespeicherte Megabyte standhalten und Rechnung tragen konnte.

Die Steigerung der Aufzeichnungsdichte war in etwa folgendermaßen:

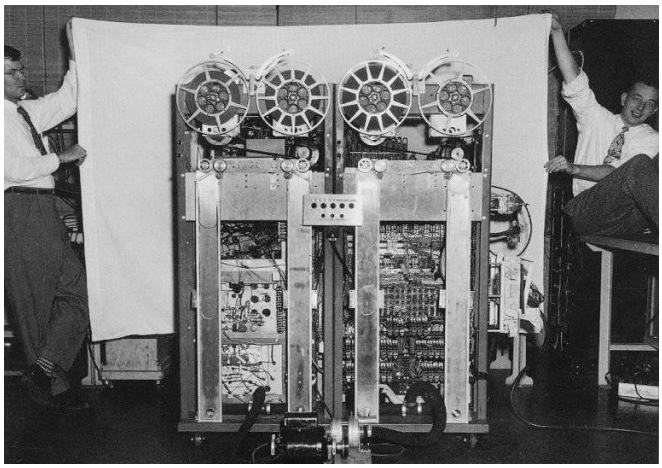
Jahr	**Produkt**	**Aufzeichnungsdichte (Millionen Bits/qcm)**
1956	RAMAC 350	0.00031
1961	1311	0.008
1964	2311	0.017
1965	2314	0.034
1970	3330	0.12
1973	3340	0.26
1975	3350	0.48
1979	3370	1.2
1980	3375	1.51
1980	3380	1.9
1985	3380-E	2.99
1987	3380-K	5.47
1991	3390-3	13.95
1994	3390-9	41.6
1996	RAMAC 3	129.2
1999	ESS	427.5
2004	ESS	2.9 Milliarden Bits/qcm
2006	DS8000	5.8 Milliarden Bits/qcm

Im Jahr 2003, mit Verfügbarkeit der 146-GB Laufwerke für ESS, wurde die magische Grenze von 1 Milliarde Bits auf dem Quadratzentimeter überschritten. Mit der heutigen Verfügbarkeit von 500-GB-SATA-Laufwerken erreicht man eine Aufzeichnungsdichte von knapp 10 Milliarden Bits/qcm. Die großen Sprünge ergaben sich speziell in den letzten 10 Jahren (siehe auch Technologie-Anhang).

Die Anfangsepoche der elektromagnetischen Speicherung

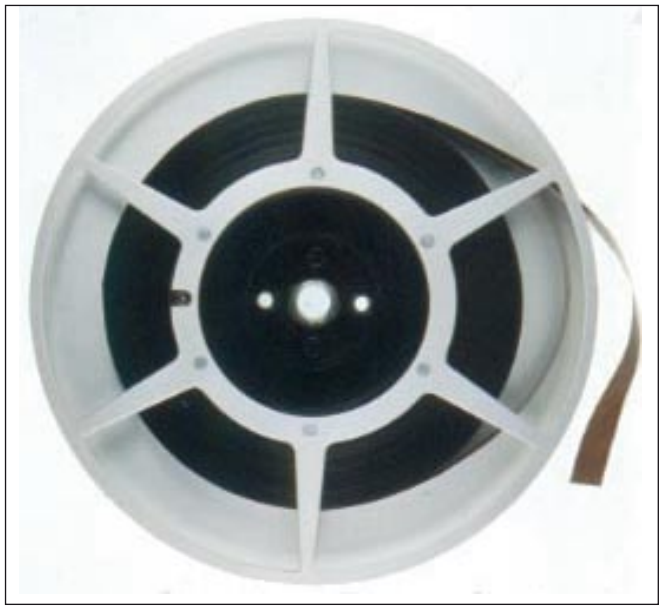
1952 bis 1961

Die Anfangsepoche der elektromagnetischen Speicherung



Aufnahme des ersten Prototypen des Magnetbandsystems IBM 726 im Jahr 1951 mit zwei Bandlaufwerken pro Einheit

Zum ersten Mal in der Speichergeschichte wurden externe Speicher mit magnetischer Aufzeichnung von IBM **1952** angeboten: die **IBM Magnetbandeinheit IBM 726** mit einer Speicherkapazität von 1.4 MB auf einem 12-Zoll-Rollenband (Durchmesser der Bandspule) mit 720 Metern Länge. Die Einheit hatte **zwei integrierte Bandlaufwerke.** Dies erwies sich allerdings als äußerst unpraktisch, weil keines der beiden Laufwerke im Fehlerfalle oder bei Wartungsarbeiten weiterhin zur Verfügung stand. Bereits ein Jahr später, **1953**, folgte **die IBM 727** mit einem Laufwerk pro Einheit, die auf einem standardisierten Rollenband von ca. 740 Metern Bandlänge eine Kapazität von 4 MB bot. Bis 1956 war ein Teil der Steuerung in der damaligen Rechnereinheit IBM 701 untergebracht. Erste dedizierte Bandsteuereinheiten gab es dann ab **1956**.



Medium-Rollenband für IBM 726

IBM Magnetbandeinheit Modell 726, Markteinführung 1952

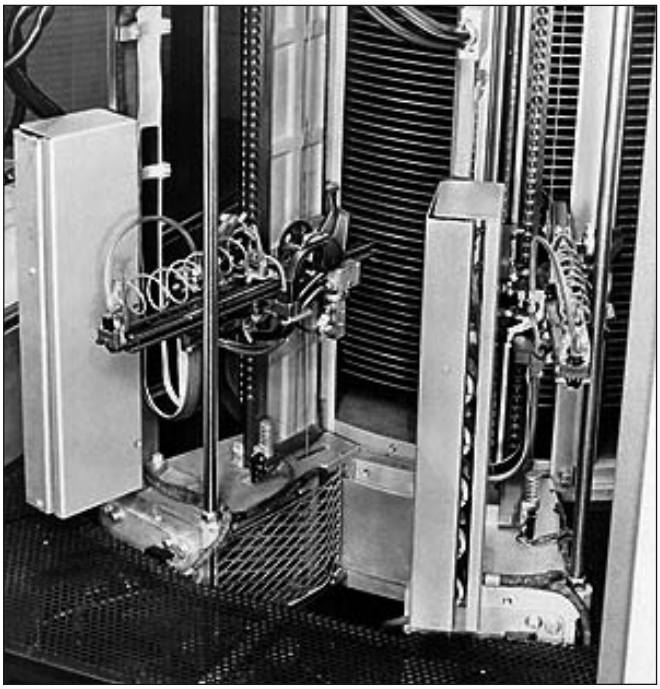
1956 kam das **System IBM RAMAC 305** auf den Markt. Dieses System hatte erstmals eine **Platten-Speichereinheit Type IBM 350**. Das IBM 305 System wurde unter dem eingetragenen Warenzeichen ‘RAMAC’ angekündigt (Random Access Method of Accounting and Control). Der Rechner des Gesamtsystems steuerte sowohl das Positionieren des Schreib-/Lesekopfes auf die gewünschte Plattenoberfläche und Plattenspur als auch das Übertragen der Daten in beide Richtungen (Schreiben und Lesen).

Die Aufzeichnungsdichte der RAMAC 350 Platte lag bei 100 Bits pro Inch (40 Bits pro cm) und der Plattenstapel rotierte mit 1200 Umdrehungen in der Minute (1200 RPM). Die durchschnittliche Suchzeit lag bei 600 ms. Der Plattenstapel bestand aus 51 Platten mit einem Durchmesser von 24 Zoll und bot eine maximale Kapazität von 5 MB (5 Millionen Characters). In Betrieb kann die RAMAC 350 im IBM Museum in Sindelfingen besichtigt werden.

IBM 350 RAMAC Platte, 51 Platten im Stapel mit 24 Zoll Durchmesser. 1956: Modell mit 5 MB (5 Millionen ‘Characters’), 1958: Modell mit 10 MB (10 Millionen ‘Characters’), Zurückziehung vom Vertrieb im Jahr 1960

IBM RAMAC 350 Plattenstapel

Damals gab es noch keine Zugriffskämme, sondern nur einen Zugriffsarm, der in die Zwischenräume des Plattenstapels mit einer Seilzugmechanik hineingesteuert wurde. Die exakte Positionierung erfolgte mit Zahnstange und Klinke und wurde per Pressluft elektronisch gesteuert. Die Köpfe schwebten über der Platte in einem Abstand von 800 Micro-Inches. 1958 wurde mit der Ankündigung des Modells 2 die Möglichkeit geschaffen, die doppelte Kapazität abzubilden (10 MB). Bei diesem Modell wurden zwei Zugriffsstationen eingesetzt, sodass eine Station schreiben oder lesen und die zweite gleichzeitig suchen konnte, um für die Verarbeitung des nächsten Satzes die Positionierung des Schreib-/Lesekopfes vorzubereiten.



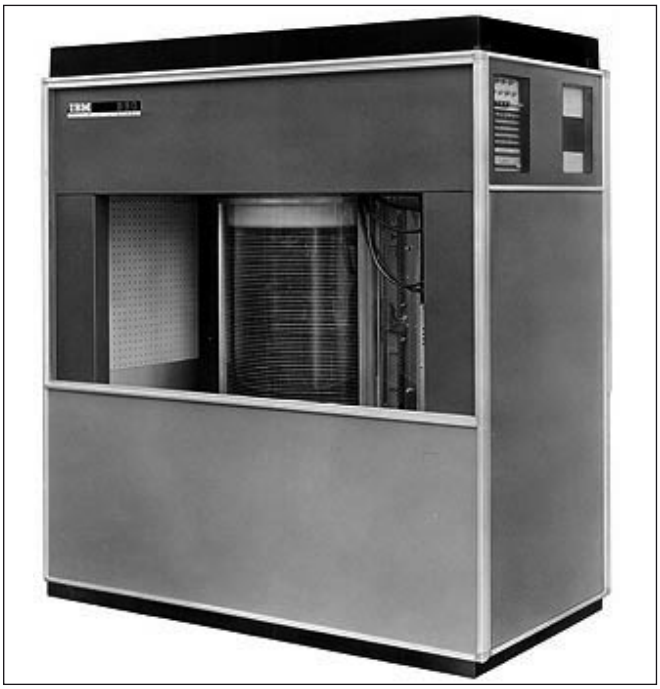
IBM RAMAC 350 Platte mit zwei Zugriffsstationen

Verkaufstechnisch wurde bereits damals mit der RAMAC derselbe Marketing-Ansatz gewählt, den man auch heute noch, nur in anderen Größenordnungen, vorfindet. Wie macht man dem Käufer und Benutzer klar, was 5 MB sind? 5 MB lagen damals über jeder normalen Vorstellungskraft. Deshalb wurden in den meisten Spezifikationsunterlagen 5 MB so definiert, dass das 60000 – 80000 Lochkarten entspreche, oder 2000 Feet Bandlänge (die Bänder gab es ja bereits 4 Jahre früher) oder 940 Seiten bedrucktem Papier.

Neben dem RAMAC Plattenstapel und der dazugehörigen Zugriffsstation enthielt die Einheit 350 eine elektronische und pneumatische Kontrolle für die Zugriffsstation und einen Luftkompressor. Die Schmierung der Plattenlager wurde durch eine Öl-Dunst-Schmierung durchgeführt.

Die Einheit 350 hatte eine Länge von 152.4 cm (60 Inches), eine Höhe von 172.72 cm (68 Inches) und eine Tiefe von 73.66 cm (29 Inches) und wog etwa eine Tonne.

Das Gesamtsystem, bestehend aus Rechner, Platteneinheit, Kartenleser, Kartenstanzer, Drucker und Konsoleneinheit, kostete 185000 Dollar und wurde auch unter Leasing zu einer monatlichen Rate von 3200 Dollar angeboten. In dem Zeitraum von 1956 bis 1960 wurden weltweit 1500 Systeme installiert. 1960 erfolgte dann die Zurückziehung aller Modelle vom Vertrieb und 1961 die Einstellung der Produktion.



IBM RAMAC 350 Einheit

Die Anfangsepoche der elektromagnetischen Speicherung

Erstauslieferung IBM RAMAC 305 in Brasilien 1956

1957 wurde der **'Data Synchronizer'** am IBM System 709 eingeführt, der Vorläufer der heutigen Übertragungskanäle.

1959 führte IBM ein neues Programmpaket, das Eingabe-/Ausgabesteuerungssystem **IOCS** für Magnetbänder und Magnetplatten (**Input/Output Control System**) ein. Es handelte sich um Makroinstruktionen zum Erstellen, Wiederauffinden und Verarbeiten von Dateien sowie zur Fehlerkorrektur. Außerdem steuerte IOCS die Datenblockbildung und die Entblockung. IOCS blieb für spätere Systeme von maßgeblicher Bedeutung.

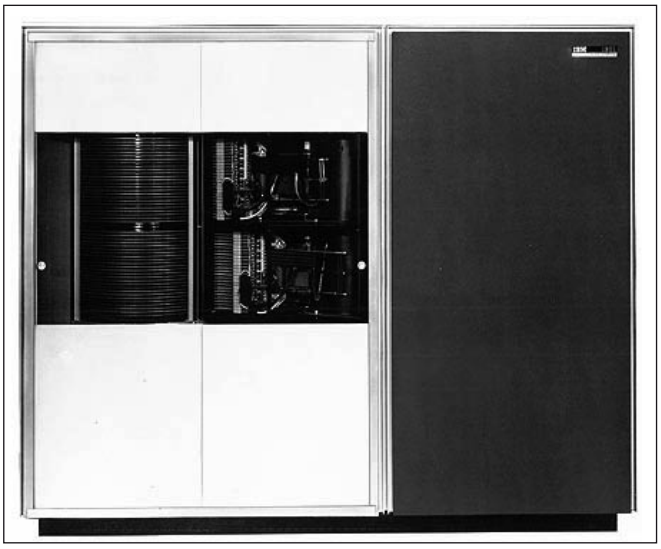
1958 verwendete IBM erstmals mit den Systemen IBM 7070 und 7090 die Technik der automatischen Programmunterbrechung (**Program Interrupt**). Der Prozessor verzweigte bei bestimmten, von Eingabe-/Ausgabeeinheiten angezeigten Bedingungen automatisch zu Routinen, die den 'Interrupt' behandelten.

1958 wurde die **Bandeinheit IBM 729** als Nachfolger des Vorgängers IBM 727 auf den Markt gebracht. Die IBM 729 Bandfamilie bot neben höheren Datenraten und höherer Speicherdichte die erste Schreibkontrolle. Mit der Einführung der IBM 729 wurde beim Schreiben auf Magnetband automatisch geprüft, ob die eben geschriebene Information gültig war. Dazu wurde Stelle für Stelle in ein Prüfregister eingelesen und geprüft.

Im Oktober 1960 wurde der **Plattenspeicher IBM 1405** angekündigt, der an den Systemen IBM 1401, 1410 und 1460 (nicht an System 7000) betrieben wurde. Dieser Plattenspeicher vervierfachte die Kapazität eines Plattenstapels im Vergleich zur RAMAC von 1956. Sowohl die Anzahl der Spuren pro 'Inch' als auch die Anzahl der Bits in einer Spur wurden verdoppelt, sodass eine vierfache Kapazität des Plattenstapels erreicht wurde. Die Plattenstapel waren in zwei Ausführungen verfügbar, in einer Einheit mit 25 Platten im Stapel und einer mit 50 Platten. Die kleinere Einheit hatte eine Kapazität von 10 MB und die größere von 20 MB. Die Aufzeichnungsdichte betrug 220 Bits pro Inch, was 40 Spuren pro Inch entspricht. Die Schwebehöhe zwischen Kopf und Platte betrug 650 Micro-Inches und die Platten drehten sich mit 1800 Umdrehungen pro Minute. Die Datenrate lag bei 17.5 Kilobytes pro Sekunde. Der IBM 1405 Plattenspeicher kam vor allem beim IBM 1401 Rechner zum Einsatz.

Ein Jahr später, im **Juni 1961**, wurde die **IBM 1301 Plateneinheit** angekündigt. Sie wurde ein Jahr später, im 3. Quartal 1962, an Kunden ausgeliefert. Die Entwicklung dieser Plateneinheit trug in größtem Maße zur Weiterentwicklung zukünftiger Plattensysteme bei und ist als **Meilenstein** in der gesamten Plattengeschichte zu betrachten. Das Plattensystem 1301 war der Wegbereiter zukünftiger zur Anwendung kommender Technologien in zwei Bereichen: die **'Air Bearing Slider'**-Technologie und **separate Schreib-/Leseköpfe für jede Plattenoberfläche** mit der Möglichkeit, alle auf einen Datenzylinder physisch eingestellten Köpfe elektronisch für die zutreffende Plattenoberfläche zu selektieren und nacheinan-

IBM 1405 Plattenspeicher, Ankündigung am 10 Oktober 1960, Zurückziehung 30. Juni 1970



IBM 1301 Plattenspeichereinheit, Ankündigung im Juni 1961, Zurückziehung aller Modelle im Oktober 1970

IBM 7340 Hypertape Drives in der Bildmitte, links die IBM 7640 Kontrolleinheit, rechts die IBM 729 Bandeinheiten

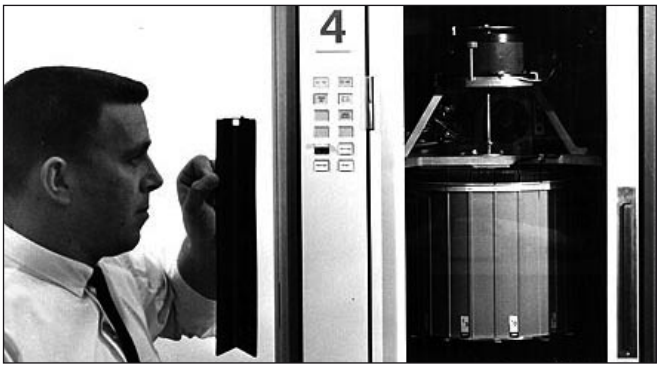
der zu bearbeiten. Damit war der Grundstein des Zylinderprinzips bei Plattenspeichern gelegt und der **'Plattenzylinder'** geboren!

Neben höherer Kapazität und Leistung bot die Maschine eine wesentlich flexiblere und höhere Übertragungsbandbreite durch das Prinzip, pro Plattenoberfläche einen Schreib-/Lesekopf zur Verfügung zu haben, und war auf die neue IBM 7000 Serie von Rechnern abgestimmt (7070, 7094, 7080 und 7090). Die Kapazitätssteigerung zur ersten RAMAC lag bei Faktor 13. Die Platten rotierten mit 1800 Umdrehungen in der Minute. Durch die Verkleinerung des Flugabstandes zwischen Kopf und Platte auf 250 Micro-Inches konnten 50 Spuren pro Inch und bis zu 520 Bits pro Inch aufgezeichnet werden. Das Modell 1 der IBM 1301 hatte ein Plattenmodul mit einer Kapazität von 28 MB, das Modell 2 arbeitete mit zwei Modulen und einer Kapazität von 56 MB. Angeschlossen an die IBM 7631 Steuereinheit konnten bis zu 10 Module (oder bis zu fünf 1301 Einheiten) zum Einsatz kommen, sie boten für die 7000 Serie eine maximale Kapazität von bis zu 280 MB. Interessant waren auch die damaligen Preise: Das Modell 1 konnte monatlich für 2100 Dollar geleast werden oder für eine Summe von 115500 Dollar gekauft werden. Das Modell 2 mit zwei Modulen lag bei 3500 Dollar monatliche Leasingrate und einem Kaufpreis von 185500 Dollar.

Die **Hyper Tape Drives IBM 7340** reflektierten die neuste Bandtechnologie, als sie **1961** auf dem Markt eingeführt wurden. Zusammen mit der Kontrolleinheit **IBM 7640** bedienten sie die IBM Rechner 7074, 7080 und 7090. Die Hyper Tapes erzielten wesentlich schnellere Lese- und Schreib-Geschwindigkeiten im Vergleich zu den bisherigen Bandeinheiten. Angeschlossen am IBM 7090 Rechner lieferten sie die doppelten Übertragungsraten im Vergleich zum IBM 729 Bandsystem, der Weiterentwicklung der IBM 727.

Am **7. April 1964** wurde mit dem **System/360** der **Magnetstreifenspeicher IBM 2321** angekündigt. Mit seiner Gesamtspeicherkapazität von 400 MB auf 1000 mit magnetisierbarer Oberfläche versehenen Plastikstreifen galt er zu dieser Zeit als Maßenspeicher. Jeder Streifen enthielt 100 Datenspuren, hatte eine Kapazität von 0.2 MB und wurde in sogenannten **'IBM 2321 Data Cell Drives'** aufbewahrt. 200 solcher Streifen konnten in einer Datenzelle untergebracht werden. Die sich auf einem drehbaren Karussell befindlichen Datenzellen wurden durch Rotation so positioniert, dass der gesuchte Speicherstreifen durch einen Greifer erfasst und um die Schreib-/Lesetrommel geführt werden konnte. Damit konnte die adressierte Datenspur gelesen oder beschrieben werden. Die aufwändige Konstruktion und der damit verbundene hohe Wartungsaufwand standen einer weiten Verbreitung entgegen.

Die Anfangsepoche der elektromagnetischen Speicherung



Prinzip und Arbeitsweise des IBM 2321 Magnetstreifenspeichers

IBM 2321, Gesamtkapazität von 400 MB, Ausbaustufen von 40 bzw. 80 MB, Zugriffszeiten von 95 bis 600 ms

Kommentar zur Anfangsepoche der elektromagnetischen Speicherung

Neben der Erfindung der ersten externen Speichereinheiten mit dem Rollenband IBM 726 und der Platteneinheit IBM RAMAC 350 war der Topmeilenstein die Einführung der Platteneinheit IBM 1301.

Die IBM 1301 war das erste Plattensystem, das mit 'Air Bearing Slidern' arbeitete und für den Zugriff einen Zugriffskamm verwendete, der jede Plattenoberfläche im Plattenstapel mit dedizierten Schreib-/Leseköpfen ansteuerte. Mit der IBM 1301 wurde die Zylinderarchitektur im Plattenstapel eingeführt, die bis heute ihren Stellenwert behalten hat.

Bemerkenswert ist auch die Tatsache, dass bereits damals nach Möglichkeiten gesucht wurde, kleine Files, die nicht so häufig gebraucht wurden, auf einem billigeren Maßenspeicher abzulegen. Für diesen Zweck wurde der Magnetstreifenspeicher IBM 2321 eingeführt.

Geprägt wurde die Epoche auch dadurch, dass bei neuen Systemankündigungen immer die dazugehörigen neuen Platteneinheiten, die dem neuen System leistungsmäßig angepasst waren, Teil der Systemankündigung waren.

Die Anfangsepoche legte viele Grundlagen für die Folge-Epochen.

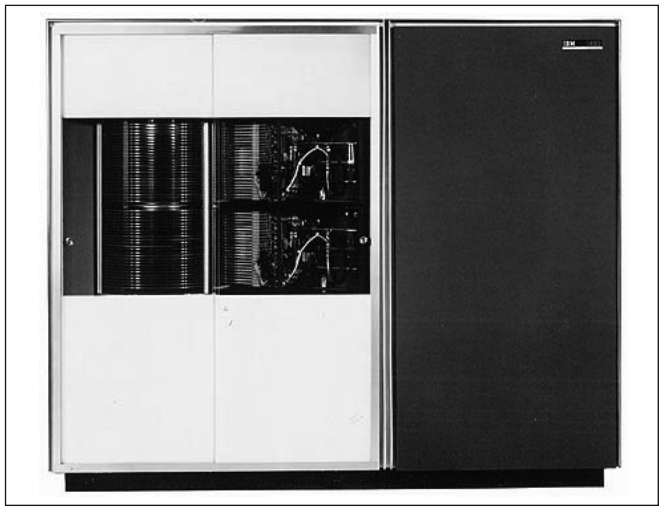
Die Epoche der Wechselplatten und die ‘Winchester’-Zeit

Die Epoche der Wechselplatten und die 'Winchester'-Zeit

IBM 1311: Ankündigung am 11. Oktober 1962, Zurückziehung aller Modelle 1971

Die **zwischen 1962 und 1964** eingeführten **Magnetplattenspeicher IBM 1311, IBM 2311 und IBM 2314** zum Anschluss an die 360er-Systeme arbeiteten mit auswechselbaren Magnetplattenstapeln. Die neue **Plattensteuereinheit IBM 2841** war **erstmals Mikroprogramm-gesteuert**. Dies war erforderlich, weil mehrere Speichertypen gesteuert werden mussten, einschließlich des neu verwendeten CKD-Formats. IBM entwickelte das Aufzeichnungsformat CKD (Count, Key, Data, frei übersetzt: Sektorkennzeichnung, Schlüssel, Daten), das als optimierte Version ECKD über das Jahr 2000 hinaus gültig blieb und bis heute, im Jahr 2006, im Mainframe-(z/OS)Umfeld eingesetzt wird. Das 'E' steht für 'Extended', also zu deutsch 'erweitert'. Wie ein Zugriff auf Daten nach dem CKD-Verfahren abläuft, ist mit dem Produkt der IBM 3330 beschrieben.

Das erste Modell des **Magnetplattenspeichers IBM 1311** mit Wechselplatten wurde im Oktober **1962 mit dem System IBM 1440** angekündigt. Der Wechselplattenstapel war aus sechs 14-Zoll-Platten aufgebaut, wog 10 Pfund und bildete eine Speicherkapazität von 2 MB ab. Im Vergleich zur IBM 1301 verdoppelte sich die lineare Speicherdichte auf 1025 Bits pro Inch. Dies wurde dadurch erreicht, dass der Abstand zwischen den Köpfen zu den Plattenoberflächen im Vergleich zur IBM 1301 um Faktor 2 auf 125 Micro-Inches reduziert wurde. Die Platten drehten mit 1500 RPMs und boten eine durchschnittliche Zugriffszeit von 150 ms.



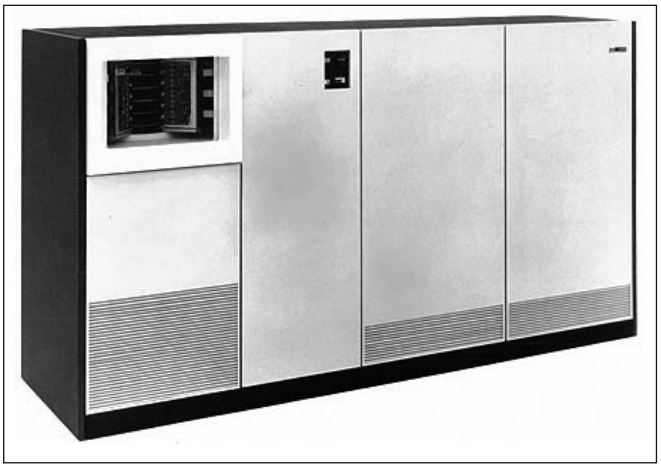
Das IBM 1302 Festplattensystem wurde am 23. September 1963 angekündigt und am 9. Februar 1965 wieder vom Vertrieb zurückgezogen

IBM 2401 Magnetbandeinheit

1963 wurde noch ein **Nachfolger der IBM 1301** mit fest eingebauten Platten angekündigt. Die **IBM 1302** sah optisch genauso aus wie die IBM 1301, bot aber auf der Plattenoberfläche 500 Datenspuren (1301 hatte 250 Spuren pro Plattenoberfläche). Wie bei der 1301 wurden 20 Platten im Stapel verwendet und die Kapazität von 28 MB auf 58.5 MB gesteigert. Damit bot das Modell 1 der 1302 eine Kapazität von 117 MB und das Modell 2 von 234 MB. Die Datentransferrate wurde von 90 KB/s auf 180 KB/s gesteigert.

1964 kam **speziell für die Rechnerfamilie System/360** das neue **Magnetbandsystem IBM 2401** auf den Markt, das mit 9 Spuren gleichzeitig auf dem Rollenband aufzeichnen konnte und eine wesentlich höhere Zuverlässigkeit im Vergleich zu den Vorgängern IBM 726, 727 und 729 bot. Dies wurde durch die Einführung eines sogenannten CRC (Cyclic Redundancy Check) erreicht, das die erste automatische Fehlerkorrektur ermöglichte (Automatic Error Capture & Correction) und Basis für die späteren ECCs (Error Correction Codes) bei Bandsystemen lieferte.

Im April **1965** – ein Jahr nach der Ankündigung des System/360 Prozessors – wurde das **Plattensystem IBM 2314** als 'IBM Direct Access Storage Facility' angekündigt. Dies war das erste Plattensystem, wo Controller und der Strang mit Plattenstapeln in einer Konfiguration bestellt und aufgebaut werden konnten. Mit der 2314 wurde ein neuer Plattenstapel mit der doppelten Anzahl an Plattenoberflächen eingeführt, der eine Speicherkapazität von 29.2 MB abbildete. Das Gesamtsystem bot eine maximale Kapazität von 233.4 MB. Im Vergleich zur 1311 war die 2314 um den Faktor 4 billiger, was das gespeicherte MB betraf. Die Datenrate lag bei 310 KB pro Sekunde. Im Januar 1969 kamen zwei neue Versionen der 2314 auf den Markt, die Modelle A1 und A2, die eine um 20 % schnellere Zugriffszeit boten, die im Bereich von 60 bis 75 ms lag. Die Kapazität blieb dieselbe. Im Dezember 1970 wurde die 2314 Plattenfamilie noch durch ein Modell B1 ergänzt, das 50 % höhere Kapazität anbot.



IBM 2305 Fixed Head Storage, bekannt als 'Zeus-Speicher', Markteinführung 1971, Zurückziehung vom Vertrieb im Januar 1980

Im **Januar 1970** wurde das Plattensystem **'IBM 2305 Fixed Head Storage'** angekündigt, das unter dem Entwicklungsnamen 'Zeus' entwickelt wurde und als **IBM Zeus** bekannter wurde als unter dem offiziellen Typennamen. 1971 erfolgten die ersten Auslieferungen. Die IBM 2305 war für große Datenbankanwendungen und für das Batch Processing bestens geeignet und wurde bei großen System/360 Rechnern der Modelle 85 und 195 und später dann bei den System/370 Modellen 155 und 165 für diese Applikationsformen eingesetzt. Mit zwar kleinen Speicherkapazitäten von 5.4 MB und 10.8 MB, jedoch mit einer durchschnittlichen Zugriffszeit von 2.5 ms bzw. 5 ms **war das System das schnellste seiner Zeit.** Die Datenrate lag bei 3 MB/s und damit 10-fach höher als bisherige Plattensysteme.

IBM 2314 Produktionslinie in Mainz 1969

IBM 2314 Plattensystemfamilie, Ankündigung Modell 1 am 22. April 1965, Modelle A1 und A2 am 10. Januar 1969, Modell B1 am 14. Dezember 1970, Zurückziehung vom Vertrieb im Jahr 1978

Die Epoche der Wechselplatten und die 'Winchester'-Zeit

1970 – 1972: Der virtuelle Speicher, der auf Magnetplatten abgebildet wurde, schaffte eine neue Qualität in der Beziehung zwischen Prozessoren und Magnetplattenspeichern. Online-Anwendungen gewannen rasch an Bedeutung. Wie die neuen Prozessoren IBM/370 trugen Subsysteme aus den Steuereinheiten **IBM 3830** und **Magnetplatteneinheiten IBM 3330** den steigenden Leistungs- und Zuverlässigkeitsanforderungen Rechnung. Neue Blockmultiplexkanäle sorgten für wesentlich höhere Datenübertragungsraten und trugen zu einer wesentlich höheren Gesamtleistung bei. Dazu einige Einzelheiten:

Die Plattensteuereinheit IBM 3830 war Mikroprogramm-gesteuert, jedoch erstmals mit einem ladbaren Mikroprogrammspeicher, der über eine Diskette geladen wurde. Mit dem Mikroprogramm war es möglich, mehr Funktionen als in den Vorgängermodellen zu realisieren. Dazu zählten das Erkennen und Beseitigen von Bitfehlern, gegebenenfalls mehrfaches Wiederholen von Kanalbefehlen vor Übergabe einer Fehlerbehandlung an den Rechner und das Sammeln von Informationen über aufgetretene Fehler für das Systemprotokoll (**LogRec**). Diese Aufzeichnungen lieferten den Bedienern Informationen über den technischen Zustand von Datenträgern und boten dem technischen Außendienst Hinweise für vorbeugende Wartung.

Bei der Empfindlichkeit magnetischer Aufzeichnung gegen elektrische und mechanische Störungen musste man von Anfang an mit Bitfehlern rechnen. Über Zusatzbits zur Paritätsprüfung entwickelte IBM immer ausgefeiltere Methoden, um Redundanzbits zu generieren, die den gespeicherten Daten angehängt wurden. Bei den **Aufzeichnungsformaten CKD und ECKD** findet diese 'Redundanz'-Aufzeichnung in den Zwischenräumen (Gaps) zwischen den Datensätzen statt. Mit ihrer Hilfe ließen sich durch entsprechende Programme in Rechnern und Steuereinheiten die meisten Bitfehler beim Übertragen ohne nennenswerte Zeitverzögerung korrigieren und durch technische Einflüsse verfälschte Daten praktisch ausschließen. Die Verfahren setzten für jede Datei die Definition sogenannter 'physischer Sätze', einer einheitlichen Länge in Bytes, voraus. Solche Sätze wurden als Blöcke bezeichnet. Bei Eingabe-/Ausgabe-Operationen wurden immer komplette Blöcke übertragen und geprüft. Meistens war es zweckmäßig, in einem Block mehrere 'logische Sätze' zusammenzufassen. Sie bestanden aus einem Ordnungsbegriff, dem in 'Feldern', die nach Lage und Länge definiert waren, die zugehörigen Angaben und Informationen folgten. Bei administrativen Anwendungen wären z. B. Artikel-, Kunden- oder Personalnummern die Ordnungsbegriffe. In den Feldern standen dann alphabetische und numerische Informationen. Die Anzahl der logischen Sätze im physischen Satz bezeichnete man als Blockungsfaktor.

IBM 3330, Kapazität 800 MB, 2 – 8 Laufwerke, Zugriffszeit 30 ms, Datenrate 806 KB/s, Ankündigung Modell 1 am 30. Juni 1970, Modell 2 am 4. Oktober 1972, Modell 11 am 17. Juli 1973, Zurückziehung vom Vertrieb am 20. Dezember 1983

Nach bisherigen hydraulischen Zugriffsmechanismen benutzten die Magnetplatteneinheiten **IBM 3330** erstmals für das 'Suchen', die **horizontale Einstellung des Zugriffskammes** mit den Schreib-/Leseköpfen, einen Linearmotor. Wie bei Lautsprechern wirkt ein durch variablen Stromfluss in einer Spule erzeugtes Magnetfeld dem statischen Feld eines Permanentmagneten entgegen. Deshalb die englische Bezeichnung **'Voice Coil Motor'** (Lautsprecherspulenmotor). Die Lage der Schreib-/Leseköpfe zu bestimmten Zeitpunkten ergab sich aus der Stärke des von der Spule erzeugten Magnetfeldes. Diese Technik setzte voraus, dass jeweils eine Oberfläche eines Plattenstapels mit vom Benutzer unveränderbaren Servospuren versehen war. Der darüber angeordnete (Nur-)Lesekopf versorgte die elektronische Steuerung mit den für gezieltes Positionieren im Rahmen eines Regelungskreislaufes notwendigen Informationen. Diese Technik reduzierte damals die durchschnittliche Dauer einer Suchoperation auf Anhieb um die Hälfte: statt 60 dann nur noch 30 ms.

Für einen verbesserten Ablauf des direkten Zugriffs auf Daten war Folgendes von Bedeutung: Die Steuereinheit IBM 3830 konnte mithilfe entsprechender Informationen von der Servoplatte als erste die Dreh- oder Winkelposition der an sie angeschlossenen Magnetplatteneinheiten erkennen (**RPS – Rotational Position Sensing**). Mit dem Plattensystem IBM 3830/3330 lief ein Routinezugriff wie folgt ab: Nach Angaben aus den Datenträgerinhaltsverzeichnissen, den sogenannten **VTOCs** (Volume Table Of Contents) und den Indizes zu den Dateien generierte der Rechner mithilfe der Programme der Zugriffsmethodensteuerung einen Kanalbefehl. Dieser enthielt die Plattenadresse, von der zu lesen oder auf die zu schreiben war, nach Einheit, Zylinder, Spur und Sektor. Die konzentrischen Spuren auf den Plattenoberflächen waren in Sektoren eingeteilt. Als Zylinder bezeichnete man die Summe der übereinanderliegenden Spuren auf einem Plattenstapel, von denen bei einer gegebenen horizontalen Einstellung eines Zugriffskammes mit ebenfalls übereinander angeordneten Schreib-/Leseköpfen gelesen oder auf die geschrieben werden konnte. Die Steuereinheit wandelte den Kanalbefehl in spezifische Befehle für die Einheiten um. Den Suchbefehl zum Einstellen des Zugriffskammes auf den korrekten Zylinder führten die Einheiten selbstständig aus und meldeten der Steuereinheit den Abschluss der Suchoperation. Dann aktivierte die Steuereinheit den richtigen Schreib-/Lesekopf, analysierte seine Position bezogen auf die Sektoren und begann mit dem Übertragen der Daten über den

IBM 3340, Maximalkapazität bis 4800 MB, Winchester-Wechselplatten Modell 35: 35 MB, Modell 70: 70 MB, bis zu 64 Laufwerke (Modelle 158, 168), Datenrate 885 KB/Sekunde, Zugriffszeit 25 ms

Kanal des Rechners, sobald der Anfang des im Kanalbefehl vorgegebenen Sektors erreicht war. Der Kanal war nur während der Datenübertragung belegt. Mehrere Platteneinheiten konnten Suchbefehle überlappt ausführen. Auch die Umdrehungswartezeit (damals bei ca. 17 bis 20 ms) belastete den Kanal nicht. Diese Methode erlaubte höhere systemeffektive Datenraten des Magnetplattensystems, die bei günstigen Konfigurations- und Anwendungsprofilen nicht sehr weit unter der Übertragungsleistung der Einheiten lag.

1973: Das an mittlere Rechnersysteme IBM/370, Modelle 115, 125 und 135, direkt anschließbare neue **Magnetplattensystem IBM 3340** benutzte als auswechselbaren Datenträger das **Datenmodul IBM 3348**. Die auswechselbaren Datenmodule 3348 boten 35 oder 70 MB Kapazität. Ein Modul enthielt in einem geschlossenen, staubgeschützten Gehäuse den Plattenstapel und einen horizontal beweglichen Zugriffskamm mit zwei Schreib-/Leseköpfen je Datenoberfläche. Das Modul war auf die Welle des Antriebsmotors im Laufwerk aufgesetzt. Nach dem automatischen Öffnen einer seitlichen Jalousie verbanden elektrische Steckverbindungen den Zugriffskamm mit dem Laufwerk und eine mechanische den Zugriffskamm mit dem Linearmotor. Der Grund für diese Konstruktion: Wenn auswechselbare Plattenstapel als Datenträger dienten, bedeutete das insbesondere für Zugriffskämme extreme Ansprüche an die Fertigungstoleranzen, weil jeder Stapel zu jedem beliebigen Laufwerk passen musste und schon minimale horizontale Abweichungen der Lage von Köpfen über Spuren zu Lesefehlern führte. Dieser Umstand setzte, der horizontalen Spurdichte Grenzen und minderte, die Sicherheit und Zuverlässigkeit der Systeme. Maximal vier Module mit einer maximalen Kapazität von 280 MB konnten über einen Direktanschluss am IBM/370 Rechner betrieben werden.

Die Epoche der Wechselplatten und die 'Winchester'-Zeit

Aufbau IBM 3340 Winchester-Wechselplattenstapel

Die IBM 3340 mit den Modellen A2, B1, B2 und C2 wurden am 13. März 1973 angekündigt. Die Modelle B1 und C2 wurden am 20. Dezember 1983 und die Modelle A2 und B2 am 1. Mai 1984 vom Vertrieb zurückgezogen.

Die Tatsache, dass beim Datenmodul der Magnetplattenspeicher IBM 3340 immer der gleiche Zugriffskamm schrieb und las, ermöglichte höhere Spurdichten bei gleichzeitig gesteigerter Zuverlässigkeit. Die Konstruktion insgesamt führte zu kürzeren Suchzeiten und etwas höherer Übertragungsleistung gegenüber der IBM 3330.

Mit den neuen **Magnetbandeinheiten IBM 3420, Modelle 3, 5 und 7**, die in 9-Spur-Technik mit einer Aufzeichnungsdichte von 1600 Bytes pro Zoll arbeiteten, konnte eine Datenrate – je nach Modell – von 120, 200 und 320 Kilobytes pro Sekunde erzielt werden. Diese Modelle wurden 1970 eingeführt.

Die neuen Magnetbandeinheiten **IBM 3420, Modelle 4, 6 und 8**, die **1973** auf dem Markt eingeführt wurden, steigerten mit dem Verfahren der 'gruppencodierten Aufzeichnung' (Group Coded Recording) die Aufzeichnungsdichte von 1600 auf 6250 Bytes per Zoll und die maximale Übertragungsrate wurde – je nach Modell – auf 470, 780 und 1250 Kilobytes pro Sekunde gesteigert. Gleichzeitig erhöhte sich die Zuverlässigkeit um den Faktor 7 (im Vergleich zu den Modellen 3, 5, und 7).

Um die Bedienung zu erleichtern, kündigte IBM **1971** parallel zum 3420 Magnetbandsystem das **IBM 3410 Magnetbandsystem** an, das von der Bauhöhe her Pultform hatte und erstmals erlaubte, die Rollenbänder horizontal von oben einzulegen. Die Kapazitäten und die Technologie waren mit der 3420 vergleichbar, allerdings war die Schreib-/Lesegeschwindigkeit wesentlich niedriger. Die 3410 wurde deshalb – neben dem 3420 Magnetbandhochleistungssystem – als kostengünstige Alternative für Benutzer der IBM System/370 und IBM System/360 angeboten.

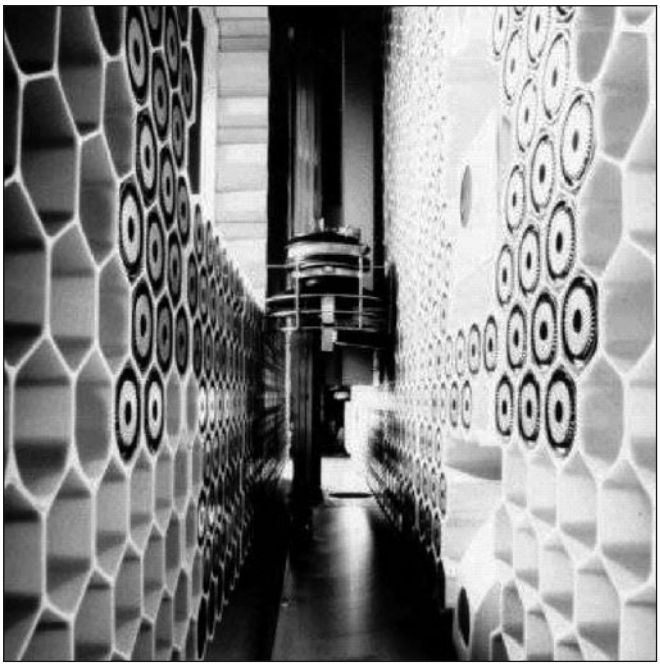
1974: Im Zuge des verstärkten Übergangs der mit dem System/370 begonnenen Direktverarbeitung mit preisgünstigen Terminals zur Programmentwicklung und Problemlösung (VM/370, APL) wuchs die Zahl der Dateien bei Benutzern großer Systeme dramatisch. Anders als bei den mit wenigen großen Dateien und ausgefeilten Sicherungsverfahren arbeitenden Datenbankanwendungen (IMS, CICS), wurde es immer schwieriger, speziell die vielen kleinen, unregelmäßig

IBM 3420 Magnetbandeinheit, 1970 Modelle 3, 5 und 7, 1973 Modelle 4, 6 und 8

IBM 3410 Magnetbandsystem mit horizontaler Bandeinlegemöglichkeit

benutzten Bestände unter manueller Verwaltung zuverlässig zu sichern und zu archivieren. Um diese Aufgabe zu lösen und um kostengünstigeres Speichern von Daten unter Systemkontrolle zu ermöglichen, führte IBM neue Hardware und Software ein.

Der Maßenspeicher **MSS (Mass Storage System) IBM 3850** bildete auf zwei speziellen zylinderförmigen Magnetbandpatronen den Inhalt eines Datenträgers des Magnetplattenspeichers IBM 3330 ab. Man kann ihn als **ersten virtuellen 'Plattenspeicher'** bezeichnen, der gleichzeitig ein automatisches Bandarchiv darstellte. Alle aktiven Patronen lagerten in einem Schrank mit einander gegenüberliegenden Regalen in wabenförmigen Zellen. Bei Anforderung durch die Programme bewegte eine sinnreiche Konstruktion einen sogenannten Picker automatisch zur richtigen Patrone. Der Picker hielt den Datenträger im Gegensatz zu späteren mechanischen Greifern elektromagnetisch fest und brachte ihn zu einer Schreib-/Lesestation, von der die gewünschten Daten zum Verarbeiten auf bestimmte Magnetplattenlaufwerke (Staging Devices) übertragen wurden oder umgekehrt. Nach Ablauf der Operation brachte die gleiche Vorrichtung die Patronen an ihren Lagerort zurück. Der Bediener war nur noch mit den Patronen befasst, die der Massenspeicher nach Programmvorgaben über entsprechende Stationen aussteuerte oder die er von außen anforderte. Der Maßenspeicher konnte bis zu 236 Gigabytes im Zugriff der Prozessoren halten. Die technische Entwicklung der Magnetspeicher – Platte, Band – ließ keinen sinnvollen Nachfolger zu, während die zugehörige Software bis heute, im Jahr 2006, hochaktuell blieb.



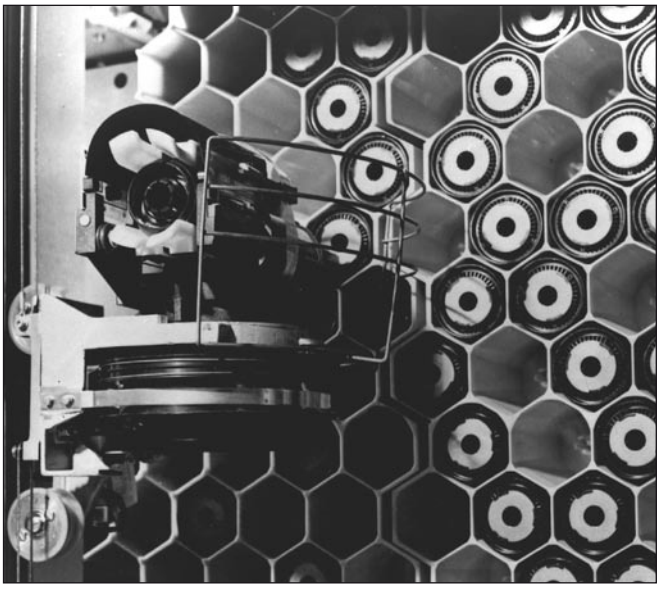
IBM Mass Storage System 3850 mit 'Honeycomb Racks' 1974

IBM MSS 3850 im Rechenzentrum im Hintergrund rechts

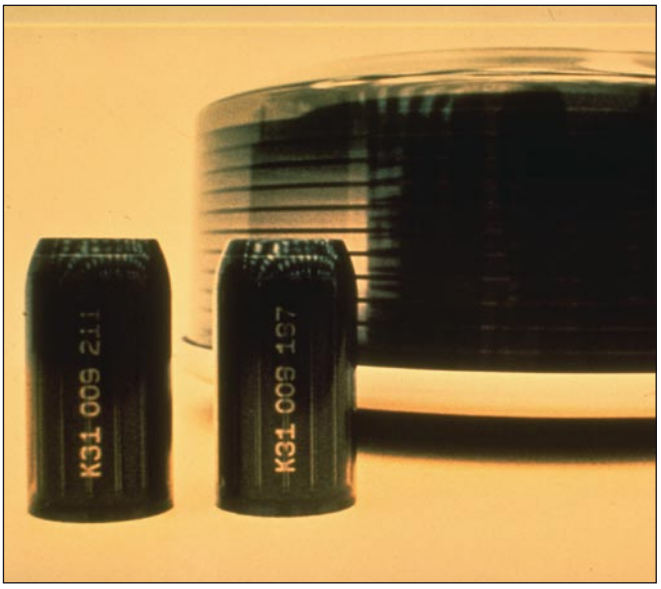
ILM, Information Life Cycle Management, das heute in aller Munde ist, wurde bereits intensiv im Jahr 1974 gelebt! Der HSM (Hierarchical Storage Manager) als Verwalter der Speicherhierarchie war damals ein Programm, das zunächst in Verbindung mit dem Massenspeicher die ihm unterstellten Dateien gemäß vorgegebener Steuerungskriterien automatisch sicherte und im Bedarfsfall wiederherstellte. Eine weitere damals bereits verfügbare Funktion war das Verlagern von Dateien vom teuren Plattenspeicherplatz auf billigere Speicher (ursprünglich auf den MSS) während der Zeiten, in denen sie nicht benötigt wurden. Um den Speicherplatz optimal zu nutzen, konnte der HSM zu archivierende Dateien komprimieren, indem er leere Bereiche auf den Datenträgern ausließ und er konnte darüber hinaus Daten mithilfe binärarithmetischer Verfahren verdichten (Compression, Compaction). Zum Verarbeiten stellte er den ursprünglichen Zustand der betreffenden Datei wieder her.

An unterschiedlichste Konfigurationen mit unterschiedlichen externen Speichern angepasst, blieb der HSM fester Bestandteil späterer Pakete von Dienstleistungsprogrammen für den Bereich externer Speicher, z. B. der IBM DF-Produkte (Data Facility), die den systemverwalteten Speicher zum Ziel hatten. So wurden die DF-Produkte als Programm-Produkte im Mainframe-Umfeld, beginnend mit DFP, 1989 eingeführt. Kurz darauf folgte der DFDSS und der DFHSM, die dann alle unter dem Programmpaket DFSMS (Data Facility System Managed Storage) 1992 zusammengefasst wurden und ein Jahr später komplett ins MVS-Betriebssystem integriert wurden. HSM spielt unter unterschiedlichen Bezeichnungen bis heute eine ganz maßgebliche Rolle.

Die Epoche der Wechselplatten und die 'Winchester'-Zeit



IBM 3850 Roboter mit Doppelgreifer (Standard) in Aktion



IBM 3850 Bandpatronen mit 19.5 Metern Bandlänge und einer Kapazität von jeweils 50 MB im Größenvergleich zum 3330 Plattenstapel mit 100 MB Kapazität

Das MSS 3850 gab es in acht Modellvarianten. Das kleinste Modell speicherte bis zu 706 Bandpatronen, bot eine Kapazität von 35.3 GB und konnte bis zur größten Variante mit bis zu 4720 Patronen und einer Kapazität von 236 GB ausgebaut werden.

Kommentar zur Epoche der Wechselplatten und der 'Winchester'-Zeit

Das starke Ansteigen der Rechnerleistungen durch neue Systeme machte es notwendig, viele neue Dinge auf der Speicherseite einzuführen, um den neuen Leistungsanforderungen dieser Rechner gerecht zu werden.

So wurden mit der IBM 2841 und IBM 3830 die ersten Mikroprogramm-gesteuerten Steuereinheiten und das bis heute gültige CKD-Format (heute ECKD bei zSeries) eingeführt. Für höhere Übertragungsbandbreiten sorgten die neu eingeführten Blockmultiplexing-Kanäle, die maßgeblich zu einer Verbesserung der Gesamtleistung von Rechner und Speicherperipherie beitrugen.

Bei den Platten selbst wurden kleinere Formfaktoren und kleinere Flughöhen zwischen Kopf und Plattenoberfläche eingeführt, um die Leistung durch schnellere Drehgeschwindigkeiten und Datenraten sowie die kapazitiven Möglichkeiten zu verbessern. Fast parallel dazu wurde die Leistung der Bandsysteme angepasst und durch neue Möglichkeiten der Mehrspurtechnik wurden die Datenraten erhöht.

Die Epoche war wie die Anfangsepoche durch neue Systemankündigungen geprägt, in denen immer das dafür passende Platten- und Bandsystem Teil der Ankündigung war. Davon losgelöste Speicherankündigungen fanden nicht statt.

Der Einsatz der damals verwendeten Wechselplatten war bei den Endbenutzern sehr beliebt, weil man diese tragbaren Plattenmodule auch für den Datenträgeraustausch zwischen unterschiedlichen Lokationen einsetzen konnte. Die 1970 und 1973 neu angekündigten Modelle der 3420 wurden ausschließlich für Backup-Zwecke eingesetzt. Erst in der Folge-Epoche, in der man wieder fest eingebaute Platten einsetzte und vom Prinzip der Wechselplatten wegging, erlebte die 3420 eine wahre Blütezeit. Das Rollenband war dann das einzige Medium, das als auswechselbarer Datenträger eingesetzt werden konnte.

Da die kapazitiven Anforderungen immer größer wurden und neben Online-Speichern neue Maßenspeicher mit höheren Kapazitäten verlangt wurden, war eines der Meilensteine dieser Epoche sicherlich das IBM 3850 Maß-Storage-System, das erste Nearline-Roboter-System der Welt mit Bandpatronen, die wie die damaligen Wechselplatten als Online-Speicher adressiert wurden. Die Tatsache, dass bei dem MSS erstmals ILM über eine HSM-Funktionalität eingeführt wurde, muss besonders hervorgehoben werden. Diese HSM-Funktionalität legte den Grundstein für die späteren DF-Produkte (Data Facility) im Mainframe-Umfeld.

Die Epoche der fest eingebauten Platten mit externen Kontrolleinheiten

Die Epoche der fest eingebauten Platten mit externen Kontrolleinheiten

IBM 3350, Kapazität bis 2536 MB, 317 MB per LW, 8 Laufwerke pro 3350 Strang, Zugriffszeit 25 ms, Datenrate 1198 KB/s, links hinten: Display-Einheit für Microfiches, rechts hinten: Steuerkonsole des Systems/370 Modell 168

IBM 3310 Direct Access Storage, Plattensystem für IBM 4331 Rechner

1975: Mit dem **Magnetplattenspeicher IBM 3350** ging die IBM wieder zu fest eingebauten Platten über. Das bedeutete einen entscheidenden Schritt zu höherer Zuverlässigkeit und Verfügbarkeit sowie höherer Aufzeichnungsdichte. Bei dieser Konstruktion waren gegenüber der mit Datenmodulen arbeitenden Type IBM 3340 potenzielle Störstellen nicht mehr vorhanden, auf die 70 % der notwendigen Technikereingriffe zurückzuführen waren. Tatsächlich registrierte der IBM Technische Außendienst bei IBM 3350 im Durchschnitt weniger als einen Eingriff pro Einheit und Jahr. IBM 3350 und die Nachfolgeprodukte setzten das mit IBM 3330 begonnene Strangkonzept fort. An eine Kopfeinheit (Master Unit) mit zwei Laufwerken konnte man bis zu drei Nebeneinheiten anschließen. Ein Strang bestand also aus acht Laufwerken. Die Kopfeinheit besaß grundsätzlich zwei Ein-/Ausgänge zur Steuereinheit IBM 3830. Die beiden Verbindungen konnten zu verschiedenen Einheiten 3830 führen. Fielen ein Kanal, eine Steuerung oder Steuerelemente in der Kopfeinheit aus, lief der Betrieb auf der zweiten Verbindung zu jedem Laufwerk weiter, wenn auch mit geringerer Leistung. Man kann in diesem neuen Konzept den **ersten Einstieg in Konzepte der Fehlertoleranz** erkennen.

Weil mit dem Magnetplattensystem IBM 3350 mit fest eingebauten Platten nun die Magnetbandrolle das einzige auswechselbare Speichermedium für Datensicherung, Offline-Archivierung und Datenträgeraustausch blieb, löste die IBM 3350 einen deutlichen Verkaufsanstieg bei Magnetbandeinheiten aus, insbesondere bei den schnellen Modellen 6 und 8 des Magnetbandsystems 3420. Fortan waren Magnetbänder unverzichtbarer Bestandteil größerer und auch mittlerer Datenverarbeitungssysteme.

1979 führte IBM ein weiteres Plattensystem ein, das **IBM 3310 Direct Access Storage**, das in einer kompakten Bauweise hohe Leistung bei einem sehr günstigen Preis bot und seinen Einsatz im Anschluss an IBM 4331, 4341, 4321, 4361 und kleine 4381 Rechner fand. Jeder Plattenstapel bot eine Gesamtkapazität von 64.5 MB und der gesamte Plattenstrang 258 MB. Bis zu vier Plattenstränge konnten an obigen Rechnermodellen betrieben werden und boten eine Gesamtkapazität von 1032 MB.

1979 – 1983: IBM 3370 und 3375 waren neu entwickelte Magnetplatteneinheiten an der neuen **Steuereinheit IBM 3880** für damals mittlere Systeme. Sie arbeiteten aus Gründen der Vereinfachung und Beschleunigung von Programmabläufen mit festen Blöcken von 512 Bytes. Der Benutzer definierte größere Blocklängen mit einem Vielfachen von 512. Damit war die erste FBA(Fixed Block Architecture)-Platte geboren.

IBM 3380 Platten und 3880 Steuereinheiten füllten damals ganze Rechenzentrumshallen

Die **neue IBM 3880 Steuereinheit** leitete ein neues Steuerungsprinzip ein, das in der Weiterentwicklung 1985 die erste Integration von Pufferspeichern, anfangs von 8 bis 16 MB, zur Folge hatte. Diese kleinen Pufferspeicher waren für das Ein- und Auslagern von Seiten (Paging, Swapping) gedacht und wirkten sich zunächst auf die Leistung älterer Konfigurationen ohne Erweiterungsspeicher aus. Die Verwendung von größeren Pufferspeichern führte dann zur Einführung neuer Caching-Algorithmen. Die zu schreibenden Daten wurden in der Steuereinheit auf Halbleiterspeicherkarten zwischengespeichert und auf einem stromunabhängigen Schreibspeicher dupliziert. Danach vermeldete die Steuereinheit dem Rechner den Abschluss der Operation, obwohl die Daten noch nicht auf die Platten geschrieben waren. Diese Operation wurde damals als 'DASD Fast Write' bezeichnet, weil durch dieses Prinzip die Antwortzeiten zwischen den Plattensystemen und dem Rechner um ein Vielfaches verbessert wurde. Dieses damals eingeführte 'Caching'-Prinzip für Plattensubsysteme wurde dann rasch weiterentwickelt und hat bis heute, im Jahr 2006, Gültigkeit.

1981: IBM 3380 waren **Magnetplatteneinheiten**, die als Neuentwicklung für die 3880 Steuereinheiten optimiert waren und das CKD-Format fortsetzten. Die IBM 3380 übertrug mit 3 MB pro Sekunde und nutzte dabei spezielle Fähigkeiten der Blockmultiplexkanäle (Data Streaming). Die neuen Laufwerke kamen gerade rechtzeitig, denn zu diesem Zeitpunkt war sehr deutlich geworden, wie notwendig es bei der engen Beziehung zwischen Prozessoren und Plattenspeichern war, die Leistung beider aufeinander abzustimmen.

1981 wurden die Standardmodelle AA4 und B04 der IBM 3380 angekündigt. **1985** folgten die erweiterten Modelle AD4, BD4, AE4 und BE4. **1987** wurden die schnellen AJ4 und BJ4 und die schnellen, großkapazitiven Modelle AK4 und BK4 angekündigt. Ebenso erfolgte **1987** die Ankündigung der 3380 CJ2, eines Modells, bei dem die Steuereinheit im Plattengehäuse integriert war.

IBM 3370, 1 Plattenstapel pro Einheit mit 2 Zugriffsarmen und 2 Adressen, Modell 1: 571 MB, Modell 2: 730 MB, jeweils per Einheit, bis zu 4 Einheiten mit 2285 – 2920 MB Kapazität, Zugriff 20 ms, Datenrate 1.86 MB/s

Die Epoche der fest eingebauten Platten mit externen Kontrolleinheiten

Das 3380 Zeitalter brachte im Antwortzeitverhalten, im Durchsatz und in den hohen Kapazitäten für die damals betriebenen IBM Rechner 3031, 3032, 3033, 3042 und /370 große Fortschritte. Die durchschnittlichen Suchzeiten bewegten sich – je nach Modell – zwischen 12 und 16 ms. Immer zwei Laufwerke waren in einer Gehäuseeinheit untergebracht. Pro Laufwerk wurden Kapazitäten von 1.26 GB (J-Modelle) und 3.78 GB (K-Modelle) abgebildet. Dies bedeutete für einen voll ausgebauten 4-pfadigen Strang von J-Modellen eine Gesamtkapazität von 20.16 GB und bei den K-Modellen wurden bis zu 60.5 GB Gesamtkapazität erreicht.

Die Platten wurden im 14-Zoll-Format gebaut. Um die gewichtigen Plattenstapel aus den Gehäuseeinheiten zu heben, wurden vom technischen Außendienst speziell dafür gebaute Hubwagen eingesetzt.

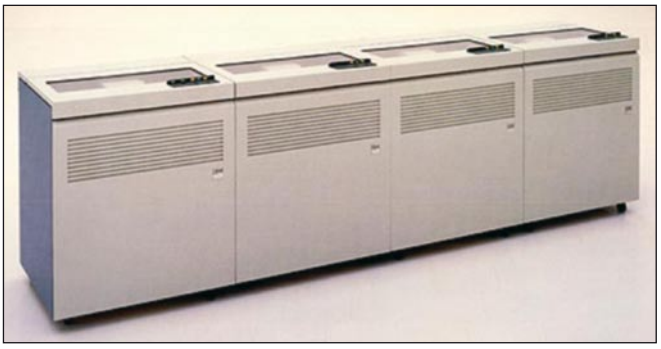
Im 3880/3380 Subsystem wurde aus Pfadredundanzgründen die Funktion **'Dynamic Path Reconnect'** eingeführt, die im Falle eines Pfadfehlers dafür sorgte, dass die Datenübertragung zu jedem HDA des Subsystems auf einem alternativen Pfad erfolgen konnte. Diese Multipfadarchitektur in Verbindung mit dem neuen Kanalsubsystem des Mainframes sorgte für einen um 30 % höheren Durchsatz im Speichersubsystem.

Physikalische Spezifikation der 3380 Plattensubsysteme:

Modelle	A/B	D	E	CJ2	J	K
Actuators per Unit	4	4	4	2	4	4
Zylinder per Device	885	885	1770	885	885	2655
Tracks per Device	13275	13275	26550	13275	13275	39825
GB per Unit (Box)	2.520	2.520	5.041	1.260	2.520	7.562

Auf der Magnetbandsubsystemseite führte IBM **1983** das letzte Rollenbandsystem mit **der IBM 3430** ein, das vor allem bei IBM 43xx- und IBM 303x Series-Prozessoren, aber auch bei dem damaligen Midrange System/38 seinen Einsatz fand. Die kompakte Bauweise erlaubte den Benutzern die Aufstellung auf wesentlich kleinerer Fläche im Vergleich zur IBM 3420. Die Rollenbänder wurden von oben horizontal eingelegt, wie bei IBM 3410 (1971).

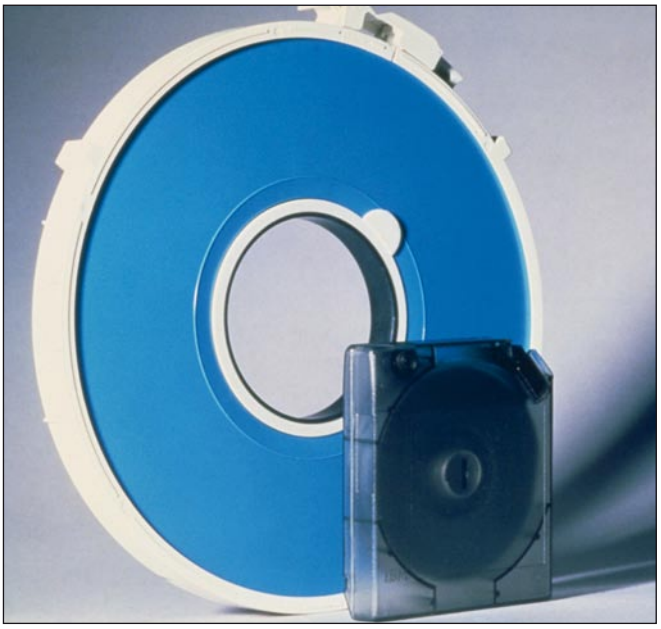
Aufgrund der ein Jahr später angekündigten neuen IBM 3480 Kassettentechnik als Ablösung der Rollenbänder fanden 3430 Einheiten nicht in großen Stückzahlen ihren Absatz, hielten sich aber dennoch ziemlich lange auf dem Markt, um der Datenträgeraustauschanforderung gerecht zu werden, die in den Folgejahren immer noch auf Rollenbändern gemacht wurde.



IBM 3430 Magnetbandsubsystem, letztes Rollenband der IBM, Markteinführung 1983

1984:

Die erreichte Dynamik in der Leistung von 3880/3380 Plattensubsystemen und /370 Rechnern erforderte auch eine Leistungsanpassung auf der Backup-Seite, **den Magnetbandsystemen**. Mit der **IBM 3480** wurde der Wechsel von den bis dahin verwendeten Rollenbandsystemen auf die Kassettentechnik eingeleitet. Das neue Magnetbandsystem IBM 3480 übertrug, gleich schnell wie die Plattenspeicher IBM 3380, mit 3 MB pro Sekunde und schloss damit die Lücke im Gesamtsystemverhalten. Die bis dahin verwendeten 10.5-Zoll-Rollenbänder der 3420 Technologie wurde durch handtellergroße, rechteckige Kassetten abgelöst. Bei einer Aufzeichnungsdichte von 38000 Bytes pro Zoll nahmen die neuen Kassetten doppelt so viel Daten auf wie herkömmliche Bandrollen, also bis zu 200 Megabyte. Mit einem Zusatz, einem Vorratsschacht (Stacker), konnten mehrere Kassetten hintereinander automatisch vorgeladen werden. Das erleichterte maßgeblich das Sichern großer Datenbestände.



IBM 3480 Magnetbandsystem

1984: Wechsel vom klassischen Rollenband auf die 3480 Kassettentechnologie

1989:

Mit der **IBM 3390 Plattentechnologie** wurde der bei IBM 3380 verwendete Formfaktor von 14 Zoll auf 10.8 Zoll reduziert. Anfangs wurde die 3390 Technologie noch mit der herkömmlichen 'braunen Beschichtung' produziert, wo Eisenoxyd-Partikel (Rostpartikel), die möglichst homogen in einer Kunstharzmasse zur gleichmäßigen Magnetisierung verteilt waren, als Magnetisierungsträger eingesetzt wurden. Die Einführung einer neuen Beschichtung, deren Entwicklung **1989** eingeleitet wurde und die **1991** zur Anwendung kam, ließ den Einsatz einer neuen Schreib-/Lesetechnik zu. Die neuen Köpfe schrieben – wie ihre Vorgänger – induktiv, erzeugten also Magnetfelder in der Beschichtung der Platten mithilfe eines winzigen Spalts im Metallkern miniaturisierter Elektromagnete. Bis zu diesem Zeitpunkt dienten die Elektromagnete auch zum Lesen. Für die Zahl der Spulenwindungen musste man einen Kompromiss für Lesen und Schreiben finden, um beide Vorgänge zu ermöglichen. Jetzt konnten sie ausschließlich für den Schreibvorgang optimiert werden, denn das Lesen erfolgt bei der 3390 Technologie über eine zweite Komponente, einen Lesefilm. Der Film war sehr klein und bestand aus dünnen Schichten verschiedener Metalle wie Eisen, Chrom und Kobalt. Kurze Zeit später wurde der Film auf Nickel und Eisen umgestellt. Wenn sich innerhalb der Spur auf der Platte die Magnetisierungsrichtung umkehrte, änderte sich der elektrische Widerstand vor allem in der mittleren Schicht des darüber befindlichen Kopfes sprunghaft. Dies bezeichnet man als magnetoresistiven Effekt, den man als Signal entsprechend verstärkt und auswertet. Diese neue Technik gewährleistete sicheres Lesen bei höchster Aufzeichnungsdichte, weil selbst kleinste Streufelder von Informationseinheiten differenziert werden konnten. Man führte – einfach gesagt – zwei Spezialisten ein, einen für das Schreiben optimierten induktiven Kopf und einen Spezialisten für den Lesevorgang (ausführliche Beschreibung im Technologie-Anhang). Beide Spezialisten zusammen bildeten das Schreib-/Leseelement, das vom Zugriffsmechanismus positioniert wurde.

Größenvergleich Magnetbandsysteme IBM 3420 gegenüber IBM 3480

Die neue Schreib-/Lesekopf-Technik der IBM 3390 benötigte eine neue Beschichtung in sogenannter Dünnfilmtechnik. Dabei beschichtete man die Oberfächen von Glasplatten aus einem speziell gehärteten Glas mit einer Legierung und unterzog sie anschließend einem Einbrennprozess. Die magnetoresistive Schreib-/Lesetechnik in Verbindung mit Dünnfilmbeschichtungen auf den Platten kam innerhalb der 3390 Baureihe mit den Modellen 3 und 9 erstmals zum Einsatz. Die seit der Einführung von Magnetplatten verwendetete Aluminiumplatte, die mit feinst gemahlenem weichmagnetischem Material in einem Kunstharz als Bindemittel beschichtet war, hatte damit ausgedient.

Im Vergleich zur IBM 3380 Baureihe lieferte die 3390 Baureihe eine Verbesserung in der Zugriffszeit von 28 % und die Datentransferrate wurde um 40 % gesteigert, weil anstelle von 3 MB/s mit 4.5 MB/s übertragen werden konnte. So konnten die Kanalgeschwindigkeiten des damals neuen 3090 Rechners genutzt werden. Ebenso konnte vom Rechner auf die Subsysteme über vier gleichzeitige Pfade übertragen werden (Four Path Data Transfer). Die 3390 Platten und auch noch die 3380 Platten wurden an der neuen IBM 3990 Steuereinheit angeschlossen, die mit höheren Plattenpuffergrößen (Cache) von 16 bis 64 MB (3990 Modelle 3) ausgestattet war.

Die Epoche der fest eingebauten Platten mit externen Kontrolleinheiten

1989: IBM 3390-1/2, 1991: IBM 3390-3, 1994: IBM 3390-9, hinten IBM 3380-K Strang mit 60 GB Kapazität, vorne 3390-3 Strang mit 90 GB Kapazität, rechts IBM 3990 Steuereinheiten

Ein voll ausgebauter 3390 Modell 3 Strang mit 1 x A38 und 2 x B3C hatte also eine Gesamtkapazität von 90 GB. Das Großkapazitätsmodell 9, das 1994 auf den Markt kam, verdreifachte nochmals die Kapazität. Allerdings kamen die Modelle 9 nur begrenzt zum Einsatz, weil die Platten aufgrund ihrer hohen Kapazität langsamer waren, und wurden für Anwendungen eingesetzt, wo die Zugriffszeiten und Datenraten ausreichten.

Physikalische Spezifikation der 3390 Plattensubsysteme

Modelle	Kapazität/ Actuator	Kapazität/ HDA	Anzahl HDAs	Kapazität/ Unit
A14	0.946	1.89	2	3.78
A18	0.946	1.89	4	7.56
B14	0.946	1.89	2	3.78
B18	0.946	1.89	4	7.56
B1C	0.946	1.89	6	11.35
A24	1.89	3.78	2	7.56
A28	1.89	3.78	4	15.1
B24	1.89	3.78	2	7.56
B28	1.89	3.78	4	15.1
B2C	1.89	3.78	6	22.7
A34	2.838	5.67	2	11.34
A38	2.838	5.67	4	22.68
B34	2.838	5.67	2	11.34
B38	2.838	5.67	4	22.68
B3C	2.838	5.67	6	34.02

1991:

In diesem Jahr wurde IBM seitig der erste Schritt in Richtung kleinerer Formfaktoren von Magnetplattenlaufwerken in Subsystemen eingeleitet. Die angebotenen **Plattensysteme IBM 9340** und IBM 0662, erstmals auf der Basis von Plattendurchmessern mit 5.25 bzw. 3.5 Zoll, bestanden aus einem Rahmen, in den die Steuereinheit zusammen mit mehreren Plattenlaufwerken eingebaut wurden. Ein Standardrahmen Typ 9309 nahm bis zu acht Platteneinschübe vom Typ IBM 9345 auf. Ein Platteneinschub enthielt zwei Laufwerke mit jeweils 1 oder 1.5 GB Speicherkapazität. Es waren 5¼-Zoll-Laufwerke neuster Technologie mit einer Datenrate von 4.4 MB/s, mittlerer Umdrehungswartezeit von nur 5.6 ms und in bewährter Count-Key-Data(CKD)-Speicherungsform. Für Anforderungen bis 24 GB gab es den Steuereinschub IBM 9341, der als 2-Pfad-Subsystem ohne Cache an die Prozessoren-Reihe 9221 oder 9370 angeschlossen war.

Größerer Speicherbedarf und noch höhere Anforderungen wurden mit dem Standardrahmen IBM 9343 mit eingebauter Steuereinheit mit 4-pfadigem Funktionsumfang und mit 9345 Einschüben abgedeckt. Die Kapazität konnte stufenweise auf 48 GB ausgebaut werden. Die IBM 9341/9343 erlaubten das unterbrechungsfreie Hinzufügen von Platteneinschüben. Hohe Leistung ohne Cache wurde im Speziellen bei Datenbankanwendungen mit hohem Randomzugriff erzielt. Erhebliche Steigerungen bei entsprechenden Anwendungen wurden später durch einen modellabhängigen 32-MB- bis 64-MB-Cache möglich.

1989/1991:

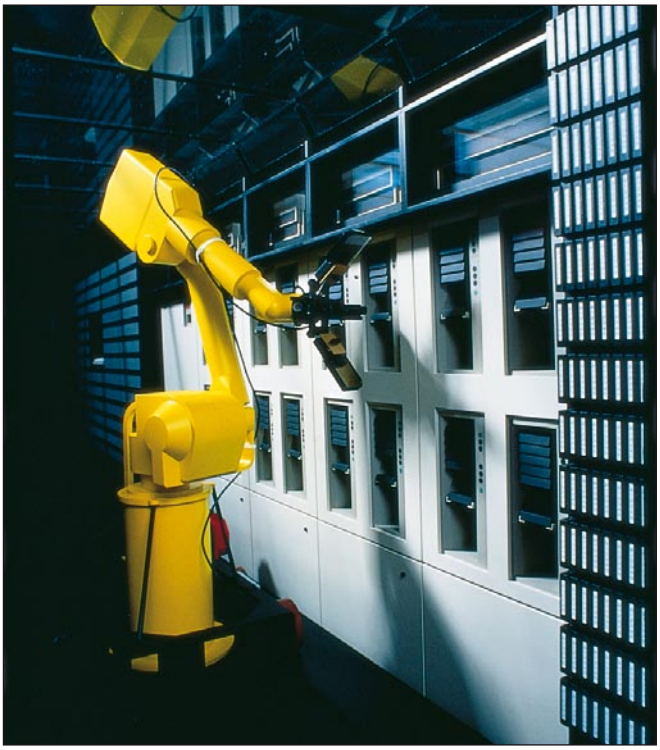
Parallel zur 3390 Plattenentwicklungsreihe wurde **1989** auf der Tape-Seite der Nachfolger des 3480 Bandsystems, die **IBM 3490,** verfügbar und zwei Jahre später die erweiterte Version in Form der **IBM 3490E**. Bei der 3490 Laufwerktechnologie wurde mit 18 Spuren, bei der 3490E Technologie mit 36 Spuren gearbeitet.

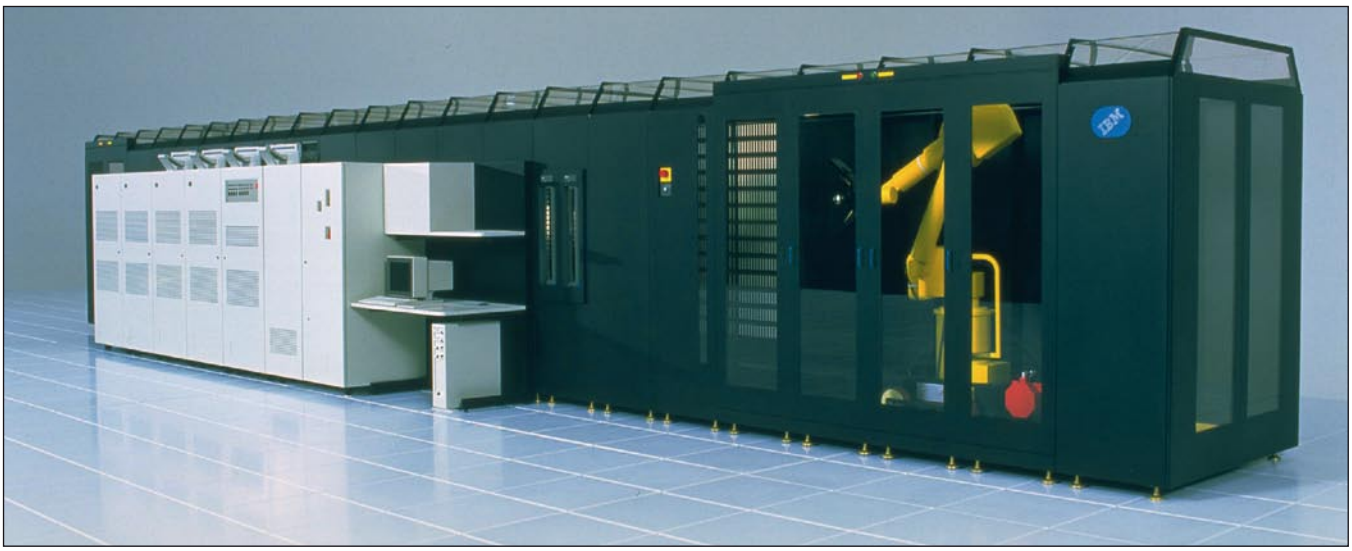
Die 3490 Technologie verwendete zur Steuerung des Bandsubsystems erstmals sehr leistungsstarke Kontrolleinheiten mit einem dynamischen Pufferspeicher von 2 bzw. 8 MB. Ebenso wurde bei der Aufzeichnung mit einem neuen Komprimierungsverfahren gearbeitet, dem IDRC (Improved Data Recording Capability), das Komprimierungsfaktoren von bis zu 3 : 1 zuließ. Die neuen 3490 Systeme konnten mit ESCON-Kanälen angesteuert werden, wie es bei den 3390 Platten der Fall war. Damit war der Grundstein für ESCON gelegt, das im Laufe der nächsten Jahre die bis dahin verwendete BMPX(Block Multiplex)-Verkabelung ablösen sollte.

Die erreichte Aufzeichnungsdichte lag bei IBM 3490 bei 38000 BPI mit Kassettenkapazitäten von 600 bzw. 1200 MB, bei 3490E wurden durch die doppelte Spurzahl 78000 BPI erreicht mit Kassettenkapazitäten von 1200 und 2400 MB.

IBM 3490 Magnetbandsystem

Die Epoche der fest eingebauten Platten mit externen Kontrolleinheiten





IBM 3495 Bandarchiv mit 5660 bis 18920 Kassetten (13.5 bis 45.5 TB), Höhe 2.38 Meter, Tiefe 2.46 Meter, Länge modellabhängig 13.4, 18.3, 23.2 oder 28.0 Meter, 4 bis maximal 64 Laufwerke IBM 3490

1992: IBM kündigte nach einigen Jahren der Kooperation im Bandarchivbereich mit den Firmen Grau und Haushahn ein eigenes automatisches **Magnetbandarchiv IBM 3495** für große Prozessoren an, die unter dem Betriebssystem MVS unterstützt waren. Bei einer Kassettenkapazität IBM 3490 von 2.4 GB konnten die Benutzer die Library-Kapazitäten von 13.5 TB auf 45.5 TB ausbauen.

Zum Vergleich: Der Maßenspeicher von 1974 erlaubte maximal etwa 0.24 TB.

Das damals verwendete Robotersystem der 3495 war tonnenschwer und wurde vom Markt bis auf wenige Kunden nicht sonderlich akzeptiert. Oft wurde an die IBM die Frage gerichtet, warum ein tonnenschwerer Roboter notwendig sei, um eine kleine 3490 Kassette zu transportieren.

1993: Etwa eineinhalb Jahre später kam dann das automatische **Magnetbandarchiv IBM 3494** in wesentlich kompakterer Bauweise und mit einem adäquaten Robotersystem, das neben der MVS-Plattform auch AS/400- und RISC-6000-Systeme unterstützte.

Das 3494 Librarysystem wurde in den Folgejahren funktional erweitert und ist bis heute noch ein aktives Librarysystem, das vor allem im Mainframe-Umfeld eingesetzt wird. Es war auch die erste IBM Tape-Library, die den Mischbetrieb von 3490 und 3490E Laufwerken und Kassetten ermöglichte und auch den Nachfolger der 3490, die IBM 3590 Laufwerktechnologie, unterstützte. Mit 3590 Laufwerken, die als Nachfolger der 3490E im Jahre 1995 verfügbar wurden, konnten mit dem ersten Modell B Kanaldatenraten von 9 MB/s realisiert werden. Die 3494 war zudem eine Library, die gleichzeitig den Mainframe und Open Systems bedienen konnte und kann. Später wurden die Modelle E und H der 3590 unterstützt.

Der Library Manager der 3494 wurde kontinuierlich weiterentwickelt und den Virtual-Tape-Server-Systemen VTS mit ihren weiterentwickelten Funktionalitäten, die mit der Library bis heute betrieben werden können, angepasst. Mit der Verfügbarkeit der ersten Jaguar-3592-Laufwerke im September 2003 (siehe Epoche Multiplattformsysteme und SAN), entwickelte IBM eine intelligente Integration dieser kleinen Laufwerke in die 3494 Library. In der exakten Größe eines Magstar-3590-Laufwerks wurden Gehäuseeinheiten, sogenannte ‘Cradles‘ gebaut, wo zwei Jaguar-Laufwerke im vorderen Teil untergebracht werden können. Nach hinten sind die Cradles offen. Dort sind zwei Netzteile untergebracht. Die beiden Jaguar-Laufwerke sind an beide Netzteile angeschlossen, um die Stromversorgung über zwei unabhängige Quellen sicherzustellen. Die IBM 3494 ist heute noch eine aktive Library. IBM gab im Jahr 2005 ein ‘SOD‘ (Statement of Direction) heraus, dass alle zukünftigen Laufwerke der Jaguar-Technologie in der 3494 Library unterstützt werden.

1993 IBM 3494

IBM 3494 Gehäuseeinheit D22 mit ‘Cradles‘ und TS1120(Jaguar)-Laufwerken

1975 bis 1993

Die Epoche der fest eingebauten Platten mit externen Kontrolleinheiten

Kommentar zur Epoche der fest eingebauten Platten mit externen Kontrolleinheiten

Die Anforderungen bezüglich Hochverfügbarkeit wurden immer höher. Deshalb ging man von den Wechselplatten auf fest eingebaute Platten über. Fest eingebaute Platten waren um ein Vielfaches zuverlässiger. Neben der Verfügbarkeit wurden auch die Ansprüche an Investitionsschutz immer höher. Um dem Rechnung zu tragen, etablierte die IBM über die gesamte Epoche das Prinzip des Wechsels für Steuereinheiten und Plattenspeicher. Dies sah so aus, dass neue Platteneinheiten an die vorhandenen Steuereinheiten angeschlossen werden konnten, dann kam eine neue, leistungsstärkere Steuereinheit, an der die alten Platten betrieben wurden. Dann wurden wieder neue Platten verfügbar, die an die vorhandene Steuereinheit angeschlossen werden konnten, danach kam wieder eine neue Steuereinheit. Dieses Wechselprinzip wurde vom Markt so positiv angenommen, dass es bis in die Folge-Epoche der RAID-Systeme gepflegt wurde.

Auf der Steuereinheiten-Seite wurden mit der IBM 3880 erstmals Cache und entsprechende Caching-Algorithmen eingeführt, um mit der Leistungsentwicklung der Rechnersysteme Schritt zu halten. Die wichtigste Funktion war hier das ‘DASD Fast Write’. Dies wurde auch mit der neuen 3990 Steuereinheit weitergepflegt und ausgebaut. Eine Funktion nach der anderen kam hinzu. Die funktionalen Erweiterungen setzten sich kontinuierlich bis zum letzten Modell der externen Steuereinheiten, der 3990 Modell 6, bis Mitte der 90er-Jahre fort.

Anfang der 90er-Jahre wurde die ‘Braune’-Platten-Produktion umgestellt und Dünnfilm-Beschichtungen wurden eingeführt. Dies führte zu wesentlich höherer Zuverlässigkeit, weil mit Grain-Strukturen gearbeitet werden konnte und Eisenoxydpartikel (Rost) als Datenträger wegfielen. Die neu entwickelte MR-(Magnetoresistive)Kopftechnik kam zum Einsatz (siehe auch Technologie-Anhang).

Im Tape-Bereich kam 1984 der große technologische Wechsel. Das Rollenband wurde durch eine kleine Kassette mit der 3480 Bandtechnik ersetzt. Dies war vor allem notwendig, um den Bandbereich in der Leistung den Platten und dem Gesamtsystem anzugleichen.

Anfang der 90er-Jahre kamen dann die ersten automatisierten Bandarchive mit den Produkten IBM 3495 und IBM 3494 mit Roboter und Greifersystemen, die die Bandkassetten transportieren konnten.

Insgesamt war die sehr lange anhaltende Epoche der fest eingebauten Platten mit externen Steuereinheiten eine Epoche, die durch konstante Weiterentwicklungen im Hinblick auf Datensicherheit und Investitionsschutz geprägt war und sicherstellte, dass die Speicher mit den neu entwickelten Rechnersystemen Schritt hielten. Oberstes Gebot war neben Datensicherheit die Gesamtleistung des Verbundes von Rechnern und Speichersystemen.

Die Epoche der RAID-Systeme

Plattensysteme mit RAID-Architekturen

IBM bot 1994 ein völlig neues Konzept externer Magnetplattensysteme unter dem erstmals seit 1956 wieder benutzten, eingetragenen Warenzeichen **'RAMAC'** an. Dieses Mal stand die Buchstaben-Reihenfolge RAMAC für '**R**AID **A**rchitecture with **M**ultilevel **A**daptive **C**ache'.

Was war der Hintergrund und was bedeutete dies konkret?

1.) Die Verarbeitungsgeschwindigkeit vergleichbarer Prozessoren war zwischen 1964 und 1994 ungefähr 400-mal schneller gewachsen als die Übertragungsrate der Zugriffsmechanismen der Plattenspeicher und etwa 30-mal schneller als die Übertragungsgeschwindigkeit der gängigen Kanäle. Am schnellsten wuchs die Kapazität pro Zugriffsmechanismus – ungefähr um den Faktor 6000 – und damit doppelt so schnell wie die Verarbeitungsgeschwindigkeit der Prozessoren. Die Ansprüche an die Datenverfügbarkeit in internationalen und firmeninternen Netzen nahmen ständig zu und, weil technisches Versagen von Komponenten nie ganz ausgeschlossen werden konnte, wuchs die Forderung nach fehlertoleranten Systemen.

2.) Parallel zur Entwicklung immer leistungsfähigerer Magnetplattenspeicher für große und mittlere Systeme hatte die Industrie – IBM eingeschlossen – für den schnell wachsenden Markt der Einzelplatzrechner (PC, Laptop, Notebook) in den Abmessungen kleine, aber auf hohem technischem Niveau stehende, leistungsfähige und zuverlässige Plattenlaufwerke auf den Markt gebracht. Damals war abzusehen, dass die Kapazitäten dieser Laufwerke rasant steigen werden. Es kam noch hinzu, dass diese kleinen Plattenlaufwerke aufgrund hoher Maßenproduktion wesentlich kostengünstiger gefertigt werden konnten als die großen Plattenfiles, die zu diesem Zeitpunkt bereits als SLEDs (Single Large Expensive Disks) bezeichnet wurden.

3.) Bereits 1987 schilderte die Universität Berkeley in Kalifornien (die Studie erfolgte im Auftrag der IBM) in einem Dokument mit dem Titel 'Eine Studie über redundante Anforderungen kostengünstiger Plattenspeicher (A Case for Redundant Arrays of Inexpensive Disks), wie man die betreffenden Plattenlaufwerke zu einem aus Sicht der Betriebssysteme einzigen adressierbaren Datenträger zusammenschalten konnte, um sie für größere Systeme zu nutzen und dabei entweder höhere Übertragungsraten oder höhere Datenverfügbarkeit oder beides zu erreichen. In diesem Papier wurden 5 RAID-Stufen (RAID-Levels) definiert, die sich in der Abwägung zwischen Übertragungsleistung und dem Grad der Fehlertoleranz unterschieden.

RAID1 beschreibt die doppelte Speicherung von Daten auf zwei Platten mit völlig identischem Inhalt (Spiegelung, Mirroring). Fällt ein Laufwerk aus, greift das System auf das andere zu.

Bei **RAID2** werden die Daten byteweise auf mehrere Platten kopiert (Mehrfachspiegelung). Auf einer weiteren Platte wird ein Fehlercode gespeichert, mit dessen Hilfe verlorene Daten rekonstruiert werden.

Bei **RAID3** werden die einzelnen Bytes ebenfalls auf mehreren Platten – allerdings abwechselnd – gespeichert und auf einer separaten Platte sogenannte Paritätsbits. Fällt eine Platte aus, lässt sich deren Inhalt über den der intakt gebliebenen Platten und die Paritätsbits wieder herstellen.

RAID4 unterscheidet sich von RAID3 dadurch, dass die Daten statt in einzelne Bytes in ganze Blöcke von mehreren Kilobytes unterteilt werden.

Bei **RAID5** erzeugt das System Paritätsbits für Blöcke von mehreren Kilobytes, die auf alle Platten so verteilt sind, dass sie immer auf einer anderen Platte stehen als die Daten, aus denen sie erzeugt wurden. Das Verfahren bietet hohe Sicherheit bei relativ schnellem Zugriff, weil durch die Verteilung parallele Schreib-Updates möglich sind. Deswegen erfuhr dieser RAID-Level die stärkste Verbreitung.

Der Vollständigkeit halber seien noch später hinzugefügte RAID-Stufen erwähnt:

RAID6 bietet gegenüber RAID5 zusätzliche Paritätsbits (zwei Parity-Schemen) und damit noch mehr Sicherheit (bis zu zwei Platten können innerhalb eines Arrays ausfallen), allerdings auf Kosten der Leistungsfähigkeit eines Arrays.

RAID7 protokolliert dazu noch sämtliche Schreibzugriffe.

RAID0 verteilt die Daten in kleinen Paketen auf mehrere Platten und ermöglicht schnellen Zugriff, mangels Redundanz jedoch keine erhöhte Sicherheit. Schließlich gibt es noch RAID10, auch als Stufe 1 + 0 bezeichnet. Sie arbeitet wie RAID0, jedoch mit zwei gleich großen Sätzen von Platten, wobei die zweite Gruppe das exakte Spiegelbild des Inhalts der ersten Gruppe enthält. Diese Lösung erfordert viele Laufwerke.

Inzwischen haben sich (bis heute, im Jahr 2008) aus diesen Basis-RAID-Stufen gänzlich neue RAID-Levels entwickelt, die jenseits der Standard-Levels liegen, teilweise unsinnig sind, aber in manchen Fällen auch eine Optimierung der Basis-Levels darstellen (wie z. B. RAID 1E, 1E0, 1.5, Matrix RAID, RAID15, 51, 53, 45, RAID5E, 5EE und RAID5DG/RAID ADG).

4.) Bereits 1992 konstituierte sich ein RAID-Beratergremium (RAID Advisory Board) aus 40 Anwender- und Herstellerfirmen, zu denen auch IBM zählte. Damit wurde RAID in den neunziger Jahren zu einer Art Industriestandard. Die Originalbezeichnung RAID wurde in der Bedeutung geringfügig korrigiert: Aus 'kostengünstigen Einheiten' (Redundant Array of Inexpensive Disks) wurden 'unabhängige Einheiten' (Redundant Array of Independent Disks).

Bereits seit 1978 besitzt IBM ein Patent über die Wiederherstellung von Daten, die auf einer ausgefallenen Einheit aus einer Gruppe von Laufwerken gespeichert wurden. Wie die gesamte Industrie bemühte sich IBM, den Herausforderungen, die sich aus den oben dargestellten Punkten ergab, wirksam zu begegnen. Die zunehmenden Veränderungen im Markt mit seiner ständig wachsenden Zahl von Anbietern, immer mehr etablierten Hardware- und Softwareplattformen, dynamischen Innovationen und den damit immer kürzer werdenden Produktzyklen machten es notwendig, die IBM Geschäftspolitik maßgeblich zu verändern und anzupassen. Bisher hatte IBM fast ausschließlich für den eigenen Bedarf zum Vertrieb über die eigene Organisation an Endbenutzer produziert und war nur vereinzelt als Unterlieferant anderer Hersteller aufgetreten (Branchen-Fachausdruck für solche Geschäfte ist OEM, Original Equipment Manufacturer). Ein Beispiel waren in den achtziger und neunziger Jahren geschlossene Verträge über Lieferungen von leicht modifizierten IBM Plattenspeichern an Siemens. IBM ließ immer Fremdprodukte aller Art innerhalb ihrer Konfigurationen und Softwareplattformen zu, ohne in den Umgebungen anderer Hersteller als Anbieter aufzutreten. Auf die bei den DV-Benutzern um sich greifende Vernetzung von Datenverarbeitungsprodukten und DV-Plattformen verschiedener Hersteller – heterogene Client-/Serverumgebungen – reagierte IBM mit Angeboten, bei denen alle für die jeweiligen Produkte relevanten eigenen und fremden Schnittstellen unterstützt wurden. Das Storage-Geschäft wurde durch das Anwerben von entsprechenden Geschäftspartnern neben dem eigentlich blauen Vertrieb zusätzlich vorangetrieben. Man arbeitete mit immer mehr Geschäftspartnern, die sich in ihrer Vertriebsorganisation erfolgreich um den Absatz von IBM Hardeware- und Softwareprodukten bemühten. Dieser Trend hat sich bis heute noch maßgeblich verstärkt.

Wie bereits eingangs dieses Kapitels beschrieben, machte IBM **1994** als erste Firma das auf **RAID5 basierende Magnetplattensystem** verfügbar, der damaligen **RAMAC 1 (IBM 9391)**. RAMAC 1 wurde im Juni 1994 angekündigt und war ab September 1994 verfügbar. Als Basis für eine RAMAC-Speichereinheit diente ein **Einschub IBM 9392** (Drawer), der mit mehreren anderen Einschüben in einen Rahmen integriert wurde. Jeder dieser Einschübe beinhaltete 4 Plattenlaufwerke auf 3.5-Zoll-Basis mit jeweils 2 GB Kapazität (Allicat-Laufwerke), zwei Netzteile, zwei Kühlgebläse, einen RISC-Mikroprozessor als Controller, einen batteriegestützten Pufferspeicher (Non Volatile) und einen Akku. Jeder Einschub stellte ein für sich geschlossenes RAID5-System dar. Bis zu 16 dieser RAID5-Einschübe konnten in den Rahmen integriert werden. RAMAC kam in zwei Ausprägungen. Das **RAMAC Array** beinhaltete nur die RAID5-Einschübe und wurde an die damalige, leistungsstarke Steuereinheit 3990 angeschlossen. Das **RAMAC Subsystem** beinhaltete die übergeordnete Controllerfunktion in demselben Gehäuse, in dem auch die Einschübe integriert waren, und war seitlich im Rahmen eingebaut. In der vollen Ausbaustufe lieferten beide Systeme eine nutzbare Plattenkapazität, RAID5-gesichert, von 90 GB aus.

Die Epoche der RAID-Systeme

RAMAC 2 und **RAMAC 3** folgten wegen der raschen Fortschritte in den kapazitiven Möglichkeiten der 3.5-Zoll-Laufwerke im Abstand von nur einem Jahr. Bei RAMAC 2 kamen die Ultrastar-XP-Laufwerke und bei RAMAC 3 die Ultrastar-2XP-Laufwerke zum Einsatz. Die rechnerische Kapazität verdoppelte sich jeweils per Subsystem.

RAMAC 2 bot mit 16 Einschüben eine nutzbare Kapazität von 180 GB und RAMAC 3 von 360 GB.

Mit RAMAC 3 in der Version des RAMAC Arrays zum Anschluss an 3990 Steuereinheiten wurde zeitgleich ein neues Modell der Steuereinheit eingeführt, der 3990 Modell 6. Diese Steuereinheit war eine Doppelsteuereinheit, die in ein Gehäuse integriert war und an der zwei RAMAC 3 Arrays betrieben werden konnten. Damit lieferten zwei RAMAC 3 Arrays an der 3990-6 als 'Doppelsubsystem' eine Gesamtkapazität von bis zu nutzbaren 720 GB aus.

Die RAID5-RAMAC-Baureihe war auf dem Markt sehr erfolgreich, konnte aber über die Jahre nicht fortgesetzt werden, weil der Kostenfaktor in der Produktion dem Preisverfall der Gigabytes in den angebotenen Subsystemen nicht standhalten konnte. Jeder Einschub als eigenständiges System und bis zu 16 Einschübe unter einer eigenen Steuereinheit, dadurch zwei Cache-Ebenen: war sehr aufwändig zu produzieren.

Dies war einer der Hauptgründe, warum **1996** eine **Kooperation mit der Firma StorageTek** bezüglich des Vertriebs und der Weiterentwicklung des verfügbaren 'Iceberg'-Plattensystems geschlossen wurde und die IBM zwei Plattensubsysteme der Firma STK in den Vertrieb nahm. Der 'Iceberg' wurde IBM seitig unter dem Namen **IBM RVA (RAMAC Virtual Array)** vermarktet, das Großkapazitätssystem von STK als **IBM RSA (RAMAC Scalable Array)**. Das RAMAC Virtual Array wurde im Juli 1996 mit sofortiger Verfügbarkeit angekündigt. Das leistungsstärkere RVA in der Turbo Version wurde knapp ein Jahr später, im April 1997, verfügbar.

Das RAMAC Virtual Array **IBM RVA** IBM 9393 erfüllte die Bedingungen von **RAID6** und bot eine sehr intelligente Lösung für das Anlegen und das Wachstum von Dateien. Normalerweise mussten Benutzer bisher für wachsende Dateien genügend Platz reservieren. Das bedeutete, einen ständigen Vorrat von leeren Spuren (allocated tracks) bzw. nicht genutzte Kapazität vorzuhalten. Wenn das System beim Ausführen eines Programms für eine wachsende Datei keine Kapazität mehr vorfand, führte das zum anomalen Abbruch der Operation. Beim RVA suchte das System automatisch freien Platz. Reservierung (Allocation) war nicht mehr erforderlich und Abbrüche wegen fehlenden Speicherplatzes ausgeschlossen. Dieser spezielle Algorithmus wurde als logisch strukturierter File-Algorithmus (Log Structured File) bezeichnet.

links: IBM RAMAC Subsystem mit integrierter Steuereinheit
rechts: IBM RAMAC Array zum Anschluss an 3990 Steuereinheiten

1996 –1998 IBM RAMAC Virtual Array RVA, IBM RAMAC Virtual Array 1996, IBM RAMAC Virtual Array Turbo 1997, IBM RAMAC Virtual Array Turbo 2 1998, Kapazitäten von 840 GB bis 1680 GB Arrays 13+2+1 RAID6 oder 2 x (5+2+1), Array Kapazität bis 420 GB, Minimalkonfiguration: 2 Arrays, Maximalkonfiguration: 4 Arrays

Waren Spuren auf den Platten wegzuschreiben, suchte sich die Kontrolleinheit den Platz auf beliebigen Platten in den verfügbaren Arrays. Die veralteten Spuren wurden in einem Space-Reclamation-Prozess wieder einfach zum Überschreiben freigegeben. Das Ganze wurde über Pointertabellen verwaltet und die Pointer einer aktuellen geänderten Datei wurden einfach entsprechend angepasst. Diese neue Log-Structured-File-Architektur als wirklich erste virtuelle Plattenarchitektur auf dem Markt erlaubte auch das sekundenschnelle Kopieren von einzelnen Dateien oder Volumes, ohne dass für die Kopie zusätzlicher Platz vorgehalten werden musste. Nur die Veränderungen in Form von neuen Spuren oder Spur-Updates benötigten wieder neuen Platz. Dabei legte das System entsprechende Kopier-Pointers an, die auf einzelner Spurbasis mit den originalen Pointers übereinstimmten. Diese Funktion wurde als 'SnapShot' bezeichnet. Durch dieses Prinzip konnte die RVA von der Plattenbelegung im Vergleich zu herkömmlichen Plattensystemen wesentlich höher genutzt werden. Beim RAMAC Virtual Array wurden im Backend IBM SSA-Platten eingesetzt und das System konnte mit 4.5-GB-, 9-GB- und später mit 18-GB-SSA-Platten konfiguriert werden.

Das zweite Produkt aus der StorageTek-Reihe, die **IBM RSA** (RAMAC Skalierbares Array) entsprach den herkömmlichen Subsystemarchitekturen, konnte aber bereits **1996** in relativ kleinen skalierbaren Schritten auf 1.4 TB ausgebaut werden. Das System wurde aber relativ kurzfristig kapazitiv von der RVA in der Turbo- und Turbo-2-Version überholt und zeigte nur eine sehr geringe Marktdurchdringung.

IBM 7133 Plattensubsystem, SSA-Architektur mit SCSI-Host-Anschlüssen, Modell T40 'Deskside' Entry, Modell D40 'Rackmountable Disk System'

Plattenart	*Kapazitäten von/bis*
18.2 GB	*72.8 – 291.2 GB*
36.4 GB	*145.6 – 582.4 GB*
72.8 GB	*291.2 GB – 1.16 TB*
146.8 GB	*587.2 GB – 2.34 TB*

Plattensysteme für Open Systems

Während sich IBM in den 90er-Jahren bei der Magnetplatten-Subsystementwicklung im Mainframebereich vornehmlich auf die RAMAC-Reihe konzentrierte, liefen parallel neue Architekturentwicklungen für Plattensubsysteme im Open-Systems-Bereich. Bis Ende der 90er-Jahre gab es kein Plattensystem, das Server aus dem Open-Systems-Bereich (AIX und andere UNIX Derivate und die Intel-Plattformen) und den Mainframe mit seinem MVS-Betriebssystem gleichzeitig bedienen konnte. Auf der Open-Systems-Seite ging IBM andere Wege. Beginnend in den 90er-Jahren wurde im IBM Labor in Hursley in England eine neue, Open-Systems-gerechte Plattenarchitektur entwickelt, die unter dem Namen **SSA** für viele Jahre zur Anwendung kommen sollte. SSA stand für **Serial Storage Architecture,** eine Architektur, die die Limitierungen der bisher an Open-Systems-Server angeschlossene SCSI(Small Computer Systems Interface)-Plattensysteme eliminierte. Bei SSA wurde für die Anbindung von Plattenlaufwerken mit einer Loop-Technik gearbeitet, die es zuließ, dass man in die Loop über neu entwickelte Adapter gleichzeitig zwei Schreib- und zwei Lese-I/Os absetzen konnte und alle Laufwerke in der Loop gleichberechtigt bedient wurden. Die bekannte Arbitrierung, wie es bei SCSI-basierenden Laufwerkssträngen der Fall war, fiel weg. Bei einem SCSI-Laufwerkstrang musste immer ein I/O fertig abgehandelt werden, bevor der nächste starten konnte. Zudem wurden Laufwerke, die ganz am Ende des SCSI-Strangs angeschlossen waren, in der Bedienung benachteiligt. Ein SCSI-Strang hatte die Limitierung, dass maximal 15 Laufwerke angeschlossen werden konnten. Diese Limitierungen überwand die neue SSA-Architektur, die **1992** erstmals in Subsystemen zum Einsatz kam.

Die Epoche der RAID-Systeme

Das bekannteste System war das IBM **7133 Serial Storage Architecture Disk Subsystem**, das an RISC Systeme/6000, Power Servers und Power Stations angeschlossen werden konnte und immer mit den neusten 3 ½-Zoll-Laufwerken ausgestattet war. Das IBM 7133 Subsystem wurde im Juli 1995 mit sofortiger Verfügbarkeit angekündigt. Es war eines der erfolgreichsten Plattensysteme im AIX-Umfeld und wurde direkt über den AIX-Verkaufskanal vertrieben. Während des Lebenszyklus der 7133 wurden über 42000 Plattensysteme weltweit in Produktion genommen, ein einzigartiger Verkaufserfolg.

Die letzten Systeme in SSA-Architektur (IBM 7133 Modelle T40 und D40) blieben bis **2003** im aktiven IBM Vertrieb und konnten damals über Gateways an bestehende FibreChannel-Infrastrukturen (SANs, Storage Area Networks) angeschlossen werden.

1999 kündigte die IBM für den Mainframe wieder ein eigenes, selbst entwickeltes Plattensubsystem unter dem Namen **IBM ESS (Enterprise Storage Server)** an, das allerdings mehr unter dem Entwicklungsnamen **'Shark'** bekannt wurde und letztendlich die RVA und RSA im IBM Vertrieb ablöste. StorageTek vertrieb danach wieder selbst die RVA unter dem Namen STK Shared Virtual Array. Die ESS war nicht nur ein Plattensystem für den Mainframe, sondern konnte auch an alle gängigen Open-Systems-Plattformen angeschlossen werden. Damit war die Voraussetzung geschaffen, mit einem Plattensubsystem sowohl den Mainframe als auch Open-Systems-Server zu bedienen, wie es auf der Bandarchivseite bereits seit Jahren möglich war. Über die Folgejahre löste die ESS dann auch im Open-Systems-Bereich teilweise das IBM 7133 Serial Disk System ab. Parallel dazu wurden FibreChannel-basierende Plattensysteme für die Open-Systems-Plattformen eingeführt.

Bandsysteme

Neben der Plattensubsystementwicklung in den 90er-Jahren führte IBM eine **neue Bandlaufwerktechnologie** als Ablösung der bis dahin verwendeten 3490 Technologie ein, die **IBM 3590**, die unter dem Namen **Magstar** bekannt wurde. Das erste Laufwerk der Magstar 3590 war das Modell B, das auf einer ½-Zoll-Kassette eine native Kapazität von 10 GB aufzeichnete. Das 3590-B Laufwerk wurde im April 1995 angekündigt und war im September des gleichen Jahres verfügbar. Bei dieser neuen Aufzeichnung wurden zum ersten

Blick in die IBM 3590 Magstar Laufwerke

Mal in der Bandtechnologie magnetoresistive Schreib-/Leseköpfe mit dem 'Track Following Servo'-Prinzip eingesetzt, wo die Leseelemente konstant kontrollieren, wie gut die Bits vom Schreibkopf geschrieben wurden. Die Kassetten wurden mit 128 Spuren vollgeschrieben und die Datenrate des Modells B der 3590 betrug damals faszinierende 9 MB pro Sekunde (native).

Die Laufwerke waren als Stand-alone-Version, also fest eingebaut in einem Gehäuse mit entsprechenden Autoloadern, zu bekommen, sie wurden aber genauso in der aktuellen Tape Library IBM 3494 eingesetzt. Voll ausgebaut konnte die IBM 3494 mit 3590-B-Laufwerken eine Kapazität von 180 TB (native) abbilden. Die Laufwerke selbst hatten zwei SCSI-Schnittstellen und konnten entweder direkt an Open-Systems-Server angeschlossen werden oder über eine Kontrolleinheit an ESCON-basierende Mainframe-Server.

Die Magstar-Entwicklungsreihe wurde dann **1999** mit dem neuen **E-Modell** erweitert, das mit 256 Spuren die doppelte Spurzahl aufwies und eine Kapazität von 20 GB auf den Kas-

IBM 3590 Magstar Tape Subsystem
1995: IBM 3590-B mit 9 MB/S Datenrate
1999: IBM 3590-E mit 14 MB/S Datenrate
2002: IBM 3590-H mit 14 MB/S Datenrate
Kassettenkapazitäten (native):

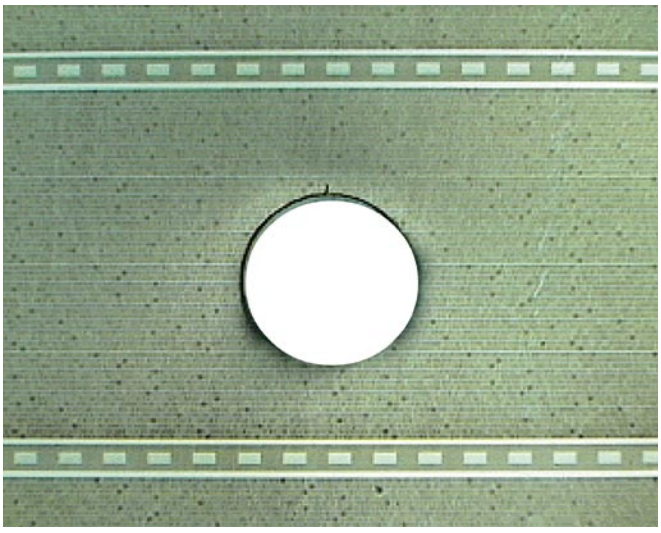
Modell	*normale Kassette*	*lange Kassette*
3590-B	*10 GB*	*20 GB*
3590-E	*20 GB*	*40 GB*
3590-H	*30 GB*	*60 GB*

setten speicherte (native ohne Komprimierung). Die Datenrate wurde von 9 MB/s auf 14 MB/s gesteigert. Parallel dazu brachte IBM eine neue Kassette, die **3590 'Extended Length Cartridge'**, die mit der doppelten Bandlänge (beschrieben mit 3590E) auch die doppelte Kapazität, nämlich 40 GB, abbildete. Damit erhöhten sich die 3494 Library-Kapazitäten auf 374 TB bei Verwendung der normalen Kassette und bis 748 TB bei der langen Kassette. Die Magstar Entwicklungsreihe war ab Mai 1999 verfügbar.

Das Modell E der Magstar-Baureihe besaß einen leistungsfähigen Controller, der neben der Aufzeichnung von Datenblöcken auch **Parity-Blöcke** aufzeichnete, um ein wesentlich höheres Error Recovery zu ermöglichen. (Bei der Aufzeichnung von 16 Spuren gleichzeitig wurde nach 7 Datenblöcken in der achten und nach weiteren 7 Datenblöcken in der 16. Spur ein Parity-Block mitgeschrieben und über das logische Spurset von acht Spuren 'gestriped' – RAID5 auf Band.)

Was Magstar in Sachen Recovery zu leisten vermag, zeigt ein Versuch mit dem Magstar E-Modell, der im Juni 1999 und im Juli 2000 im IBM Labor in Tucson, Arizona, durchgeführt wurde. Zu dem Versuch wurden alle maßgeblichen Consultants wie Gartner Group und IDC eingeladen. Man hat mit dem E-Modell eine Kassette mit 256 Spuren voll beschrieben und ein sauberes Read Verify erhalten. Dann wurde die Kassette aus dem Laufwerk entladen. In die Mitte des Bandes wurde dann ein 1.25 mm großes Loch gestanzt. Dieses Loch spiegelte ca.100 000 Bits wieder, die dann fehlten. Die Kassette mit dem Loch wurde wieder ins Laufwerk geladen und die Leseoperation wurde erfolgreich ohne Fehler durchgeführt. Alle fehlende Datenbits konnten durch den ECC (Error Correction Code) über die Bitblock-Verteilung und die Parity-Bit-Blöcke 'recovered' werden. Das war zu dieser Zeit mit Sicherheit mit keiner anderen Bandlaufwerkstechnologie durchführbar!

Das Laufwerk hatte zwei SCSI-Schnittstellen. Ab Mai 2000 waren die Magstar-Laufwerke auch mit zwei FibreChannel-Schnittstellen verfügbar. Damit konnten die Laufwerke direkt an ein FibreChannel-basierendes SAN-Netzwerk angeschlossen werden. Etwas zeitversetzt kamen auch die entsprechenden Controller auf den Markt, welche den Anschluss der Laufwerke auch an FICON-basierende Mainframe-Rechner ermöglichten.



Lochstanzversuch mit Magstar 3590 Bändern 1999 und 2000 im IBM Labor in Tucson

1994 bis 1998

Die Epoche der RAID-Systeme

Das letzte Modell der 3590 Magstar-Reihe kam im Jahr 2002 mit dem 3590 Modell H, das auf denselben Kassetten 384 Spuren aufzeichnete und so die Kassettenkapazitäten um 50 % steigerte. Das Modell H wurde am 25. Juni 2002 angekündigt und war bereits am 28. Juni 2002 verfügbar. Die Datenrate der 3590-H blieb auf dem Stand des Vorgängers 3590-E, bei 14 MB pro Sekunde. Die 3494 Library-Kassettenkapazitäten erreichten jetzt 1 Petabyte native Kapazität in der vollen Ausbaustufe (ohne Kompression).

Im September **1996** wurde von IBM ein **neues Kassettenbandsystem** vorgestellt, das neben hoher Streaminggeschwindigkeit extrem schnelle Zugriffszeiten realisierte. Die **IBM Magstar MP 3570** bot eine Datentransferrate von 15 MB pro Sekunde und Zugriffszeiten von nur wenigen Sekunden. Dies wurde mit einem besonderen Kassettendesign ermöglicht, wo die Grundstellung des Bandes in einem Zwei-Spulen-Design in der Mittenstellung lag. So konnte beim Laden des Bandes sowohl nach links als auch nach rechts gespult werden, um schnell an eine File zu gelangen. Diese neue Konzeption allerdings ermöglichte nur kleine Kassettenkapazitäten von 5 MB und später 10 MB, die vom Markt nicht in der breiten Akzeptanz angenommen wurde, wie man es ursprünglich gedacht hatte. Deshalb war der Lebenszyklus der Magstar MP auch nur von kurzer Dauer und auf etwa 3 Jahre beschränkt.

Mit der Verfügbarkeit der ersten Magstar Modelle **3590-B 1995** ergab sich im Mainframe-Umfeld eine **neue Problemstellung**, denn die jetzt hohen Kassettenkapazitäten von 10 GB native und die hohen Datenraten von 9 MB/s native wurden nur zu einem kleinen Teil genutzt. Eine weltweite Studie ergab, dass der Nutzgrad im Mainframe-Umfeld bei etwa 200 – 300 MB pro Kassette lag und die Datenraten bei durchschnittlich 2 – 3 MB/s lagen und bei vielen Backup-Jobs sogar unter 1 MB/s. Um dieses Problem zu beheben, fing IBM an, **Tape-Virtualisierungslösungen** für den Mainframe-Bereich zu entwickeln, um die dahinter angeschlossene physikalische Ressource sowohl in der Kapazität der Kassetten als auch in der Datenrate von Laufwerken maximal nutzen zu können. Aus einer anfänglichen Software-basierenden Volume-Stacking-Lösung entstand der **3494 Virtual Tape Server VTS**, der im Juni **1997** mit dem ersten **Modell B16** auf den Markt kam. Dabei wurden die Tape Volumes, die es zu sichern galt, auf einem Plattenpuffer zwischengespeichert und dabei indexiert. Sobald so viele Volumes in den Plattenpuffer geschrieben waren, dass sie eine 3590 Kassette füllen konnten, wurde eine Kopie der gesamten Ansammlung von Platten auf die schnellen Laufwerke geschrieben (Premigration) und dabei die Kassetten maximal aufgefüllt (Stacked Volume Prinzip). Das Original verblieb im Plattenpuffer. Dadurch konnte der Nutzgrad der Kassettenkapazitäten und der Laufwerkgeschwindigkeiten maximiert werden. Recalls von Volumes, die noch im Plattenpuffer standen, wurden in Plattengeschwindigkeit durchgeführt und beim Backup konnten viele Jobs parallel angestoßen werden, da viele virtuelle Laufwerke zur Verfügung standen. IBM implementierte eine sogenannte ‘Outboard’-Lösung, die völlig transparent zum Host und dem MVS-Betriebssystem war und dieselbe Transparenz gegenüber dem Bandverwaltungssystem und dem Katalogeintrag zeigte. Emuliert wurden 3490E Laufwerke mit Kassettenkapazitäten von 400 bzw. 800 MB und der Host registrierte gar nicht, dass er keine physischen Laufwerke im Zugriff hatte.

IBM Magstar MP 3570 Bandsubsystem

Der erste Virtual Tape Server **IBM 3494 B16** war damals direkt als Gehäuseeinheit in das 3494 Bandarchiv integriert. Das B16-Gehäuse beherbergte den RS/6000 Server, einen RAID-5-abgesicherten Plattencache mit SSA-Platten, der bis auf 72 GB ausgebaut werden konnte, sowie die für den Host notwendigen ESCON-Anschlüsse und 400 Kassettenstellplätze. Später arbeitete man mit externen Einheiten, wie die Weiterentwicklung dieses Konzepts zeigt. Virtuelle Tape-Lösungen sind in unserer heutigen Zeit nicht mehr wegzudenken, weil es die einzige Möglichkeit ist, über diesen Weg Tape- und Tape-Library-Ressourcen maximal zu nutzen.

Das Prinzip, den VTS (Virtual Tape Server) direkt in das 3494 Bandarchiv zu integrieren, erwies sich als äußerst ungünstig. Zudem war die Durchsatzstärke der damaligen B16 sehr klein und erreichte nur 6 – 8 MB/s, weil ohne Kompression gearbeitet wurde. Deshalb kündigte die IBM bereits ein Jahr später, im Juni **1998**, mit einer Verfügbarkeit im August 1998, ein neues **externes Modell B18** an, das als separate Einheit neben der 3494 Library betrieben werden konnte. Diese neue Einheit bot Kompression an, die direkt in den Host-Adaptern, damals auf ESCON-Basis, durchgeführt wurde. Das hatte den Vorteil, dass alle Daten im Plattenpuffer bereits komprimiert waren und der Nutzgrad im Plattenpuffer durchschnittlich um Faktor 3 vergrößert wurde. Durch den Einsatz von SSA-Laufwerken mit 9 und 18 GB Kapazität wurde die Plattenpuffergröße auf 288 GB und 576 GB gesteigert. Später wurden mit 36-GB-SSA-Files Plattenpuffer von bis zu 5.2 TB (nutzbar) möglich. Dadurch wurde die Verweilzeit der Volumes im Plattenpuffer um ein Vielfaches vergrößert und Recalls direkt vom Plattenpuffer als virtuelle Mounts in Plattengeschwindigkeit durchgeführt. Die damaligen B18-Einheiten erreichten, je nach Ausbaustufe, Durchsatzmöglichkeiten von 30 bis 40 MB/s.

Viele andere Firmen sprangen sehr schnell auf diesen Geschäftszug auf und es wurden eine ganze Reihe von Bandvirtualisierungslösungen, HW-basierend, aber auch reine SW-Stacking-Lösungen auf dem Markt verfügbar. Viele verschwanden aber wieder sehr schnell vom Markt.

Das Modell B18 wurde in den Folgejahren funktional erweitert. Import-/Export-Möglichkeiten von logischen Volumes waren möglich. Mit der Einführung von Folgemodellen im August **2001**, der Modelle **B10 und B20**, wurden neben den ESCON-Anschlüssen auch FICON-Anschlüsse verfügbar und eine große Palette von neuen Funktionen kam zum Einsatz, die auch noch für das alte Modell B18 in eingeschränkter Form zur Verfügung stand.

IBM Virtual Tape Server VTS, 1997: Modell B16, 1998: Modell B18, 2001: Modell B10 und B20

Diese neuen Erweiterungen, als **'Advanced Function'** bezeichnet, erlaubten über ein **APM (Advanced Policy Management)** neue Funktionalitäten am IBM VTS. Sie sind bis heute aktiv im Einsatz und bilden die **Verzahnung mit SMS (System Managed Storage)** als integralem Bestandteil des z/OS-Betriebssystems.

Die Funktionalität **'Logical Volume Pool Assignment'** bietet die Möglichkeit, logische Volumes einzeln pro Nummernkreis in unterschiedlichen physischen Kassetten-Pools zu verwalten. Damit wurde der IBM VTS mandantenfähig. Mit **'Selective Dual Copy'** können zwei Kopien eines logischen Volumes in zwei unterschiedlichen physischen Kassettenpools erzeugt werden. Die Funktion **'Peer to Peer Copy Control'** adressiert einen PtP-VTS-Komplex und erlaubt die Steuerung, welche logischen Volumes in einem PtP-Komplex im Immediate Copy Mode und welche logischen Volumes im Deferred Copy Mode gespiegelt werden.

Mit **'Tape Volume Cache Management'** (diese Funktion war bereits vorher verfügbar, allerdings Host-gesteuert) können logische Volumes einer Cache Management Preference Group zugeordnet werden. Es stehen zwei Preference Groups zur Verfügung. Die Preference Group steuert, wie die logischen Volumes im Plattenpuffer behandelt werden.

Die wohl wichtigste Einführung im August 2000 war aber das **Prinzip der Spiegelung von ganzen VTS-Systemen**, die auch heute noch als Peer to Peer VTS bezeichnet wird und sich im Markt, vor allem bei Banken und Versicherungskunden, schnell durchsetzte. Dabei bietet ein Peer-to-Peer-VTS-Komplex **synchrone Spiegelmöglichkeit** (bis heute die einzige synchrone Spiegelung auf dem Markt) und eine asynchrone Spiegelung an. Über eine weitere Funktion, die **2004** eingeführt wurde (**Selective Peer to Peer Copy**), können auch einzelne Volumes oder ganze Storage-Gruppen sowohl gespiegelt als auch ungespiegelt in einem VTS-Peer-to-Peer-Spiegelkomplex verwaltet werden.

Die VTS-Geschichte, die 1997 begann, war für die IBM vor allem im Mainframe-Umfeld eine einzigartige Erfolgsgeschichte und IBM entschloss sich, diese Erfolgsgeschichte mit neuen Systemen und neuen Funktionalitäten über die folgenden Jahre fortzusetzen.

Linear Tape Open LTO Ultrium

In der **ersten Hälfte der neunziger Jahre** hatte sich im Open-Systems-Umfeld vor allem im Mittelstandsbereich und bei kleineren Betrieben eine **Tape-Technologie** durchgesetzt, die diesen Markt beherrschte: **DLT** (Digital Linear Tape), eine günstige Alternative zur High End-½ Zoll-Technologie, die für kleine Unternehmen bezahlbar war und bis Mitte 1995 eine Marktdurchdringung von 90 % hatte. DLT-Hersteller und -Anbieter war Quantum. Der kleinere Teil des Marktes ging an Helical-Scan-basierende Technologien und die QIC(Quarter Inch Cartridge)- und Mini-QIC-Technologien.

Mit dem **Zusammenschluss der Firmen IBM, HP und Seagate** in einem neuen Tape-Entwicklungsgremium, dem sogenannten **LTO(Linear Tape Open)**-Gremium, wurde eine Initiative gestartet, dem dominierenden umd marktbeherrschenden DLT von Quantum eine Alternative entgegenzusetzen. Das LTO-Entwicklungsgremium fing 1996 mit seiner Arbeit an.

Die Spezifikationen von LTO wurden von **1997 bis 1999** entwickelt, um als **erster Bandtechnologie-Standard** auf den Markt zu kommen und den damals herrschenden Wildwuchs an unterschiedlichen Technologien und Medien zu stoppen. LTO ist sowohl von der Technologie als auch vom Medium ein offizieller ISO-Standard, der es erlaubt, LTO-Medien, unabhängig von welchem Hersteller, auf jedem LTO-Laufwerk, auch unabhängig von welchem Hersteller, zu verarbeiten. Heute haben über 30 Firmen die Fertigungslizensen von LTO erworben. Hinzu kommen zusätzlich die OEM-Abkommen, die die Entwicklungsfirmen IBM, HP und Seagate eingegangen waren. LTO hatte also eine große Chance, sich langfristig gegenüber DLT auf dem Markt durchzusetzen.

Wenn man heute, 2008, die Marktanteile betrachtet, ging diese Rechnung auf. DLT ist heute nahezu vom Markt verschwunden.

Als die LTO Group **1997** anfing, die Spezifikationen für die Standardisierung zu entwickeln, hatte sie zuerst alle vorhandenen Bandtechnologien auf ihre Plus- und Minus- Punkte bewertet, um das Gute in LTO miteinzuarbeiten und das Negative auszulassen. Schluss mit dem Wildwuchs, Kompatibilität war gefragt! Ob Sony AIT oder 4 mm, 8 mm Helical Scan Technologie von Exabyte, Quantums DLT bis hin zum Super DLT, Philipps NCTP, Quarter Inch Cartridge oder Mini Quarter Inch Cartridge: alles Technologien, die keine Kompatibilität vorwiesen und daher Schwierigkeiten hatten, sich weiter dauerhaft auf dem Markt zu behaupten!

Bei der Aufzeichnungstechnik entschied man sich aufgrund der hohen Zuverlässigkeit für die IBM Magstar-Lineare-Serpentinen-Aufzeichnung, mit der das Band bei einer Datenrate von 15 MB/s mit 384 Spuren (LTO1) und 35 MB/s mit 512 Spuren (LTO2) beschrieben wird. Dabei wurden immer 8 Spuren gleichzeitig geschrieben. Die Kassetten erreichten bei diesen Spurdichten native Kapazitäten von 100 GB (LTO1) und 200 GB (LTO2). Bei dieser hohen Spudichte war es notwendig, eine möglichst exakte Führung der Schreib-/Lese-Elemente sicherzustellen.

Zu diesem Zweck wurde ein neues, **zeitbasierendes Spurnachführungssystem** verwendet, das mit fünf vorgeschriebenen Servobändern mit jeweils acht Servobereichen arbeitete. Von der AIT(Advanced Intelligent Tape)-Technik von Sony wurde das in der Kassette integrierte Memory Chip übernommen, allerdings mit größerer Speicherkapazität (4 KB).

LTO1 Laufwerk mit LTO1 Ultrium Kassette

Kassettenladevorgang beim IBM LTO1 Laufwerk

IBM Autoloader und Library-Angebot mit LTO-Laufwerken, Ende 2000, links IBM 3584 Library, rechts unten IBM 3583 Library, darüber der Autolaoder IBM 3581 und das LTO Stand Alone Laufwerk IBM 3580

Dort wird das festgehalten, was z. B. bei Magstar in der Tape-Volume-Control-Region auf Band gespeichert wird: Herstellerangaben als Read-Only-Information, das Inhaltsverzeichnis des Bandes und Angaben wie viele Schreib-/Lesefehler auf dem Band vorkommen, um aus Sicherheitsgründen bei einem bestimmten Schwellwert die Kassette auf 'Read Only'-Status zu setzen. Ebenso war Platz für anwenderspezifische Informationen vorgesehen. Neben der Speicherung auf dem Chip wurden aus Sicherheitsgründen alle Informationen auch in der Tape Volume Control Region des Bandes abgespeichert.

Passend zur LTO-Laufwerkentwicklung entwickelte IBM **eine neues Library-Konzept**, das mit einem Greifersystem arbeitet (bis heute), das die Features der standardisierten LTO-Kassetten optimal handhaben konnte. Das Robotersystem mit diesem Spezialgreifer wurde **Ende 2000** mit der **neuen IBM 3584 Library** realisiert. Die IBM 3584, zum damaligen Zeitpunkt ausschließlich an Open Systems betreibbar, stellte das passende Gegenstück zur IBM 3494 Library dar, allerdings mit wesentlich moderneren Möglichkeiten und einer Robotergeschwindigkeit, die bis heute unerreicht ist.

1994 bis 1998

Die Epoche der RAID-Systeme

Mit der Einführung der LTO-Bandtechnologie auf dem Markt änderte die IBM ihr bisher verfolgtes Konzept, alles selbst herzustellen. In dem damals bereits hart umkämpften Open-Systems-Markt als Einzelhersteller zu bestehen, der alle Arten von Lösungsformen abildet und fertigungstechnisch baut, war unmöglich. Deshalb schloss IBM bereits 1999 mit der Firma ADIC ein OEM-Abkommen, das so aussah, dass ADIC LTO-Laufwerke von IBM bekam und im Gegenzug IBM die kleinen Libraries von ADIC, die dann unter IBM Logo mit IBM LTO-Laufwerken vermarktet wurden.

So stellte ADIC die damaligen 'Midrange Libraries', die **IBM 3583** und später die **IBM 3582** (bei ADIC Skalar 100 und Skalar 24), zur Verfügung. Der einzige Unterschied zu den ADIC-Produkten war, dass IBM in beide Libraries eine eigene FibreChannel-basierende Steuereinheit integrierte, um den Libraries über die dadurch geschaffene Multipfadarchitektur die Möglichkeit zu geben, logische Partitionen als Sub-Libraries anzulegen und um direkt FibreChannel-LTO-Laufwerke einzubauen. ADIC verwendete SCSI-Laufwerke, die dann über ein Gateway an SANs anschließbar waren. FC-Laufwerke wurden bei ADIC erst Anfang 2005 verwendet und eingebaut.

Die heute aktuellen Autoloader, Midrange Libraries und die Besonderheiten des 3584 Bandarchivs sind in der Folge-Epoche näher beschrieben.

IBM 3583 Ultrium Scalable Tape Library mit geöffneter Tür

IBM 3582 Display – Mini-Library mit bis zu 23 LTO-Kassetten

Kommentar zur Epoche der RAID-Systeme

Im Vergleich zur Vorgänger-Epoche, wo über Jahre dieselben Steuereinheiten und Platten aktuell blieben, war die Epoche der RAID-Systeme mit sehr schnellen Entwicklungszyklen sehr hektisch. RAMAC 1, 2 und 3 kamen hintereinander im Abstand von nur einem Jahr.

In dieser Epoche herrschte ein enormer Preisdruck, der durch neue Mitbewerber und Player im Plattenumfeld hervorgerufen wurde.

RAMAC 3 war der letzte 'Rolls Royce', den die IBM baute, dann liefen die Kosten davon und die Produkte waren viel zu teuer im Vergleich zum Mitbewerb. Das führte zum Zusammenschluss von StorageTek und IBM und zum IBM Vertrieb des RAMAC Virtual Arrays (ehemals STK Iceberg). Plötzlich musste der IBM Vertrieb ein Produkt verkaufen, von dem man sich bisher immer distanziert hatte. Ein massives Umdenken war im Speichervertrieb angesagt. Log Structured File, die bisher von IBM negativ kommentierte 'Garbage Collection – das kann doch nichts sein' musste nun positiv dargestellt werden. Die Verbrüderung mit STK im Plattenumfeld musste erst einmal verdaut werden.

Im Open-Systems-Bereich glich die erfolgreiche Einführung der SSA-Architektur und der erfolgreiche Verkauf des Plattensystems IBM 7133 den Einbruch im Mainframe-Umfeld entsprechend aus.

Ebenso hektisch waren die Entwicklungsjahre im Tape-Umfeld geprägt. Neben Magstar 3590, Magstar MP und der Entwicklung des Virtual-Tape-Servers (VTS) wurde die LTO(Linear Tape Open)-Bandtechnologie zusammen mit Hewlett Packard und Seagate entwickelt und spezifiziert. Im Herbst 1999 war der Standard der LTO1-Technologie fertig. Danach begann ein Prestige-Wettrennen zwischen IBM, HP und Seagate. Jeder wollte der Erste sein, der mit LTO1-Laufwerken auf den Markt kommt.

Seagate kündigte als erste Firma, bereits im Frühjahr 2000, ihre LTO1-Laufwerke an. Die Seagate-Laufwerke wurden aber dann doch als letzte verfügbar. IBM kündigte im September 2000 mit sofortiger Verfügbarkeit an. HP folgte mit Ankündigung und Verfügbarkeit einen Monat später.

Die Epoche war geprägt durch Vertriebskämpfe vieler Mitbewerber im Platten- und Tape-Umfeld und war deshalb so hektisch, weil enorme Preis- und Prestigekämpfe stattfanden.

Die Epoche der Multiplattform-Systeme und des FibreChannel SAN und NAS

Die Epoche der Multiplattform-Systeme und des FibreChannel SAN und NAS

1999 – 2006: IBM SAN, Switches und Direktoren der Hersteller Brocade, McData (CNT) und Cisco

SAN, NAS, iSCSI

Die Epoche der RAID-Systeme war bis 1999 so gestaltet, dass Plattensysteme für den Mainframe-Bereich zur Verfügung standen, während der Open-Systems-Bereich eigene, damals direkt angeschlossene Plattensysteme verwendete. Nur Tape Libraries waren in der Lage, beide Welten zu unterstützen. Das Datenwachstum in dieser Zeit war dramatisch, oft über 60 % pro Jahr für Rechenzentrumsbetriebe. Die IP-Netze und die damit verbundene Infrastruktur, damals Ethernet auf 10-Mbit-Basis, erlaubten kaum, riesige Datenmengen über IP-Infrastrukturen hin und her zu bewegen. Der Ruf nach **separaten Speichernetzwerkinfrastrukturen** wurde immer lauter, um der Datenflut entgegenzuwirken. 1998 kamen von der Firma Brocade die ersten FibreChannel Switches auf 8-Port-Basis mit 1-Gbit-Ports auf den Markt. Eine **neue Infrastruktur, das SAN (Storage Area Network)** auf FibreChannel-Basis sollte in den kommenden Jahren eine wahre Blütezeit erleben.

Die Anfänge der **ersten SAN-Infrastrukturen** fanden allerdings schon zwei Jahre früher statt. Sie kamen bereits **1996** in **Hochleistungsrechenzentren** der Universitäten zum Einsatz und wurden von der Firma Ancor zur Verfügung gestellt. **1998** wurde der erste 8-Port-Switch von der Firma Brocade im kommerziellen Rechenzentrumsumfeld bei der IT Austria in Wien eingesetzt und in den Folgejahren gingen viele Firmen dazu über, in den unterschiedlichsten Unternehmensbereichen SAN-Infrastrukturen zu implementieren. Nicht nur im Open-Systems-Bereich erlebte die Glasfaser eine neue Blütezeit, auch die Mainframe-Umgebungen wechselten, beginnend mit dem Jahr 2001, von der bisherigen ESCON(Enterprise Storage CONnection)-Umgebung mit den bis dahin verwendeten ESCON-Direktoren auf Glasfasernetze, um zusammengefasste ESCON-Pakete über die Glasfaser zu übertragen. Das Ganze wird heute als FICON bezeichnet (Fibre CONnection). Dadurch konnten die hohen Übertragungsraten der Monomode-Glasfaser mit entsprechenden Multiplexingverfahren auch im Mainframe-Bereich genutzt werden.

Bereits im Jahr **2003** waren Direktoren auf dem Markt, die sowohl FibreChannel- als auch FICON-Ports abbilden konnten. Leider kann bis heute noch kein FC-Port als FICON-Port oder umgekehrt genutzt werden und es wird sicherlich noch einige Zeit dauern, bis SAN-basierende Infrastrukturen sowohl von der Open-Systems-Umgebung als auch von den Mainframes gleichzeitig genutzt werden können. Die FibreChannel SANs entwickelten sich rasch weiter. **2002** kamen die ersten **Switches mit 2-Gbit-Ports** auf den Markt. Im **Jahr 2006** waren alle Hersteller im Begriff von 2-Gbit- auf 4-Gbit-Technologie umzustellen und bis Ende 2006 war die Umstellung soweit, dass 'End to End'-SAN-Implementierungen auf 4-Gbit-Basis realisiert werden konnten. Sowohl die Host Adapter aller Server-Hersteller als auch die Host-Ports aller Storage-Anbieter waren auf 4-Gbit-Technologie umgestellt. Heute, im Jahr 2008, wiederholt sich dieses Spiel nur auf 8-Gbit-Technologie Basis. Seit Anfang 2008 sind die ersten 8-Gbit-fähigen Direktoren auf den Markt gekommen.

In der Anfangszeit dieser Epoche waren **SAN-Implementierungen sehr kostenintensiv**. Nicht jede Firma, vor allem im Mittelstandsbereich, konnte sich SANs finanziell leisten, um der neuen Bandbreitenmöglichkeit Rechnung zu tragen. Hinzu kam die Hochpreispolitik der Glasfaserhersteller, die vor allem die für das Multiplexing erforderliche Monomode-Faser über die Anfangsjahre sehr teuer gestaltete. Die Server-Adapter-Hersteller schlossen sich dieser Hochpreispolitik an. Das führte dazu, dass die Entwickler von Multiplexing-Verfahren auf der Glasfaser kein großes Umsatzpotenzial mehr erkennen konnten und sich deshalb neuen Multiplexing-Verfahren auf Ethernet-Basis zuwandten. So kommt es, dass heute IP-Infrastrukturen auf Ethernet-Basis wesentlich größere Übertragungsbandbreiten bieten (2-Gbit- und 10-Gbit-Ethernet) im Vergleich zu FibreChannel-Netzen.

Die Hochpreispolitik führte auch dazu, dass neben dem SAN eine neue Technologie, das NAS (Networking Attached Storage), Einzug hielt. NAS-Technologie in Form von File-Server-Geräten mit eingebauten Platten kamen ebenso zum Einsatz wie Gateways, die prinzipiell nichts anderes als einen File-Server darstellten, aber die Möglichkeiten hatten, Plattenkapazitäten innerhalb eines SANs für File-Serving-Zwecke zu nutzen.

Heute sind SANs selbst für sehr kleine Betriebe bezahlbar. Der Preisverfall von 2003 bis 2006 liegt in einem Bereich von nahezu 96 %. Dieser Preisverfall wurde maßgeblich durch die neuen Bandbreitenmöglichkeiten der IP-Netze auf Ethernet-Basis erzielt und stellt durchaus, vor allem heute (10-Gbit-Ethernet), die Notwendigkeit von SANs infrage, wenn es um Engpässe in der Datenübertragungsmöglichkeit geht. Trotzdem, vor allem aufgrund der Managementmöglichkeiten, ist davon auszugehen, dass SAN-Netze von 4 Gbit in Richtung 8 Gbit (oder 12 Gbit) weiterentwickelt werden, die zu den bestehenden Netzen (Autosensing) kompatibel bleiben.

Neben den Möglichkeiten und Vorteilen von SANs bot IBM in den Jahren 2000 bis 2004 auch Lösungen und Produkte auf dem Sektor **Networking Attached Storage (NAS)** an. NAS-Systeme sind vorkonfigurierte Fileserver. Sie bestehen bis heute aus einem oder mehreren internen Servern mit vorkonfigurierter Plattenkapazität und werden über Ethernet an das LAN angeschlossen, wo sie ihren Plattenspeicher als Fileserver oder als HTTP-Server zur Verfügung stellen. NAS-Server sind speziell für das File Sharing entwickelt. Der Vorteil von NAS im Vergleich zu klassischen Fileservern bestand im geringen Installations- und Wartungsaufwand.

Die IBM NAS-Modelle 200 und 300 waren Lösungen speziell für Kunden, die in einem Windows-Umfeld Speicher konsolidieren wollten. Alle IBM NAS Appliances wurden mit einem vorkonfigurierten Microsoft-Windows-Betriebssystem ausgeliefert. Dieses Windows wurde speziell für Fileserving angepasst. Routinen, die überflüssig waren, wurden aus dem Betriebssystem entfernt, um die Leistung und Zuverlässigkeit zu erhöhen. Dadurch boten die IBM NAS-Modelle sehr gute Performance in einer Windows-Umgebung (CIFS). Auch in einem gemischten Windows- und UNIX(NFS)-Umfeld zeigten die IBM NAS-Modelle eine gute Leistung.

Für klassische Block-I/O-Anwendungen, wie Datenbanken, waren solche Fileserver nicht vorgesehen, weil diese Anwendungen direkt auf einen formatierten Plattenplatz ohne ein darüberliegendes Filesystem zugreifen. Dafür waren Lösungen wie Direct Attached Storage (DAS), Storage Area Networks (SAN) und iSCSI-Produkte besser geeignet.

IBM NAS Gateway 500, verfügbar Februar 2004, pSeries-basierend, bis 224 TB Kapazität im SAN

Um die Speicherressourcen im SAN für Fileserving zu nutzen, bot die IBM, damals am Anfang als einziger Vendor, die Möglichkeit, vorhandene IP-Netzwerkstrukturen über ein **NAS Gateway** mit dem SAN zu verbinden. Das NAS Gateway war ein dedizierter Fileserver. Dieser war mit dem SAN über FibreChannel verbunden und so in der Lage, Kapazitäten z. B. eines Enterprise Storage Servers ESS im SAN optimal für Fileserving einzusetzen. Durch die Nutzbarmachung eines SAN-Speicherpools über Gateways konnten Fileserver in ein SAN konsolidiert werden. Ebenso wurde durch die Nutzung des Speichers im SAN eine Server-unabhängige Skalierung bezüglich der Kapazitäten möglich. Das **NAS Gateway 300G** wurde später, im Jahr 2004, durch das wesentlich leistungsstärkere Gateway 500 ersetzt.

iSCSI ist eine gemeinsame Entwicklung der Firmen **IBM und Cisco**. Dabei wird das SCSI-Protokoll über TCP/IP übertragen. iSCSI verfolgt damit einen ähnlichen Ansatz wie SAN, mit dem Unterschied, dass bei iSCSI eine TCP/IP-Verbindung das SCSI-Kabel ersetzt, und stellte damals in Bereichen mit niedrigen oder mittleren Leistungsanforderungen eine kostengünstigere Alternative zu SANs dar. Der Vorteil bestand darin, dass bereits vorhandene IP-Netzwerke direkt genutzt werden konnten und nicht ein separates Glasfasernetz in Form eines SANs aufgebaut werden musste. Die Implementierung von iSCSI-Lösungen im Vergleich zu SANs war wesentlich einfacher und erforderte nicht den hohen IT-Skill, der bei einem Aufbau eines Storage Area Network (SAN) notwendig war. Die IBM Modelle **IP Storage 200i (Modelle 200 und 225)** verwendeten im Gegensatz zu den NAS und NAS Gateways Linux als Betriebssystem (Kernel) und boten eine Kapazität von 108 GB bis 1.74 TB an.

Die iSCSI-Lösungen setzten sich allerdings auf dem Markt nicht wirklich durch, zumal kurze Zeit später der enorme Preisverfall der Monomode-Glasfaser einsetzte, der die Implementierung von SANs auch im Mittelstand und für Kleinbetriebe bezahlbar machte.

Plattensysteme

Im Juli 1999 wurde das erste IBM multiplattformfähige Plattensystem, der Enterprise Storage Server ESS, angekündigt. Unter dessen Entwicklungsname **'Shark'** fand das System allerdings weit mehr Verbreitung als unter dem Begriff ESS. Die Typenbezeichnung war 2105 und die **1999** angekündigten **Modelle** waren die **E10 und E20** mit einer Kapazität von 420 GB bis 11.2 TB. Dabei wurden Plattenlaufwerke von 9 GB, 18 GB und 36 GB als SSA-Platten verwendet, die über vier sogenannte Device Adapter Pairs in SSA-Loop-Technik an den Rechner angebunden waren. Am Anfang konnten die Plattentypen nicht gemischt werden, später allerdings war es möglich, unterschiedliche Platten in die Arrays einzubauen. Cache-Größen von bis zu 6 GB waren konfigurierbar und ein Strom unabhängiger Schreibspeicher (NVS Non Volatile Storage) von 384 MB stand zur Verfügung. Die Arrays waren am Anfang ausschließlich unter RAID5 betreibbar. Für den Mainframe standen die 3380- und 3390-Spurformate zur Verfügung, an Open Systems (UNIX, Windows NT und AS/400) wurden entsprechende LUNs emuliert (Logical Unit Number).

Vom Vorgänger der ESS, dem Versatile Storage Server, einem Kurzläufer von wenigen Monaten, wurde der **IBM Data Path Optimizer** für AIX und Windows NT als integraler Bestandteil für die Multipfadfähigkeit der Maschine übernommen.

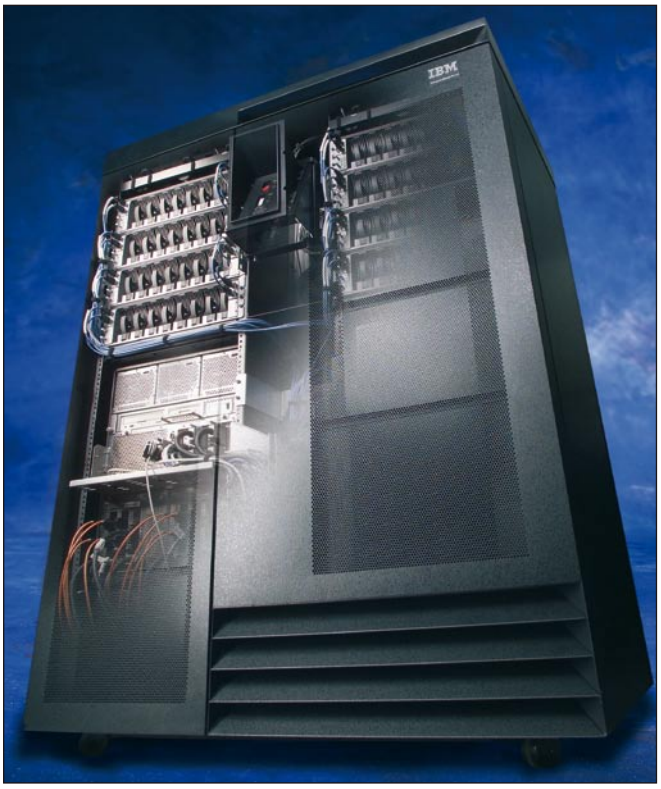
Funktional war die Maschine am Anfang noch schwach auf der Brust und es standen nur begrenzt Copy Services als Funktionen zur Verfügung, was die Markteinführung nicht gerade vereinfachte. Später kamen dann die Mainframe-Funktionen **'Concurrent Copy'** und **XRC**, 'eXtended Remote Copy' als asynchrone Spiegelmöglichkeit ganzer Systeme für S/390-Server (mit konsistentem Datenbestand auf der Sekundärseite) und wiederum zeitversetzt die Funktion **PPRC 'Peer to Peer Remote Copy'** als synchrone Spiegelung sowohl für den Mainframe als auch den Open-Systems-Bereich. **Auch FlashCopy (Instant- oder Point-in-Time-Copy)** wurden auf der Maschine implementiert, um in Sekundenschnelle Kopien erzeugen zu können, mit denen man sofort arbeiten konnte.

Knapp zwei Jahre nach der Markteinführung bot die Maschine dann allerdings im Mainframe-Umfeld eine wesentlich höhere Funktionsvielfalt im Vergleich zu den auf dem Markt befindlichen Mitbewerbersystemen. Die Funktion **PAV (Parallel Access Volume)** erlaubt im Mainframe-Umfeld multiple I/O-Operationen auf denselben logischen Volume. **Multiple Allegiance** lässt den parallelen Schreib- und Lesezugriff auf dasselbe logische Volume von verschiedenen S/390-Servern zu. **I/O Priority Queuing** verwendet die Priorisierungsinformationen des OS/390-Workloadmanagers im Goal Mode, um die Sequenz der einzelnen I/Os optimal zu steuern und Workloads mit unterschiedlicher Wichtigkeit entsprechend ihrer Priorisierung optimal zu behandeln.

Das System kam mit neuen Management Tools, der sogenannten **'StorWatch'**-Familie auf den Markt, die es erlaubte, die Maschinen über das Web zu administrieren. Standardmäßig wurde die Maschine mit dem **ESS Specialist** ausgeliefert, der vor allem für das Konfigurieren und für spätere Modifizierungen notwendig war. Als separates, zusätzlich käufliches Produkt wurde der **ESS Expert** zur Verfügung gestellt, der ausführliche Informationen über genutzte Kapazitäten, Performance, Antwortzeiten, Cache Hit Ratios und mehr lieferte, um ein optimales Tuning auf der Maschine für die unterschiedlichen Workloads durchführen zu können.

Im Mai **2001** kamen die **Folgemodelle F10 und F20**, die erstmals native FICON-Anschlüsse boten, um mit FibreChannel im Mainframe-Umfeld zu arbeiten. Auch die bisher verwendeten ESCON- und FC-Adapter waren verbessert worden und boten 100 MB/s Durchsatz pro Port in FICON und FibreChannel Speed. Sie lieferten Entfernungsmöglichkeiten von bis zu 10 km (native) und bis zu 100 km über entsprechende SAN-Fabric-Komponenten aus. Die neuen Modelle konnten mit bis zu 16 FICON-Anschlüssen und einer zusätzlichen Cache-Option von 24 GB, also maximal 32 GB, ausgestattet sein. Damit verbesserte sich die Leistung zu den Vorgängermodellen um ca. 40 %.

Funktional stellten die neuen Modelle die Funktion PPRC und FlashCopy auch der Systemplattform AS/400 und iSeries zur Verfügung. Im **März 2002** wurde auch die **FICON-Unterstützung** für die bei Airline-Buchungssystemen verwendete TPF(Transaction Processing Facility)-Anwendung verfügbar.



Enterprise Storage Server ESS (Shark), 1999: Modelle E10/E20 bis 11 TB, 2001: Modelle F10/F20 bis 35 TB, 2002: Modelle 800 und 800 Turbo II bis 56 TB, April 2004: Entry-Modell 750 bis 5 TB (Brutto-Kapazitäten)

Der Wechsel von ESCON- auf FICON-Infrastrukturen brachte damals enorme Infrastrukurverbesserungen und Infrastruktureinsparungen mit sich, weil in der Regel vier ESCON-Kanäle mit 18 MB/s durch einen FICON-Kanal mit 100 MB/s ersetzt werden konnten. Allerdings dauert die Penetration noch bis in die heutige Zeit an.

Im August 2002 kam die letzte Modellreihe der **ESS mit den Modellen 800 und 800 Turbo II** auf den Markt, die dann noch durch ein **Einstiegsmodell 750** im **Frühjahr 2004** ergänzt wurde. Die neuen Modelle waren leistungsstärker und verwendeten zwei 2-, 4- oder 6-way-SMP-Rechner mit bis zu 64 GB Cache.

Dadurch waren ca. 42 % mehr Transaktionen bei Cache Standard Workloads möglich. Die neuen Modelle konnten im Mischbetrieb Laufwerke mit 18 GB, 36 GB, 73 GB und 146 GB ausgestattet werden und boten als Brutto-Kapazität von 582 GB bis 56 TB. Die Laufwerke mit 18 GB, 36 GB und 73 GB gab es mit Umdrehungsgeschwindigkeiten von wahlweise 10000 Umdrehungen/Minute und 15000 Umdrehungen/Minute, während das großkapazitive 146-GB-Laufwerk nur mit 10000 Umdrehungen/Minute zur Verfügung stand.

Wahlweise konnten die Arrays mit RAID0, 1, 5 und 10 konfiguriert werden und alle Platten in einem Array wurden mit einem Striping-Verfahren (alle Laufwerke arbeiten gleichzeitig, wenn eine Spur in kleinen Stücken parallel auf alle Laufwerke des Arrays geschrieben oder gelesen wird) angesteuert, um eine optimale Schreib- und Leseperformance sicherzustellen. Für die Array-Anbindung wurden SSA-160-Adapter verwendet, die in einer Doppelloop einen Gesamtdurchsatz von 320 MB/s bei acht gleichzeitigen I/Os erreichen konnten. Der Gesamtdurchsatz (interne Bandbreite) lag beim Turbo-II-Modell bei 3.2 GB/s. Die maximale Cache-Größe lag bei bis zu 64 GB.

Das RAID-Management war komplett in die Adapterkarten verlagert und bei sequenziellen Schreib-Workloads war die Maschine in der Lage, vom RAID5- in den RAID3-Modus umzuschalten, bei der bis zu 64 gleichzeitige physische I/Os auf die Platten möglich waren.

Alle Maschinen waren von Beginn an auf 2-Gbit-Hostanbindungungstechnologie, also 2 Gbit FICON und 2 Gbit FibreChannel, ausgelegt. Daneben konnte die Maschine immer auch noch mit ESCON- und Ultra-SCSI-Adaptern konfiguriert werden.

Die **Modelle 800 und 800 Turbo II** waren die Basis, um ein komplett **neues Remote-Copy-Verfahren** zu entwickeln, das in seiner Leistungsfähigkeit bis heute einmalig im Markt ist und bei den Folgeprodukten der DS6000 und DS8000 maßgeblich seinen Einsatz findet. Dieses neue Remote-Copy-Verfahren wurde den neuen ESS-Modellen im April 2004 zur Verfügung gestellt und ermöglichte durch eine bidirektionale Implementierung über eine FibreChannel-Verbindung eine synchrone Spiegellast (PPRC) von bis zu 16000 I/Os in der Sekunde (ESS), die sich später bei den neuen DS8000 auf bis zu **23000 I/Os** pro Sekunde steigerte – und das über einen einzigen FibreChannel Link. Bei dieser neuen Implementierung können extrem kurze Antwortzeiten, auch bei synchroner Spiegelung über lange Entfernungen, erzielt werden. Bei den Modellen 800 der ESS wurden Antwortzeiten von 2 ms bei einer synchronen Spiegelung von über 75 km erzielt.

Ebenso wurden für die neuen Modelle die Management-Möglichkeiten erweitert. So wurden erstmals offene APIs als Industrie Standard Interface (SNIA-SMI-S) zur Verfügung gestellt. Die Maschinen konnten mit Scripting Tools über das CLI (Command Line Interface) automatisiert werden. Ebenso stand eine Web-basierende grafische Benutzeroberfläche (GUI) für einfache Administration zur Verfügung.

Für alle ESS-Modelle wurden RS/6000 basierende POWER5-Prozessoren eingesetzt. Die E-Modelle verwendeten eine H50 als 4-Way-SMP-Version mit einer Taktrate von 340 MHz. Die F-Modelle benutzten eine H70 als 4-Way-SMP mit 540 MHz. Die letzten Modelle der 800er Reihe verwendeten einen H80 Rechner. Das Einstiegsmodell 750 benutzte eine H80 als 2x2-Way mit 600 MHz, die Modelle 800 eine H80 als 2x4-Way mit 600 MHz und die leistungsstärkste ESS 800 Turbo 2 eine H80 als 2x6-Way mit 750 MHz. Für die interne Übertragung in den Maschinen wurde mit PCI-Bussen gearbeitet.

Im Januar 2006 zog IBM alle ESS-Modelle mit Wirkung vom April 2006 vom Vertrieb zurück.

Parallel zur ESS wurde bei IBM bereits 1999 der Startschuss gegeben, an Server-basierenden Speicherarchitekturen zu arbeiten, die die Basis für die heutige, erste Server-basierende Speicherarchitektur legten und heute mit dem Produkt der DS8000 zur Verfügung stehen. Diese neue Architekturentwicklung fand parallel zur Weiterentwicklung der ESS statt und wurde unter größter Geheimhaltung vorangetrieben.

Das erfolgreiche, SSA-basierende Plattensystem IBM 7133, das vor allem in RS/6000- und pSeries-Umgebungen eingesetzt wurde, bekam 2002 auch die Möglichkeit, in FibreChannel SANs zum Einsatz zu kommen. Mit dem Adapter 7140 konnte das Plattensystem an SANs angeschlossen werden und stellte eine neue Funktion über diesen Adapter, ‘InstantCopy’, zur Verfügung. Damit konnte neben einem RAID1-Spiegel eine weitere Kopie erzeugt werden, die dann für asynchrone Backup- oder Testzwecke verwendet werden konnte. Da die hier eingesetzte SSA-Technologie auch im Enterprise Storage Server ESS (Shark) eingesetzt wurde, bezeichnete man die Kombination von 7133 und dem Adapter 7140 auch als ‘Hammershark’. Die Nähe zur ESS (Shark) wurde auch durch eine Aufrüstoption unterstrichen, bei der 7133 Kapazität zur Erweiterung einer Shark genutzt werden konnte.

Nach der Markteinführung der ESS als multiplattformfähiges Enterprise-Plattensystem erkannte IBM sehr schnell, dass selbst Einstiegskonfigurationen einer Shark für den Mittelstandsbereich zu teuer waren. Deshalb wurde bereits im **Jahr 2000** mit der Firma **Compaq** ein **OEM-Vetrtriebsvertrag** unterzeichnet, der IBM erlaubte, Compaqs aktuelles Plattensystem mit FibreChannel-Anschlüssen zu vertreiben. Das Produkt MSS Modular Storage Server unterstützte heterogene Serverplattformen im Windows-NT- und UNIX-Umfeld und wurde über Hubs und/oder Switches an FC-Server angeschlossen.

Die Zeit des MSS Modular Storage Servers war aber nur von sehr kurzer Dauer und es wurden nur vereinzelte Systeme installiert. Mit der Übernahme von Compaq durch Hewlett Packard HP war die Allianz mit Compaq beendet.

Daraufhin griff IBM auf eine seit Anfang der 90er-Jahre bestehende **OEM-Allianz** mit der **Firma LSI** zurück die es erlaubte, dass IBM unter IBM Logo Plattenprodukte von LSI speziell im Intel-Umfeld (NetFinity) über den IBM xSeries-Kanal vertrieb. Bereits 1993 wurden die LSI-Produkte IBM 7135-110 RAIDiant Array, in den Vertrieb aufgenommen, 1995 das IBM 7135-210 RAIDiant Array, 1998 der IBM 3526 FC RAID Controller, 2000 die IBM 3552 FAStT500 und die IBM 3542 FAStT200 und im Jahr 2001 die IBM 1742 FAStT700.

Im **Jahr 2001** wurde diese Allianz zwischen IBM und LSI auf den gesamten Storage-Vetrieb erweitert, eine Allianz, die sehr erfolgreich werden sollte und bis in die heutige Zeit hochaktuell ist. Das erste Produkt, das IBM über diese Allianz vertrieb, war ein Entry-Plattenprodukt, die **FAStT200**, damals auch als IBM 3542 bekannt, das mit bis zu zwei FibreChannel Ports mit der Grundeinheit (10 eingebaute Festplatten) und zwei Erweiterungseinheiten EXP500 (pro EXP500 zusätzliche 15 Platten) eine maximale Kapazität von bis zu 2.1 TB anbot. Das System stellte eine preisgünstige, performante Plattenlösung für dezentrale Abteilungen und Arbeitsgruppen mit bis zu vier Servern dar. Das System bot als Entrysystem damals bereits eine ganze Reihe von Sicherheitsoptionen, wie Dual Active RAID Controller, RAID 0, 1, 10, 3 und 5, redundante Netzteile und Batterie sowie die Möglichkeit, alle Komponenten als 'Hot Swaps' zu tauschen. Im System konnten im Mischbetrieb Platten mit 18 GB, 36 GB und 73 GB eingesetzt werden.

Die Abkürzung **'FAStT'** steht für **'Fibre Array Storage Technology'**. Der ursprüngliche Name sollte FAST – für Fibre Array Storage Server – lauten, war aber bei der Ankündigung der LSI-Produkte bereits vergeben und konnte nicht verwendet werden. Deshalb entschloss man sich für FAStT mit der Intention, dass das kleine 't' nicht ausgesprochen wird. Trotz allem setzte sich die Aussprache als 'fast_T' in ganzer Breite durch.

	FAStT900	FAStT700	FAStT600 Turbo	FAStT600	FAStT200	FAStT200
Host Interface	2 Gbps FC	2 Gbps FC	2 Gbps FC	1 Gbps FC	2 Gbps FC	2 Gbps FC
SAN attachments (max)	4 FC-SW	4 FC-SW	4 FC-SW	4 FC-SW	4 FC-SW	4 FC-SW
Direct attachments (max)	8 FC-AL	8 FC-AL	4 FC-AL	4 FC-AL	2 FC-AL	4 FC-AL
Redundant drive channels	Four 2 Gb FC	Four 2 Gb FC	Four 2 Gb FC	Four 2 Gb FC	Four 1 Gb FC	Four 2 Gb FC
Drive types supported	FC, SATA	FC	FC, SATA	FC, SATA	FC	FC, SATA
Max drives	224	224	112	56	66	56
Max physical capacity with FC	32 TB	32 TB	16.4 TB	8.2 TB	9.6 TB	–
Max physical capacity with SATA	56 TB	–	28 TB	28 TB	–	14 TB
XOR technology	ASIC	ASIC	Integrated	Integrated	Integrated	Integrated
Subsystem Cache	2 GB	2 GB	2 GB	512 MB	256 MB	512 MB

IBM FAStT-Plattensubsystemfamilie, Spezifikationsübersicht

Die Epoche der Multiplattform-Systeme und des FibreChannel SAN und NAS

Die Reihe der FAStT-Plattensysteme wurde im **Oktober 2001** um die leistungsfähigeren Produkte **FAStT500 und FAStT700** erweitert, die mit bis zu 4 FibreChannel-Anschlüssen höhere Kapazitäten boten. Im Hochleistungsbereich kam im **Februar 2003** die **FAStT900** dazu, um die Baureihe auch im oberen Leistungsbereich abzurunden. Im **April 2003** wurde die FAStT500 duch die **FAStT600 und FAStT600 Turbo** ersetzt. Neben den Plattenerweiterungseinheiten EXP500 und EXP700 (EXP steht für Expansion), die mit FibreChannel-Platten bestückt waren, kündigte IBM im **Oktober 2003** die Erweiterungseinheit EXP100 an, die mit preislich günstigeren SATA-Platten bestückt war. Der dazu passende Controller wurde im **Mai 2004** mit der **FAStT100** angekündigt. **SATA-Platten** (Serial Advanced Technology Attached) sind günstige IDE-Platten mit seriellem Interface, die bisher auschließlich im Heim-PC-Bereich ihre Anwendung fanden. Mit Ausnahme der schon länger verfügbaren FAStT200 waren alle Systeme jetzt mit 2 Gbit FibreChannel ausgestattet, und das sowohl Host-seitig als auch in der Anbindung der Plattenloops an den jeweiligen FAStT Controller. Über 90 % des Vertriebs dieser LSI-Plattensubsystem-Serie lief über IBM oder IBM Geschäftspartner und war im FibreChannel-Umfeld ein riesiger Erfolg, der bis heute anhält.

Je nach Controllerstärke wurde bei den FAStT-Plattensystemen eine Vielzahl von neuen Funktionalitäten eingeführt, die dann vor allem in den oberen Modellen zum standardmäßigen Einsatz kamen und bis heute in den Systemen verwendet werden. Zum besseren Verständnis werden diese Funktionen im Folgenden beschrieben.

DCE und DVE: Dynamic Capacity Expansion und Dynamic Volume Expansion ermöglichen, zugewiesene Speicherbereiche im laufenden Betrieb zu vergrößern. Innerhalb des FAStT-Systems werden physische Laufwerke zu einzelnen Array-Groups gebunden. Die Laufwerke können sich dabei über verschiedene EXP-Erweiterungseinheiten verteilen. Die Selektion der Laufwerke kann dem System überlassen werden (dann wird stets die performanteste Konfiguration gewählt) oder vom Administrator manuell definiert werden. Für die erstellte Array-Group wird ein RAID-Level definiert der für die gesamte Array-Group Gültigkeit hat. Innerhalb dieser Array-Group können verschiedene Volumes definiert werden. Diese Volumes werden den einzelnen Servern zugewiesen. Für den Fall, dass der gewählte Plattenzusammenschluss (Array-Group) mehr Kapazität benötigt oder eine höhere Performance erforderlich wird, gibt es die Möglichkeit, dieser Group eine oder mehrere physische Laufwerke hinzuzufügen (DCE). Die in dieser Group befindlichen Volumes nutzen diese neue Kapazität, die Disk wird automatisch in die Volume-Verteilung aufgenommen.

Sollte innerhalb eines zugewiesenen Volumes mehr Kapazität benötigt werden, so ist im laufenden Betrieb auch hier eine Kapazitätserweiterung möglich (DVE). Voraussetzung hierfür ist, dass entsprechende Kapazitäten innerhalb der Array-Group zur Verfügung stehen (in diesem Fall kann über DCE eine beliebige, undefinierte Platte ergänzt werden). Diese Funktionalitäten ermöglichen es dem Anwender flexibel auf Kapazitätsanforderungen einzugehen und diese auf einfachste Weise umzusetzen. Alle Funktionen arbeiten im laufenden Betrieb mit Data in Place und werden über den FAStT Storage Manager initialisiert.

DRM: Dynamic RAID Migration. Genauso wie zugewiesene Kapazitäten sind ebenfalls zugewiesene RAID-Level nicht statisch, sondern auch im laufenden Betrieb veränderbar. Sollten sich durch Änderungen im Lastprofil, der Kapazitäts- oder Sicherheitsanforderung Änderungswünsche am eingesetzten RAID-Level ergeben, so kann dieser im laufenden Betrieb je RAID-Array angepasst werden. Es stehen die RAID-Level 0, 1, 1+0, 3 und 5 zur Verfügung. Es kann von jedem RAID-Level zu jedem anderen RAID-Level gewechselt werden. Voraussetzung ist, dass entsprechende Plattenkapazitäten (z. B. bei einem Wechsel von RAID5 zu RAID1) verfügbar sind (DCE). Auch dies ist im laufenden Betrieb mit Data in Place möglich. Die RAID-Berechnung wird im Controller über HW-Funktionalität durchgeführt.

DAE: Dynamic Array Expansion. Sollte innerhalb einer FAStT weitere physische Kapazität benötigt werden, so gibt es die Möglichkeit, im laufenden Betrieb die angeschlossenen Erweiterungseinheiten mit weiteren FC-Laufwerken in verschiedenen Kapazitätsgrößen und mit unterschiedlichen RPMs zu bestücken (bis zu 14 Laufwerke je EXP700). Sollten in den angeschlossenen EXPs keine freien Slots zur Verfügung stehen, so besteht die Möglichkeit, im laufenden Betrieb eine weitere EXP700 an den FAStT-Controller auszuschließen. So lassen sich einzelne Laufwerke oder ganze Erweiterungseinheiten im laufenden Betrieb hinzufügen.

Neben diesen Standard-Funktionalitäten sind weitere, optionale Funktionen verfügbar:

- *FlashCopy: Die Funktion FlashCopy (Point-in-Time-Copy) ermöglicht bis zu 4 virtuelle Kopien eines bestehenden Volumes. Diese Funktion kommt vor allem für die Erstellung von Testumgebungen und die Erhöhung von Onlinezeiten bei Backups zum Einsatz. Wie funktioniert nun FlashCopy bei den FAStT-Systemen?*
  Über dem FAStT Storage Manager (FSM) wird der FlashCopy-Befehl über die GUI-Oberfläche abgesetzt. Über das GUI wird in diesem Schritt ebenfalls ein Repository angelegt, ein physischer Datenbereich für die Änderungen zwischen der Kopie und T0, dem Zeitpunkt des Spiegelungsbeginns. Die Größe des Repositorys ist abhängig von der Anzahl der Änderungen am Originalvolume. Die Erfahrung zeigt, dass eine Repositorygröße von 20 % des Originalvolumes sinnvoll ist. Sollte bei längerer Kopievorhaltung oder überdurchschnittlich vielen Änderungen der Source mehr Repository-Kapazität benötigt werden, so kann dieser Bereich flexibel vergrößert werden (DVE). Die Erstellung einer FlashCopy ist innerhalb von wenigen Sekunden abgeschlossen. Zur weiteren Benutzung einer FlashCopy ist es wichtig, zu T0 auf einen konsistenten Datenbestand zurückzugreifen, da ansonsten eine inkonsistente Kopie entsteht. Die erstellte virtuelle Kopie (nach wenigen Sekunden) kann z. B. dem Backup-Server zur Datensicherung zugewiesen werden. Wird während der Nutzung einer FlashCopy (Target) eine Änderung an der Source (Block A) vorgenommen, so wird das Original vor der Änderung in das Repository übernommen. Wird das Target geändert, so wird diese Änderung ebenfalls im Repository gespeichert. Wurden an der Source keine Änderungen vorgenommen, so greift man bei der Nutzung der Copy (Target) automatisch auf die Source zurück.
- *VolumeCopy: Ergänzend zur Funktion FlashCopy ist für FAStT-Systeme mit FSM 8.4 auch die Funktion VolumeCopy verfügbar. VolumeCopy erstellt im Gegensatz zu FlashCopy eine physische Kopie. Es muss also für die Erstellung einer VolumeCopy die gleiche physische Kapazität wie die der Source zur Verfügung stehen. Es besteht die Möglichkeit, diese Kopie nach Wunsch zu priorisieren. Dies bedeutet, dass I/O-Tätigkeiten zum Host bevorzugt, die Erstellung der Kopie nur zweitrangig behandelt wird. VolumeCopy ist eine Funktion, die besonders für Testumgebungen (z. B. zur Durchführung von Releasewechseln etc.) oder für Online-Backups geeignet ist.*
- *Remote Mirroring: Während die Kopierfunktionen FlashCopy und VolumeCopy Datenkopien innerhalb eines FAStT-Systems zur Verfügung stellen, erstellt Remote Mirroring eine Kopie von Daten über eine SAN-Infrastruktur von einer FAStT auf eine weitere. Diese Funktion ist für die FAStT-Systeme 700 und 900 verfügbar. Remote Mirroring ist eine bidirektionale, synchrone Kopie. Über dedizierte FibreChannel Ports können bis zu 32 Volumes auf eine zweite FAStT gespiegelt werden. Remote Mirroring wird für Katastrophenvorsorge und High-Performance Copys eingesetzt.*
- *Partitioning: Um unterschiedliche Speicherbereiche auf der FAStT voneinander abzugrenzen, steht die Funktion Partitioning zur Verfügung. Die Abgrenzung bewirkt, dass definierte Server nur auf den ihnen zugewiesenen Speicherbereich zugreifen können. Auf Speicherbereiche von 'fremden' Servern kann somit kein Zugriff erfolgen. Dies erhöht die Sicherheit der einzelnen Speicherbereiche. Besonders in Microsoft-Umgebungen ist diese Funktion sehr wertvoll, da die Betriebssysteme keine derartige Funktionalität bieten.*

FAStT RAID Controller Familie

Laufwerkerweiterungseinheiten, Management SW und das Einbau-Rack

EXP100
Serial-ATA-Technologie

EXP700
2-Gb-FC-AL-Technologie

Laufwerkeinheiten

FAStT Storage Manager (FSM)

Das Rack

Mit der Ankündigung der ersten **Server-basierenden Speicherarchitekturen** mit den Plattenprodukten DS6000 und DS8000 unternahm IBM im **September 2004** ein generelles, einheitliches Re-Branding der Plattenprodukte. DS steht für Disk System. Die Zahl dahinter soll die Leistungsfähigkeit des jeweiligen Systems reflektieren: je höher die Zahl, desto höher die Leistung.

Neben der Vereinheitlichung der Plattenproduktnamen wurden auch die Namen der jeweiligen Funktionalitäten entsprechend angepasst. Vergleichbare Funktionen der DS6000 und DS8000 wurden bei den FAStT-Systemen mit denselben Begriffen bezeichnet. Dies betraf vor allem die Remote-Copy-Spiegelverfahren.

Das synchrone Spiegeln von Plattensystemen wird bei der DS6000 und DS8000 als Metro Mirroring bezeichnet, das asynchrone Spiegelverfahren als Global Mirroring. Die neu eingeführten Begriffe sollen dem neu entwickelten Spiegelverfahren, das auf der ESS Modell 800 als Plattform entwickelt wurde und heute das leistungsfähigste Spiegeln im Markt darstellt, gebührend Rechnung tragen.

Für die FAStT-Plattenfamilien wurden im Einzelnen folgende Rebrandings durchgeführt:

Naming Prior to Sept 7, 2004	New naming as of Sept 7, 2004
IBM TotalStorage FAStT Storage Server	IBM TotalStorage DS4000
FAStT	DS4000
FAStT Family	DS4000 series
FAStT Storage Manager vX.Y (example FSM v9.10)	DS4000 Storage Manager vX.y (example 9.10)
FAStT100	DS4100
FAStT600	DS4300
FAStT600 with Turbo Feature	DS4300 Turbo
FAStT700	DS4400
FAStT900	DS4500
EXP700	DS4000EXP700
EXP100	DS4000EXP100
FAStT FlashCopy	FlashCopy for DS4000
FAStT VolumeCopy	VolumeCopy for DS4000
FAStT Remote Volume Mirror (RVM)	Enhanced Remote Mirroring for DS4000
FAStT Synchronous Mirroring	Metro Mirroring for DS4000
FAStT Asynchronous Mirroring (New Feature) w/o Consistency Group	Global Copy for DS4000
FAStT Asynchronous Mirroring (New Feature) with Consistency Group	Global Mirroring for DS4000

Die **jetzt neu bezeichnete DS4000-Plattensubsystemfamilie** wurde im **Mai 2005** durch ein neues **Hochleistungsmodell DS4800** ergänzt, das eine doppelt so hohe Leistungsfähigkeit im Vergleich zur DS4500 auslieferte. Die DS4800 war vom Aufbau her bereits beim Verfügbarwerden eine 4-Gbit-Maschine und somit das erste 4-Gbit-Plattensystem auf dem Markt. Die Cachegrößen am Anfang konnten je nach Modell mit 4 GB, 8 GB und später 16 GB konfiguriert werden. Die DS4800 hatte zudem noch eine andere positive Eigenschaft in der Bauweise: Die Maschine war **RoHS-konform** und ent-

sprach bereits 2005 der neuen EU-Richtlinie, die seit 1. Juli 2006 in Kraft ist.

Im Jahr 2003 bildete sich in den USA ein Gremium, das es sich zur Aufgabe machte, Plattensubsysteme unterschiedlicher Hersteller nach realen und produktiven Workloads auszutesten und diese Ergebnisse zu publizieren. Dieses heute fest etablierte Gremium nennt sich **'Storage Performance Council' bzw. SPC.**

Die neue DS4800 übertraf als Midrange-Plattensystem in der Leistung alle anderen Hersteller mit großem Abstand und erreichte eine Spitzenpunktzahl von 42254. Damit ist die DS4800 heute das schnellste Plattensystem im Midrange-Markt.

Da die älteren DS4000-Produkte mit Ausnahme der neu verfügbar gewordenen DS4800 nicht der RoHS-Richtline entsprachen, wurden im Mai 2006 alle nicht RoHS-konformen Produkte durch entsprechend neue ersetzt.

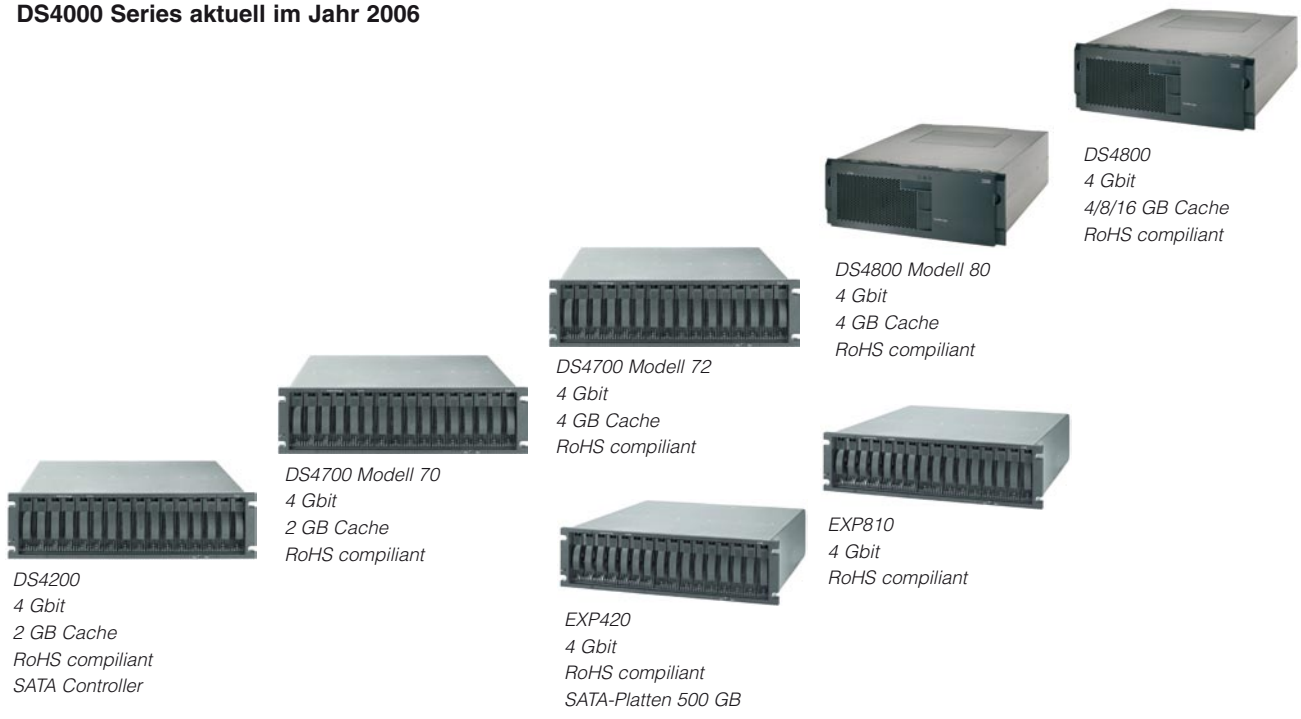
RoHS steht für 'Restriction of Hazardous Substances' und stellt ab dem 1. Juli 2006 eine gesetzlich kontrollierte EU-Richtlinie dar, die genau vorschreibt, welche Materialien, Legierungen, Barium-Derivate etc. in Neuprodukten eingesetzt werden dürfen. Diese neue Richtlinie gilt EU-weit und betrifft alle Neuprodukte in der Elektro- und IT-Industrie, die ab dem 1. Juli 2006 vertrieben werden.

Die DS4100 als SATA Controller wurde durch den neuen **DS4200 SATA Controller** ersetzt, die EXP100 mit 250-GB- und 400-GB-SATA-Platten durch die EXP420 mit 500-GB-SATA-Platten.

Die DS4300 Dual Controller wurden durch die **DS4700 Modell 70** und die DS4300 Turbo durch die **DS4700 Modell 72** ersetzt. Ebenso wurde die EXP710 mit FibreChannel-Platten durch die EXP810 mit FC-Platten auf 4-Gbit-Basis ersetzt.

Die DS4500, ehemals FAStT900, wurde durch ein neues Einstiegsmodell, das **Modell 80 der leistungsstarken DS4800**, ersetzt. Bei diesem Einstiegsmodell 80 wurde die interne Übertragungsbandbreite leicht verringert und die Prozessoren der Controller wurden mit einer 100-MHz-Taktung (im Vergleich zu einer 133-MHz-Taktung) versehen. Damit schließt das Modell 80 der DS4800 die Leistungslücke zwischen den Hochleistungsmodellen der DS4800 und dem Midrange-Modell 72 der DS4700.

DS4000 Series aktuell im Jahr 2006



Aktuelle Modelle der IBM DS4000-Plattenfamilie im Jahre 2006

DS4000 Erweiterungen 2007 und 2008

Im Jahr 2007 stellte IBM für die **DS4000 Plattenfamilie** eine ganze Reihe von Erweiterungen vor. Im **Mai 2007** wurden für die DS4000 **750 GB SATA Platten** und **300 GB Fibre-Channel Platten mit 15000 Umdrehungen** in der Minute eingeführt. Damit skaliert eine DS4800 auf bis zu 68 TB bzw. 168 TB Speicherkapazität (je nach Verwendung des Plattentyps). Im Oktober 2007 kamen viele funktionale Erweiterungen für alle DS4000 Systeme dazu. Für die Systeme DS4200 und DS4700 steht als neue RAID-Option RAID6 zur Verfügung. RAID6 sichert den Ausfall von zwei Platten im RAID-Array ab, da mit zwei Parity-Schemen gearbeitet wird (siehe auch unter RAID). Die Anzahl der Mirror Relations wurde verdoppelt, ebenso die Anzahl der Flash Copies. Mit Verfügbarkeit im Februar 2008 können für alle Systeme LUNs konfiguriert werden, die größer als 2 TB sind.

Die **leistungsstarke DS4800** bekam die Erweiterung, dass bis zu 512 Server and das Plattensystem angeschlossen werden können. Es stehen jetzt vier aktuelle Modelle der DS4800 zur Verfügung, die sich hauptsächlich in der Cache-Größe unterscheiden. So sind, je nach Modell, die Cache-Größen 4 GB, 8 GB und 16 GB verfügbar.

Bedient werden die RAID-Level 0, 1, 3, 5 und 10 (der neue RAID6-Level steht ausschliesslich der DS4200 und DS4700 zur Verfügung). Die Berechnung der RAID-Level bzw. die **XOR-Kalkulation** in der DS4800 erfolgt in einem **eigens enwickelten ASIC**. Dies ist ein spezieller Prozessor, der speziell für diese Workload konzipiert wurde. Dies macht die DS4800 frei von Latenzen, weil weder der Cache noch andere durchsatzrelevanten Teile der Maschine damit belastet werden.

Ansicht und Aufbau des Hochleistungsmodells DS4800

Die DS4800 hat alle Festplatten in der Expansion Unit EXP810 konfiguriert. Weitere Expansion Units, sowie auch einzelne Festplatten, können im laufenden Betrieb hinzugefügt und in Betrieb genommen werden.

Alle Komponenten der DS4800 sind redundant ausgelegt, z. B. Controller, Stromzufuhr und Kühlung. Ebenfalls wird der Cache gespiegelt und ist über eine Batterie bis zu 72 Stunden vor Verlust geschützt. Zusätzlich ist bei der DS4800 die Midplane, die als Interconnect Module bezeichnet wird, im laufenden Betrieb tauschbar. Dies ist bislang einmalig bei Plattensystemen im Midrange-Umfeld.

Ebenfalls einmalig ist die Performance der DS4800. Die Maschine hält bis heute die Führungsposition im Benchmark des Storage Performance Councils. Dies ist ein Random I/O-orientierter Benchmark, bei dem reale Workloads hinterlegt werden. Die genauen Testdaten, sowie die Beschreibung findet man unter www.storageperformance.org.

Der DS4000 Storage Manager wird für alle Systeme kostenfrei mit jedem System mitgeliefert. Ebenso ist die Multipathing Software sowie die Call Home-Funktionalität ohne zusätzliche Kosten einsetzbar. Als Premium Features stehen die Funktionen FlashCopy, VolumeCopy und eRVM zur Verfügung. Diese sind nicht kostenfrei, jedoch wird immer nur eine Lizenz pro System und Funktion herangezogen.

Sowohl bei Kapazität und Leistung als auch beim Preis bietet die aktuelle Modellreihe der DS4000 für jeden Endbenutzer eine maßgeschneiderte Plattenlösung im FibreChannel-Umfeld an.

SAN Virtualisierung

Die **Epoche der Multiplattform-Systeme** war stark durch die Einführung von FibreChannel-Netzen und den Aufbau von SANs geprägt. Damit wurde das **Thema Speichervirtualisierung im SAN** zu einem der aktuellsten Themen im Storage-Umfeld – wobei das Konzept von Virtualisierung nicht neu ist. Speichervirtualisierungskonzepte sind im Mainframe (z.B. DFSMS) oder im UNIX-Bereich in Form von 'Logischen Volume-Managern' schon lange im Einsatz. Der Einsatz von Storage Area Networks (SAN) hat die Entwicklung hin zur Speichervirtualisierung beschleunigt. Ebenso die Komplexität heutiger heterogener Infrastrukturen hinsichtlich der Server, Speichernetze und Speichersubsysteme.

Der primäre Ansatz von Speichervirtualisierung war die Entkoppelung der physischen Speicherressourcen von der direkten Zuordnung zu Serversystemen. Diese SAN-basierte Lösung legt eine Virtualisierungsebene zwischen Server- und Speichersysteme. Vorrangiges Ziel ist die gemeinsame Nutzung von Speicher quer über die gesamte Speicherhardware sowie alle Serverplattformen und Betriebssysteme. Virtualisierung im Speichernetzwerk ermöglicht es, Speicherressourcen plattformunabhängig zu integrieren, aufzurüsten, zu migrieren, zu replizieren und zu verteilen.

Um diese enorme Flexibilität innerhalb eines SANs zu bekommen, entwickelte die IBM in Hursley, UK, das **Produkt SAN Volume Controller, auch SVC** genannt, das im Juni **2003** angekündigt und im September 2003 verfügbar wurde.

Der IBM **SAN Volume Controller** wurde für einen Einsatz entwickelt, bei dem Kapazitäten mehrerer heterogener Speichersysteme zu einem einzigen Speicherreservoir, das von einem zentralen Punkt aus verwaltet werden kann, zusammengefasst werden. Er ermöglicht Änderungen an physischen Speichersystemen mit minimalen oder keinen Beeinträchtigungen für Anwendungen, die auf den Hosts ausgeführt werden, und er minimiert Ausfallzeiten durch geplante oder ungeplante Ereignisse, Wartungsmaßnahmen und Sicherungen. Zudem erhöht der SVC die Auslastung der Speicherkapazitäten, die

IBM SAN Volume Controller SVC

Onlineverfügbarkeit sowie die Produktivität und Effizienz von Administratoren. Darüber hinaus bietet er zur weiteren Vereinfachung des Betriebs die Möglichkeit, erweiterte Kopierservices systemübergreifend für Speichersysteme vieler verschiedener Anbieter einzusetzen. In seinem vierten Release wurde der SAN Volume Controller auf die Verwaltung noch größerer und verschiedenartiger Speicherumgebungen ausgelegt. Mit der jetzt erweiterten Unterstützung für zahlreiche Speichersysteme anderer Anbieter, wie zum Beispiel EMC, HP und HDS, erlaubt der SAN Volume Controller die Einrichtung einer mehrstufigen Speicherumgebung, sodass jede Datei aufgrund ihres Stellenwerts auf dem entsprechenden Subsystem abgespeichert werden kann. Die **neueste Version des SAN Volume Controller**, die im **Juli 2006** verfügbar wurde, beruht auf **4-Gbit-Technologie und ist RoHS-konform**.

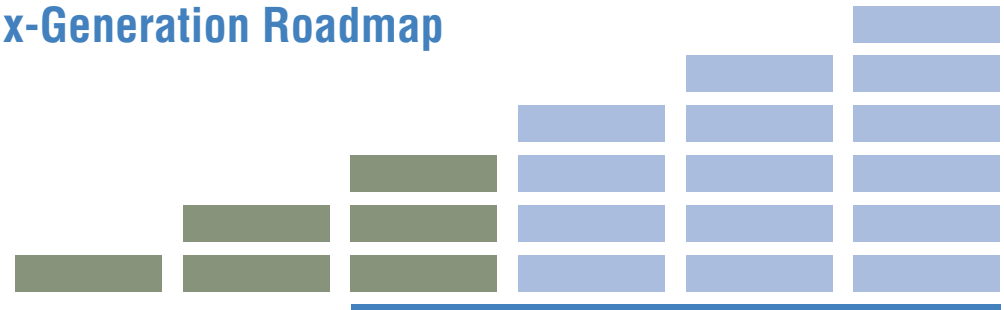
Bandsysteme

Neben der Plattensubsystem-Weiterentwicklung und den Fortschritte auf dem Gebiet von SAN-Glasfasernetzen blieb auch die Weiterentwicklung der Tape-Technologien spannend. Nach Einführung der **LTO-Bandtechnologie 2000** (Ultrium 1) mit 100 GB nativer Kapazität auf der Kassette, wurde bereits im Februar **2003** die Generation 2 (Ultrium 2) der LTO-Laufwerke mit einer Kassettenkapazität von 200 GB native eingeführt. Im Februar **2005** kam die **Drittgeneration (Ultrium 3)**, wieder mit einer Kapazitätsverdoppelung auf 400 GB native pro Kassette. Auch die Geschwindigkeit der Laufwerke wurde massiv verbessert. LTO2-Laufwerke arbeiteten bereits mit 35 MB/s mit 8-Spur-Technik. Bei **LTO3** wurde auf 16-Spur-Technik und eine Datenrate von 80 MB/s umgestellt.

Wie vorgesehen wurde die Schreib-/Lese-Rückwärtskompatibilität eingehalten. So können LTO3-Laufwerke immer noch Kassetten der Generation 1 lesemässig verarbeiten und Kassetten der Generation 2 sowohl schreiben als auch lesen.

Mit der Einführung der LTO3-Laufwerkgeneration im Februar **2005** wurden neben den bisher verwendeten überschreibbaren Kassetten auch sogenannte **WORM-Kassetten (Write Once Read Many)** eingeführt, die durch eine Mehrfachkennzeichnung als Überschreibschutz keine Möglichkeit mehr bieten, abgespeicherte Daten zu verändern oder zu löschen. Die WORM-Kennzeichnung ist auf der linken Seite der Kassette in einem eingebauten Transponder Chip hinterlegt, das mit Radiofrequenz ausgelesen wird. Die Kennzeichnung steht zudem im eingebauten Memory Chip der

Ultrium LTO Six-Generation Roadmap



	Generation 1	Generation 2	Generation 3	Generation 4	Generation 5	Generation 6
Native Capacity	100 GB	200 GB	400 GB	800 GB	1.6 TB	3.2 TB
Native Transfer Rate	up to 20 MB/s	up to 40 MB/s	up to 80 MB/s	up to 120 MB/s	up to 180 MB/s	up to 270 MB/s

Standardisierte LTO-Roadmap bis Generation 6

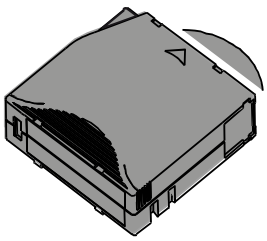
Kassette und in der Tape Volume Control Region am Bandanfang. Um absolut manipuliersicher zu sein, wird bei der Herstellung dieser WORM-Kassetten die Kennzeichnung noch auf die vorgeschriebenen Servobänder gebracht: eine Vierfach-Kennzeichnung, die sicherstellt, dass LTO3-WORM-Kassetten nicht mehr überschrieben werden können. LTO-WORM-Kassetten sind zweifarbig gekennzeichnet. Die eine Hälfte ist schwarz, und die Andere in weißer Farbe.

Mit der Verfügbarkeit von LTO3 wurde die offizielle Roadmap um zwei weitere Generationen mit LTO5 und LTO6 erweitert. Es ist davon auszugehen, dass alle 2 bis 2 ½ Jahre eine neue LTO-Generation gemäß dieser Roadmap auf dem Markt verfügbar wird.

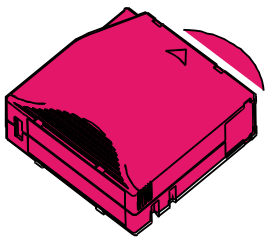
Die etwas eigenwillige Farbwahl der standardisierten LTO-Kassetten kommt aufgrund einer wichtigen Anforderung zustande. Farbenblinde müssen in der Lage sein, die Kassetten optisch zu unterscheiden.

Lag der Marktanteil 1999 mit 90 % noch auf Quantums DLT-Technologie, hat sich das Bild seither um 180 Grad gedreht und die LTO-Technologie dominiert heute mit über 80 %.

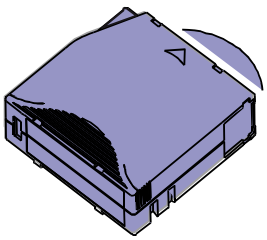
Der LTO-Massenmarkt mit billigen LTO-Laufwerken in halbhoher (half high) Bauweise wurde bisher ausschliesslich von Hewlett Packard HP abgedeckt. Mit einer Ankündigung im Oktober 2006 steigt auch IBM in dieses Segment ein und bietet **halbhohe LTO3-Laufwerke** im Niedrigpreisbereich an. Die Laufwerke sind mit einer Ultra 160 SCSI Schnittstelle ausgestattet und arbeiten mit einer Daten-Transfer-Rate von 60 MB/Sekunde (native).



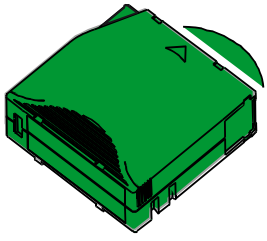
LTO1-Kassette,
100 GB Kapazität (native)



LTO2-Kassette,
200 GB Kapazität (native)



LTO3-Kassette,
400 GB Kapazität (native)

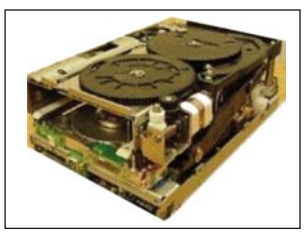


LTO4-Kassette,
800 GB Kapazität (native)

Im **April 2007** kündigte IBM mit sofortiger Verfügbarkeit als erster Hersteller **LTO4-Laufwerke und -Kassetten** an.

Das IBM TS1040 Laufwerk ist ein LTO4-Laufwerk in voller Bauhöhe (Full High) und für den Einbau in IBM Libraries vorgesehen. Das Laufwerk arbeitet mit einer Datenrate von bis zu 120 MB/s (native) und die Kassettenkapazität der neuen LTO4-Kassetten beträgt 800 GB (native).

Das Laufwerksinterface bietet einen Anschluss über einen 4 Gbit FibreChannel, SCSI oder 3 Gbps Dual Port SAS. Von der TS1120 (3592) Technologie wurde der Surface Control Guiding-Mechanismus zur optimierten Bandführung übernommen. Ebenso bietet das Laufwerk vorausschauende Fehlerbehandlung SARS (Statistical Analysis and Reporting System), einen verbesserten ECC (Error Correction Code) und Digital Speed Matching. Ein interner Pufferspeicher von 256 MB (LTO3 hatte nur 128 MB) sorgt dafür, dass der Datenstrom nicht abreißt. Insgesamt passt sich das Laufwerk an 6 Geschwindigkeiten des Servers an (30, 48, 66, 84 103 und 120 MB/s) und reduziert so die Start/Stop-Aktivitäten. Adaptive Kompression von Daten dient zur optimalen Nutzung der Bandkapazitäten. WORM-Kassetten sind neben den Standard-Kassetten verwendbar und die Laufwerke sind als erste LTO-Generation encryption-fähig. LTO4 von IBM zeichnet sich durch sehr geringen Energiebedarf mit Powermanagement-Funktion, u. a. Sleeping-Modus bei Nichtgebrauch, aus (maximaler Strombedarf 27 Watt). IBM setzt auf hohe Qualität, z. B. den Einsatz von Metall und vermeidet Plastik (weniger Störungen). Diese Merkmale führen zu einem stabileren Betrieb im Vergleich zu den alten LTO-Generationen. Es können Laufwerks-/Zugriffsstatistiken geführt und ggfs. Fehler vorhergesagt werden, um so Laufwerke präventiv auszutauschen, bevor ein Fehler auftritt. Viele Innovationen, die im Enterprise-Bandlaufwerk IBM System Storage TS1120 stecken, wurden in die IBM LTO4-Laufwerke integriert. Unter Nutzung der Verschlüsselungstechnik der TS1120 Technologie können die LTO4-Bandlaufwerke Daten komprimieren und danach verschlüsseln mit praktisch fast keiner Auswirkung auf die Leistung des Laufwerks (ca. 1%). Die Unterschiede der Verschlüsselungstechnik liegen im 'Key-Handling' und sind ausführlich im Kapitel 'Encryption' beschrieben.



IBM TS1040 LTO4-Laufwerk ohne Gehäuse

Die IBM TS1040 Laufwerke sind in den Autoloader TS3100 und in die IBM Libraries TS3200, TS3310 und TS3500 einbaubar. Neben den TS1040 Laufwerken für die Libraries kündigte IBM im **April 2007** auch **Stand Alone LTO4-Laufwerke** an, die auch in ein 19 Zoll Rack eingebaut werden können. Die Stand Alone Laufwerke werden unter dem Begriff TS2340 vermarktet und weisen dieselben Spezifikationen wie TS1040 auf. Auch hier handelt es sich um Laufwerke mit voller Bauhöhe (Full High). Die TS2340 haben wahlweise einen Anschluss über 3 Gbps Dual Port SAS oder SCSI und ein LED Display zur Administration.

LTO4-Laufwerke können LTO1-Kassetten nicht mehr verarbeiten. Sie können LTO2-Kassetten auslesen und LTO3-Kassetten im LTO3-Mode beschreiben und lesen. Mit den LTO4-Kassetten bietet ein LTO4-Laufwerk das Maximum an Kapazität (800 GB auf der Kassette) und die hohe Datenrate von bis zu 120 MB/Sekunde.

Im November 2006 kündigte IBM in der LTO4-Reihe die **IBM TS2240** als externes LTO4-Laufwerk mit halber Bauhöhe (Half High) als weitere Option zu Full High Laufwerken an. Das halbhohe LTO4-Laufwerk ist als 'Stand Alone' Laufwerk oder 'Rackmounted' erhältlich. Im Gegensatz zu LTO3 erzielen die halbhohen LTO4-Laufwerke dieselben Datenraten (120 MB/s) wie die Laufwerke mit voller Bauhöhe. Für die Produktion der halbhohen LTO4-Laufwerke nahm IBM Mitte 2006 ein neues **Fertigungswerk in Singapur** in Betrieb. Viele Elemente aus der Produktion der Laufwerke mit voller Bauhöhe wurden dort übernommen, sodass sichergestellt ist, dass die halbhohen Laufwerke in Qualität und Sicherheit einem extrem hohen Maßstab gerecht werden.

Im **Februar 2008** gab IBM bekannt, dass die **halbhohen LTO4-Laufwerke** im **TS3100 Autoloader** und in der **TS3200 Library** integriert werden können. Anstelle von einem ganz hohen Laufwerk können zwei halbhohe Laufwerke eingebaut werden. Damit fasst der Autoloader TS3100 bis zu zwei halbhohe LTO4-Laufwerke und die Library TS3200 bis zu vier. In der TS3200 ist auch der Mischbetrieb von ganz hohen und halbhohen Laufwerken unterstützt (ein ganz hohes und zwei halbhohe Laufwerke).

Die Epoche der Multiplattform-Systeme und des FibreChannel SAN und NAS

Die **Magstar-3590**-Bandentwicklung als 1/2-Zoll-Format lief parallel zur LTO-Entwicklung, allerdings wurde dabei das ausgereifte Aufzeichnungsverfahren mit Parity-Informationen beibehalten. Die Magstar-Entwicklungsreihe fand im **Juni 2002** mit der **Generation 3, den 3590-H Modellen**, ihren Abschluss. Dabei wurde die Lese-Rückwärtskompatibilität der vorangegangenen Generation, also für Kassetten, die mit dem B- oder E-Modell beschrieben wurden, sichergestellt.

Mit dem neuen H-Modell schrieb man 384 Spuren in noch verdichteterer Form auf die Kassette. Für die Extended Length Cartridge bedeutete das eine Kapazität von 60 GB unkomprimiert und 180 GB komprimiert (3 : 1). Die Datenrate von 14 MB/s wurde beibehalten. Damit erreichte Magstar im Jahr 2002 dieselbe Spurdichte wie die LTO-Generation 1 im Jahr 2000.

Mit dem von IBM im **September 2003** neu angekündigten **Kompaktlaufwerk 3592** wurde ein **neues 'Tape-Zeitalter'** eingeleitet, dessen Dimension nur wenigen bis heute bewusst ist! Schaut man aber genauer auf die integrierte Technologie und auf die neue Beschichtungsart der 3592 Kassetten, kann man diesen Entwicklungsschritt mit 1984 vergleichen. 1984 leitete die IBM den Wechsel vom Rollenband auf die Kassettentechnologie ein. Mit der 3592 Technologie, die unter dem Entwicklungsnamen 'Jaguar' entwickelt wurde, ergaben sich technologische Möglichkeiten für Bandkassetten und den Einsatz von Bandlaufwerken, von denen bis dahin nur geträumt werden konnte.

Das Laufwerk selbst hat neben der neuen **Flat-Lap-Kopf-Technik**, dem neuen, angepassten Guiding System mit Roller Bearing und dem zeitgesteuertem Spurnachführungssystem mit Servobändern (alle drei Elemente sind auch im LTO2- und LTO3-Laufwerk integriert) **erstmals PRML Encoding** (Partial Response Maximum Likehood – siehe auch Technologie-Anhang) implementiert, das eine Bit-Abbildung von 1 : 1 auf Band erlaubt.

Das **PRML Encoding** hatte auf der Platte bereits 1995 Einzug gehalten, aber es war bisher nicht möglich, dieses Verfahren auf Band einzusetzen. Grundvoraussetzung für den Einsatz von PRML ist die Erzeugung von extrem gut durchmagnetisierten Bits auf dem Datenträger. Die Beschichtung der 3592 Kassette im Zusammenhang mit den Flat-Lap-Köpfen mit 10-fach höherer Induktionsstärke macht es möglich, solch hochqualitative Bits zu erzeugen. Die 3592 Kassette ist auch das erste Medium, bei dem mit starker Induktion gearbeitet werden kann, ohne negative Erscheinungen wie z. B. weniger Kapazität zu bekommen. Damit kann man ca. 50 % mehr in der Datenspur unterbringen und so 50 % mehr an Kapazität auf Kassetten mit der gleichen Bandlänge und der gleichen Spurzahl realisieren.

Die Flat-Lap-Kopf-Technik verbessert das Schreib-/Lesesignal zwischen Kopf und Band, weil eine wesentlich höhere Induktion ermöglicht wird. Vor und hinter der Kopfreihe wird zusätzlich ein Unterdruck erzeugt, der Staubpartikel, die sich – aus welchen Gründen auch immer – ansammeln, entsprechend absaugt (**'Self Cleaning Machines'**). Diese Kopftechnologie kam erstmals in vollem Umfang bei LTO2 zur Anwendung und wurde in weitergehender Form im neuen 3592 Laufwerk implementiert. Dadurch ist die Qualität der erzeugten Bits wesentlich höher und die Möglichkeit, die Daten wieder zu lesen, ist ziemlich unübertroffen in der Industrie. Ebenso wird dadurch sichergestellt, dass selbst bei Langzeitlagerung von Bändern nahezu kein Impulsverlust auftritt.

Das 'Surface Control Guiding System' ermöglicht eine ganz exakte Führung der Schreib-/Leseelemente über Servobänder. Die Repositionierung der Elemente auf den Servospuren erfolgt über eine Zeitsteuerung, die über vorgeschriebene Analogsignale (Analogspuren) in den Servobändern und eine Zeitmessung durchgeführt wird. Dadurch wird die Oberfläche des Bandes für die Feinführung genutzt und nicht der Randbereich des Bandes, wo bei klassischen Aufzeichnungen Servospuren aufgebracht waren. Damit lassen sich die Fehlerquellen vermeiden, die aufgrund der Bandspannung im Außenbereich auftreten können. Flat-Lap-Technik und das neue Guiding System wurden erstmals in den Produkten LTO2 und 3592 implementiert. PRML Encoding ist nur in der 3592 realisiert, weil die neue Beschichtung der 3592 Kassette Voraussetzung für dieses neue Encoding-Verfahren ist.

Mit 40-MB/s-Datenrate und unglaublich schnellen Spul- und Fast-Search-Zeiten entpuppte sich das IBM 3592 Laufwerk in dieser Zeit als das schnellste und zuverlässigste Laufwerk auf dem Weltmarkt. Eine weitere Besonderheit der Jaguar-Technologie sind neue, funktionale Möglichkeiten, die sich aufgrund der 3592 Beschichtung implementieren ließen, weil ein 3592 Medium keine Limitierungen in der Benutzerhäufigkeit hat. Mit der Funktion **'Virtual Back Hitch'** können viele Rücksetz-Zeiten beim File Retrieval und alle Start-/Stopp-Positionierungszeiten bei einzelnen File-Transfers, die nicht sequenziell verarbeitet werden können, auf breiter Basis eliminiert werden.

Klassisch arbeitet ein Laufwerk so, dass immer über den Pufferspeicher auf Band geschrieben wird. Das Laufwerk bekommt also Daten in den Pufferspeicher und über den sogenannten Flush-Buffer-Befehl die Instruktion, die bisher erhaltenen Daten auf Band rauszuschreiben. Ist dies geschehen, wird der Flush-Buffer-Befehl als Tapemark in der Spur hinterlegt. Sind keine zu schreibenden Daten mehr vorhanden, stoppt das Laufwerk. Kommt wieder etwas in den Pufferspeicher mit dem entsprechenden Flush-Buffer-Befehl, kann der Befehl nicht sofort ausgeführt werden, weil das Laufwerk ja gestoppt hat. Das Laufwerk muss jetzt zurücksetzen, um an dem Tapemark, der zuletzt geschrieben wurde, in der Streaminggeschwindigkeit zu sein, in der der Flush Buffer Befehl durchgeführt werden kann. Dieses Zurücksetzen wird als Backhitch bezeichnet. Ein Backhitch ist also nichts 'Gutes', da er Zeit benötigt und das Band zusätzlich (vor allem im Außenbereich) stresst. Mit Virtual Backhitch werden in der 3592 solche Positionierungsvorgänge auf den minimalsten Level eingegrenzt!

Wie funktioniert nun Virtual Backhitch: Stellt die Control Unit aufgrund der Pufferspeicherauslastung fest, dass nicht mehr im Streamingmodus gearbeitet werden kann, schreibt das Laufwerk in die momentane Spur eine sogenannte RABF-Marke. RABF steht für Recursive Accumulative Backhitchless Flush. Die RABF-Marke verweist einfach in ein Spurset, das vor einem liegt und eigentlich zum Schreiben noch gar nicht vorgesehen ist. In diesem neuen Spurset streamt das Laufwerk einfach weiter, d. h., tröpfelt etwas in den Pufferspeicher rein, wird zeitgleich eine Kopie auf Band in dem Spurset erzeugt, das noch nicht zum Schreiben vorgesehen ist. Da einfach weitergestreamt wird, unabhängig davon, ob Daten kommen oder nicht, können natürlich regelrechte 'Löcher' zwischen den tatsächlich weggeschriebenen Daten entstehen. Den ganzen Vorgang könnte man auch als **'Nonvolatile Caching auf Tape'** bezeichnen, da parallel auf Band und in den Pufferspeicher geschrieben wird. Das Band stellt den stromunabhängigen NVS-Schreibspeicher dar und der Pufferspeicher den Cache. Geht der Cache aus irgendwelchen Gründen kaputt, kann jedes Laufwerk trotzdem die auf Band geschriebenen Daten verarbeiten, da der Hinweis auf das Spurset über die RABF-Marke sichergestellt ist.

Ist der Pufferspeicher nun zu 50 % voll oder kommt das Laufwerk ans Bandende, dreht das Laufwerk um und schreibt zurück. Ist der Pufferspeicher voll, wird der gesamte Pufferspeicher nun sequenziell in das richtige Datenspurset herausgeschrieben. Danach wird der Verweis der RABF-Marke aufgelöst, so, als ob nichts geschehen wäre. Virtual Backhitch oder Nonvolatile Caching auf Tape ist eine geniale Tape-Funktionalität, weil Repositionierungszeiten eingespart werden und zusätzlicher Bandstress vermieden wird.

Das **3592 Laufwerk** selbst hat **noch nie dagewesene Schock- und Vibrationseigenschaften** und kann selbst bei hohen Erschütterungen noch sauber schreiben und lesen. Dies wurde durch eine selbstkalibrierende Aufhängung des Laufwerks im Gehäuse realisiert.

Die Epoche der Multiplattform-Systeme und des FibreChannel SAN und NAS

Im **November 2005** wurde die zweite Generation der Jaguar-Laufwerke verfügbar. Die **3592 Generation 2**, die Marketing-seitig als TS1120-Laufwerk (TS steht für Tape System) bezeichnet wird, war und ist das erste Bandlaufwerk, das mit **zwei 4-Gbit-FibreChannel-Schnittstellen** angesteuert wird und dadurch auf komprimierter Basis eine Datenrate von 260 MB/s realisiert. Die bisher verwendete 8-Spur-Technik wurde auf 16-Spur-Technik umgestellt und die Induktion der Köpfe wurde nochmals verdoppelt, sodass noch 'bessere' Bits auf dasselbe Medium geschrieben werden können. Die mit kurzer Bandlänge gestaltete 60-GB-Kassette wird beim Beschreiben mit der zweiten Generation zur 100-GB-Kassette und die 300-GB-Kassette wird zur 500-GB-Kassette. Jaguar 2 schreibt mit einer unglaublichen Datenrate von 100 MB/s 896 Spuren auf die 3592 Kassetten. Aufgrund der extrem guten erzeugten Streufelder von Bits und Tapemarks konnte die Streaming-Geschwindigkeit beim High Speed Search auf 10 ms erhöht werden. Das heißt, das Band spult bei einer Fileanforderung mit 36 km in der Stunde vor. Diese hohe Geschwindigkeit ist einmalig im Markt und erlaubt trotz hochkapazitiver Kassetten durchschnittliche Zugriffszeiten von 27 Sekunden (100-GB-Kassette) und 46 Sekunden (500-GB-Kassette). Auch der Stromverbrauch konnte auf 46 Watt reduziert werden. Vergleichbare Laufwerke benötigen heute das Doppelte an Strom.

Auch die einmalige Funktion des Virtual Backhitch wurde durch die Verwendung eines 512 MB großen Pufferspeichers (Generation 1 hatte 128 MB Pufferspeicher) und durch zwei eingebaute Control Units weiter optimiert.

Im Vergleich zu LTO bietet die Jaguar-Technologie neben überschreibbaren Kassetten auch WORM(Write Once Read Many)-Kassetten an. **3592 WORM-Kassetten** sind hochpreisiger im Vergleich zu LTO3-WORM-Kassetten, bieten aber neben allen WORM-Sicherheiten den Vorteil, dass das **Fortschreiben immer möglich** ist, auch wenn in dem Bereich, wo fortgeschrieben werden muss, fehlerhafte Spuren oder Spurbereiche vorkommen sollten. Dies wird dadurch sichergestellt, dass bei 3592 Kassetten auf den Servobändern nicht nur die WORM-Kennzeichnung aufgebracht wird, sondern zusätzlich sogenannte **Appendflags**, die für die Fehlerkorrektur herangezogen werden. LTO3-WORM-Kassetten arbeiten ohne Appendflags, d. h., gibt es beim Fortschreiben der Kassette fehlerhafte Bereiche, dann lässt sich die LTO-WORM-Kassette einfach nicht mehr weiterschreiben.

Ende Oktober 2006 kündigte IBM eine **neue 700 GB Kassette** für die TS1120 Laufwerke an. Damit stehen jetzt drei Kassetten (überschreibbar und als WORM) zur Verfügung: Die 100 GB Kassette mit 120 Meter Bandlänge, die 500 GB Kassette mit 609 Meter Bandlänge und die neue 700 GB Kassette mit 825 Meter Bandlänge.

Mit der Markteinführung von LTO im Jahr **2000** kündigte IBM ein neues Bandarchiv, die IBM 3584, für den unternehmensweiten Einsatz im Open-Systems-Bereich an. Im Laufe der Folgejahre wurde die 3584 Library maßgeblich erweitert. Heute stellt sie die **strategische Library-Plattform der IBM** dar. Seit **Juni 2005** wird neben den Open-Systems-Plattformen auch die **zSeries** (Mainframe) als Serverplattform unterstützt. Sie machen aus der 3584 ein Bandarchiv, das unternehmensweit für alle Rechnerplattformen eingesetzt werden kann und das aufgrund der hohen Leistungsfähigkeit und Flexibilität alle vergleichbaren Archivprodukte auf dem Markt in den Schatten stellt.

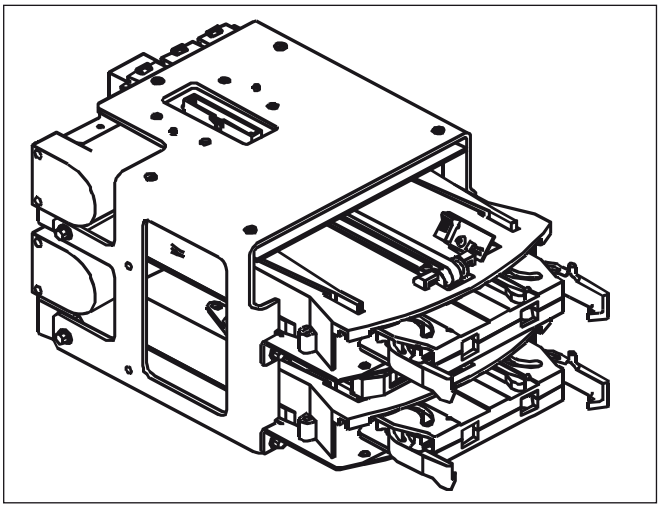
IBM TS1120(3592 Generation 2)-Bandlaufwerk der Spitzenklasse

IBM 3592 WORM-Kassetten in platingrauer Farbe und WORM-Labels

IBM 3584 Grundeinheit mit Ein-/Ausgabestation

Auf der CeBIT in Hannover **sorgte die 3584 immer wieder für viele Menschentrauben**. Viele können es nicht fassen, dass es technisch möglich ist, einen Roboter mit dieser wahnsinnig schnellen Geschwindigkeit zu betreiben. Bei **einer Roboter-Service-Zeit von unter 3 Sekunden** in einer 2-Frame-Konfiguration (Kassette holen, in ein Bandlaufwerk laden, Kassette aus einem anderen Bandlaufwerk entladen und ins Fach zurückbringen) ist es fast nicht möglich, die bewegten Kassetten mit dem bloßen Auge zu sichten. Schneller geht es wirklich nicht mehr!

Diese Geschwindigkeit ist möglich, weil sich **ein neu entwickelter Greifer (IBM Patent)** in die LTO- und 3592 Kassetten in dort angebrachte Einkerbungen einhakt, die Kassetten in einen Schacht zieht und der Greifer solange verhakt bleibt, bis die Kassetten in ein Laufwerk oder an einen Stellplatz zurückgebracht werden. Die Kassette kann also auch bei höchsten Geschwindigkeiten oder schnellsten Drehungen nicht verloren gehen. Alle Roboterbewegungen sind Servo-kontrolliert, um diesen Geschwindigkeiten Rechnung zu tragen.



IBM 3584 Doppelgreifer

In dem 3584 Archiv können sowohl LTO-Laufwerke und -Kassetten als auch 3592 Laufwerke und Kassetten betrieben werden. Ebenso ist der Mischbetrieb in getrennten Frames möglich.

Wie schon bei der älteren 3494 Library früher eingeführt, wurde das System im **Frühjahr 2005** so erweitert, dass es mit **zwei Robotersystemen** betrieben werden kann, wobei der zweite Roboter immer im Active Mode betrieben wird. Dabei werden Zahlen von weit über 1000 Mounts in der Stunde erzielt. Des Weiteren kann die 3584 mit zwei Robotersystemen nahezu unterbrechungsfrei mit Frames erweitert werden (max. 60 Sekunden, wobei kein Roboter Command verloren gehen kann).

Seit Mai 2005 kann das **3584 Archiv auch an zSeries-Server** angebunden werden. Dies gilt ausschließlich für den Betrieb von 3592 Laufwerken und wurde dadurch realisiert, dass an der 3584 **jetzt mehrere Library Manager 3953 L05 und J70 ESCON/FICON Controller** betrieben werden können. Im Gegensatz zur 3494 Library sind die Library Manager und J70 Controller nicht mehr in die Library Frames integriert, sondern werden in externen Frames 3953 F05 installiert und an die Library angeschlossen. Dies bietet eine wesentlich höhere Konfigurationsflexibilität. Die 3584 unterstützt bis zu vier Library Manager und damit die Anbindung von bis zu acht Virtual-Tape-Servern (3494-B10/B20).

Bei VTS-Spiegelungen (Peer to Peer VTS) kann jetzt auf der primären Seite eine 3494 Library und auf der sekundären Seite eine 3584 Library oder umgekehrt stehen, d. h., im VTS-Peer-to-Peer-Spiegel-Betrieb ist der Mischbetrieb beider Libraries möglich.

Für den Betrieb von Native-Laufwerken an zSeries-Systemen gibt es fast keine Limitierungen, da in der 3584 bis zu 64 x J70-Controller mit bis zu 192 Laufwerken konfiguriert werden können.

Das 3584 Archiv bietet in jeglicher Hinsicht maßgebliche Vorteile im Vergleich zu herkömmlichen Libraries. Eine **Multipfad-Architektur** mit direkter Anbindung von LTO- und 3592 FibreChannel-Laufwerken ermöglicht das logische Partitionieren der Library (bis zu 192 Partitionen), Ebenso stehen die Optionen **'Control Path Failover' und 'Data Path Failover'** zur Verfügung. Mit der Funktion **ALMS (Advanced Library Management System)** können die logischen Partitionen dynamisch vergrößert und verkleinert werden, unabhängig davon, wo die Laufwerke oder die Kassetten in der Library untergebracht sind.

Die **WWN(World Wide Name)**-Adresse ist den Laufwerkschlitten zugeordnet und erlaubt einen Laufwerkaustausch, ohne dass neu 'gebootet' werden muss.

Mit der Anbindung der 3584 an zSeries-Server signalisierte IBM klar, dass die 3584 mit ihrem einzigartigen Robotersystem die strategische Library-Plattform für zukünftige Weiterentwicklungen darstellt. Von dieser Weiterentwicklung profitiert aber auch die 3494 Library, die alle Erweiterungen erhält, die am Library Manager entwickelt werden. Ebenso wird die 3494 neben der 3584 alle geplanten neuen Laufwerkgenerationen des 3592 Laufwerks integrieren (SOD vom Mai 2005).

Im **Mai 2006** wurde die **IBM 3584** im Zuge der Frameumstellungen auf **4-Gbit-FibreChannel-Technologie in TS3500 (TS steht für Tape System)** umbenannt und es stehen entsprechende neue Frames zur Verfügung. Die Library IBM 3584 bzw. TS3500 wird von IBM selbst produziert. Mit der Umstellung auf 4-Gbit-Technologie wurde auch die **RoHS-Konformität** des TS3500 Library-Produktes bekanntgegeben. Ebenso wurden die bis dahin verwendeten **J70-Steuereinheiten**, die für die Anbindung von TS1120 Laufwerken an den Mainframe benötigt wurden, durch neue RoHS-konforme leistungsstärkere **C06-Steuereinheiten** mit 4-Gbit-FICON-Anschlüssen ersetzt. RoHS steht für 'Restriction of Hazardous Substances' und stellt ab dem 1. Juli 2006 eine gesetzlich kontrollierte EU-Richtlinie dar, die genau vorschreibt, welche Materialien, Legierungen, Barium-Derivate etc. in Neuprodukten eingesetzt werden dürfen. Diese neue Richtlinie gilt EU-weit und betrifft alle Neuprodukte in der Elektro- und IT-Industrie, die ab dem 1. Juli 2006 neu vertrieben werden.

Im mittleren und unteren Systemsegment produzierte IBM keine eigenen Libraries, sondern kaufte die 'nackten' Libraries über OEM-Verträge ein, um sie dann mit IBM LTO-Laufwerken auszustatten.

IBM 3583 Modelle L18, L36 und L72, Kapazität mit LTO1-Kassetten bis 7.2 TB (native), ADIC vertrieb unter ADIC Logo das Produkt als Scalar 100

Mit der Ankündigung der ersten LTO-Laufwerke im Jahr **2000** wurden auch **OEM Libraries** und Autoloader angekündigt, die nach dem Umbau/Ausbau als IBM Logo-Produkt weitervertrieben wurden. So wurden **1999** ein Autoloader der Firma ADIC **als IBM 3581** Produkt und ein mittleres Archivsystem auch von ADIC **als IBM 3583** eingeführt. Der Autoloader war mit einem LTO-SCSI-Laufwerk (LVD oder HVD) ausgestattet und konnte bis zu 8 LTO-Kassetten verwalten. Das Archivsystem 3583 konnte mit 6 LTO-Laufwerken ausgestattet werden und je nach Modell 18 Kassetten, 36 und 72 Kassetten verwalten. Am Anfang wurden ausschließlich SCSI-Laufwerke eingebaut. Um die Library an ein FibreChannel-Netz anschließen zu können, wurden über ein eingebautes FC-AL Gateway die bis zu sechs SCSI-Laufwerke 'daisy chained' angeschlossen.

Diese FibreChannel-Anschlussweise der IBM 3583 verursachte bei bestimmten Host-FibreChannel-Adaptern Adressierungsprobleme. Deshalb wurde im Mai 2003 in der 3583 ein eigener IBM Controller eingebaut, der durch eine Multipfad-Architektur erlaubte, direkt FC-Laufwerke in der Library zu betreiben und die Library zu partitionieren. Bis zu drei logische Partitionen waren in der 3583 möglich.

Im **Mai 2004** wurde dieses Midrange-Library-Portfolio ergänzt durch eine Mini-Library, die **IBM 3582**. Auch ist ein OEM-Produkt der Firma ADIC, sie ist ausgestattet mit dem IBM Controller für die Anbindung von FC-LTO-Laufwerken und bietet die Möglichkeit, sie mit zwei logischen Partitionen zu betreiben.

In die Mini-Library wurden 1 bis 2 LTO-Laufwerke eingebaut und es konnten bis zu 23 Kassetten verwaltet werden. Neun Kassetten hatten fest eingebaute Stellplätze, die anderen 14 Stellplätze wurden über zwei herausnehmbare 7er-Kassettenmagazine abgedeckt.

Im **Mai 2006** wurde der OEM-Autoloader von ADIC durch einen neuen Autoloader der **Firma BDT IBM 3581** ersetzt, der auch bis zu 8 Kassetten verwalten konnte, aber in Flachbauweise konstruiert war und dadurch sehr platzsparend in ein Standard-19-Zoll-Rack integriert werden konnte.

Auch bei der Midrange-Library-Reihe musste bis 30. Juni 2006 die Umstellung auf die neue, ab 1. Juli 2006 geltende EU-RoHS-Richtlinie durchgeführt werden.

Die Library 3583 wurde bereits im **Oktober 2005** durch eine **neue Library TS3310** ersetzt. Die TS3310 stammt auch aus dem Hause ADIC, hatte vorbereitete 4-Gbit-Technologie integriert und war zum Verfügbarkeitszeitraum **RoHS-konform**. Am Anfang konnte die Library nur mit der Grundeinheit, einer Erweiterungseinheit und bis zu sechs LTO3-Laufwerken und 122 Kassettenstellplätzen konfiguriert werden, im **Mai 2006** wurden dann die Erweiterungen bis zum Maximalausbau mit bis zu vier Erweiterungseinheiten und damit bis zu 18 LTO3-Laufwerken und bis zu 398 Kassettenstellplätzen bekannt gegeben. IBM verwendet bei dieser neuen Library wiederum einen eigenen FC-Controller für die Anbindung der Laufwerke. Damit ist die Library partitionierbar und es können bis zu 18 logische Partitionen betrieben werden. Die Zuordnung von Laufwerken und Stellplätzen auf logische Partitionen ist wesentlich flexibler als beim Vorgänger 3583.

IBM 3581 Modelle L28 und F28, Autoloader der Firma BDT

IBM 3582 Mini-Library mit bis zu 23 LTO-Kassetten.
ADIC vertrieb unter ADIC Logo das Produkt als Scalar 24

Die neue IBM Midrange Library IBM TS3310 mit bis zu 18 LTO3-Laufwerken und maximal 398 Kassettenstellplätzen, hier mit einem Erweiterungsframe. ADIC vertreibt unter ADIC Logo das Produkt als ADIC 500i

Im Zuge der gesamten RoHS-Umstellung wurde sowohl der Autoloader von BDT IBM 3581 als auch die Mini-Library von ADIC IBM 3582 durch neue Produkte der Firma BDT ersetzt. Die 3581 wurde durch die **IBM TS3100** ersetzt, ein neuer Autoloader mit einem 4-Gbit-Fibre-LTO3-Laufwerk und bis zu 22 Kassettenstellplätzen. Die Mini-Library 3582 wurde durch die **IBM TS3200** ersetzt, die bis zu zwei **4-Gbit-LTO3**-Laufwerke und bis zu 44 Kassetten aufnehmen kann. Die TS3200 besitzt dieselbe Partitioniermöglichkeit wie ihr Vorgänger.

Damit sind alle Library-Produkte und -Bandlaufwerke, die IBM vertreibt, **durchgängig auf 4-Gbit-Technologie** umgestellt und entsprechen **der EU-RoHS**-Richtlinie, die am 1. Juli 2006 für Neuprodukte in Kraft getreten ist.

Tape-Virtualisierung im Open-Systems-Umfeld

Tape-Virtualisierung war bisher nur im Mainframe-Umfeld sinnvoll (siehe VTS Virtual-Tape-Server). Mit der Zunahme der Geschwindigkeit und der Datenrate, wurde es immer schwieriger, die Geschwindigkeit solcher Laufwerke auch maximal auszunutzen. Ein LTO3-Laufwerk arbeitet heute mit 80 MB/s, ein Jaguar-Laufwerk TS1120 sogar mit 100 MB/s (native). Um diese hohen Laufwerkgeschwindigkeiten auch maximal nutzen zu können, werden zunehmend Plattenpuffer als Zwischenspeicher eingesetzt. Um eine geeignete, sinnvolle Backup-Konzeption zu etablieren, muss zwischen dem klassischen Backup über das IP-Netz (**LAN-Backup**) und dem **LAN-free-Backup**, wo die Daten über ein Fibre-Channel-Netz (SAN) auf die Bandlaufwerke übertragen werden, unterschieden werden.

Wird der Backup klassisch über das LAN gefahren, ist die sinnvollste Art einer Virtualisierung, den IBM **TSM (Tivoli Storage Manager)** einzusetzen, über den TSM-Server auf einen Plattenpuffer zu sichern und unter Kontrolle des TSM vom Plattenpuffer auf Band zu migrieren. Diese Funktionalität hat der TSM schon seit vielen Jahren. Auf diese Weise können die hohen Geschwindigkeiten der Laufwerke genutzt werden. Der Vorteil dieser Lösung liegt in der Automation, weil man alle Policy-Regeln des TSM entsprechend nutzen kann.

Im LAN-free-Bereich macht eine Virtualisierung dann Sinn, wenn viele LAN-free-Clients betrieben werden, denn jeder LAN-free-Client benötigt ein dediziertes Bandlaufwerk. Betreibt ein Rechenzentrum den Backup mit vielen LAN-free-Clients, benötigt man dieselbe Anzahl an physikalischen Laufwerken, und das kann teuer werden. Hier machen virtuelle Bandarchive Sinn, weil man viele virtuelle Laufwerke zur Verfügung hat. Um diesem Umstand gerecht zu werden, kündigte IBM im **Oktober 2005** das **virtuelle Bandsystem TS7510** für Open-Systems-Umgebungen an.

IBM TS3100

IBM TS3200

Bandvirtualisierung für Open Systems (Virtual Tape Library VTL)
Das virtuelle Bandarchiv **TS7510** besteht aus mehreren wichtigen Komponenten und ist Server-basierend aufgebaut. Als Server werden xSeries-Rechner mit einem Linux Kernel eingesetzt. Die Maschine kann mit einer (Single) oder zwei xSeries (Dual-Server-Konfiguration) ausgestattet werden. Die Dual-Server-Konfiguration bietet die Möglichkeit des 'Active Failover und Failback'. Die eingesetzten Server stellen die virtuelle Einheit dar und emulieren virtuelle Bandlaufwerke. Bis zu 512 virtuelle Laufwerke werden pro Rechnereinheit emuliert. Bei einer Dual-Server-Konfiguration werden bis zu 1024 virtuelle Laufwerke emuliert. Abgebildet werden LTO2-, LTO3- und 3592(Jaguar 1)-Laufwerke. Es stehen bis zu 128 virtuelle Libraries und bis zu 8192 virtuelle Volumes zur Verfügung, also pro Rechnereinheit 64 virtuelle Libraries und 4096 virtuelle Volumes.

Geschrieben wird in einen Plattenpufferspeicher. Hierzu sind in dem Gehäuse zwei Platten-Controller und entsprechende Plattenerweiterungseinschübe auf Basis der DS4000 eingebaut. Das erste Frame wird mit einem zweiten Frame erweitert, wenn größere Kapazitäten notwendig werden. Die Minimalkonfiguration erlaubt 5 TB nutzbaren Plattenplatz und kann auf die maximale Konfiguration von bis zu 46 TB ausgebaut werden. Insgesamt stehen acht 2-Gbit-Ports zur Verfügung, vier für den Plattenzugriff und vier für den Host und die physikalischen Bandlaufwerke. Spiegelungen der Maschine können über zwei 1-Gbit-Ethernet-Ports über das IP-Netz durchgeführt werden (Remote-Replikation). Dabei kann mit Compression und Encryption als Zusatzoptionen gearbeitet werden.

Vituelle Tape Library IBM TS7510

Archivierungslösungen

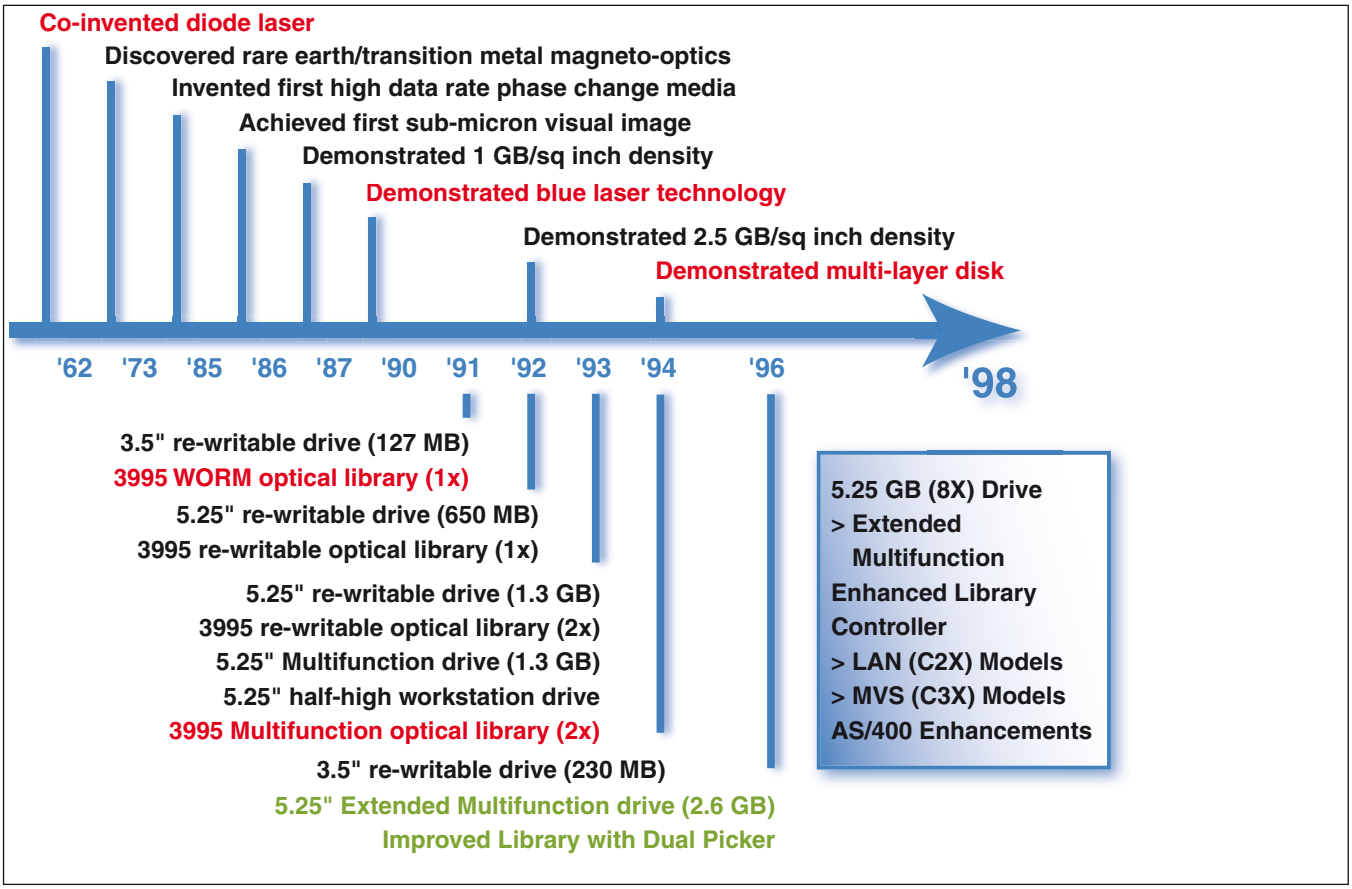
Archivierung von Informationen ist schon ein sehr altes Thema, denkt man an die Höhlenmalereien der Grotte Chauvet im Vallon Pont d'Arc in Südfrankreich, die auf ca. 31.000 Jahre geschätzt werden – die wohl erste Überliefung von Informationen aus Menschenhand.

Die Techniken, der Nachwelt Informationen zukommen zu lassen, haben sich seitdem stets verbessert. Zunächst wurden Information in Stein gemeißelt und in Ton gebrannt, bis die Ägypter ca. 4000 Jahre v. u. Z. das Papyrus – den Vorgänger der heutigen Papiers – entdeckten und dazu auch eine einfache Technik, Informationen darauf festzuhalten. Mit fortschreitender Technik war es nun auch möglich, mehr Informationen zu speichern und der Nachwelt zu hinterlassen. Denken wir nur an das Alte Testament, das einen Zeitraum von ca. 2000 Jahren vor Christi Geburt umfasst und aus insgesamt 12 Büchern besteht.

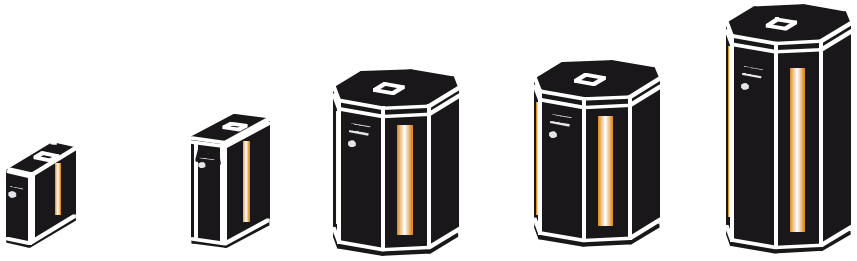
Mit der Erfindung des Buckdruckes durch Johannes Gutenberg aus Mainz, Mitte des 15. Jahrhunderts, war es dann auch möglich, Informationen einfach zu vervielfältigen, was zu vielen interessanten Überlieferungen beigetragen hat und natürlich auch zum Wachstum an Informationen.

Bereits **1982** gründeten IBM und Sony eine Entwicklungsallianz mit dem Ziel, **optische Technologien** gemeinsam weiterzuentwickeln. So enstanden im IT-Umfeld Anfang der neunziger Jahre optische Archivierungsmöglichkeiten, sogenannte 'Juke'-Boxen mit optischen Platten und entsprechenden Schreib-/Lesegeräten, die sich in dieser Zeit auch als Archivierungseinheiten zur Langzeitarchivierung durchsetzten. Es etablierten sich drei unterschiedliche optische Medien, **die WORM-Platte (Write Once Read Many)**, die **magneto-optische Platte MO** und die 'Schein-WORM'-Platte, die sogenannte **CCW(Continuous Composite WORM)-Platte**, eine magneto-optische Platte, die bei der Herstellung eine Kennzeichnung bekommt, die sicherstellt, dass das Medium nicht versehentlich überschrieben wird.

Bereits **1992** begann man die für die IT-Branche entwickelten Formate WORM, MO und CCW, basierend auf der Rote-Laser-Technik, zu standardisieren. Ein ISO-Standardisierungsgremium bildete sich und selbst der deutsche DIN bildete den NI23-Arbeitskreis, einen Ausschuss des deutschen DIN zur Normierung von optischen Datenträgern, der der internationalen ISO direkt zuarbeitete und seine entsprechende Position abgab. Die als 1 x Standard verabschiedete ISO-Norm reflektierte eine 650-GB-Platte in allen drei Formaten, mit dem



Entwicklung der optischen Technologien in den neunziger Jahren



MODEL	CX0	CX2	CX4	CX6	CX8
Cartridges	20	52	104	156	258
Drives	1 or 2	2	2 or 4	4 or 6	4 or 6
Capacity	104 GB	270 GB	540 GB	810 GB	1.34 TB
Attachments: X = 6 X = 2 X = 4 X = 3 X = 1 (Expansion)	 SCSI LAN AS/400 -- --	 SCSI LAN AS/400 S/390 S/390	 SCSI LAN AS/400 S/390 --	 SCSI LAN AS/400 S/390 S/390	 SCSI LAN AS/400 S/390 S/390
Model Upgrades	--	--	From	To	--

IBM 3995 optisches Archivierungssystem mit den einzelnen Modellvarianten mit 8 x Standard

2 x Standard kam die 1.3-GB-, mit dem 4 x Standard die 2.6-GB-Platte. 1998 wurde der letzte Standard als 8 x Standard verabschiedet, der immer noch auf dem roten Laser basierte. Der 8 x Standard bot auf allen drei Formaten 5.2 GB Kapazität pro Platte.

Um Rückwärtskompatibilität zu gewährleisten, war der Standard so gestaltet, dass Medien des 1 x Standards sowohl lese- als auch schreibmäßig von Schreib-/Lesegeräten des 2 x Standards verarbeitet werden konnten und selbst der 4 x Standard noch in der Lage war, Medien des 1 x Standards zu lesen. Mit dem 8 x Standard kam der Bruch, da nicht mehr vorgesehen war, Medien des 1 x Standards auf Schreib-/Lesegeräten des 8 x Standards verarbeiten zu können.

Mitte der neunziger Jahre wurde die WORM-Platte – vom Gesetzgeber anerkannt für viele Archivierungsanforderungen – mit einer Haltbarkeit von 300 Jahren spezifiziert. Viele Unternehmen gingen dazu über, Langzeitarchivierung auf diesen optischen Juke-Boxen zu betreiben. IBM bot auf diesem Gebiet das **IBM 3995 optische Archivsystem** an.

Nach der Verabschiedung des 8 x Standards passierte viele Jahre nichts mehr bezüglich einer sinnvollen Standardisierung. Dies lag darin begründet, dass ein 16 x Standard technisch nicht realisierbar war, da der rote Laser in diesem Bereich eine Streuung zeigte, die es nicht zuließ, auf einem nach 8 x Standard erzeugten Pit durch Optimierung der Laserfrequenz 4 Pits unterzubringen. Auch trug der technologische Wechsel zur Blaue-Laser-Technik, (auch 'Blue Ray' genannt) dazu bei, dass ein noch realisierbarer Standard in Rote-Laser-Technik als 12 x Standard nicht weiterverfolgt wurde.

Hier noch ein wichtiger Hinweis, der zeigt, wie tief IBM in der Entwicklung von optischen Technologien engagiert war. Der heutige **blaue Laser ist ein IBM Patent**, das aus einer gemeinsamen Entwicklung von IBM und Sony hervorging.

Im **Jahr 2000** ergab sich in Deutschland eine massive Änderung für die Langzeitarchivierung. Die **GDPdU**-Richtlinie wurde verabschiedet und im Jahr **2001 nochmals genauer verifiziert**. Diese Richtlinie nahm Abschied von der Vorgabe, bei bestimmten aufzubewahrenden Daten optische WORM-Medien zu verwenden. Damit war der Weg frei, neue Lösungskonzeptionen zu entwickeln, die nicht unbedingt auf optischen Technologien aufgebaut sind.

1999 bis 2005

Die Epoche der Multiplattform-Systeme und des FibreChannel SAN und NAS

Wenn die Archivierung gesetzlichen Bestimmungen und Anforderungen unterliegt, spricht man auch von revisionssicherer Archivierung. In vielen Bereichen erkennt der Gesetzgeber die digitale Archivierung als revisionssicher an, stellt aber gleichzeitig auch Anforderungen an die Art und Weise der Archivierung. In Deutschland gibt es zum Beispiel die **Grundsätze zum Datenzugriff und zur Prüfbarkeit digitaler Unterlagen (GDPdU)**. Viele Gesetze und Bestimmungen stellen allgemeine Anforderungen an die revisionssichere Archivierung. Zumeist werden die Aufbewahrungszeiträume von archivierten Daten vorgeschrieben. Weiterhin muss sichergestellt sein, dass die Daten im digitalen Archiv nicht verändert oder gelöscht werden können – in einer Welt mit 2 – 3 neuen Computerviren pro Tag gar keine so einfache Sache. Einige Bestimmungen schreiben auch vor, Kopien der Originaldaten in getrennten Räumen zu erstellen. Revisionssicherheit bedeutet natürlich auch, dass die Integrität der Daten zu jeder Zeit nachweislich gewährleistet ist. D . h., es muss auch anhand von Protokollen nachgewiesen werden, dass die Daten dem Original entsprechen. Die eingesetzte Technologie zur Archivierung wird aber heute von fast keinem Gesetz oder einer Bestimmung vorgeschrieben.

In einigen Fällen, z. B. im medizinischen und pharmazeutischen Bereich, müssen Daten 30 Jahre und länger archiviert werden. Eine Frage, die sich daraus ergibt, ist: Gibt es ein digitales Archiv, das auch in 30 Jahren noch Bestand hat? Eine zeitgemäße Antwort darauf ist, dass die Informationen von Zeit zu Zeit auf neue Systeme und Technologien überführt werden müssen. Überführung von Informationen und informationsverarbeitenden Systemen auf neue Systeme und Technologien – nachfolgend auch Migration genannt – ist heutzutage der einzige Weg, um die Daten auch in 30 Jahren noch lesen zu können.

Aufgrund der veränderten Bedingungen für die Langzeitarchivierung kündigte IBM das **System DR450** im **Oktober 2003** an. Das System wurde mit Standardkomponenten aufgebaut und als Langzeitarchivierungslösung für den Open-Systems-Bereich zur Verfügung gestellt. Die Lösung bestand aus zwei pSeries-p615-Servern, die in einem hochverfügbaren HACMP-Cluster (AIX-Betriebsystem) zusammengefasst waren. Auf diesen Servern war der IBM Tivoli Storage Manager für Data Retention aufgesetzt, der sicherstellte, dass die archivierten Dateien und Dokumente innerhalb der Aufbewahrungsfrist nicht gelöscht oder modifiziert werden. Die Daten wurden auf SATA-Platten (siehe auch Technologie-Anhang) einer FAStT600 mit EXP100-Erweiterungseinheiten abgespeichert. Kapazitäten von 3.5 TB bis 56 TB konnten konfiguriert werden. Das Magnetplattensystem war über ein redundant ausgelegtes, FibreChannel-basierendes SAN an die Server angeschlossen. Optional konnten an diesem SAN auch IBM 3592 Bandlaufwerke mit entsprechenden WORM-Kassetten und/oder überschreibbaren Kassetten betrieben werden.

Bereits im **Jahr 2005** kam die **Weiterentwicklung der DR450 mit der DR550** auf den Markt, die im Jahr 2006 mit allen seinen Komponenten auf RoHS-Konformität umgestellt wurde.

Das IBM DR550-System besteht heute aus den IBM Standard Software- und Hardware- Komponenten AIX, SSAM, pSeries p52A, den SAN-Komponenten 2005-B16, Disk System DS4700 und Disk-Erweiterungseinheiten EXP810. Optional kann man ein hochverfügbares DR550-System bestellen. Dabei beinhaltet das System zusätzlich noch die Cluster-Software HACMP. Der Vorteil dieses Konzepts liegt auf der Hand: Der Anwender erhält ein System, dessen Komponenten schon lange im Markt erprobt sind. Kernkomponente von DR550 ist der System Storage Archive Manager, ein Derivat vom Tivoli Storage Manager for Data Retention, der jegliche Veränderung oder Löschung von Informationen verhindert und somit die revisionssichere Speicherung erlaubt.

In der Produktfamilie IBM System Storage™ **DR550 gibt es heute drei Modelle**:

- *Das DR550-Express-System ist eine Einstiegslösung, die aus einer IBM pSeries Model 52A besteht mit internen SCSI-Festplatten, die als RAID5 konfiguriert sind. Mit dem DR550 Express Model wird ein Monitor-Kit mit Keyboard und Maus geliefert sowie ein SAN Switch 2005-B16, der für den Anschluss von Tape oder einer Disk-Erweiterungseinheit vorgesehen ist. Ein DR550-Express-System erhält man mit einer Einstiegskapazität von 1.1 TB (brutto). Das System kann um 4 TB oder 8 TB (brutto) erweitert werden, durch den Anschluss eines DS4700-RAID5-Systems. Ein entsprechender Einbauschrank kann optional mitbestellt werden.*
- *Das DR550-Single-Node-System besteht aus einem Einbauschrank, in dem eine IBM pSeries p52A, ein SAN Switch 2005-B16, ein Disk System DS4700 und optional ein oder mehrere EXP810 eingebaut und fertig konfiguriert sind. Ein Single-Node-System kann man mit einer Festplatten-Kapazität von 8 TB oder 16 TB bestellen und bis auf 112 TB (brutto) ausbauen.*
- *Das DR550-Dual-Node-System besteht aus den gleichen Komponenten wie das Single-Node-System, mit dem Unterschied, dass alle Komponenten redundant (doppelt) ausgelegt sind. D. h., in den Einbauschrank sind zwei IBM p52A und zwei SAN Switches 2005-B16 eingebaut und redundant konfiguriert. Ein Dual-Node-System kann man mit einer Festplatten-Kapazität von 8 TB oder 16 TB bestellen und bis auf 112 TB (brutto) ausbauen. Es handelt sich hierbei um ein hochverfügbares System*

Die Informationen werden innerhalb der DR550 auf Festplatten gespeichert – optional auch verschlüsselt – und erlauben schnelle Zugriffzeiten. Die Performance aus Anwendersicht kann mithilfe der Multiobjekt-Transaktion noch gesteigert werden, insbesondere, wenn viele Objekte mit einem Male gelesen oder geschrieben werden. Dabei werden innerhalb einer Transaktion mehrere Objekte gespeichert oder gelesen.

Mandantenfähigkeit lässt sich mit SSAM auch realisieren. Das erlaubt die logische Trennung von Speicherbereichen für verschiedene Klienten und auch das Reporting von benutzter Speicherkapazität. Dabei ist sichergestellt, dass ein Klient nur auf die Daten zugreifen kann, die in seiner Partition gespeichert sind.

IBM DR550 Express

Die Anbindung des Archivsystems IBM DR550 an die Anwendung erfolgt über das Tivoli Storage Manager for Data Retention API (Application Programming Interface). Typische Anwendungen, die Daten auf einem DR550-System zu archivieren, sind Dokumenten-Management-Systeme, wie z. B. IBM Content Manager oder Enterprise-Content-Management-Systeme, wie z. B. Opentext Lifelink Enterprise Archive Server. Das TSM API steht dem Anwender frei zur Verfügung. Alle Anwendungen, die das TSM API implementiert haben – gleichgültig auf welcher Plattform diese Anwendung betrieben wird – können Datenobjekte auf der DR550 archivieren und lesen.

IBM DR550-Dual-Node-System

DR550 File System Gateway

Im **Mai 2007** kündigte die IBM für die DR550 ein **File System Gateway** an. Über das DR550 Gateway kann ein Dateisystem aufgebaut werden, in dem die Daten vor dem Überschreiben geschützt sind. Das Gateway wird vorkonfiguriert ausgeliefert. Nach außen wird ein CIFS- oder NFS-Dateisystem ausgegeben. Die Gateways können 'geclustered' werden, um eine HA- (High Availability) fähige Lösung aufzubauen.

Im **August 2007** stellte IBM der **DR550 750-GB-SATA-Platten** zur Verfügung. Mit diesen neuen großen Platten kann eine DR550 auf bis zu **168 GB Kapazität** ausgebaut werden.

Seit **August 2007** stehen mit der DR550 **Komplettlösungspakete** für die **E-Mail-Archivierung** zur Verfügung. Diese Komplettlösungen für Lotus Domino und Microsoft Exchange adressieren die Anforderungen kleinerer und mittelständischer Unternehmen. Das Lösungspaket aus einer Hand enthält aufeinander abgestimmte, skalierbare Software- und Hardware-Komponenten der IBM und kann rasch implementiert werden. Basierend auf unternehmensspezifischen Aufbewahrungs- und Zugriffsprofilen ermöglicht die Lösung eine sichere Verwaltung und Archivierung von E-Mails einschließlich Dateianhängen in deren gesamten Lebenszyklus. Dabei wird sowohl geschäftlichen Anforderungen als auch nationalen oder internationalen Compliance-Vorschriften Rechnung getragen und ebenso den Wünschen vieler Unternehmen nach einer Optimierung des Speicherbedarfs und der Reduzierung von Administrationskosten. Das Komplettpaket zur E-Mail-Archivierung für den SMB- (Small and Medium Business) Bereich ist jederzeit zu einer umfassenden Archivierungslösung erweiterbar, die sämtliche unstrukturierte Unternehmensinformationen wie zum Beispiel digitalisierte Korrespondenz, Office-Dokumente, Faxe, Präsentationen, Audio- und Videodateien etc. verwalten kann. Ebenso ist eine Anbindung an SAP zur Dokumenten- und Datenarchivierung möglich. Auf diese Weise kombiniert das Komplettpaket modernste Technologie mit einem raschen ROI sowie höchster Investitions- und Zukunftssicherheit.

Das Komplettpaket besteht aus der benutzerfreundlichen und leistungsfähigen E-Mail-Archivierungslösung IBM CommonStore für Lotus Domino und Microsoft Exchange sowie dem IBM Content Manager als Basis-Repository. Hinzu kommt ein IBM System x3650 Server mit Intel Xeon Quadcore-Prozessoren, der speziell für anspruchsvolle Aufgaben im Unternehmenseinsatz wie zum Beispiel Enterprise Content Management (ECM), Virtualisierung, Enterprise Resource Planning (ERP) oder Datenbankanwendungen entwickelt wurde. Als Speicher-Komponente dient das IBM System Storage DR550, das eine leistungsfähige Funktionalität zur Ablage relevanter Dokumente gemäß gesetzlicher Aufbewahrungsfristen und -vorschriften auf magnetischen Speichermedien bietet. Die DR550 unterstützt dabei eine nicht löschbare und nicht wieder beschreibbare Datenspeicherung.

Im **Februar 2008** machte IBM die **DR550 Lösung als Maschine** mit zwei Modellen DR1 und DR2 in der **Version 4.5** verfügbar. Dies zeigt deutlich, dass IBM stark in das Information-Retention-Segment investiert. Durch die erweiterte Nutzung des System Storage Archive Managers (SSAM) für policy-basierten Information-Retention-Betrieb ermöglicht die DR550 eine transparente und automatisierte Bewegung archivierter Daten zwischen verschiedenen Speicherklassen. Dadurch können Kosten eingespart werden, ohne die Sicherheit der archivierten Daten zu gefährden. Das inzwischen preisgekrönte System ist jetzt als Maschine verfügbar. Es stehen zwei Modelle zur Verfügung: Die DR1 besteht aus einem 25U-Einheiten großen Rack, ist vorintegriert und eignet sich besonders für mittelständische Kunden. Die DR2 wurde speziell für Großunternehmen entwickelt und ist in einem größeren 36U-Einheiten-Rack untergebracht. Die DR2 bietet Single-oder Dual-Node-Konfigurationsoptionen für höhere Verfügbarkeit und Skalierbarkeit.

Die physikalische Anbindung der DR550 an die Serversysteme erfolgt über Ethernet-Schnittstellen und basiert auf dem TCPIP-Protokoll. Standardmäßig wird ein Ethernet Interface benutzt. Wahlweise können aber auch 2 oder mehr Ethernet Interfaces angeschlossen werden, was eine Skalierbarkeit des Datendurchsatzes erlaubt.

Das TSM API im Zusammenwirken mit SSAM bietet der Anwendung verschiedene Möglichkeiten zur Kontrolle der Aufbewahrungszeit. So kann eine Anwendung unter Benutzung der Ereignis-basierenden Aufbewahrungsregel Datenobjekte mittels Event löschen. Natürlich nur unter der Bedingung, dass die konfigurierbare Mindestaufbewahrungszeit für das Datenobjekt bereits abgelaufen ist. Mithilfe der chronologischen Aufbewahrungsregel sorgt SSAM für die Löschung der Daten nach Ablauf einer festgelegten Aufbewahrungszeit. Die Aufbewahrungsregeln werden dabei in sogenannten Management-Klassen definiert, die Anwendung weist ein Objekt dann nur noch einer Management-Klasse zu, wodurch dem Objekt die entsprechende Aufbewahrungszeit zugeordnet wird. Mit dem zusätzlichen Löschschutz, einer weiteren Option des TSM API, kann die vordefinierte Aufbewahrungsregel für Objekte außer Kraft gesetzt werden. Damit kann verhindert werden, dass ein Objekt nach Ablauf der normalen Aufbewahrungsfrist gelöscht wird.

Die DR550 bietet auch den Anschluss anderer externer Speichertechnologien, wie z. B. WORM Tape oder optische Speicher. Generell wird empfohlen, dass ein externes Gerät eine Native-WORM-Funktionalität besitzt, wenn es an ein DR550-System angeschlossen wird.

Der Anschluss von Bandlaufwerken an die DR550 erfolgt über das SAN. Die Anbindung von WORM Tape hat zwei entscheidende Vorteile für den Anwender:

1. *Kopien der Daten können auf WORM Tape geschrieben werden und in einem anderen Brandabschnitt katastrophensicher aufbewahrt werden.*
2. *Wenn die Daten auf Festplatte in der DR550 'altern' und somit die Zugriffe seltener werden, können sie auf WORM Tape ausgelagert werden. Die Tapes benötigen weniger Strom, Wartung und Austausch und sind somit viel kostengünstiger als Festplatten.*

IBM 3996 optische Archive – 960 GB bis 10.4 TB

Im Herbst **2003** wurde endlich ein neuer Standard, basierend auf dem blauen Laser, auf dem Markt verfügbar. Der neue 1 x Standard reflektierte eine 30-GB-Platte in den bisher klassischen Formaten WORM, MO und CCW. Die Firma Plasmon, die auch sehr aktiv in der Standardisierung der neuen Blaue-Laser-Technik mitwirkte, bot als erste Firma im Jahr **2004** Juke-Boxen mit den neuen optischen Platten an. **Im Herbst 2005** kam ein **OEM-Vertrag zwischen Plasmon und IBM** zustande, der es IBM erlaubt, diese Juke-Boxen unter IBM Logo zu vermarkten. Am Anfang war der Verkauf zum Anschluss an iSeries-Server beschränkt, seit **Juni 2006** können die Juke-Boxen auch an pSeries-basierende Server angeschlossen werden. Mit den Modellen 32, 80 und 174 bietet die **IBM 3996** Kapazitäten von 960 GB bis 5.2 TB an. Im **August 2007** kündigte IBM für die 3996 optischen Archivsysteme die Verwendung der neuen **optischen Platten mit 60 GB** in den Formaten WORM, MO und CCW an. Die neue Plattengeneration reflektiert den 2 x Standard basierend auf der blauen Laser-Technologie. Damit skaliert die IBM 3996 auf eine Kapazität von bis zu 10.4 TB. Die kapazitiven Möglichkeiten liegen also deutlich unter den Möglichkeiten einer DR550.

Es bleibt abzuwarten, ob basierend auf der Blaue-Laser-Technik ein klassischer optischer 4 x Standard verabschiedet wird, da inzwischen andere Lösungsoptionen für die Langzeitarchivierung auf dem Markt etabliert sind.

Hinzu kommt noch die Tatsache, dass, ebenfalls auf Blaue-Laser-Technik basierend, der Standard der ersten holografischen Platte mit einer Kapazität von 150 GB im XY-Format im Jahre 2003 verabschiedet wurde, Anfang 2005 bereits der 2 x Standard mit einer 500-GB-Platte und Ende 2005 ein 2 x Standard als Zusatzstandard für die 'Consumer'-Industrie in Form einer 300-GB-Platte. Holografische CDs und CD-ROMs lassen sich aufgrund der verwendeten Polymerbeschichtung (Kunststoff) wesentlich kostengünstiger produzieren als z. B. klassische DVDs oder im IT-Umfeld verwendete WORM-, MO- oder CCW-Platten.

Neue NAS-Produkte

Das **Jahr 2005** war durch eine weitere Überraschung geprägt. Am 4. April 2005 wurde eine enge Kooperation und Allianz zwischen **IBM** und der Firma **Network Appliance** bekannt gegeben. Diese Allianz ermöglicht IBM, alle NetApp-Produkte als IBM Logo-Produkte zu vertreiben. Da IBM bisher auf dem Gebiet des NAS (Network Attached Storage) nur begrenzt aktiv war, ermöglicht diese Partnerschaft IBM, auf dem Gebiet des 'Fileservings' ein breites Storage-Lösungsportfolio anzubieten. Am Anfang waren nur die kleinen NetApp-Geräte unter IBM Logo verfügbar, seit Ende 2005 die Geräte mittlerer Leistungsklasse, seit Juni 2006 entsprechende Gateways und seit August 2006 die Hochleistungsgeräte und damit nahezu die gesamte Produktpalette von NetApp.

IBM Nseries N5000 und Nseries N7000

Die **IBM Nseries** bietet Antworten auf die Fragen der Vereinfachung der Storage-Infrastrukturen und des zentralen Datenmanagements, des Backups der Daten und deren einfachster Wiederherstellung.

IBM läutet mit der Nserie die Konvergenz der Speicherwelten ein. Stand NAS für 'Einfachheit' und Funktionsvielfalt, so wurden beim Thema SAN Argumente wie High Performance, Skalierung und Ausfallsicherheit genannt. Mit der Nseries werden nun diese Unterschiede aufgehoben.

Es stehen dabei generell zwei Modelltypen zur Verfügung. Systeme mit internen Diskdrives (Filer) und Systeme, die externe SAN-Disk-Ressourcen verwenden, die sogenannten NAS-Gateways.

Die Skalierung der beiden Systemreihen reicht von Entry- über Midrange- hin zu Enterprise-Storagekapazitäten und Anschlussmöglichkeiten. Als herausragendes Merkmal sei erwähnt, dass alle Systeme dasselbe Betriebssystem Data ONTAP nutzen.

Die Nseries-Produkte bieten eine große Möglichkeit, Server im Netzwerk mit ihren spezifischen Zugriffsprotokollen anzuschließen. Diese umfassen die NAS-File-I/O-Protokolle (CIFS, NFS) sowie die Block-I/O-Protokolle iSCSI und FCP.

Die Nseries bietet eine enorme Flexibilität in Bezug auf die Ausbaustufen und die Möglichkeit, verschiedene Disk-Technologien für die jeweilige Lösung zusammenzustellen. Fiber Channel Diskdrives und SATA Diskdrives können gemischt werden.

Eine Nseries, bestückt mit FibreChannel Disks, kann so für ein Mission-Critical-, High-Performance- und Transaktionsorientiertes Umfeld eingesetzt werden. Nseries bestückt mit SATA Diskdrives, kann die ideale Wahl für Kunden sein, die eine Disk-to-Disk-Backup-Möglichkeit suchen oder sich für Archivierung interessieren.

Alle Nseries-Systeme nutzen ein einziges Betriebssystem mit einer enormen Vielfalt von weiteren, zusätzlichen SW- Optionen, angefangen vom Storage- und System-Management, über Inbound- und Outbound-Copy-Services bis hin zur kompletten D/R-Lösung mit integrierten Backup-Funktionen.

IBM N3700

Da das Thema 'gesetzeskonforme Archivierung' einen immer wichtigeren Stellenwert einnimmt, bietet die Nseries WORM-Funktionen an. Damit können Daten als nicht 'lösch-/veränderbar' abgespeichert werden, um den entsprechenden Richtlinien Rechnung zu tragen. Es können bestimmte Speicherbereiche im System oder auch die gesamte Nseries einfach als 'WORM'-Bereiche definiert werden. Eine breite Unterstützung der namhaften Application-Management-SW-Hersteller ist gegeben.

Die Nseries bietet SW-Funktionen an, die dem Systemadministrator das Management seiner Microsoft Exchange-, Microsoft SQL-, IBM DB2- und Oracle-Datenbanken erleichtern. Mit 255 SnapShots (Point-in-Time-Kopien) können Applikationen bei Fehlern leicht wiederhergestellt werden. Ein patentierter RAID-DP-Algorithmus sorgt für hohe Datenverfügbarkeit und schützt vor dem Datenverlust bei Ausfall von zwei Platten in einer RAID-Gruppe. Dies ist vor allem beim Einsatz von SATA-Platten sinnvoll.

Das Einstiegssystem **N3700** bietet eine Kapazität von bis zu 16 TB, die mittleren Systeme **N5200** bis zu 84 TB (bis zu 168 LUNs) und **N5500** bis zu 168 TB (bis zu 336 LUNs). Wahlweise können Platten mit 72 GB, 144 GB und 300 GB konfiguriert werden. Bei den mittleren Systeme werden zwei 2.8-GHz-Xeon-Prozessoren verwendet. Der sogenannte 'Kopf' der Filer (der Begriff hat sich für den Controller der Appliance-Lösung eingebürgert) ist bei den mittleren Systemen auch als Gateway verfügbar und bedient sich der verfügbaren Plattenkapazitäten in einem SAN.

Die Epoche der Multiplattform-Systeme und des FibreChannel SAN und NAS

Die Hochleistungssysteme Nseries 7000 bestehen aus der **N7600**, die kapazitiv bis 420 TB (bis zu 672 LUNs) skaliert, und der **N7800**, die auf bis 504 TB (bis zu 672 LUNs) ausgebaut werden kann. Bei der N7600 werden vier 2.6-GHz-AMD-Opteron-Prozessoren, bei der N7800 acht 2.8-GHz-AMD-Opteron-Prozessoren verwendet. Bei beiden Hochleistungssystemen können neben den anderen Platten auch 500-GB-SATA-Platten eingebaut werden. Auch die 'Köpfe' der Hochleistungssysteme sind als Gateways verfügbar.

Um das Portfolio der Nseries Produkte zwischen N5000 und N7000 in der Skalierbarkeit abzurunden, kündigte IBM im November 2006 das **Modell N5600** als neues leistungsstärkstes Modell der N5000-Reihe mit einer Verfügbarkeit ab dem 8. Dezember 2006 an. Die N5600 bietet auf Basis einer 64-bit-Architektur Kapazitäten von bis zu 252 TB und eine 30–40 % höhere Leistung im Vergleich zur N5500 und schließt damit die Lücke zwischen N5500 und N7600.

Als Ergänzung der N5600 Appliance kündigte IBM im **Mai 2007** das NAS **Gateway N5600** an. Zum selben Zeitpunkt, mit Verfügbarkeit im **Juni 2007,** wird die Nseries Reihe durch die neuen Modelle **N5300** und das NAS **Gateway N5300** ergänzt. Die N5300 integriert eine 64-Bit-Maschine und skaliert kapazitiv auf bis zu 126 TB. Das Gateway ist durch die 64-Bit-Maschine bestens für 4-Gbit-SANs geeignet. Im **August 2007** kommen zwei neue Einstiegssysteme hinzu, das Entrysystem **N3300**, das bis auf 24 TB Kapazität ausbaubar ist, und die **N3600**, die auf bis zu 69 TB Kapazität ausgebaut werden kann. Die alte N5500 und das N5500 Gateway wurde im **Oktober 2007** vom Vertrieb zurückgezogen.

Im **Februar 2008** kündigte IBM eine **neue Generation der Nseries** für große Rechenzentrumsbetriebe an.

Die nächste Generation der N7000series ist sowohl als Appliance als auch als Gateway verfügbar. Die neuen Modelle **N7700** und **N7900** ermöglichen eine höhere Skalierbarkeit für den Rechenzentrumsbetrieb in großem Maßstab. Die neuen Systeme bieten eine Skalierbarkeit mit **bis zu 1176 TB** Kapazitätsunterstützung für speicherintensive Anforderungen in Rechenzentren. Die N7000series ermöglicht es IT-Betreibern, SAN- und NAS-Speicheranforderungen auf einem einzigen System zu konsolidieren.

Für die Entry Systeme stehen neue Plattenlaufwerke zur Verfügung. Für die **N3300** und **N3600** werden nun auch SATA- oder SAS-Laufwerke im Controller unterstützt. Die N3300 verfügt über eine erweiterte Skalierbarkeit von bis zu 68 TB und die N3600 von bis zu 104 TB. Zusätzlich wurde bei folgenden Modellen die **Kapazität aufgestockt:** N5300 (von 168 auf 336 TB), N5600 (von 252 auf 504 TB), N7600 (von 420 auf 840 TB) und N7800 (von 504 auf 1008 TB). Der SnapManager für Office SharePoint-Server von Microsoft ist jetzt auf allen Systemen der Nseries verfügbar.

Im **Oktober 2007** wird den Nseries-Produkten eine neue Management Software, der **Virtual File Manager VFM** zur Verfügung gestellt. Dieses Software-Produkt wurde von der Firma Brocade entwickelt und bietet die Möglichkeit der File Virtualisierung.

Der **VFM** stellt dem Enduser eine logische Sicht auf seine Daten, unabhängig vom Speicherort, in Form eines **Global Namespace** zur Verfügung. So wird eine einheitliche, zentrale und logische Sicht auf einen Dateienpool über physische Systeme und Lokationen hinweg möglich. Die 'Übersetzung' der logischen Sicht auf den physikalischen Speicherplatz übernimmt der VFM, der auf einem dedizierten Server läuft. Machen wir ein Beispiel: Die User einer Engineering-Abteilung haben Sicht und Zugriff auf die Verzeichnisse 'Engineering' mit den entsprechenden Unterverzeichnissen. Die Daten selbst liegen aber verstreut auf File Servern und Servern, die beispielsweise in Ulm, Dallas oder einem anderen Ort stehen können.

Der VFM verbessert den Nutzungsgrad der Speichersysteme, indem Daten, die bestimmten Regeln unterliegen, transparent auf günstigere Speicher verlagert werden können (ILM). Storage-Erweiterungen (z. B. neue File Server) erfolgen mittels VFM ohne Impact auf den User. Der VFM hilft auch bei der Datenmigration von existierenden Windows/Linux-File Servern auf die Nseries. VFM unterstützt Nseries, Windows-Server (CIFS), UNIX-Server (NFS) sowie NetApp-Systeme.

VFM ist in zwei Varianten verfügbar: eine Migration und eine Enterprise Edition. Die Migration Edition ist nur für die Datenmigration bestehender Systeme auf die Nseries nutzbar. Die Enterprise Edition hat den vollen Funktionsumfang.

Nseries unterstützt eine Vielzahl von Servern und Betriebssystemen via FCP, CIFS, NFS und iSCSI. Die angeschlossenen Platten-Arrays können in RAID4 oder RAID-DP (Double Parity), das dem RAID6-Level entspricht, konfiguriert werden.

Mit der Nseries-Familie stellt die IBM maßgeschneiderte Plattenspeicherlösungen im Filer- und NAS-Umfeld zur Verfügung.

Es stehen für die IBM Nseries zurzeit 40 verschiedene SW-Funktionen zur Verfügung. Im Anhang sind die wichtigsten beschrieben.

SnapShot:
SnapShots sind lesbare, konsistente Sofort-Kopien (Point-in-Time Copy) eines Datenbestandes, genauer eines Files oder einer LUN. Es können bis zu 255 SnapShots pro File/LUN erzeugt werden. Das Besondere daran ist, dass sie platzsparend sind, d. h. dass sie beim Erzeugen keinen zusätzlichen Plattenplatz benötigen. Dies geschieht dadurch, dass nur der Verweis auf die Daten (Pointer) gespeichert wird. SnapShots können manuell angestoßen oder auch automatisiert prozessiert werden – so können z. B. alle 5 Minuten Point-in-Time Kopien erzeugt werden. Durch die hohe Anzahl der SnapShots (255) ist damit ein sehr granulares und zeitnahes Backup/Recovery möglich. Da SnapShots in einem speziellen Verzeichnis abgelegt werden, ist es auch für den User selbst sehr einfach (per Drag&Drop) seine gelöschten Dateien selbst wiederherzustellen.

SnapRestore:
SnapRestore nutzt die zeitnahen SnapShot Kopien um eine File, ein File-System oder eine LUN wieder herzustellen. Sind SnapShots nur lesbare Kopien und müssen für die Wiederherstellung der Daten in das aktive Filesystem kopiert werden, kann mit SnapRestore sehr leicht das aktive Filesystem auf eine SnapShot Kopie aufgesetzt werden. Dies bedeutet, dass für eine Datenwiederherstellung keine Daten umherkopiert werden. Ein solches Restoring geht mit einem einzelnen Kommando sehr schnell vonstatten! Eine 300 GB große Datenbank kann mit dieser Technik in ca. 2-3 Minuten restored werden.

SnapManager:
SnapManager wurde speziell für Datenbanken (Oracle und MS SQL) und Mail-Applikationen (MS Exchange) entwickelt. Basierend auf SnapRestore bietet der SnapManager ein hochintegratives SW-Paket für den Systemadministrator. Statt aufwändiger Scripts kann er hiermit per GUI einfache Backup-Tasks aufsetzen, Restoring-Prozesse anstoßen oder per 'Klick' Datenbank-Clones für Testzwecke erzeugen. Es besteht sogar die Möglichkeit, einzelne Mailboxen (mit Kalendereinträgen, Kontakten, Attachments) für das MS Exchange-Umfeld nach einem Fehlerfall einfach wiederherzustellen.

SnapMirror:
Sind SnapShot und SnapRestore Backup/Recovery-Funktionen, ist mit SnapMirror eine Funktion entwickelt worden, die für die Spiegelung der Daten von einer Nseries auf eine andere Nseries eine D/R Funktion liefert. SnapMirror spiegelt Daten – vereinfacht gesagt – synchron oder asynchron über SAN oder LAN auf eine Disaster-Recovery Site. Bei Ausfall der Produktiv-Seite stehen die Daten trotz Ausfalls zur Verfügung. Ist die Produktiv-Seite wieder operational, werden die Daten von der DR-Seite wieder zurückkopiert.

SnapVault und SnapLock:
SnapVault und SnapLock sind Lösungen für das Backup und die Archivierung von Daten. SnapVault ermöglicht ein Backup einer Nseries auf eine andere (oder von vielen auf eine). Durch den Einsatz von günstigen SATA Disks kann auf das Backupsystem schnell und kostengünstig gesichert werden. Da ein Restoring von Disk erfolgt, sind die Daten auch schnell wiederherstellbar. SnapLock bietet die Möglichkeit, die gesamte Nseries oder einen Teil der Disks als WORM- (Write Once Read Many) Bereich zu nutzen. Daten, die hier abgelegt werden, können erst nach dem Ablauf des Aufbewahrungsdatums gelöscht und/oder verändert werden.

Cluster Failover
Cluster Failover (CFO) ist eine Standardkomponente bei jedem Dual-Controller (A20 oder G20) der Nseries. Bei Ausfall eines Controllers übernimmt automatisch der zweite Controller die Server-I/Os und der Datenzugriff bleibt erhalten.

MetroCluster

Erweitert man die Cluster-Failover Funktion (CFO) über Gebäudegrenzen hinweg, erhält man den MetroCluster.

Der MetroCluster repliziert die Daten vom Controller 1 des primären RZs zum Controller 2 des sekundären RZs und garantiert Datenkonsistenz und deren Verfügbarkeit. Fällt die primäre Site aus, erlaubt MetroCluster einen Failover mit einem einzigen Befehl des Administrators. MetroCluster ist im Gegensatz zu SnapMirror keine DR- sondern eine Business Continuity-Lösung. Basierend auf dem Design, in dem die Daten bereits auf beiden Seiten synchron und im sofortigen Zugriff liegen, ist eine sofortige Weiterführung der Geschäftsprozesse möglich. Diese Übernahme ist soweit automatisiert, dass der Administrator nur noch einen einzigen Befehl absetzen muss. Stretch MetroCluster bietet einen Schutz von bis zu 300 m zwischen zwei Nseries Systemen. Fabric MetroCluster bietet darüber hinaus einen Schutz von bis zu 100 km mit SAN-Switchen.

A-SIS (Advanced Single Instance Storage)

Um die gespeicherte Datenmenge auf einem Backup- oder Archivsystem zu reduzieren, kann Data-Deduplication eingesetzt werden. Bei Nseries heißt diese Funktion A-SIS (Advanced Single Instance Storage Deduplication). Identische Daten werden nur einmal gespeichert, was bei bestimmten Applikationen bis zu 70 % Ersparnis in der Speicherkapazität ausmachen kann. Auf Systemebene analysiert die Nseries identische Blöcke im Hintergrund (für die Applikation transparent) und speichert diese nur einmal ab. A-SIS ist ideal für Archive oder Backup-Systeme.

SMBR (Single Mailbox Recovery für Exchange)

Im Gegensatz zu Lotus Domino, bei dem jede Mailbox separat als eigene Datenbank gespeichert wird (und damit eine einfach wiederherstellbare Einheit darstellt), legt Exchange verschiedene Mailboxen zusammen und in .edb- und .stm-Files ab. Diese Files werden im Laufe der Zeit sehr groß. Das erschwert das Restoring, weil man mit sehr großen .edb- und .stm-Files arbeiten muss, wenn eine einzelne Mailbox wiederhergestellt werden muss.

Die Lösung hierfür ist SMBR, eine SW, die diese großen Files durchsucht, um die gewünschte Mailbox (inklusive Anhängen, Foldern etc.) wiederherzustellen. SMBR setzt auf den SnapShots auf und muss nicht auf dem Produktionsserver laufen. Eine ‘Content Analyse’ ist eine Option der SMBR: mit ihr kann der Inhalt von E-Mails, Anhängen etc. analysiert und protokolliert werden.

SnapDrive

SnapDrive ist die SW, die es ermöglicht, aus dem Applikationserver heraus Volumes zu verwalten, zu vergrößern, konsistente SnapShots auszulösen und LUNs zu clonen (ohne FlexClone Feature).

Müssen in einem klassischen Storage-Subsystem die Volumes für eine Datenbank vergrößert werden, muss der Serveradministrator spätestens dann einen Prozess anstoßen, um mehr Speicherplatz zu beantragen.

Mit SnapDrive ist das jetzt sehr einfach: Aus dem GUI der Applicationserver heraus wählt man das entsprechende Volume an und gibt die neue größere Kapazität ein. Im Hintergrund (also unterbrechungsfrei) wird das Volume vergrößert. SnapDrive kommuniziert mit der Nseries, um diese Aktionen durchzuführen. SnapDrive spart Zeit und Kosten und hilft Prozesse zu vereinfachen, da diese via GUI erledigt werden können. Darüber hinaus kann SnapDrive auch neue Drives anlegen und diese ‘mounten’ (d. h. den Servern zur Verfügung stellen). SnapDrive unterstützt MSCS (Microsoft Cluster) und VSS und sollte für iSCSI und FC immer eingesetzt werden. Mit dieser Funktion werden Microsoft Windows sowie Unix/Linux Derivate unterstützt.

SyncMirror

SyncMirror ist ein Standardfeature, bei dem jeder I/O faktisch auf zwei getrennte Diskpools geschrieben wird. Dies entspricht quasi einem RAID1 innerhalb eines Systems. SyncMirror lässt sich am besten mit dem LVM (Logical Volume Manager) unter AIX vergleichen.

RAID-DP

Herkömmliche Single-Parity RAID-Technologien bieten Schutz beim Ausfall eines Disk-Drives. Es wird erwartet, dass kein anderer Fehler in der Daten-Rekonstruktionszeit auftritt. Sollte dennoch ein Fehler in dieser Phase auftreten, kann es zu Datenverlusten führen. Die ist besonders im Hinblick auf die immer größer werdenden SATA Drives der Fall, die inzwischen 1 TB Datenvolumen fassen können. RAID DP (Double Parity) sichert den Ausfall von zwei Disk-Drives. RAID DP schreibt 16 Stripes, die vorher kalkuliert werden (Data und Parity), gleichzeitig auf das Disk-Backend.

Spares Drives greifen bei Ausfall einer Disk. RAID-DP entspricht also RAID6.

FlexVol

FlexVol ist eine Standard SW, die für das LUN Management entwickelt wurde. FlexVol bedient sich dabei aus dem gesamten Disk-Storagepool der Nseries, um daraus die einzelnen LUNs zu formen. Es werden dabei alle verfügbaren 'Spindeln' im Backend ausgenutzt, ohne dass I/O Engpässe durch dedizierte Disk-Zuteilungen entstehen.

FlexShare

FlexShare priorisiert den Service für wichtige Workloads durch Zuordnung von Prioritätsstufen von 1–5 für jedes Volume. So kann z. B. der I/O einer Datenbank gegenüber Fileservices bevorzugt bedient werden. FlexShare ist ein Standardfeature der Nseries.

FlexClone

FlexClone ermöglicht, mehrfache Clones von FlexVols zu erzeugen (z. B. für Testzwecke, QA, DWH etc.). FlexClones verbrauchen zum Zeitpunkt der Erstellung keinen zusätzlichen Speicherplatz. Sie setzen auf SnapShots auf und sind in Sekunden erzeugt- und schreibend veränderbar. Sie haben die gleiche Performance wie FlexVols inklusive aller Eigenschaften eines FlexVols (dynamisches Vergrößern und Verkleinern). Eine Einschränkung ist zu berücksichtigen: 'Parent FlexVol' und Base SnapShot können nicht gelöscht werden, solange ein davon abhängiges FlexClone Volume existiert. Über 'Split' kann das FlexClone Volume von seinem 'Parent Volume' getrennt werden und unabhängig weiter existieren. Dafür wird freier Platz im Aggregat benötigt, um die 'Shared Blocks' zu kopieren.

Kommentar zur Epoche Multiplattform-Systeme und FibreChannel SAN und NAS

Die Schnelligkeit und Hektik nahm in dieser Epoche im Vergleich zur Vorgänger-Epoche noch weiter zu. Hinzu kamen neue Komplexitäten durch neue FibreChannel-Netze (SANs) und die Möglichkeiten im Networking-Attached-Storage-Umfeld. Storage wurde durch viele neue Produkte im SAN- und NAS-Umfeld zwar vielseitiger aber auch komplexer und schwieriger.

In dieser Epoche lösten sich die Speichersysteme von ihrer bisherigen Abhängigkeit von den Servern. Multiplattform-Systeme wie die ESS konnten jede Server-Plattform bedienen. Neben diesen Multiplattform-Systemen wurden parallel für die Open-System-Plattformen FibreChannel-Systeme gebaut, die sich an den entstehenden und immer weiter verbreiteten SANs anschließen ließen. SANs und SAN-Strategien beherrschten immer mehr den Markt. Das gilt auch heute noch! Auch die zSeries stellte von ESCON auf FICON um, um ESCON-Protokollpakete über den FibreChannel zu transportieren. SAN-Virtualisierung war das Fokusthema, dem IBM mit dem SAN Volume Controller antwortete. SAN-Management, vor allem zentrales Management , wurde zu einer der Hauptanforderungen von Rechenzentren, die SANs betrieben. Dazu entwickelte IBM das Programmpaket IBM TPC (Total Productivity Center). Nach 1-Gbit-SANs kamen 2-Gbit- und 4-Gbit-SANs in Zeitabständen von etwa drei Jahren zum Einsatz. Seit 2006 sind 4-Gbit-FibreChannel-SAN-'End to End'-Lösungen möglich.

Im NAS-Umfeld brachte IBM hauptsächlich Lösungen in Form von Gateways, die sich allerdings während der gesamten Epoche nicht durchsetzen konnten. IBM hatte zuwenig Fokus auf die NAS-Möglichkeiten und konzentrierte sich hauptsächlich auf die SANs. Am Ende der Epoche, im Jahr 2005, kam es zu einer Allianz der Firmen IBM und Network Appliance, die auf dem Gebiet NAS sehr erfolgreich waren und bis heute erfolgreich sind. Diese Allianz ermöglicht IBM, auch auf dem Gebiet NAS entwickelte Lösungen anzubieten und das Gesamtspeicher-Portfolio auszubauen.

Auf der Bandentwicklungsseite war neben der Einführung von LTO2 und LTO3 eines der Meilensteine die Einführung der sensationellen Jaguar-Bandlaufwerktechnologie mit dem 3592 und später TS1120 Laufwerk. Jaguar ist das 'Vehicle', um im Tape-Umfeld ganz neue Wege beschreiten zu können. Jaguar eröffnet technische Möglichkeiten, die wir uns vielleicht heute noch gar nicht vorstellen können.

Für die Langzeitarchivierung entstanden neben optischen Archiven neue und bessere Lösungsansätze, weil der Gesetzgeber nicht mehr vorschrieb, welches Medium benutzt werden muss. Dadurch wurden bessere und sinnvollere Lösungen möglich. IBM entwickelte die Lösung DR450 (Data Retention), die durch die DR550 ersetzt wurde und heute erfolgreich im Langzeitarchivierungsbereich eingesetzt wird.

Die verfügbaren Produkte am Ende dieser vielseitigen Epoche werden mit absoluter Sicherheit die Folge-Epoche mit entsprechenden Weiterentwicklungen begleiten und Lösungen für die durch SAN und NAS komplex gewordene Speicherwelt bieten. Insbesondere werden vor allem die Bereiche des DS4000-Plattensystems, des San Volume Controller, der gesamte Band- und Bandarchivbereich, Band-Virtualisierungslösungen und der Bereich der Langzeitarchivierung mit der DR550 schnell vorangetrieben werden.

Die Epoche der Server-basierenden Speichersysteme mit neuen Infrastrukturmöglichkeiten

Die Epoche der Server-basierenden Speichersysteme mit neuen Infrastrukturmöglichkeiten

Was wird uns die neue Epoche der Server-basierenden Speicherarchitekturen mit neuen Infrastrukturmöglichkeiten bringen? Was erwartet uns technologisch in den nächsten Jahren?

Der Autor erlaubt sich, eine persönliche und durchaus realistische Abschätzung der nahen Zukunft in dieses Storage-Kompendium einzuarbeiten.

Server-basierende Plattensysteme

Die Ansätze von Server-basierenden Speicherarchitekturen sieht man bereits in der Vorgänger-Epoche in den Produkten DR550 (pSeries-Server) für die Langzeitarchivierung und TS7510 (xSeries-Server mit Linux Kernel) für die Bandvirtualisierung im Open-Systems-Bereich. Auch die SAN-Virtualisierung mit dem SAN Volume Controller muss dazugezählt werden.

Allerdings bedeuten Server-basierende Speicherarchitekturen weit mehr. Systeme auf dieser Basis sind nicht nur Maschinen, auf denen Daten gespeichert werden, sondern auch gleichzeitig Server, die Applikationen durchführen können. Konsolidiert man heute Server auf der Serverebene und Speichereinheiten auf der Speicherebene, lassen Server-basierende Speicherarchitekturen eine Vertikalkonsolidierung von Servern und Storage zu. Applikationen können von der Serverebene auf die Storageebene verlagert werden. Dies ist vor allem für Applikationen von wesentlichem Vorteil, die viele I/Os zwischen einer Servereinheit und einer Storageeinheit produzieren. Diese I/Os können alle eingespart werden, wenn die Applikation von einer Speichereinheit durchgeführt wird ('No I/O-is the best I/O').

Das erste Produkt in dieser Server-basierenden Speicherarchitektur ist das Plattensystem DS8000, das am **12. Oktober 2004** von IBM angekündigt wurde. Die Ankündigung erfolgte zweifelsfrei viel zu früh und es dauerte fast noch ein ganzes Jahr, bis die Maschine den Stabilitätsgrad hatte, um produktiv eingesetzt zu werden. Heute hat die Maschine eine noch nie dagewesene Stabilität und eine Leistungsfähigkeit, die kein anderes vergleichbares System auf dem Markt ausliefert. Mit der DS8000 wurde am 12. Oktober auch der kleine Bruder, die DS6000, angekündigt.

IBM DS8100 2-Wege-System

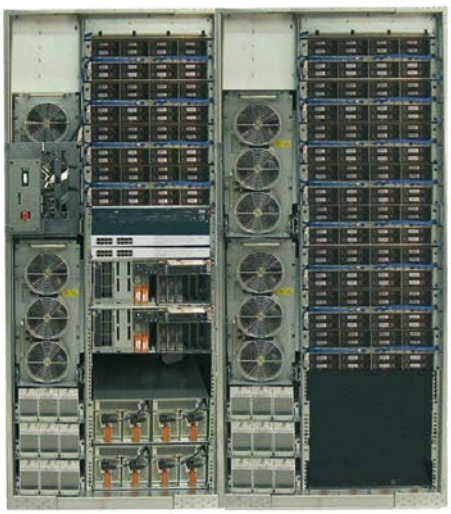
IBM DS8300 4-Wege-System

Die DS8000 ist das derzeit leistungsfähigste Plattensystem auf dem Markt. Mit 3,4 Millionen I/Os pro Sekunde ist sie unschlagbar im Vergleich zu anderen Plattensystemen. IBM trat mit dem Speichersystem DS8000 aber nicht an, um irgendwelche Spitzenpositionen auf irgenwelchen Ranglisten zu erzielen, sondern um mit weniger, leistungsfähiger Hardware die Datenspeicherung günstiger zu gestalten. Es geht um die Wirtschaftlichkeit und Spitzentechnologie hilft dabei, die Total Cost of Ownership (TCO) günstiger zu gestalten. TCO setzen sich aus Hardware-Anschaffungskosten und Betriebskosten zusammen. Mit dem Einsatz von DS8000 wird beides reduziert. DS8000 führt zu geringeren Anschaffungskosten und wesentlich günstigeren SAN-Infrastrukturkosten, weil weniger Komponenten benötigt werden. Gleichzeitig reduziert sich der Management- und Tuning-Aufwand, weil die Hardware leistungsfähiger ist und weniger überwacht und optimiert werden muss. Durch Architekturprinzipien wie ein Stripe-all-Design und selbstlernende und sich an veränderte Workloads automatisch anpassende Caching-Algorithmen erreicht die DS8000, dass die Hardware optimal genutzt wird.

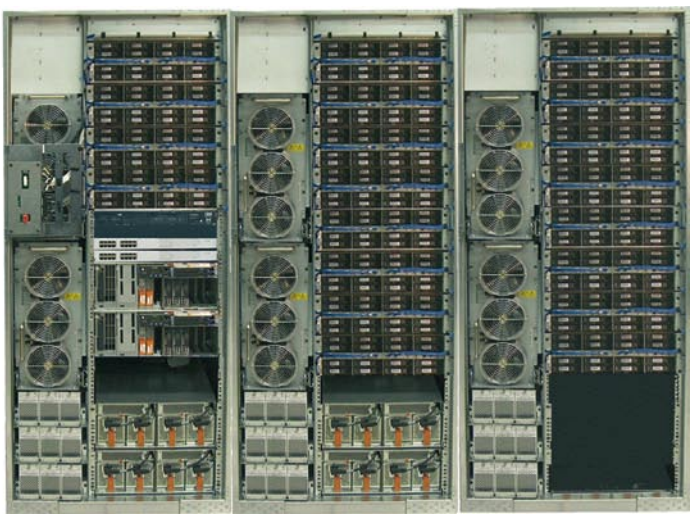
Die folgende Tabelle gibt einen Überblick über die Skalierbarkeit des Speichersystems DS8000. Die Systeme wurden im Hinblick auf nahezu lineare Skalierbarkeit konzipiert. Dies geht aus der unten stehenden Tabelle hervor. Der Einstieg in die DS8000-Welt bietet das 2-Wege-System DS8100, das bis 116 TB auf Basis von 300-GB-Laufwerken skaliert. Steigen die Kapazitätsanforderungen, erhöhen sich bei dem System alle zur Leistungssteigerung erforderlichen Komponenten wie Anzahl der Prozessoren, Cache, FC/FICON-Adapter und Disk-Adapter. Dies gilt auch für die zukünftig geplanten Erweiterungen. Steigen die Kapazitätsanforderungen über 192 GB hinaus, sind zukünftig 8-Wege- und 12-Wege-Systeme bis hin zum 32-Wege-System ohne Architekturänderung machbar. Bei einer nahezu verdreifachten Kapazität von über 500 TB skaliert das System nahezu linear mit, da dann auch alle wesentlichen Komponenten zur Leistungserbringung wie Cache, Prozessoren, FC/FICON-Adapter und Disk-Adapter beim Upgrade vom 4-Wege- auf das 12-Wege-System verdreifacht werden. Die zukünftigen Systeme sind Bestandteil einer lang angelegten Roadmap. Die 8-Wege- und 12-Wege-Systeme werden realisiert werden, sobald sich der Bedarf nach dieser extrem hohen Leistungsfähigkeit auf dem Markt abzeichnet.

	2-way	4-way	8-way	12-way
Server Processors	2-way POWER5	4-way POWER5	8-way POWER5	12-way POWER5
Cache	16 to 128 GB	32 to 256 GB	64 to 512 GB	96 to 768 GB
FICON (2 Gb/s) (4 ports per adapter)	8 to 64	8 to 128	up to 256	up to 384
FibreChannel (2 Gb/s) (2 ports per adapter)	8 to 64	8 to 128	up to 256	up to 384
ESCON (2 Gb/s) (4 ports per adapter)	4 to 32	8 to 64	up to 128	up to 192
Device Ports	8 to 32	8 to 64	8 to 128	8 to 192
Drives 73 GB (15K RPM), 146 GB, (15K RPM) 146 GB (15K RPM), 300 GB (10K RPM)	16 to 384	16 to 640	up to 1792	up to 1792
Physical Capacity	1.2 to 115 TB	1.2 to 192 TB	up to 538 TB	up to 538 TB
Number of Frames	1 to 2	1 to 3	2 to 8	2 to 8

Die Epoche der Server-basierenden Speichersysteme mit neuen Infrastrukturmöglichkeiten



IBM DS8100 Innenaufbau



IBM DS8300 Innenaufbau

Die DS8100 besteht aus einer Basiseinheit und maximal einer Erweiterungseinheit. An die DS8300 können bis zu 2 Erweiterungseinheiten angeschlossen werden. **Seit Oktober 2006** können bei den DS8000 Turbo Modellen bis zu vier Erweiterungseinheitenangeschlossen werden. Damit bietet die DS8000 Turbo die Möglichkeit, über 300 TB FC-Platten und über 500 TB FATA-Platten zu betreiben (maximal 1024 Plattenlaufwerke).

Die **Speichersysteme DS8000** sind die ersten Speichersysteme im Markt mit echten Storage-LPARs. Wie bei einem Server können auf einer physischen Einheit logische LPARs mit eigenem Prozessor, eigener Bandbreite, eigenen HBAs, eigenem Disk-Adapter, eigenen Laufwerken und eigenem Mikrocode gebildet werden. Die LPARs sind robust voneinander isoliert. Selbst ein Crash in einer LPAR beeinflusst die zweite LPAR nicht. Das LPAR-Konzept eignet sich daher besonders gut für die Sicherstellung eines bestimmten Service-Levels für bestimmte Anwendungsbereiche. Produktion und Test können auf einer Maschine gefahren werden, ohne dass die Testaktivitäten die Produktion beeinflussen. zSeries und Open Systems Workload können ebenfalls ohne Bedenken auf der gleichen Maschine betrieben werden, da jede LPAR mit ihren eigenen Ressourcen ausgestattet ist.

Zum heutigen Stand bietet die DS8300 2 LPARs, die jeweils 50 % der Ressourcen erhalten. In Kürze wird die Aufteilung flexibler werden und Sub-Prozessor-Allokationen werden möglich sein.

In der Zukunft wird es neben den Storage-LPARs auch **Anwendungs-LPARs** geben. Da die DS8000 eine pSeries integriert hat, können alle Möglichkeiten eines Servers für die Datenspeicherung genutzt werden. Es ist insbesondere daran gedacht, Anwendungen mit einer hohen Affinität zu Storage (viele I/Os) direkt auf dem Speichersystem laufen zu lassen und die internen Bandbreiten der DS8000 für optimale Performance zu nutzen.

Ein wesentliches Merkmal der DS8000 sind die **überlegenen Copy-Services**, die einen unterbrechungsfreien RZ-Betrieb ermöglichen. Dazu gehören die Funktionen synchrones und asynchrones Kopieren (Metro Mirror/Copy, Global Mirror/Copy) sowie Point-in-Time Copy (FlashCopy). Die Spiegelfunktionen der DS8000 sind heute wegen ihrer Leistungsfähigkeit (Entfernung, Anzahl IOs über eine Leitung, Anzahl gleichzeitiger Spiegel) und ihrer Funktionalität (Konsistenzgruppen, Suspend-/Resume-Möglichkeit, Umdrehen der Spiegelungsrichtung, 3 Spiegel gleichzeitig etc.) einzigartig im Markt. Derzeit gibt es **keine leistungsfähigere Remote-Copy-Implementierung** auf dem Markt.

Während DS8000 auf Basis der POWER5-Architektur ganz klar den High End-Speichermarkt adressiert, gibt es noch ein riesiges Marktpotenzial mit geringeren Anforderungen an Leistung und Skalierbarkeit. Dieses Marktpotenzial adressiert IBM mit einem modularen Speichersystem auf Basis der PowerPC-Prozessortechnologie. Wichtig für dieses Marktsegment ist ein kompaktes, modulares Design, das eine hohe Leistung auf kleinstem Raum ermöglicht, keine besonderen Anforderungen an die RZ-Infrastruktur stellt und sich perfekt in die vorhandene IT-Landschaft mit Servern in 19-Zoll-Racks integrieren lässt.

Diesen Designanfordrungen entspricht das **Speichersystem DS6000**. Das System ist voll und ganz für den Maßenmarkt und seine spezifischen Anforderungen konzipiert. Das System basiert auf modularen Einschüben mit 3U-Bauhöhe, die sich in vorhandene 19-Zoll-Racks einbauen lassen, ohne dabei Kompromisse in der Leistungsfähigkeit einzugehen. Das Speichersystem skaliert von 288 Gigabyte bis auf 38.4 TB.

DS6000 stellt für IBM bei Plattensystemen einen Durchbruch im Platzbedarf dar. Während der bisherige Einstieg in die Enterprise-Welt mit der ESS 750 nicht ohne mindestens einen Quadratmeter Stellfläche plus zusätzliche Servicefläche sowie einen stabilen Doppelboden, der eine Tragfähigkeit von über einer Tonne gestatten musste, zu ermöglichen war, erreicht die DS6000 völlig neue Dimensionen. Der maßstabsgerechte Vergleich unten zeigt den großen technologischen Fortschritt. Der Einstieg in die Enterprise-Welt ist jetzt mit weniger als 5 % des Volumens bisheriger Speichersysteme möglich. Damit einher gehen verringerte Anforderungen bzgl. der Servicefläche, der RZ-Infrastruktur und des Strombedarfs.

Die DS6000 entspricht im logischen Aufbau der DS8000. Der Unterschied besteht in der Plattform, auf der das System betrieben wird. HA- und DA-Code sind im wesentlichen identisch, da es sich um autarke Einheiten handelt. Die Funktionalität des Speichersystems ist ebenfalls identisch. Der große Unterschied ist die Prozessor-Plattform. Da der PowerPC-Prozessor über keine Partitionierungsmöglichkeiten verfügt, entfällt diese Komponente. Zudem wird der Prozessor nicht durch AIX, sondern durch Linux gesteuert.

IBM ESS 750 mit 5 TB Kapazität

IBM DS6000 mit 4.8 TB Kapazität

Die Epoche der Server-basierenden Speichersysteme mit neuen Infrastrukturmöglichkeiten

Ansicht IBM DS6800 Plattensystem

Wesentliche Änderungen betreffen auch den Abstraction Layer, der alle HW-Spezifikationen von der Funktionsebene isoliert. DS6000 ist also ein reines Speichersystem, das nicht für Applikationszwecke in der Zukunft eingesetzt werden kann.

Die Maschine besteht im Prinzip aus 5 Komponenten, sogenannten 'Customer replaceable Units' (CRUs). Dies sind Laufwerke, Stromversorgung, Controller, Batterie und ein 'Light-Path-Diagnose-Modul'. Wann immer eines dieser Teile defekt ist, erhält der Kunde ein Ersatzteil zugeschickt, das er in Minuten mit wenigen Handgriffen selbst einsetzen kann.

Das Gleiche gilt für den Mikrocode. Wann immer eine neue Version verfügbar ist, wird der Kunde davon informiert und kann sich aus dem Internet den neuesten Mikrocode herunterladen und unterbrechungsfrei einspielen.

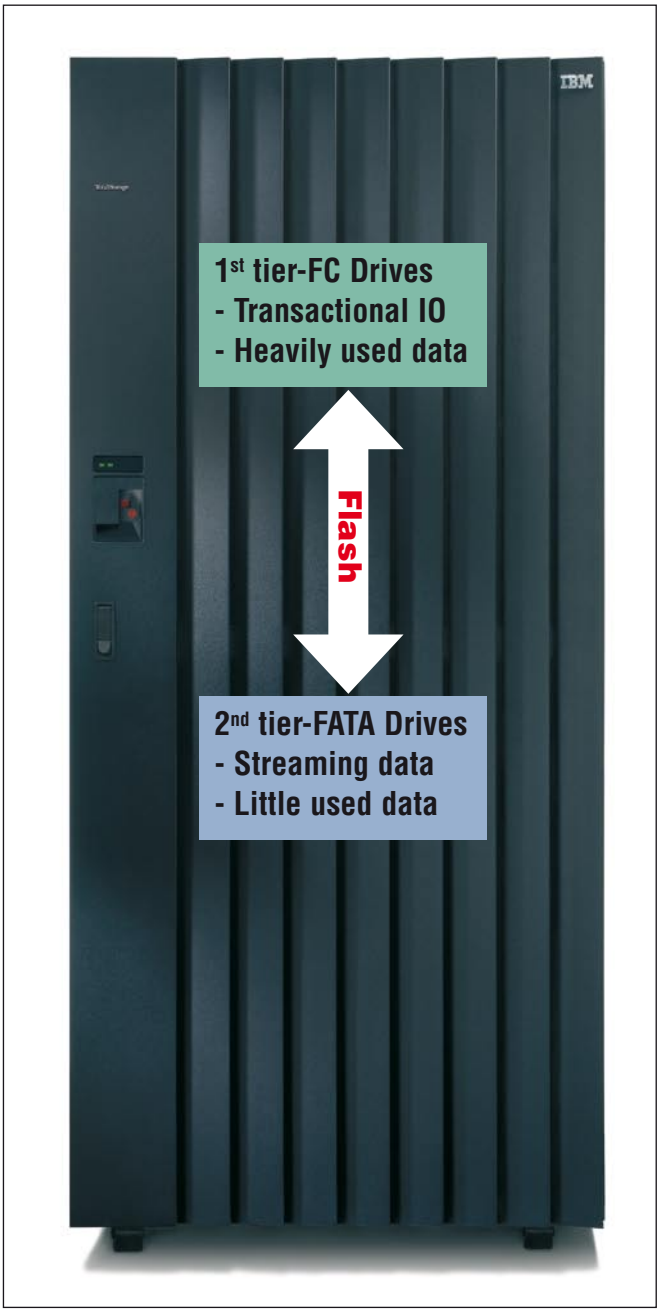
Unterstützt wird man bei der Fehleranalyse durch Light Path Diagnostics, SNMP-Meldungen und ein GUI. Die Fehleranalyse ist voll vom System gestützt. Interaktive Hilfefunktionen vereinfachen die Reparatur. Zur einfachen Administration und schnellen Implementierung verfügt das Speichersystem über einen Express Configuration Wizzard, der das erstmalige Konfigurieren oder Umkonfigurieren stark vereinfacht.

Die **DS6000 und DS8000** sind die ersten Plattensysteme von IBM, die **mit vier Jahren Gewährleistung** angekündigt wurden. Aufgrund dieser langen Gewährleistung erkennt man bereits, dass diese neue Architektur sehr lange Bestand haben wird.

Am 22. August 2006 mit Verfügbarkeit im September 2006 kündigte IBM die neusten Erweiterungen für die DS6000 und DS8000 an. Beide Plattensysteme können jetzt neben der Auswahl an FibreChannel-Platten auch mit günstigeren Fibre-ATA**(FATA)-Platten** bestückt werden. Die Platten haben eine Kapazität von **500 GB** pro Laufwerk. Das bietet für die Plattensysteme die Möglichkeit, innerhalb des Systems mit einer **Zwei-'tier'-Storage-Hierarchie** zu arbeiten.

So können die **FC-Platten** für die klassische Transaktionsverarbeitung und hoch frequentierte Daten eingesetzt werden und die günstigeren, aber langsameren **FATA-Platten** für sequenzielle Datenverarbeitung und nur wenig benutzte Daten.

Damit bietet eine DS8000 voll ausgebaut mit FATA-Platten eine Bruttokapazität von 320 TB und eine kleine DS6000 bis zu 64 TB an Bruttokapazität.



Für die DS8000 wurden neue Modelle angekündigt, die **POWER P5+ Prozessortechnologie** integriert haben und damit 15 Prozent höhere Leistungsfähigkeit ausliefern. Die I/O-Cages wurden auf **4-Gbit-FibreChannel- und FICON-Ports** umgestellt. Damit ist die DS8000 das erste Plattensystem auf dem Markt, das mit 4-Gbit-FICON-Ports am zSeries-Host arbeiten kann. Die DS8000-Systeme mit P5+ Prozessoren werden als **DS8000-Turbo-Systeme** bezeichnet.

Ein neue Einrichtung, die als **'Synergy Feature'** bezeichnet und im November 2006 verfügbar wird, verbessert maßgeblich die Gesamtleistung der Komponenten DS8000 Turbo, DB 2 und AIX und erzielt eine wesentlich effektivere Kommunikation zwischen Rechner und Speichersystem für DB2-Anwendungen. Dies wird durch eine gezielte 'End to End'-Prioritätensteuerung ermöglicht.

Teil der Ankündigung vom 22. August 2006 war zudem die Möglichkeit, bei DS8000-Turbo-Systemen mit einem dritten Spiegel bei Remote-Copy-Verfahren zu arbeiten. Diese Lösung wird als **'3 Site Disaster Recovery Solution'** bezeichnet. Damit wird auch im Disaster-Fall ein konstanter Zugriff auf die Daten gewährleistet. Diese Möglichkeit war bisher nur über RPQ (Request for Price Quoting) gegeben.

Das **dreifache Remote-Spiegelverfahren** setzt sich aus den Komponenten 'Metro Mirror' und 'Global Mirror' (ehemals PPRC XD – Peer to Peer Remote Copy eXtended Distance – und FlashCopy) zusammen. Von der Lokation A wird mit der DS8000-Turbo synchron zur Lokation B gespiegelt. Die Lokation B spiegelt dann asynchron auf die Lokation C weiter.

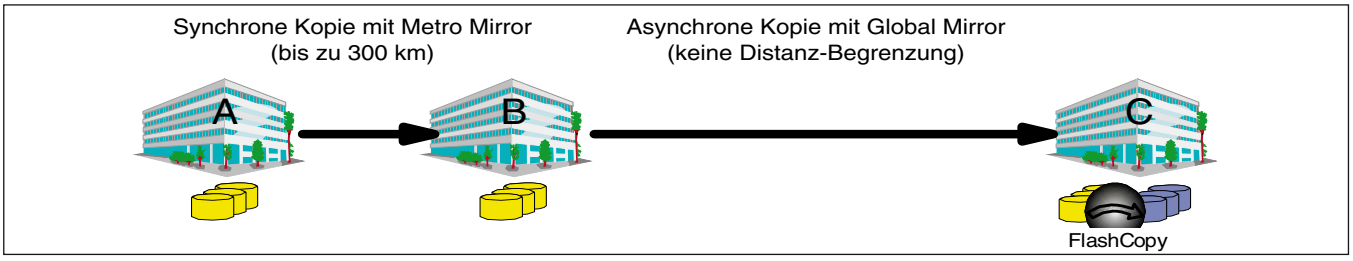
Fällt die Lokation A aus, wird die Anwendung auf die Lokation B umgeleitet und Global Mirror läuft ohne Unterbrechung zwischen den Lokationen B und C weiter. Kommt Lokation A wieder online, wird von der Lokation B nach A auf inkrementeller Basis resynchronisiert. Nach Abschluss der Resynchronisation kann die Anwendung von B nach A zurückverlagert werden.

Fällt die Lokation B aus, stellen inkrementelle Resynchronisationsverfahren die Verbindung zwischen A und C her und gewährleisten damit weiterhin unterbrechungsfreien Spiegelbetrieb. Steht die Lokation B wieder zur Verfügung, wird über ein inkrementelles Verfahren von A nach B resynchronisiert. Nach Abschluss der Resynchronisation werden die Lokationen B und C wieder über Global Mirror verbunden und so die Ausgangssituation wieder hergestellt.

Fällt Lokation C aus, ist die Anwendung A nicht betroffen und die synchrone Spiegelung zwischen A und B bleibt erhalten. Steht die Lokation C wieder zur Verfügung, wird Global Mirror inkrementell wieder aufgebaut und die Ausgangssituation wieder hergestellt.

Parallel Access Volume (PAV): Die Funktion, von mehreren Rechnern und Anwendungen gleichzeitig auf diesselbe Plattenadresse zu schreiben und von ihr zu lesen, wurde im z/OS Umfeld mit dem Enterprise Storage Server (siehe auch ESS) eingeführt und für die DS8000 weiter optimiert. Der Workload Manager (WLM) im z/OS steuerte die dynamische Zuordnung und Umordnung der dafür notwendigen 'Alias'-Adressen, was als dynamisches PAV bezeichnet wird. Im November 2006 kündigte IBM für die DS8000 **Hyper PAV** an. Dabei wurde die Funktion so geändert, dass z/OS im Zusammenspiel mit der DS8000 für jeden einzelnen I/O-eine Alias-Adresse aus einem Pool nehmen kann, ohne dass eine Koordination zwischen z/OS Systemen notwendig ist. Damit kann sofort auf sich ändernde Workloads reagiert werden. Hyper-PAV benötigt bei derselben Workload nur etwa die Hälfte an Alias-Adressen und bietet im Vergleich zum dynamischen PAV bei derselben Anzahl an Alias-Adressen die Möglichkeit, 50 % mehr I/O's zu prozessieren.

Am 22. August 2006 wurde als neues Software-Produkt IBM **TotalStorage Productivity Center (TPC)** für Replikation angekündigt. TPC für Replikation erlaubt ein zentrales Management der Copy-Services-Funktionen Metro Mirror, Global Mirror und FlashCopy für die Produkte DS6000, DS8000, DS8000 Turbo, SAN Volume Controller SVC und ESS Modell 800.



2006 bis 2008

Die Epoche der Server-basierenden Speichersysteme mit neuen Infrastrukturmöglichkeiten

Im Jahr 2007 stellte IBM für die DS8000 Plattensysteme erhebliche Erweiterungen zur Verfügung.

Im **Februar 2007** kündigte IBM mit Verfügbarkeit März für alle DS8000 Systeme neben den bisherigen Plattentypen **300-GB-FibreChannel-Platten mit 15 000 Umdrehungen** an. Die DS8000 war damit das erste System auf dem Markt, das mit diesen schnellen großkapazitiven Platten ausgestattet werden konnte. Im Mai 2008 kamen erhebliche Kapazitätserweiterungen dazu. Konnte bisher ein 4-Way-System mit bis zu 640 Platten (bei 2-Way sind es bis zu 384 Platten) in einer Konfiguration von drei Einheiten ausgestattet sein, war es seit Juni 2007 möglich, ein 4-Way-System DS8300 **mit bis zu 1024 Platten** zu konfigurieren. Dazu werden an die Basiseinheit vier zusätzliche Erweiterungseinheiten angeschlossen, insgesamt also eine Konfiguration mit fünf Gehäuseeinheiten. Bei Verwendung der 500-GB-FATA-Platten steigerte sich so die maximale Kapazität auf bis zu 512 TB pro System.

Seit **November 2007** stehen der DS8000 weitere leistungsoptimierende Features zur Verfügung. Die Funktion **Storage Pool Striping mit Rotate Extends** reflektiert einen neuen Default-Algorithmus, der neue logische Platten in 1-GB-Schritten über das Backend verteilt und so die Leistung ohne spezielles Tuning optimiert. **AMP (Adaptive Multistream Prefetching)** stellt eine neue Caching-Technologie von IBM Research zur Verfügung, die, auf Workloads bezogen, ein selbstoptimierendes Prefetching durchführt. AMP entscheidet dynamisch, was und wann in den Cache vorgeladen wird. Das kann den Durchsatz von sequentiellen und Batch-Workloads dramatisch verbessern, d. h. Laufzeiten können erheblich verkürzt werden. AMP verbessert den Lese-Durchsatz aus einem RAID5-Array nahezu um Faktor zwei! AMP kann Hot Spot-Situationen verhindern, wenn sehr hohe Anforderungen an sequentielle Workloads gestellt werden. **IBM z/OS Global Mirror Multiple Reader** bietet im zSeries-Umfeld einen deutlich höheren Durchsatz bei Remote-Spiegelungen.

Neben den neuen Performance Features wurden im November 2007 für die Maschinen auch neue funktionale Erweiterungen verfügbar. **Space efficient FlashCopy** kann durch Reduzierung der benötigten Plattenkapazität für Kopien signifikant die Kosten senken. Dadurch können gleichzeitig die Strom- und Klimaanforderungen gesenkt werden. Flash-Copies benötigen nur Plattenplatz, wenn Veränderungen am Original durchgeführt werden. **Dynamic Volume Expansion** vereinfacht das Management durch 'Online'-Vergrösserung von logischen Laufwerken bei Datenwachstum. Neben diesen Erweiterungen wird das TPC (Total Productivity Center) bei Auslieferung von neuen DS8000 Systemen standardmäßig als neu umbenanntes **SSPC (System Storage Productivity Center)** mitgeliefert und bietet damit einen einheitlichen Zugang zum Management von IBM und anderen Speichersystemen. Damit stehen dem IT-Benutzer eine einheitliche Management-Oberfläche und eine gemeinsame Konsole für DS8000 Systeme und die SVCs (San Volume Controller) zur Verfügung.

Am **26. Februar 2008** wurden zusammen mit der Ankündigung des neuen Mainframe Systems z10 neue Erweiterungen für das DS8000 Plattensystem speziell im z/OS-Umfeld angekündigt. Mit der Funktion **z/OS Metro/Global Mirror Incremental Resync** entfällt bei einer dreifachen Remote Spiegelung die Notwendigkeit, eine volle Kopie für die Resynchronisation nach einer Hyper Swap-Situation zu erzeugen. Die Erweiterung **Extended Distance FICON** reduziert den Bedarf von Kanalerweiterungseinheiten (Channel Extenders) bei z/OS Global Mirror- (zweifache Remote-Spiegelung) und z/OS Metro/Global Mirror- (dreifache Remote-Spiegelung) Konfigurationen durch eine höhere Parallelität der Leseoperationen. Die neue Möglichkeit des **z/OS Basic Hyper Swap** erlaubt die Konfiguration von Disk Replication Services über das grafische User Interface (GUI) der DS8000 Plattensysteme. Diese Möglichkeit kann auch für DS6000 und ESS-Systeme genutzt werden, bei denen das GUI auch zur Verfügung steht. Die Möglichkeit des z/OS Basic Hyper Swap ist auch ohne GDPS (Geographically Dispersed Parallel Sysplex) benutzbar. z/OS Basic Hyper Swap ist eine einzigartige Funktion, die auch ohne GDPS einen automatischen Failover auf das Remote-Plattensystem vornimmt, wenn die primäre Site nicht mehr zur Verfügung stehen sollte. Dies ist aber ausschließlich bei den IBM Systemen DS8000, DS6000 und ESS möglich. Plattensysteme anderer Hersteller werden nicht unterstützt.

DS3000 Platten Entry System

Im **Februar 2007** überraschte IBM den Markt mit der Ankündigung eines Entry Plattensystems, der **DS3000**. Das System wird wie die DS4000 Reihe von der Firma LSI gebaut und als IBM Logo-Produkt vertrieben. Am Anfang war das System nur über den IBM System x-Vertriebskanal beziehbar. Als die ersten Performance Benchmarks ergaben, in welche ansehnliche Leistungsklasse die Maschine skaliert, wurde das System im April 2007 in den gesamten Storage-Vertrieb der IBM integriert.

Die DS3000 Familie zeichnet sich im allgemeinen durch ein hervorragendes Preis-Leistungs-Verhältnis aus und bietet einen extrem günstigen Einstiegspreis. Es stehen drei Modelle zur Verfügung.

Die **DS3200** zeichnet sich durch 6 SAS-Hostanschlüsse aus. Die **DS3300** verfügt maximal über 4 x 1-Gbit-Ethernet-iSCSI-Hostanschlüsse. Die **DS3400** zeichnet sich durch maximal 4 FibreChannel Ports aus, die in den Geschwindigkeiten 4, 2 und 1 Gbit/s arbeiten können. Die Platten- und Backendtechnologie bietet in allen drei Modellen SAS die Möglichkeit, auch SATA-Platten anzubinden. Anfangs standen ausschließlich 300-GB-SAS (Serial Attached SCSI)-Platten zur Verfügung, bei einer maximalen ausbaubaren Kapazität auf bis zu 14.4 TB. Seit Oktober 2007 können auch 750-GB-SATA-Platten integriert werden. Damit skalieren die Systeme auf eine maximale Kapazität von 35.2 TB. Der Mischbetrieb von SAS-Platten und SATA-Platten wurde sichergestellt.

Allgemein zeichnen sich die Modelle durch eine geringe Bauhöhe von nur 2U aus, sowie eine Skalierbarkeit auf bis zu 48 Festplatten. In der Kopfeinheit können bis zu 12 Festplatten untergebracht werden. Der Controller und damit das Gesamtsystem ist über die Erweiterungseinheiten EXP3000 (bis zu 3 weiteren Einheiten) ausbaubar.

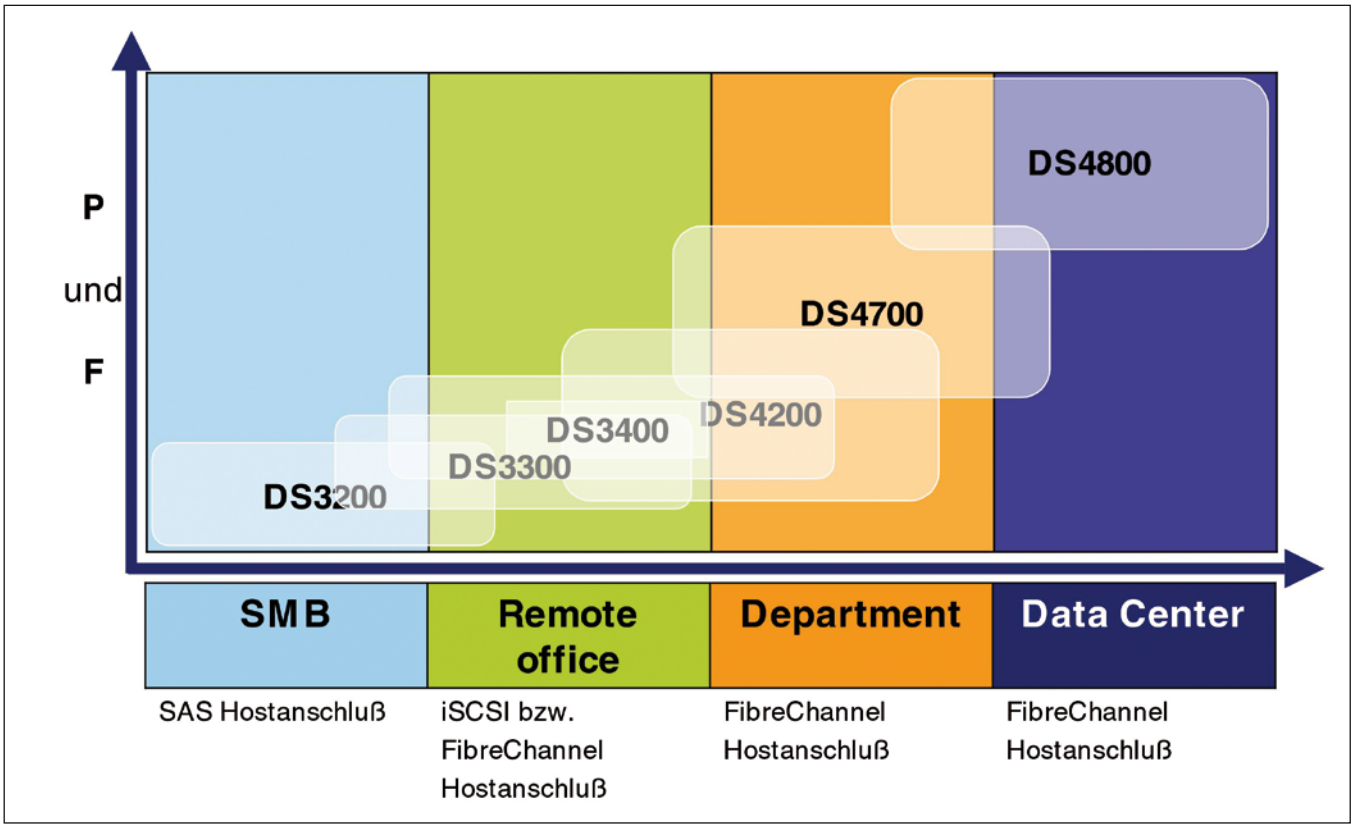
Seitens der Betriebssysteme hat die DS3000 Familie eine geringeres Portfolio an anschließbaren Betriebssystemen. Hier werden die Systeme Windows, Linux und Novell standardmäßig angeboten. Zusätzlich bietet die Maschine je nach Modell die Anschlußmöglichkeit von VMware, SVC (SAN Volume Controller) oder AIX.

Jedes Modell besitzt einen Ethernet-Zugang für das Management. Zusätzlich kann die Maschine auch über die Host Connectivity administriert werden. Mit 512 MB RAM pro Controller steht ausreichend Speicher für das Caching zur Verfügung. Dieser Speicher kann mit weiteren 512 MB pro Controller erweitert werden. Dies ergibt eine maximale Cache-Größe von 2 GB pro System.

Obwohl es sich um ein ‘Entry System’ handelt, wird bei der DS3000 auf redundante Komponenten nicht verzichtet. Die DS3000 hat auch bei einer Single Controller-Variante redundant ausgelegte Lüfter und Netzteile.

Rück- und Frontansicht der DS3000

Die Epoche der Server-basierenden Speichersysteme mit neuen Infrastrukturmöglichkeiten



Positionierung der Entry Systeme DS3000 zu den bestehenden DS4000 Systemen (P= Performance, F=Funktionalität)

Der **DS3000 Storage Manager** ist analog zum DS4000 Manager ebenfalls intuitiv zu bedienen und leicht zu installieren. Der initiale Setup der Maschine erfolgt in sechs im Dialog geführten Schritten und sollte innerhalb von 45 Minuten vollzogen sein, d.h. nach spätestens 45 Minuten sind die ersten Volumes den angeschlossenen Servern bekannt. Der DS3000 Storage Manager unterstützt die RAID-Level 0, 1, 3, 5 und 10.

Durch den **Recovery Guru** und durch automatische Benachrichtigungen erhält man proaktiv, auch schon vor Crash Situationen, Meldungen über eventuelle Fehler oder Überschreitungen von Schwellwerten. Die Benachrichtigung erfolgt über SNMP und/oder per E-Mail.

Im **Februar 2008** kündigte IBM für alle drei DS3000 Modelle **1000-GB- (1 TB) SATA-Platten** an. Damit bekommt das Entry System DS3000 als erstes Plattensystem der IBM die Möglichkeit, Platten mit 1 TB an Kapazität zu verwenden. Mit diesen Platten kann jede Erweiterungseinheit mit 12 TB Plattenplatz konfiguriert werden. Das Gesamtsystem, bestehend aus einer Kopfeinheit und drei Erweiterungseinheiten, liefert mit den 1-TB-SATA-Platten eine maximale Kapazität von 48 TB.

Für die DS3300 steht ein neues Modell 32T zur Verfügung, das speziell für Telco-Umgebungen ein 48-V-basierendes doppeltes Netzteil auf Gleichstrombasis in der Controller-Einheit integriert hat (iSCSI DC-Powered Controller). Damit kann die DS3300-32T problemlos in Telco-Umgebungen eingesetzt werden.

Virtualisierung

SAN Volume Controller SVC – Virtualisierung von Plattensystemen im SAN-Umfeld

Die ersten Geräte des SAN Volume Controllers SVC waren im September 2003 verfügbar und sind in der Epoche der Multiplattform-Systeme und des FibreChannel SAN und NAS beschrieben. Die Konzeption des SVC und die dazugehörige Software wurde über die Jahre kontinuierlich weiterentwickelt und der heutige Produktplan des SVC sieht für die nächsten Jahre viele neue Erweiterungen bezüglich der Hardware und Software vor. Deshalb ist es notwendig, die Konzeption und den SVC selbst besser zu verstehen.

Die Administration von Speicherkapazitäten stellt nach wie vor einen hohen Anteil der IT-Kosten dar. Virtualisierte Speicherlandschaften bieten hier eine mögliche Lösung, indem sie Speichersysteme sowohl unterschiedlicher Bauarten und Hersteller gemeinsam verwalten und flexibel den Servern zuweisen können. Eine Speichervirtualisierung mit dem SAN Volume Controller begegnet diesen Herausforderungen mit einer skalierbaren Architektur. Viel Aufwand haben viele Unternehmen in den letzten Jahren betrieben, um sich von einer dezentralen Speicherlandschaft mit direkt angeschlossenem Speicher (Direct Attached Storage, DAS) hin zu einer zentralen Speicherlösung mit einem Storage Area Network (SAN) zu entwickeln. Wohl existieren Storage Area Networks nicht flächendeckend, vielfach bilden sie aber bereits den Standard. In der jetzigen Phase geht es darum, sowohl Serverumgebungen als auch Speichersysteme weitgehend zu virtualisieren. Die derzeit existierenden SAN-Lösungen sind oft recht starre Umgebungen und bilden meist Einzellösungen für Plattformen bestimmter Hersteller und Fachbereiche. Folglich sind Änderungen aufwändig und unterbrechen meist den Anwendungsbetrieb. Falls ein IT-Betreiber unterschiedliche Speichersysteme – unter Umständen mehrerer Hersteller – installiert hat, benötigt er jeweils unterschiedliche Administrations-Tools. Auch ist die hardwarebasierende Datenspiegelung zwischen Speichersystemen unterschiedlicher Bauart nicht möglich, das Gleiche gilt für Speichersysteme unterschiedlicher Hersteller.

Administrationsarbeiten geraten damit sehr aufwändig, weil Plattformwissen und die zugehörige Pflege gleich mannigfach vorzuhalten ist. Weiterhin lasten heutige Speicherlösungen ihre installierten Speicherkapazitäten schlecht aus. Weltweite Kundenumfragen ergaben, dass die effektive Nutzung bei etwa 50 Prozent liegt. Auch beim Austausch von Speichersystemen im Rahmen von Business-Continuity-Maßnahmen war hoher Aufwand notwendig, und zumeist wurde der Anwendungsbetrieb für die gesamte Migrationsdauer unterbrochen. Der IBM SAN Volume Controller (SVC) erleichert dies erheblich, indem er **Migrationen bei laufendem Betrieb** ermöglicht.

Eine Speichervirtualisierung mit IBM SVC bietet gegenüber traditionellen Speicherlösungen den Vorteil, dass **Speicherplatz weitgehend herstellerunabhängig zugeordnet** werden kann. Hieraus folgt eine einfachere Administration, da sich die virtuelle Oberfläche unabhängig von den jeweils installierten Speichersystemen konfigurieren lässt. Die Grundidee besteht darin, schnell und flexibel Plattenspeicher dort zuordnen zu können, wo gerade Bedarf entsteht. Dem stand bisher im Wege, dass am SAN angeschlossene Server nur Zugriff auf die ihnen zugeordneten Speicherbereiche hatten. Eine virtualisierte Speicherlösung wie der SVC löst dieses starre Muster auf, da die Server hier nur noch Zugriff auf virtuelle Speicherbereiche haben. Die SVC-Software entscheidet nach entsprechenden Vorgaben, wo die Daten physisch abgelegt werden. Dabei bietet die Lösung Schnittstellen zu allen marktüblichen Betriebssystemen, Hardwareplattformen und Speichersystemen und ermöglicht ein Management über eine zentrale Konsole.

Einige wichtige Hardwarefunktionen verbleiben nach wie vor im Speichersystem und sind unabhängig vom SVC. So werden die Systeme vor dem Anschluss an den SVC wie in der Vergangenheit eingerichtet. Es werden RAID-Arrays und anschließend die LUNs (Logical Units) gebildet. Dies erfolgt im Zuge der Installation mittels der vom Hersteller mitgelieferten Administrierungssoftware. Die so definierten LUNs werden dann dem SVC zugeteilt, der sie in sogenannte ‘Managed Disks’ (MD) umwandelt. Eine MD hat eine feste Größe und ist nicht mehr veränderbar. Danach werden eine oder mehrere MDs in sogenannte ‘Managed Disk Groups’ zusammengefasst. Für die angeschlossenen Server werden im SVC virtuelle LUNs definiert, worauf der jeweilige Server zugreift. Die Definition legt auch fest, welcher Managed Disk Group die virtuelle LUN zugeordnet wird. Dies bestimmt, wo letztlich die Daten physisch abgelegt werden.

Einfache Datenmigration

Eine Datenmigration von LUNs zwischen unterschiedlichen Speichersystemen führt der SVC im laufenden Arbeitsbetrieb durch. Verliert eine Applikation, deren Daten sich beispielsweise auf einem High-Performance-Speicher befinden, für das Unternehmen an Priorität – etwa aufgrund hoher Betriebskosten für den Fachbereich – lassen sich deren Daten online auf günstigere Low-Performance-Speicher migrieren. Dies erfolgt durch eine Änderung der Zuordnung einer virtuellen LUN zur Managed Disk Group.

Neben der Datenmigration zum Speichersystemaustausch können nun auch Pflegearbeiten während der Arbeitszeit stattfinden. Speicherinhalte der Subsysteme lassen sich im laufenden Betrieb auf andere, freie Bereiche eines anderen Subsystems verlagern. Alle Datenverlagerungen finden unter Kontrolle des Speicheradministrators statt und müssen von diesem eingeleitet werden. Der SVC bietet hierbei keine regelbasierte Datenverschiebungen aufgrund von Zugriffshäufigkeiten an. Dies können Analyse-Tools übernehmen, die etwa im IBM Total Storage Productivity Center eingebettet sind. Es bietet ein komplettes SAN-Management inklusive Speicherplatz- und Zugriffsanalysen und kann regelbasierte Kopiervorgänge in einem hierarchischen Speichersystem (HSM) anstoßen. Dabei lassen sich alle IBM Speicherprodukte, wie auch Tape Libraries, integrieren. Eine Virtualisierung von Tape Libraries mit dem IBM SVC ist allerdings nicht möglich.

SVC Redundanz

Der redundante Aufbau des SVC stellt eine höchstmögliche Verfügbarkeit der Speicherumgebung sicher. Die Linux-Cluster-Lösung besteht aus mindestens zwei IBM System x-Servern. Alle kritischen Elemente sind doppelt ausgelegt, was das Risiko von Systemausfällen weitgehend minimiert. In jedem der zum Cluster gehörigen Server läuft ein Linux-Kernel, das von IBM auf Virtualisierungsbedürfnisse angepasst wurde und nicht von außen modifizierbar ist. Notwendige Veränderungen werden wie ein Firmware-Update behandelt, das der Kunde selbst oder ein Dienstleister durchführt. Die Virtualisierungssoftware läuft genau wie optionale Copy-Service-Routinen unter Kontrolle dieses Linux-Kernels.

Insgesamt stellt der SVC eine **vollständige Appliance-Lösung** dar, die Hardware, Software und Managementkonsole beinhaltet. Jeder der zwei zu einem so genannten Node-Paar gehörigen IBM System x-Server ist mit einem 2-Prozessorsystem **mit 2 x 2.4-GHz-Intel-Prozessoren** ausgestattet, hat 8 GB Cache und vier 4-Gbit/s-FibreChannel-Ports (2145-8G4). Diese neuen Prozessoren stehen seit Mai 2007 zur Verfügung und erhöhen die Leistungsfähigkeit von 160000 I/Os auf bis zu 276000 I/Os. Über die Managementkonsole können alle Administrationen ausgeführt und im Fehlerfall auch notwendige Analysedaten ausgelesen werden. Enthalten sind auch zwei Komponenten zur unterbrechungsfreien, batteriegepufferten Stromversorgung (UPS).

Die Implementierung des SVC vollzieht sich wie bei anderen Geräten, die neu ins SAN eingebunden werden. Anpassungen im SAN-Zoning und die Zuweisungen der Datenpfade dürften für einen erfahrenen SAN-Administrator nichts Neues bedeuten. Zur Datenaufnahme in der virtuellen Ebene führt der Administrator zunächst einen Imagemode mit jeder einzelnen LUN durch. Dies bedeutet, dass die bisherigen LUNs eines Servers dem SVC zugeordnet und von hier als virtuelle LUN an den jeweiligen Server weitergereicht werden. Im nächsten Schritt können diese LUNs einer anderen ‘Managed Disk Group’ zugeordnet werden. Das hat zur Folge, dass die Daten auf die dieser Gruppe zugeordneten MDs verlagert werden. Dies erfolgt im laufenden Betrieb und ist für den jeweiligen Server transparent.

SAN Volume Controller SVC: 3 Knoten im Cluster

Ein Highlight beim SAN Volume Controller besteht in der **hohen Performance**. Der SVC erfüllt alle Skalierungsanforderungen, indem ein Cluster heute auf bis zu acht Knoten erweitert werden kann. Falls ein Cluster mit einem Node-Paar an Performance- oder Kapazitätsgrenzen stößt, ist im laufenden Betrieb eine Erweiterung um jeweils ein weiteres Node-Paar (bis zu einem Maximum von vier Node-Paaren pro Cluster) möglich. Jeder Knoten des IBM SVC hat einen 8 GB großen Hauptspeicher, der fast vollständig als Cache genutzt wird. Bei **bis zu acht Nodes pro Cluster** ergibt sich somit eine **Cache-Größe von bis zu 64 GB**. Die Nutzung des Cache im IBM SVC bringt Performance-Vorteile gegenüber traditionellen Speicherlösungen, da sämtliche Speicherzugriffe gegen den Cache des SVC laufen. Damit verliert die Performance der angeschlossenen Speichersysteme an Bedeutung, weil ein Großteil der I/Os durch den vorgeschalteten Cache des SVC bedient wird.

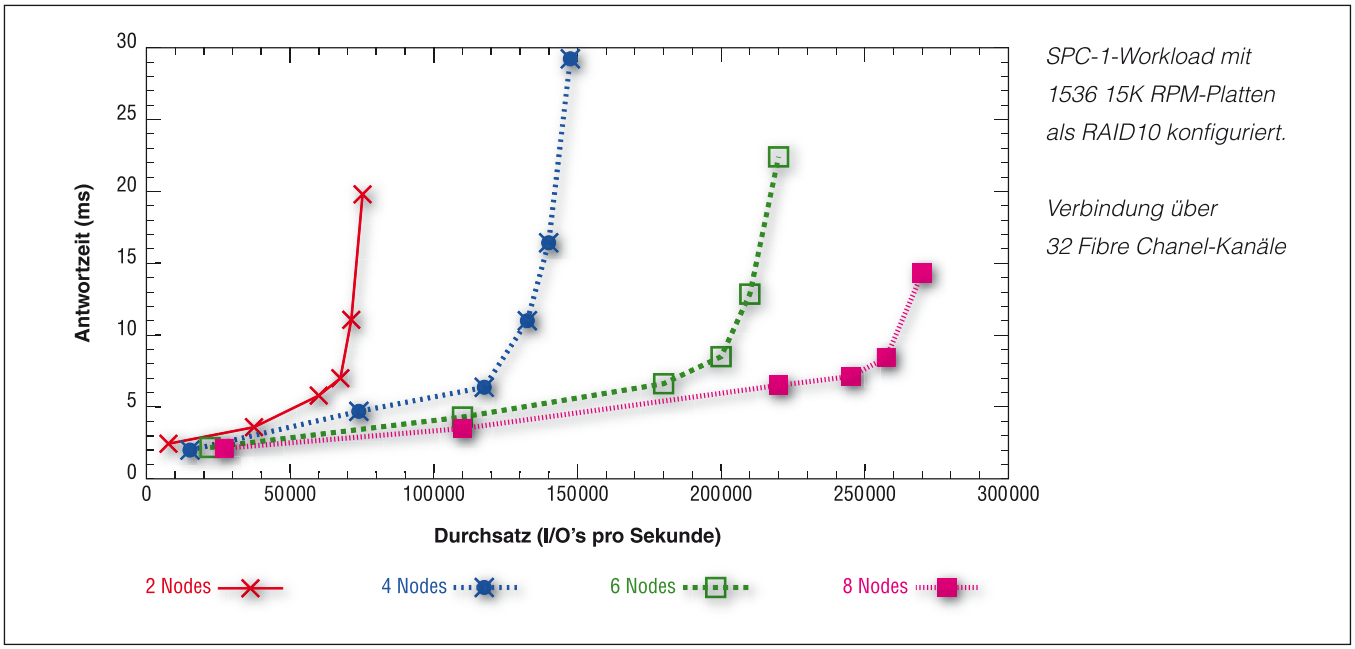
Damit erfüllt der IBM SAN Volume Controller alle Performance-Anforderungen von marktgängigen Systemen. Unterstrichen wird dies vom SPC- (Storage Performance Council) Benchmark-Test der gleichnamigen Herstellervereinigung, den alle Anbieter von Speichersystemen freiwillig durchführen können.

SVC Standortübergreifende Spiegelungen

Eine hardware-basierende Datenspiegelung (synchron oder asynchron) als typische Funktion zur Erfüllung von Business-Continuity-Vorgaben ist normalerweise herstellerübergreifend nicht möglich. Der SVC führt eine Datenspiegelung in der Virtualisierungsschicht durch. In einer virtualisierten Speicherumgebung ist damit die Datenspiegelung nun auch zwischen Systemen unterschiedlicher Bauart und Hersteller möglich. Im Desasterfall lässt sich schnell vom primären auf den sekundären Speicher umschalten. Dies erfordert in der Regel einen Eingriff durch den Administrator, was durch vorher erstellte Skripts unterstützt oder automatisiert werden kann.

SVC Spiegelungsoptionen

Insgesamt laufen beim SVC alle Spiegelungsaktivitäten zwischen verteilten Standorten zumeist über Glasfaser ab. Bei großen Entfernungen können SAN-Router zum Einsatz kommen, die auf der einen Seite das FC-Protokoll in ein IP-Protokoll und auf der Gegenseite wieder in ein FC-Protokoll umwandeln. Alternativ bieten einige Hersteller hierfür auch die DWDM- (Dense Wave Division Multiplexing) oder CWDM- (Coarse Wave Division Multiplexing) Technologie an. Mit diesen Geräten ist eine Bündelung von verschiedenen Protokollen (etwa IP, FC, FICON, ESCON) auf die Mindestzahl von physischen Verbindungen (in der Regel Dark Fibre) möglich.



SPC-1-Workload mit 1536 15K RPM-Platten als RAID10 konfiguriert.

Verbindung über 32 Fibre Chanel-Kanäle

*Diese Grafik beschreibt eindrucksvoll die nahezu lineare Skalierung eines SVC Clusters von 2 Knoten über 4, 6 bis 8 Knoten. Es waren bei diesen Tests 1536 15K RPM Laufwerke in RAID10 notwendig, um einen 8 Knoten SVC in die Sättigung zu bringen. Damit hält der **SVC den Rekord im SPC-1 Benchmark**. Die Benchmark Ergebnisse können unter **www.storageperformance.org** abgerufen werden.*

SVC Kopierfunktionen (FlashCopies)

FlashCopies eignen sich zur Backupunterstützung von Datenbanken und Filesystemen. Ein FlashCopy wird als Point in Time Copy ausgeführt und erzeugt im Moment seiner Aktivierung eine Festplatte, die eine exakte Kopie der Quellfestplatte zum Startzeitpunkt ist.

Mit dem Software Release **SVC Version 4.2**, das im **Mai 2007** zur Verfügung gestellt wurde, kam die neue Funktion **Multi Target FlashCopy** hinzu, die es erlaubt, 16 Kopien einer Festplatte zu erzeugen, d. h. von einer Source 16 Targets abzubilden.

Mit dem **Software Release SVC Version 4.2.1 vom Oktober 2007** kündigte IBM die Funktion **des Incremental FlashCopy** und des **Cascaded FlashCopy** an. Incremental bedeutet, dass eine bestehende FlashCopy auf den aktuell Stand gehoben werden kann. Es werden dann nur die bis dahin angefallenen Änderungen der Ursprungsfestplatte kopiert. Cascaded bedeutet, dass man auch von der Kopie weitere Kopien erzeugen kann. Das ist insbesonders dann wichtig, wenn man eine inkrementelle FlashCopy speichern möchte, bevor man sie fortführt. In Summe können bis zu 16 Kopien einer Ursprungsfestplatte erzeugt werden. Konsistenzgruppen werden in allen Modi unterstützt.

Die Version SVC Version 4.2.1 bietet zudem die Möglichkeit, Copy Service-Kapazitäten von 256 TB per I/O-Gruppe auf 1024 TB zu steigern und eröffnet jetzt die Möglichkeit, eine maximale Kapazität von **8 PetaByte** (vorher war es nur 2 PB) per SVC Cluster abzubilden.

Installationsbasis Oktober 2007

Weltweit wurden bis zum Oktober 2007 mehr als **10000 SVC Knoten** ausgeliefert. Diese Knoten sind in mehr als 3400 SVC Systemen im produktiven Einsatz.

Backup-Konzepte

In den letzten Jahren war der Preisverfall von Plattenlaufwerken sehr groß und mit den Technologien SATA, FATA und SAS sind neben hochperformanten FibreChannel Platten Billigvarianten auf dem Markt, die viele glauben lassen, dass es aus Performance-Gründen sinnvoll ist, Backups von Online-Platten von Tape Libraries auf diese günstigen Plattenvarianten zu verlagern. Doch hier ist Vorsicht geboten! Eine Platte ist ein völlig anderes Medium als eine Bandkassette. Platten haben Vor- und Nachteile und auch Tapes haben Vor- und Nachteile!

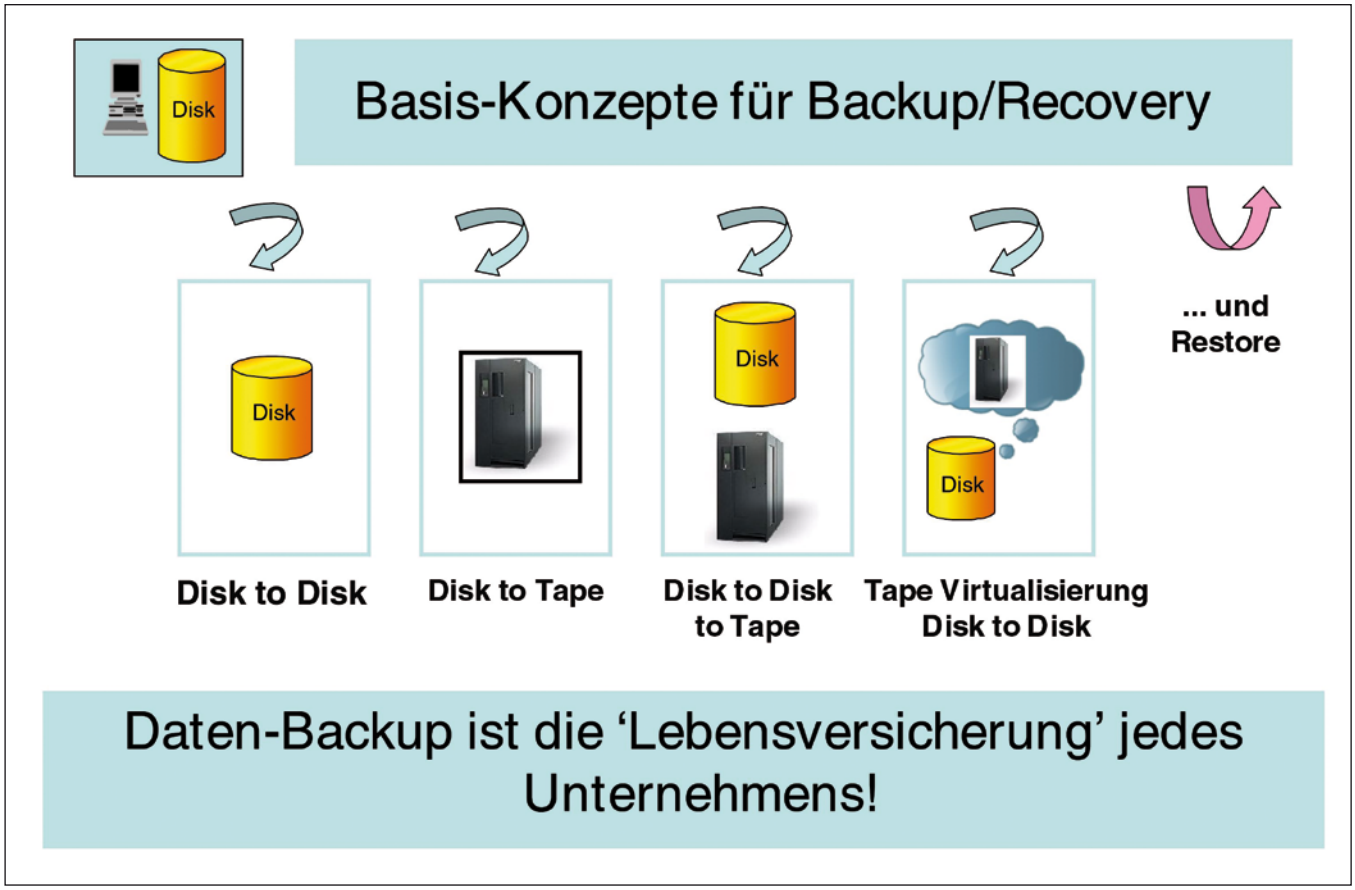
Eine Platte bietet direkten Zugriff und Vorteile beim Single File Restore und Filesystem Restore. Ebenso bietet sich die Möglichkeit, mit ‘multiple Streams’ auf der Platte zu arbeiten, um so die Backup-Zeitfenster zu reduzieren. Platten haben aber genauso Nachteile! Sie bieten keinen Vorteil bei großen File Backups oder Full Image Backups (hier ist das Tape meistens wesentlich schneller). Platten können nicht bewegt, also nicht als bewegliche Datenträger eingesetzt werden, sie benötigen Strom und kosten bei entsprechend hohen Kapazitäten mehr als Tape-Lösungen. Auch stellt sich die fundamentale Frage: Welche Plattentypen setzt man für das Backup ein? Sollen es ATA-, SATA-, FATA-, SAS- oder FC-Platten sein? Was ist das richtige Preis-Leistungs-Verhältnis in Relation zur Zuverlässigkeit und Sicherheit?

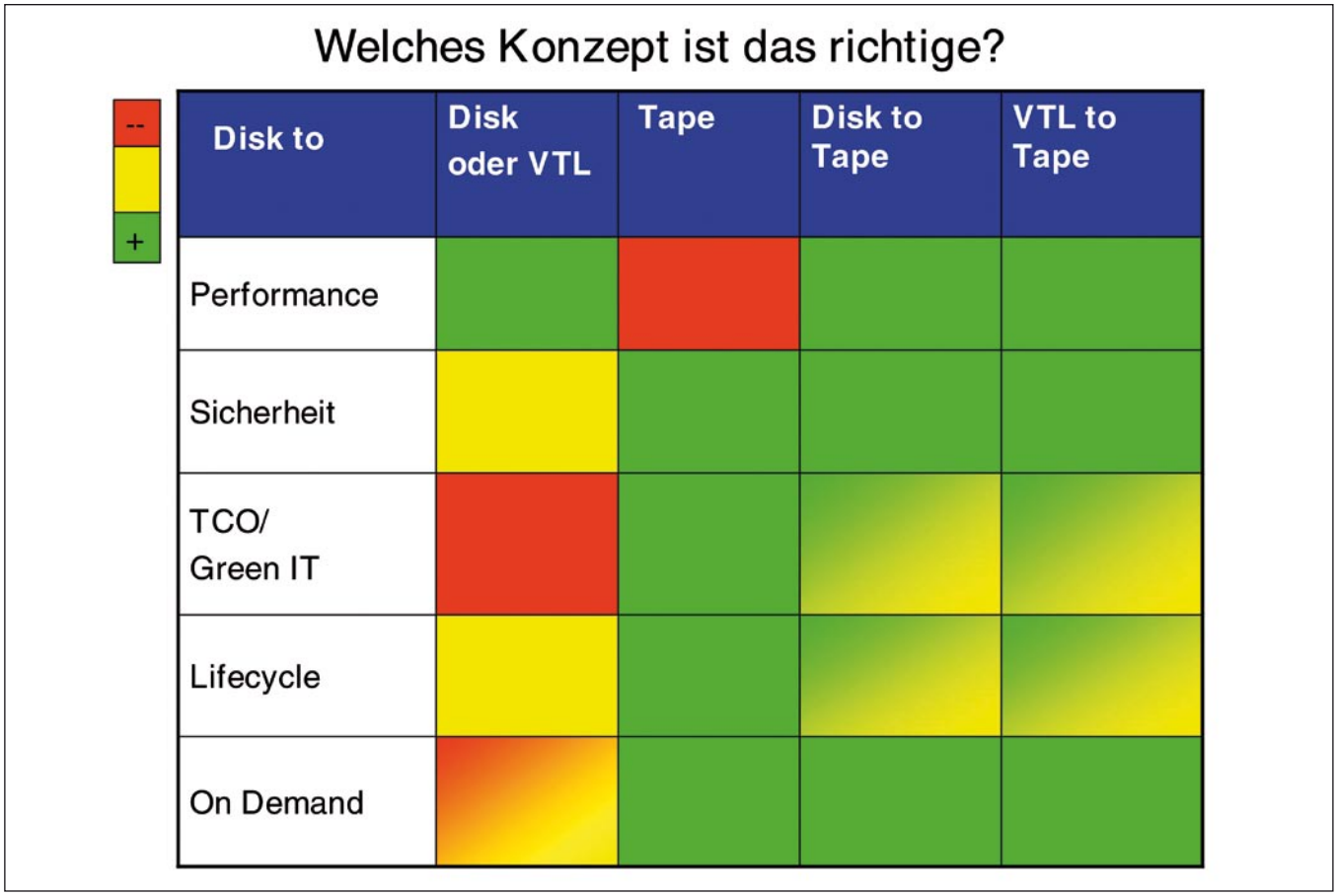
Tapes dagegen sind sequentielle ‘Streaming Devices’ und heute in der Datenrate meistens schneller als Platten. Deshalb ist es in der Regel performanter, große Files oder große Datenbanken direkt auf native Tapes zu sichern. Tapes sind ‘On Demand’, skalieren also in Performance und Kapazität, indem einer Tape Library entweder zusätzliche Laufwerke oder zusätzliche Kassetten zugeführt werden. Tapes sind austauschbare Datenträger, d. h. sie können in Sicherheitsräume ausgelagert oder zwischen unterschiedlichen Lokationen ausgetauscht werden. Tapes benötigen keinen Strom, wenn sie einmal beschrieben sind, und sind wesentlich länger haltbar (LTO als ECMA Standard bis zu 30 Jahren). Neben überschreibbaren Kassetten stehen heute nahezu in allen Tape-Technologien auch WORM-Bänder (Write Once Read Many) zur Verfügung, die sich besonders im Umfeld der Langzeitarchivierung einsetzen lassen.

Da beide Technologien Vor- und Nachteile aufweisen, bietet es sich an, in Backup-Konzepten die Stärken beider Technologien zu bündeln und richtig zu nutzen. Dies ist insbesondere deshalb sinnvoll, weil heutige Bandlaufwerke so extrem hohe Datenraten erreichen, dass deren Ausnutzung nur bei der Übertragung von sehr großen Files möglich ist. Durch einen den Bandlaufwerken vorgeschalteten Plattenpuffer in Form eines entsprechenden Plattensystems oder einer VTL (Virtual Tape Library, die prinzipiell aus einem Plattensystem besteht) ergibt sich die Möglichkeit, die vorhandene Ressource Bandlaufwerke maximal auszunutzen, auch wenn es sich um kleine zu sichernde Files handelt. Die Kombination von Platte und Band kann sehr individuell an die Anforderungen im Backup-Bereich sowie an die vorhandene Infrastruktur angepasst werden, sodass effektive Backups und Restores sichergestellt sind. In diesem Konzept werden die Disks als Backup-Puffer (Staging Area) benutzt, um kürzlich zurückliegende Backups für eine gewisse Zeit online zu halten und dann nach Ablauf einer gesetzten Frist auf Tape auszulagern. Dies gilt auch bei Virtualisierungslösungen von Tape: Hier werden Tape-Laufwerke auf Disk emuliert, d. h. Disks stellen sich wie ein Bandlaufwerk dar! Integrierte Tape-Virtualisierungslösungen kombinieren Hardware und Software zu einer einfach bedienbaren Virtuellen Tape Library (VTL). Erst die Möglichkeit, auch auf physische Tapes zu migrieren, bietet eine operative, kostengünstige und effiziente Lösung.

Die Grundkonzepte solcher Backup-Lösungen sind Disk to Disk (D2D) Backup, klassisch Disk to Tape (D2T), Disk to Disk to Tape (D2D2T), Disk to VTL und Disk to VTL to Tape, wie in der Grafik aufgezeigt. Darüber hinaus können solche grundlegenden Backup-Konzepte variiert und miteinander kombiniert werden. Welches Konzept dann die höchste Effektivität hat, ist in vielen Fällen von der vorhandenen Infrastruktur abhängig und wie das Backup durchgeführt werden soll. Denn hier gibt es auch zwei fundamentale Grundkonzeptionen, zum einen das klassische ‘Backup über das LAN’ und zum anderen ‘LAN-free Backups’, bei denen die zu sichernden Daten über ein SAN (Storage Area Network) übertragen werden.

Wird mit einer Kombinationslösung aus Plattte und Band gearbeitet, stellt sich die Frage: Wie groß muss das vorgeschaltete Plattensystem von der Kapazität her sein, wie muss die Tape Library dahinter konfiguriert werden und wie implementiert man die Migration von Platte auf Band sinnvoll!





Es gibt VTLs auf dem Markt, die keine Tape-Anbindung haben. Da stellt sich diese Frage natürlich nicht! Allerdings ist von solchen Konzeptionen dringend abzuraten, da es in eine technologische Einbahnstraße führt, aus der man nur mit sehr viel Aufwand und Kosten wieder herauskommt. VTLs ohne Tape-Anbindung mit sehr günstigen ATA- oder SATA-Platten müssen in der Regel nach drei Jahren aus Sicherheitsgründen komplett ausgetauscht werden, um Plattenausfälle und Datenverlust zu vermeiden. Sind hier sehr große Kapazitäten betroffen, kommt es in der Regel zu einer regelrechten Kostenexplosion!

Welches Konzept bezüglich den Anforderungen Leistung, Sicherheit, Kosten, Haltbarkeit und einfache Erweiterbarkeit die beste Effektivität bietet, ist in der Grafik 'Welches Konzept ist das richtige?' dargestellt. Die Farbe rot reflektiert geringe Effektivität, gelb ist neutral und grün steht für sehr gut. Man sieht sehr deutlich, dass Kombinationslösungen aus Platte und Tape die höchste Effektivität erreichen, also die Konzepte von 'Disk to Disk to Tape' und 'Disk to VTL to Tape'. Die beiden grün/gelben Felder beider Konzepte (TCO und Lifecycle) können durch sinnvoll verwendete kleinere Plattenkapazitäten mehr in Richtung grün gebracht werden oder durch eine zu große unvernünftige Auslegung des Plattenpuffers auch in Richtung gelb und gar rot tendieren.

Konzeptionelle und neutrale Beratung gewinnt maßgeblich an Bedeutung. Nur ein der Infrastruktur, den Backup-Mengen und den zu erwartenden Recall-Aktivitäten angepasstes Sizing bringt das Ergebnis, welche Konzeption die sinnvollste und effektivste ist. Diese wichtige neutrale Beratung kann in der Regel nur von Firmen durchgeführt werden, die alle konzeptionellen Möglichkeiten im Angebotsportfolio haben.

Für alle genannten Lösungen (Backup to Tape, Backup to Disk und Backup Disk to Disk to Tape und Tape Virtualisierung) bietet IBM System Storage in Kombination mit IBM Tivoli Produkten eine breite Auswahl an Backup-Tools und Strategien für jede Unternehmensgröße, Komplexität und Ansprüche an das Backup. Vom kleinen Autoloader bis zum Information Lifecycle Management-Konzept mit automatischem Backup und – ggf. Speicherung auf WORM- (Write Once Read Many) Medien zur Erfüllung gesetzlicher Richtlinien zur Archivierung – kann über das IBM Portfolio alles realisiert werden. Damit ist IBM als Lösungsanbieter einmalig!

Virtualisierung von Bandsystemen.

Neue Server-basierende Tape-Virtualisierung für System z

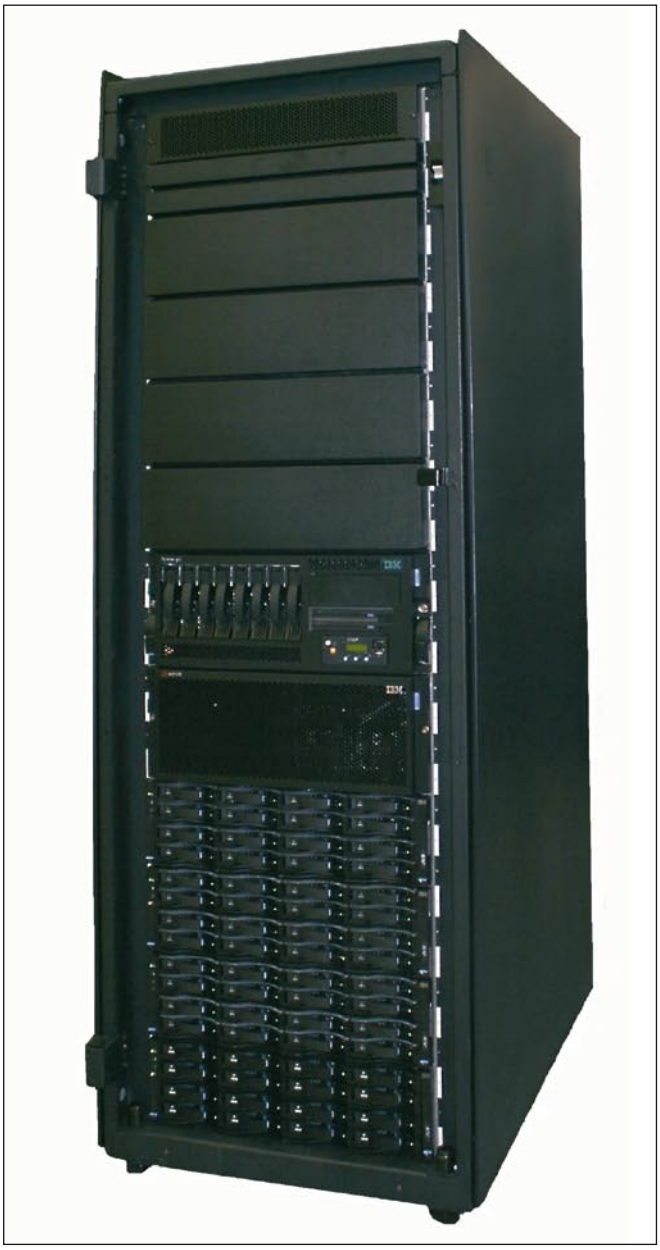
Am **29. August 2006** kündigte die IBM eine neue Virtualisierungslösung für die z/OS-Umgebung an. Die **TS7700** wird als Nachfolger der IBM 3494 B10 und B20 Virtual-Tape-Server Ende September 2006 verfügbar. Dies war insbesondere auch deshalb notwendig, weil die alten Einheiten B18, B10 und B20 nicht der ab 1. Juli gültigen EU-Richtlinie RoHS entsprachen. Die neue TS7700 Serie für die Bandvirtualisierung im Mainframe-Umfeld bildet eine **neue, langfristige Basis** durch eine modulare und skalierbare Server-Architektur. Man spricht von einer **'Distributed Node Architecture'**.

Die TS7700 besteht aus dem 3592 Gehäuse F05, der TS7740 verteilten Node-Architektur mit unterschiedlichen Nodes (siehe unten), die entsprechende Funktionen wahrnehmen, einschließlich zwei aktiv geschalteter 'I/O-Drawer', die entsprechende FICON-Anschlüsse zum Host, Fibre-Channel-Anschlüsse zu den Platten und Ethernet-Verbindungen zum Library Manager zur Verfügung stellen, dem TS7740 Cache Controller als RAID-Steuereinheit und entsprechenden RAID-basierenden Platteneinschüben, die den TVC (Tape Volume Cache) als Plattenpuffer reflektieren.

Für die Plattenpufferkapazitäten werden RAID Arrays mit 146-GB-Platten eingesetzt. Der RAID Controller arbeitet mit 16 integrierten Platten, eine TS7740 Cache-Erweiterung beinhaltet ebenfalls 16 Platten. In die Gehäuseeinheit sind neben dem RAID Controller bis zu drei Plattenerweiterungseinheiten integrierbar. Damit können bis zu 64 Platten in die Grundeinheit eingebaut werden. Die Gesamtkapazität liegt bei 6 TB (native) und 18 TB (komprimiert).

Nodes sind funktionale Aufgaben, die als Applikation auf einem aus zwei Dual Core 64-bit, 1,9 GHz Prozessoren bestehenden pSeries-Server laufen. Das vNode (v steht für Virtual) stellt dem Host virtuelle Laufwerke zur Verfügung und ist für die Host-Kommunikation verantwortlich. Das hNode (h steht für Hierachical Storage Management) ist für die Verwaltung der logischen Volumes im Plattenpuffer zuständig. Es managed auch die Migration von Platte auf Band und ist für die gesamte Verwaltung der physischen Band-Ressourcen einschließlich des Tape-Reclamation-Prozesses zuständig. Die Kombination aus vNode und hNode wird als gNode (g steht für general) bezeichnet und reflektiert die Virtualisierungseinheit des Systems. Dieser Aufbau erlaubt einfache zukünftige Skalierung bezüglich der Kapazitäten und der Leistung, indem mehrere gNodes oder auch nur vNodes im Cluster miteinander gekoppelt werden.

Die Ankündigung enthält ein 'SOD' (Statement of Direction), nachdem ein Cluster mit zwei gNodes verfügbar sein wird. Das sorgt für erhöhte Verfügbarkeit, erhöhte Performance sowie mehr virtuelle Devices innerhalb des Clusters und ermöglicht unterbrechungsfreie Code-Updates. Bei späteren Erweiterungen können dann vNodes und hNodes auf unterschiedlichen Systemen laufen und so horizontal innerhalb des Clusters skalieren. Zukünftig vorstellbar ist, dass die vNodes dann unterschiedliche Aufgaben wahrnehmen, z. B. Virtualisierung für Open Systems oder Datenarchivierung.



IBM TS7700 Virtual Tape Server für zSeries

Zwei Cluster können heute in einem **Dual Cluster Grid** (oder auch Multi Cluster Grid) zusammengeschlossen werden und dienen dazu, entsprechende Disaster-Recovery-Lösungen aufzubauen. Sie sind in der Funktionsweise dem Peer to Peer VTS mit den Modellen B10 und B20 vergleichbar (synchrone und asynchrone Spiegelung). Dazu ist das neu angekündigte TS7700 Grid Communication Feature notwendig.

Seit August 2007 sind Konfigurationen von drei Clustern in einem Grid möglich. Die Architektur selbst erlaubt langfristig bis zu acht Cluster in einem Grid und wird in einer späteren Ankündigung bekannt gegeben.

Der Unterschied der Spiegelung liegt darin, dass bei der TS7700 dedizierte Ethernet-Verbindungen (IP) eingesetzt werden. Mit zwei 1-Gbit-Ethernet-Verbindungen wird die Spiegelung durchgeführt und sie ist wesentlich leistungsfähiger im Vergleich zu Peer to Peer VTS mit FICON. Das senkt die Kosten, nutzt offene Standards und vereinfacht die Infrastruktur.

Die Einheit bietet neben dem 6-TB-Plattenpuffer (18 TB komprimiert) einen maximalen Durchsatz von 600 MB/s (Peak) und damit knapp das Doppelte einer 3494-B20. Es stehen bis zu vier 4-Gbit-FICON-Anschlüsse in dieser Basiseinheit zur Verfügung. Unterstützt ist die TS3500 Library (3584 Library) über die Anbindung des IBM 3953 Library Managers Modell L05. Die Einheit unterstützte im Backend TS1120 Laufwerke im Emulation-Mode der 3592 J1A. Seit Februar 2007 ist auch die ‚Native'-Unterstützung verfügbar. Bis zu 16 physische Bandlaufwerke TS1120 können an einer TS7700 betrieben werden.

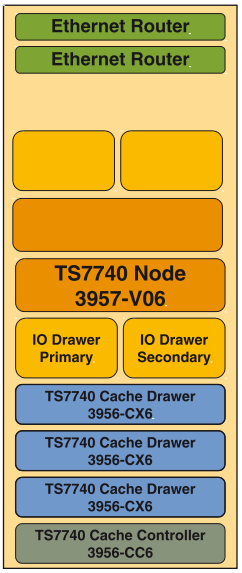
Zum z/OS-Host werden wie bei den alten VTS-Modellen 3490-E Laufwerke abgebildet. Funktional bietet die neue Einheit Funktionen, die bei dem Modell 3494-B20 zur Verfügung standen. Dies sind vor allem die Funktionen des APM (Advanced Policy Management) wie 'Volume Pooling', 'Dual Copy', 'Logical Volume Sizes' und 'Multiple Reclamation Policies' (Beschreibung der Funktionen unter 3494-B20), die bei TS7700 als Outboard Policy Management bezeichnet werden.

Die TS7740 Node emuliert zum Host viele 3490-E Laufwerke, die für den Host wie physikalisch vorhandene Laufwerke sichtbar sind. Die TS7740 ist also völlig transparent

TS7700 Aufbau

- **System pSeries Server**
  POWER5 Prozessor (2 Dual Core CPUs)
- **Zwei I/O Expansion Drawer**
- **RAID Disk Controller und drei Disk Expansion Drawer**
- **Redundante Network Router**
  Stellen interne Netzwerkverbindungen zum Library Manager und zur Konfiguration des Disk Controllers zur Verfügung

  Stellen geschütztes NAT Interface für den Zugriff auf die Dienste des Management Interfaces, die auf dem Controller laufen, zur Verfügung
- **Platz für einen zweiten 3957 Controller und I/O Drawer (Statement of Direction)**



IBM 3952 Frame Modell F05

zum Betriebssystem, zum Bandverwaltungssystem und zum Katalogeintrag. Die emulierten Laufwerke bezeichnet man als virtuelle Laufwerke. Die TS7700 unterstützt bis zu 256 virtuelle Laufwerke sowie bis zu 100 000 virtuelle Volumes und dies erhöht sich später durch die Bildung von Multi Node Clusters entsprechend. In einem Dual Cluster Grid (Spiegelung) stehen 512 virtuelle Laufwerke zur Verfügung.

Geschrieben wird in den verfügbaren Plattenpuffer, den TS7740 Cache (ebenso wird bei Recalls aus dem Plattenpuffer gelesen). Ist ein Volume im Plattenpuffer gespeichert, wird dieses als virtuelles Volume bezeichnet. Durch einen Algorithmus, der als Premigration bezeichnet wird, werden, etwas zeitversetzt, Kopien der virtuellen Volumes auf die physischen Bandlaufwerke migriert und auf physische 3592 Kassetten geschrieben. Die Kopie des virtuellen Volumes, das jetzt auf einer physikalischen Kassette gespeichert ist, wird als logisches Volume bezeichnet. Viele logische Volumes auf einer Bandkassette nennt man 'Stacked Volumes'. Ein Stacked Volume ist also eine physikalische Kassette, die viele logische Volumes enthält.

Durch den Premigration-Prozess werden die Laufwerkgeschwindigkeiten der physischen Laufwerke voll genutzt. Ebenso werden die Kassetten in ihrer Kapazität maximal ausgenutzt, weil viele logische Volumes auf einer Kassette gespeichert werden. Kommt ein Recall des Hosts auf ein logisches Volume, wird der Recall aus dem Plattenpuffer bedient, solange das Original des virtuellen Volumes dort

noch verfügbar ist. Ist das virtuelle Volume im Plattenpuffer nicht mehr verfügbar, wird die physische Kassette, welche die Kopie des logischen Volumes enthält, in ein Bandlaufwerk geladen.

Das angeforderte logische Volume wird dann in den Plattenpuffer zurückgeladen und der Recall wird aus dem Plattenpuffer bedient.

Sollte aufgrund einer großen Schreib-Workload der Plattenpufferplatz (TS7740 Cache) knapp werden, werden im Plattenpuffer die virtuellen Volumes, die am wenigsten benutzt wurden (LRU Least Recently Used Algorithmus), zum Überschreiben freigegeben, nachdem deren Kopie bereits auf Band 'pre-migriert' wurde.

Der Reclamation-Prozess der physischen Kassetten findet im Hintergrund von Laufwerk zu Laufwerk statt, ohne dass der Plattenpuffer belastet wird. Es erfolgt kein Zurückladen der noch validen logischen Volumes auf der Kassette in den Plattenpuffer, sondern diese werden direkt von Laufwerk zu Laufwerk auf eine neue Kassette geschrieben. Die alte Kassette geht dann automatisch in den vorhandenen Scratch Pool zurück.

Die TS7700 bildet die neue Basis eines Mainframe-basierenden Virtual Tape Servers. Es sind viele neue Funktionalitäten, Verbesserung in Leistung und Kapazität und größere Konfigurationen in den nächsten Monaten vorgesehen, die die TS7700 zum absoluten Tape-Virtualisierungsspezialisten im Mainframe-Umfeld machen, der mit keiner anderen Lösung vergleichbar ist. Das war bereits mit den Modellen B10 und B20 der Fall, die als einzige Systeme über die APM-Funktionalität eine Verbindung ins SMS (System Managed Storage) als Teil des z/OS-Betriebssystem hatten. Dabei werden die SMS-Konstruktnamen, wie Data Class, Storage Class und Management Class sowie die Storage Group-Namen einschließlich der VolSer-Zuordnung automatisch in die Library Manager Data Base übernommen. Das bietet kein anderes System auf dem Markt.

Sowohl für die 3494-B10- und B20-Modelle als auch für die neue TS7700 ist eine noch stärkere Integration in das SMS des z/OS-Betriebssystems längerfristig vorgesehen.

Grid-Spiegelungen mit TS7700 Systemen

Bei einer Multi Cluster Grid-Spiegelung ist es erforderlich zu definieren, wie jede Storage Group eines jeden Clusters im Grid zu behandeln ist. Das Setzen einer Policy auf dem LM definiert den Consistency Point für jede definierte Site im Subsystem. Aus den Policies wird ein Array aus Consistency Point-Werten, wobei jedes Element des Arrays die Policy für eine bestimmte Site repräsentiert. Wenn z. B. für Site A eine Kopie zum Zeitpunkt des Unloads und für Site B eine Kopie zu einem späteren Zeitpunkt gewünscht wird, sieht das Array so aus: Site A hat RD und Site B hat RD.

Die Replication Policies können auf den LMs verschieden sein. Die Policy Settings auf jedem LM diktieren, welche Aktionen ausgeführt werden, wenn ein Volume auf einem virtuellen Laufwerk, das zu diesem LM gehört, gemountet wird. **Consistency Policies bestimmen** die Site, welche die Daten als erste erhält (TVC Selektion), welche Sites eine Kopie der Daten erhalten und zu welchem Zeitpunkt.

Consistency Policies werden über die Management Class konfiguriert. Dabei werden Policies einem Management Class-Namen zugeordnet. Policies werden über den ETL Specialist auf dem Library Manager gesetzt.

Eine neue Definition ist die des **'Copy Consistency Points'**, der bestimmt, wann ein Target Volume auf einer TS7700 mit dem Source Volume konsistent ist. Bei den **Consistency Policy Optionen** werden für jede Site Consistency Points gesetzt. Dabei stehen drei Definitionen zur Verfügung:

RUN (R) – diese Site benötigt eine valide Kopie des logischen Volumes, bevor Device End an den Rewind Unload- (RUN) Befehl vom Host geliefert wird (entspricht weitgehend dem Immediate Mode beim PtP VTS).

Deferred (D) – diese Site benötigt eine valide Kopie zu einem späteren Zeitpunkt, nachdem der Job beendet wurde (ähnlich wie Deferred Mode beim PtP VTS).

No copy (N) – diese Site soll keine Kopie von Volumes in dieser Management Class erhalten.

Durch diese neue Möglichkeit können extrem flexible Spiegelvarianten in Multi Cluster Grid-Konfigurationen realisiert werden.

TS7700 Erweiterungen 2007

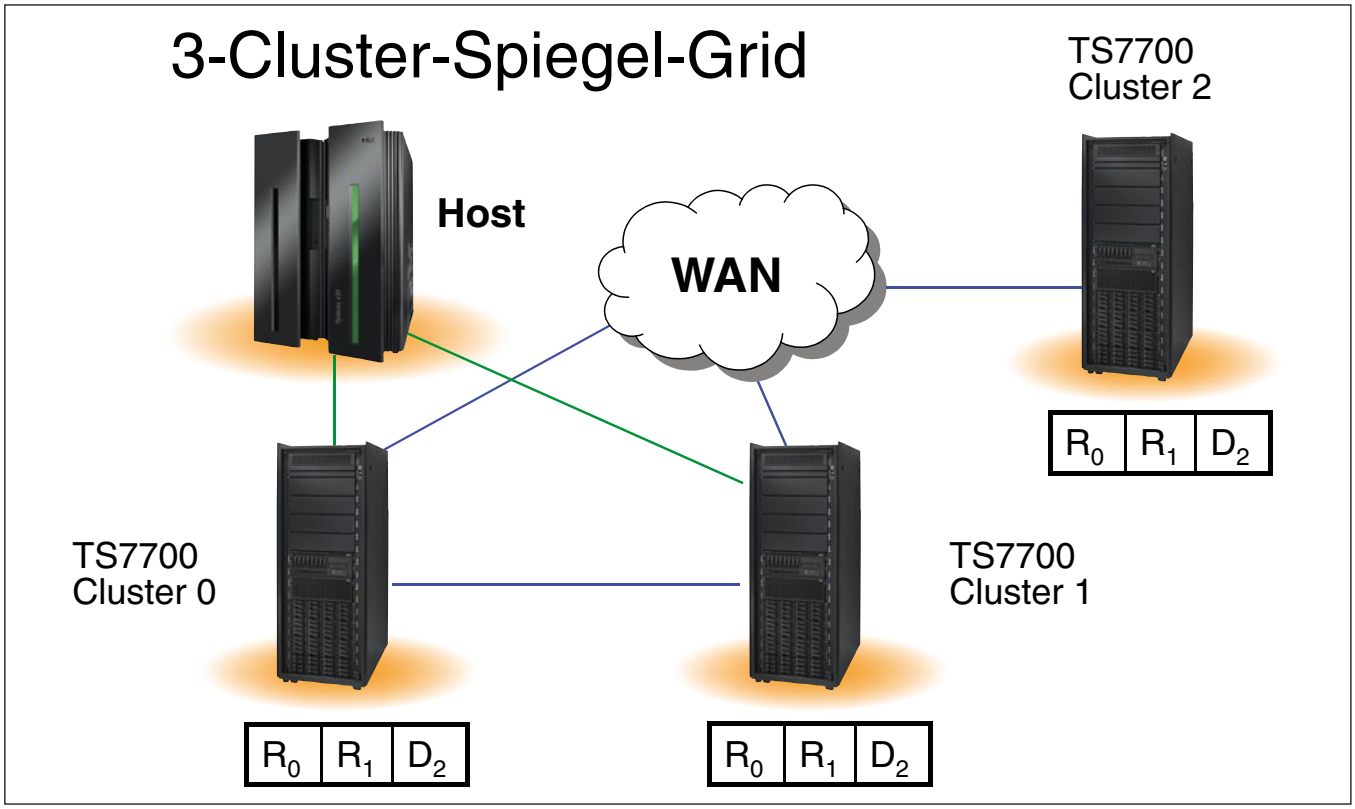
Bei der Ankündigung der TS7700 machte IBM die Aussage, dass mit vier zusätzlichen Releases im Jahr 2007 in die TS7700 Systeme eine Vielfalt von neuen funktionalen Erweiterungen integriert werden, die weit über den Rahmen des Vorgängersystems B20 hinausgehen. So wurde jedes Quartal ein neuer Release verfügbar, R1.1 im 1.Q., R1.2 im 2.Q., R1.3 im 3 Q. und R1.4 im 4.Quartal 2007. Die verfügbaren Funktionen werden in AF-basierende Funktionen (Advanced Functions) und Nicht-AF-basierende Funktionen unterschieden.

Advanced Functions (AF)

Die Funktionalität 'Logical Volume Pooling' bietet die Möglichkeit, logische Volumes einzeln pro Nummernkreis in unterschiedlichen physischen Kassetten-Pools zu verwalten. Damit ist die TS7700 mandantenfähig.

Mit **'Tape Volume Cache Management'** können logische Volumes einer Cache Management Preference Group zugeordnet werden. Es stehen zwei Preference Groups zur Verfügung. Die Preference Group steuert, wie die logischen Volumes im Plattenpuffer behandelt werden und beeinflusst die Verweildauer von virtuellen Volumes im Tape Volume Cache (Plattenpuffer). Mit **'Dual Selective Copy'** können zwei Kopien eines logischen Volumes in zwei unterschiedlichen physischen Kassettenpools erzeugt werden. Die TS7700 speichert dann zum Schutz vor einem Mediendefekt ein Duplikat eines logischen Volumes auf einer zweiten Kassette. **'Cross-site Replication'** wird in einer Grid-Konfiguration verwendet, um eine Kopie in einer zweiten Site zu erzeugen. Je nach Einstellung erfolgt dies auf synchroner oder asynchroner Basis. Es können aber auch ungespiegelte Volumes in einer Grid-Konfiguration verwaltet werden. **'Logical Virtual Volume Sizes'** dient der Auswahl von Volumegrössen, die die Standard-Volumegrößen von 400/800 MB übersteigen. Zur Wahl stehen 1000, 2000 und 4000 MB.

Secure Data Erasure mit Encryption Shredding reflektiert die Wiederkehr der bereits aus dem VTS bekannten Funktion zur vollautomatisierten 'Zerstörung' von Datenbeständen auf Leerkassetten (nach Reklamation). Das ist ein genialer Trick, wenn TS1120 Tape Encryption im Einsatz ist: 'nur' der Schlüssel wird auf der Kassette gelöscht, nicht die gesamte Band-Datenstruktur! Bei einem **3-Site-Grid** wird ein TS7700 Grid so erweitert, dass bis zu 3 Cluster für noch besseren Business Continuity Support zur Verfügung stehen (Dreifach-Remote-Spiegelung). Jetzt können in einem 3-Site-Grid zwei TS7700 zum Beispiel lokal synchron gespiegelt werden und beide Systeme dann asynchron in großer Entfernung über WAN in eine dritte TS7700.



Dreiseitiger TS7700 Spiegel-Grid, Cluster 0 und 1 sind synchron gespiegelt und Cluster 0 und 1 asynchron über eine WAN Strecke auf Cluster 2

Copy Export (Stand Alone-Systeme in R1.3, Grid-Systeme in R1.4) erlaubt es, eine Gesamt-Kopie der Daten einer TS7700 zu exportieren, z. B. für Disaster Recovery-Zwecke, während die originalen Daten weiterhin in der TS7700 verbleiben und für die Produktionssite verfügbar sind.

Copy Export ist eine neue TS7700 Funktion, zur Unterstützung des Transfers von Daten (auf stacked Tapes) in eine regionale/überregionale Sicherheitszone für Disaster Recovery-Zwecke. Dabei kann mit der Funktion Dual Selektive Copy eine Kopie von Volumes innerhalb einer TS7700 erzeugt werden, die in einem separaten Storage Pool und damit separaten Kassetten abgespeichert wird. Copy Export exportiert dann eine Kopie von ausgewählten Daten (log. Volumes), die primäre Kopie verbleibt dabei allerdings in der produktiven TS7700.

Die 'Copy Exported physical Volumes' werden weiterhin von der produktiven TS7700 verwaltet, d. h. die Volumes werden weiterhin in der TS7700 Datenbank geführt. Die TS7700 managed dabei auch den 'Offsite' Volume Reclamation-Prozess (Recycling der ausgelagerten Kassetten). Informationen über die exportierten Volumes können über einen Host Console Request und einen BVIR Request initiiert werden.

Ein 'Copy Export Recovery' wird durch das TS7700 Management Interface unter dem Service- und Troubleshooting-Menü oder per User-Id und Password mit 'role-based' Zugriff durchgeführt. Die TS7700 ist 'offline' zu den Hosts während der Recovery. Dabei wird die VolSer des Volumes benötigt, von welchem die Datenbank restored werden soll.

Es besteht auch die Kunden-Option zur Bereinigung bestehender Datensätze in einer TS7700 für den Copy Export Recovery-Vorgang (mfg cleanup). Während des Recovery-Prozesses ist das einzig angezeigte Panel im Management Interface das Status Panel zum Copy Export Recovery. Wenn die Recovery abgeschlossen ist, kann die TS7700 wieder 'online' zu den angeschlossenen Hosts genommen werden – sobald dort TCDB und TMS ebenfalls wieder restored sind.

Zusätzliche Funktionen (Nicht-AF)

Autonomic Ownership Takeover

Ownership bedeutet, dass jedes Volume in einer Grid-Konfiguration einen Owner hat (Cluster). Um ein Volume auf einem bestimmten Cluster zu 'mounten', muss der Cluster Owner sein. Ist er es nicht, requestiert die TS7740 auf dieser Clustersite den Ownership-Transfer von der anderen Clustersite. Im Falle, dass die andere Site nicht reagiert, wird über die TSSC (Total Storage Service Console) geprüft, was mit der TS7740 auf der anderen Site los ist. Ist die TS7740 tatsächlich nicht mehr verfügbar, wird der Autonomic Ownership Takeover initiiert. Der Ownership-Transfer kann auch immer manuell angestoßen werden.

Tape TS1120 Encryption Support

Mit der TS7740 wird eine Verschlüsselung auf den angeschlossenen TS1120 Bandlaufwerken unterstützt (ab Release 1.2). Die virtuellen Laufwerke haben damit nichts zu tun. Encryption wird durch die exisitierenden Storage Pools (bis zu 32 Pools) auf der TS7740 gemanaged. Der Host verwendet die existierenden Storage-Konstrukte der Storage Groups und der Management Class, um festzulegen, welche logischen Volumes auf welchen Pool physischer 3592 Kassetten kommen. Ist ein Pool definiert, wo mit Encryption gearbeitet werden muss, wird er entsprechend behandelt. Dabei kann für jeden Pool definiert werden, welche 'Public/Private' Key-Paare verwendet werden, um verschlüsselte Schlüssel auf der Kassette abzuspeichern, die mit Verschlüsselung berarbeitet wurden.

Host Console Request

Diese Option erlaubt einem MVS Console Operator das Durchführen von Abfragen und eine einfache Problemanalyse zum TS7700 Zustand, ohne Zugriff auf das Web Interface (MI) zu haben.

Enhanced Automated Read Only Recovery (im Backend)

Hier handelt es sich um die Wiederkehr der bereits im Vorgänger VTS bekannten Funktion zur vollautomatisierten Recovery von fehlerauffälligen 3592 Backend Kassetten, auch unter Nutzung der 2. Kopie in einem Grid-System.

Remote Read and Write Pipelining

Diese Funktion bringt erhebliche Performance Verbesserungen beim Lesen und Schreiben in den 'remote' Cache einer TS7700 im Grid.

FICON

Die bisherigen 128 logischen Kanäle per FICON-Kanal, die in sehr komlexen Konfigurationen mit vielen LPARs auf der System z Site zum Engpass werden konnten, wurden auf 256 logische Kanäle erweitert. Damit dürften sich keine Engpässe mehr bezüglich der Konfigurationen ergeben.

Die TS7700 ist ein Tape-Virtualisierungsprodukt für das Mainframe-Umfeld, für das aufgrund seiner neuen Architektur in den kommenden Jahren sowohl kontinuierliche Erweiterungen als auch die Integration von völlig neuen Entwicklungen vorgesehen sind.

Neue Server-basierende Tape-Virtualisierung für Open Systems

Bereits im Oktober 2005 stellte IBM mit der Virtualisierungslösung **TS7510** eine virtuelle Tape Library für den Open Systems-Bereich zur Verfügung, die sich allerdings von den anderen virtuellen Tape Libraries auf dem Markt in der Funktionalität nur wenig unterschied.

Im **April 2007** kündigte die IBM die neue **Virtuelle Tape Library TS7520 (VTL)** als Nachfolger der bisherigen TS7510 an. Die TS7520 bietet neben höherer Kapazität und Leistung einen ganz neuen Funktionsumfang, der speziell für die Anforderungen im Open Systems-Umfeld in die Maschine integriert wurde. Bevor das neue Produkt TS7520 detailiert beschrieben wird, ist es wichtig, den **Trend bezüglich VTLs im Open Systems** Umfeld näher zu beleuchten.

Nachdem Festplatten in den letzten Jahren immer günstiger wurden, besonders durch die Einführung von neuen Technologien wie beispielsweise SATA, werden Disk-basierende Backup-Lösungen immer mehr propagiert. Viele Marktbeobachter und Kunden vergleichen oft Disk-Lösungen mit alten Tape-Technologien und finden so Vorteile bei den Disk-Lösungen, wie schnelleres Backup und Restore, oder auch stabilerer und zuverlässigerer Betrieb als bei älteren Bandtechnologien. Aber die Zuverlässigkeit mit reinen Disk-Lösungen wollte sich doch nicht so einstellen, wie von vielen gewünscht! Besonders beim Einsatz der günstigen SATA-Laufwerke gibt es immer wieder Disk-Drive-Fehler bis hin zum Datenverlust, auch unter RAID-Schutz.

Tape ist ein eigenständiges Medium und bietet die Vorteile von sehr schnellem Lesen und Schreiben von großen Datenmengen, lange Aufbewahrungssicherheit und braucht außerdem keinen Strom **(Tape ist 'cool')**. Daneben kann Tape auch als auswechselbarer Datenträger eingesetzt werden. Disk ist ebenso ein eigenständiges Medium und bietet den Vorteil von schnellem direktem Zugriff mit kurzen Antwortzeiten und der Möglichkeit der Parallelität. Optimale maßgeschneiderte Lösungen bieten sich daher in der Kombination von beiden Technologien.

Heutzutage werden Disk-Backup-Lösungen hauptsächlich als Plattenpuffer für hochkapazitive und schnelle Tape Drives eingesetzt. Dabei werden Sicherungsdaten kurzfristig (ein bis mehrere Tage) auf diesen Disk-Systemen gespeichert und zeitnah oder später auf Band migriert oder kopiert. Disk-Backup-Systeme kommen auch zum Einsatz, um den Restore-Vorgang zu optimieren, sinnvollerweise allerdings nur dort, wo Platten Vorteile gegenüber Tape besitzen, also bei kleinen Files und Filesystemen, aber nicht bei Datenbanken oder Image-Backups.

Aufgrund der Klimaveränderung, der damit verbundenen Diskussion um die Verringerung von CO_2 (Green-IT) und der steigenden Energiekosten, sollten Disk-Backup-Systeme nur noch mit Bedacht eingesetzt werden und nur dort, wo es wirklich nötig ist bzw. Disk echte Vorteile gegenüber Tape bietet. Dies gilt gerade für kleinere und mittelständische Unternehmen. Disk-Backup-Systeme locken mit vermeintlich günstigen Kosten, neigen aber dazu, dass die Strom- und Klimakosten über einen Zeitraum von sechs Jahren deutlich höher sind als die Investitionskosten. Dazu kommen Kosten für Datenmigration, die bedingt durch den kurzen Lebenszyklus von Disk-Einheiten entstehen.

Neben der schon lang bestehenden Möglichkeit, den **TSM (Tivoli Storage Manager)** so einzusetzen, dass ein **LAN-Backup auf einen Plattenpuffer** läuft und später die Migration auf Tape erfolgt, bietet die IBM nun auch eine neue VTL (Virtual Tape Library) in Form des Produkts **TS7520** an. Der große Vorteil der TS7520 liegt vor allem im LAN-Free-Bereich, weil viele LAN-Free-Clients direkt auf die TS7520 sichern können, ohne dass viele physikalische Bandlaufwerke benötigt werden (ein LAN-Free-Client benötigt immer ein direktes Bandlaufwerk, sei es nun physikalisch oder virtuell!). Dabei stellt prinzipiell die TS7520 den Plattenpuffer und eine Emulationssoftware zur Verfügung, die Bandlaufwerke emuliert. Geschrieben wird aber direkt in den Plattenpuffer.

Die Migration vom TS7520 Plattenpuffer auf Band kann durch die Backup Software (z. B. TSM) erfolgen oder auch direkt – allerdings mit entsprechenden Performance-Einschränkungen – durch die TS7520 selbst.

Zunächst sind VTLs nichts anderes als Disk-Backup-Lösungen. Allerdings stellen sich VTLs als Tape Libraries mit Bandlaufwerken dar. Somit bieten VTLs Vorteile, wenn die Backup-Applikation kein natives Disk-Backup unterstützt oder die Lizenzkosten für ein solches Disk-Backup zu hoch sind.

VTLs können auch dort sinnvoll eingesetzt werden, wo viele langsame LAN-free-Sicherungen vorhanden sind, denn hier lassen sich sonst erforderliche Tape Drives einsparen. VTLs sind aber meist auch integrierte Lösungen, die den Anwender von jeglichen administrativen Tätigkeiten, die sonst bei nativen Disk-Systemen notwendig wären, frei hält.

IBM Virtual Tape Library TS7520 für Open System-Umgebungen

Ein großer Vorteil von VTL-Systemen ist die Möglichkeit, die Daten zu komprimieren und somit die Disk-Kapazität besser auszunützen. Allerdings sinkt in den meisten Fällen die Performance der Systeme bei Einsatz von Kompression. Ein weiterer Vorteil ist, dass die Daten direkt von der VTL auf Tape migriert oder kopiert werden können und dies transparent für den Backup-Server passiert.

Die IBM Tape Virtualization Engine TS7520 stellt eine **integrierte Lösung aus Hardware und Software** dar, die Bandvirtualisierung für Open Systems-Server über physische FibreChannel (FC)- und iSCSI-Verbindungen bereitstellt. Durch **System x-basierende Server (1-4)** werden Bandlaufwerke emuliert. Die Emulation der TS7520 unterstützt die Laufwerksformate LTO2, LTO3 und LTO4 sowie die 'High End'-Technologien 3592 und TS1120. Der **Plattenpuffer** wird durch 500-GB- und/oder 750-GB-SATA-Platten reflektiert. Dabei sind die Platten in einem Einschub als **RAID5 Array** abgebildet. Die 16 Platten im Einschub reflektieren dabei ein 6+P und ein 7+P RAID5 Array (P= Parity Platte) sowie eine Hot Spare-Platte, die für beide Arrays im Fehlerfall als Rebuild-Platte einspringt. Die Kapazitäten können mit **500-GB-Platten auf bis zu 884 TB** und mit **750-GB-Platten auf bis zu 1.3 PB** (native Kapazitäten) ausgebaut werden. Die kleinste kapazitive Einstiegsgröße beginnt bei 6.5 TB. Erweiterungen können – je nach Plattenart – in 6.5-TB-Schritten oder 9.75-TB-Schritten erfolgen.

Die TS7520 bietet eine **einmalige Leistungsfähigkeit**, indem bis zu 4800 MB/s Durchsatzmenge ermöglicht werden (4800 MB/s beim Lesen vom Plattenpuffer und bis zu 4500 MB/s beim Schreiben in den Plattenpuffer). Sowohl von der Leistungsfähigkeit als auch von den kapazitiven Möglichkeiten gibt es derzeit kein anderes VTL-System auf dem Markt, das diese Möglichkeiten bietet.

Die TS7520 lässt sich so konfigurieren, dass bis zu vier Server 'geclustered' werden können. Dabei werden bis zu 512 virtuelle Libraries, bis zu 4096 virtuelle Bandlaufwerke und bis zu 256000 virtuelle Kassetten emuliert. Zu den Hosts können bis zu 32 FC-Anschlüsse auf Basis von 4 Gbit-Fibre-Channel konfiguriert werden. Für die Tape-Anbindung stehen bis zu 16 FC-Anschlüsse, auch auf Basis von 4 Gbit-Fibre, zur Verfügung.

Die Epoche der Server-basierenden Speichersysteme mit neuen Infrastrukturmöglichkeiten

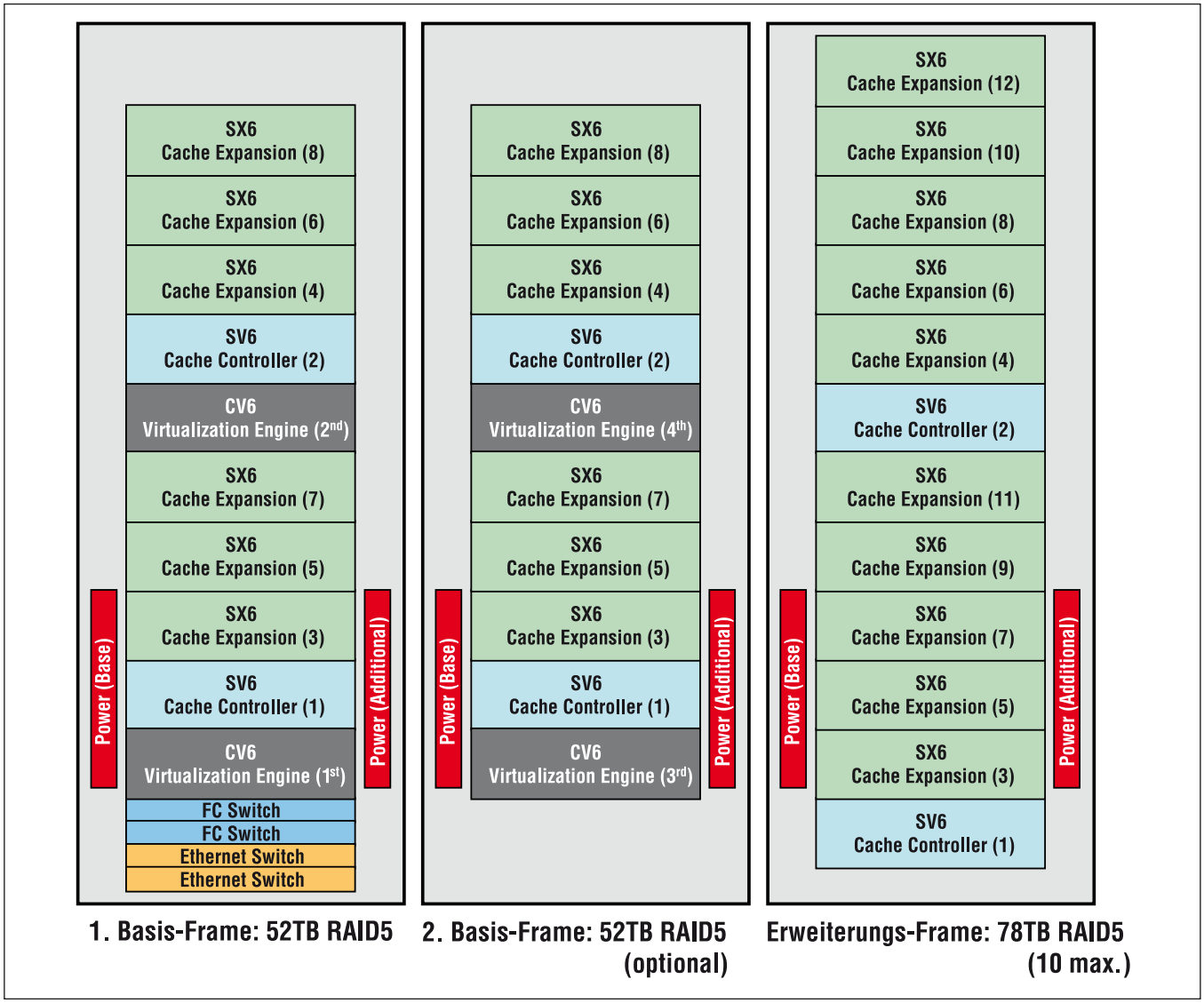
Neben der hohen Flexibilität, Leistungsfähigkeit und Kapazität bietet die TS7520 einen Funktionsumfang, die derzeit keine andere VTL auf dem Markt übertrifft. Auf diese Funktionalität wird später noch ausführlich eingegangen.

Bei einer maximalen Konfiguration der TS7520 werden bis zu **12 Gehäuse-Einheiten** benötigt. Wichtig dabei ist die Tatsache, dass selbst bei kleinen Konfigurationen wegen Sicherheits- und Leistungsanforderungen immer mit zwei Platten-Controllern (SV6) gearbeitet wird (Enterprise Edition). Sowohl die Platten-Controller als auch die Virtualization Engines (CV6) sind über ein integriertes SAN auf Basis von 4-Gbit-FibreChannel miteinander verbunden. Dieser hochredundante Aufbau zeigt sich auch in der Stromzufuhr. Jede Gehäuseeinheit ist mit zwei Stromversorgungen ausgestattet.

Als Einstiegsversion steht die **TS7520 Limited Edition** zur Verfügung. Diese günstige Einstiegsoption steht für Einstiegskonfigurationen mit einer Virtualization Engine (CV6) und nur einem Platten-Controller (SV6) zur Verfügung und unterstützt Plattenkapazitäten bis zu 29.25 TB (maximal). Die Limited Edition enthält standardmässig nicht den ganzen Funktionsumfang. Tape Caching ist z. B. ein kostenpflichtiges Feature.

Die **Enterprise Edition** unterstützt jede Art von TS7520 Konfiguration bis zum Maximalausbau und bietet Tape Caching standardmässig an.

Die IBM TS7520 wurde als komplette Appliance entwickelt und alle Einzel-Komponenten aufeinander abgestimmt. Das komplette System wurde mit allen Funktionen getestet und ausführlichen Performance-Messungen unterworfen.



Aufbau einer TS7520 in einer Konfiguration mit drei Gehäuseeinheiten

Für IT-Betreiber stellt sich die IBM TS7520 als eine oder mehrere 'virtuelle' Tape Libraries dar und man muss sich nicht um die Anbindung sowie die Konfiguration des Disk-Storage-Systems kümmern. Die IBM TS7520 VTL ist als eine Einheit zu betrachten und auch der Support wird dementsprechend durchgeführt, d. h. es werden keine Einzelkomponenten, sondern immer die gesamte Appliance betrachtet und gewartet (z. B. für Firmeware-Updates). Somit kümmert sich auch nur ein Supportteam um die IBM TS7520. Eine zentrale Stelle nimmt die Kundenanfragen entgegen. Eine Voranalyse des Problems durch den Kunden, wie es bei Implementierungen, die aus Einzelkomponenten zusammengestellt wurden der Fall ist, entfällt.

Mit **bis zu 48 x 4-Gbit-FC-Anschlüssen** bietet die IBM TS7520 die höchste Anzahl von 'Connectivity'-Möglichkeiten.

End-to-End-4-Gbit-Support bedeutet, dass die Maschine in ihrer gesamten Anbindung zu den Hosts als auch zum Disk und Tape Backend eine ausbalancierte Leistung für Front End, Disk Cache und Tape Backend realisiert. Dabei werden bis zu 256 physische Backend Tape-Laufwerke (IBM TS1120 J1A/E05 Drives und/oder LTO2, LTO3 und LTO4 Drives) sowie bis zu 32 physische Backend Tape Libraries (IBM TS3500 Library, IBM TS3310 Library, IBM TS3200 Library, IBM 3582 Library, IBM 3583 Library und IBM 3494 Library) unterstützt.

Während manche andere VTL-Anbieter für die Anzahl an emulierten Libraries, Drives und Volumes extra Lizenzkosten berechnen, ist bei der TS7520 immer alles inklusive.

Die IBM TS7520 bietet eine selbstständige Call-Home-Funktionalität zum IBM Service-Team. Vergleichbare VTLs bieten nur eine 'E-Mail'-Call-Home-Funktionalität an, so dass nur intern, ohne Benachrichtigung des Herstellers, über Fehler informiert wird.

Als großes Alleinstellungsmerkmal bietet die IBM TS7520 als einziges VTL-System auf dem Markt einen **Multipath Device Driver für FC Failover und Loadbalancing**. Die Maschine kann problemlos in ein Dual-Fabric integriert werden. Einen Dual-Fabric-Support bieten alle auf dem Markt verfügbaren VTLs an, doch nur die IBM TS7520 bietet zusätzlich Failover und Loadbalancing. Dies wird durch den Multipfadtreiber gewährleistet.

Ohne einen Multipath Device Driver können die einzelnen virtuellen Tape Drives zwar mühsam und manuell über die Frontend FC Ports der VTL verteilt werden, doch ein Loadbalancing kann dadurch nicht erreicht werden, weil die Backup-Applikation, wie z. B. TSM, von der Verteilung der virtuellen Tape Drives über mehrere FC Ports keine Informationen bekommt. Alle Backup-Applikationen verteilen die Workload auf die verfügbaren Tape Drives nach 'Round-Robin'-und/oder 'last recently used'-Algorithmen. Dadurch kommt es aber in Verbindung mit VTLs vor, dass einzelne FC Links overloaded sind, und gleichzeitig manche Links ohne Last sind.

Besonders das **Loadbalancing** ist für eine ausgewogene und hohe Performance bei VTLs enorm wichtig. Bei Disk-Systemen ist ein Loadbalancing schon seit langem eine Standardfunktion. Bei VTLs, die eigentlich ein Disk-System sind bzw. Disk-Backup anbieten, ist dies nicht der Fall, obwohl die meisten VTLs mehrere FC Ports im Frontend anbieten. Der **IBM Multipath Device Driver** für Tapes verteilt die Workload auf die verfügbaren FC Links, so dass die einzelnen Links eine **ausgewogene Workloadverteilung** haben. Zusätzlich ermöglicht der IBM Multipath Device Driver ein automatisches Failover, falls ein Link oder eine Fabric ausfällt oder auch nur für Wartungsarbeiten offline ist.

Spezielle **Disk Caching Algorithmen** beschleunigen den Recall von Tape, wenn das logische Volume sich nicht mehr im Plattenpuffer befindet. Werden VTLs eingesetzt, schreibt die Backup-Software die einzelnen Files in einen Plattenpuffer. Im Falle einer Migration von Platte auf Band wird das logische Volume in einer 1:1-Kopie auf eine Kassette gespeichert. Im Falle eines Restores einer einzelnen File, die sich nicht mehr im Plattenpuffer befindet, müssen viele VTLs das komplette logische Volume in den Plattenpuffer zurückladen, um von dort den Host-Request zu bedienen. Das kann je nach emulierter Kassette (man bedenke: eine LTO4-Kassette faßt 800 GB Kapazität) sehr lange dauern, oft bis zu zwei Stunden und länger! Der **Caching Algorithmus der TS7520** arbeitet in solchen Fällen wesentlich intelligenter! Anstatt des Zurückladens des gesamten logischen Volumes in den Plattenpuffer wird die zu 'restorende' File direkt von Tape ausgelesen (ohne Zurückladen) und dem Host zur Verfügung gestellt. Das spart wertvolle Zeit bei Restore-Aktivitäten.

Es besteht die Möglichkeit, mehrere TS7520 VTLs **über IP-Verbindungen asynchron zu spiegeln**. Dabei wird nach dem 'demount' einer Kassette, zu einem bestimmten Zeitpunkt oder intervallmässig, eine Kopie des logischen Volumes über IP zu einer anderen TS7520 übertragen. Bevor auf erzeugte kopierte Volumes zugegriffen werden kann, müssen diese Volumes durch einen administrativen Eingriff freigegeben werden, damit aktiv mit ihnen gearbeitet werden kann. IP Replication kann nicht gleichzeitig mit Tape Caching durchgeführt werden.

Die TS7520 unterstützt die **Encryption-Möglichkeiten für Backend Tape Drives** (TS1120 oder LTO4). Bei dieser HW-Encryption, die in diese Laufwerke integriert ist, sind keine Leistungsbeeinträchtigungen zu erwarten. Für Tape-Laufwerke, die nicht mit Verschlüsselungstechnik arbeiten, bietet die TS7520 die Möglichkeit, die Daten auch über Software Encryption-Methoden zu verschlüsseln, bevor sie auf das physikalische Tape geschrieben werden.

Die TS7520 ermöglicht neben FC-Anbindungen auch iSCSI-Anbindungen, wenn Kunden iSCSI im Einsatz haben oder haben wollen.

Über die Funktionalität des **Hosted Backups** besteht die Möglichkeit, die Backup-Software auch direkt auf der TS7520 zu installieren. Dies ist allerdings nur zu empfehlen, wenn keine allzu großen Leistungsanforderungen an den Backup-Server gestellt werden.

Die Weiterentwicklung beider Systeme, TS7700 für den Mainframe und TS7520 für Open Systems, ergibt die Möglichkeit, die Mainframe-Umgebung mit einem Virtualisierungsspezialisten individuell mit den vorhandenen Möglichkeiten im z/OS optimal zu bedienen sowie der heterogenen Open Systems-Umgebung und deren Anforderungen optimal gerecht zu werden. Beide Spezialisten können an derselben Library TS3500 (3584) betrieben werden. Seit August 2007 ist auch die ältere 3494 Library unterstützt.

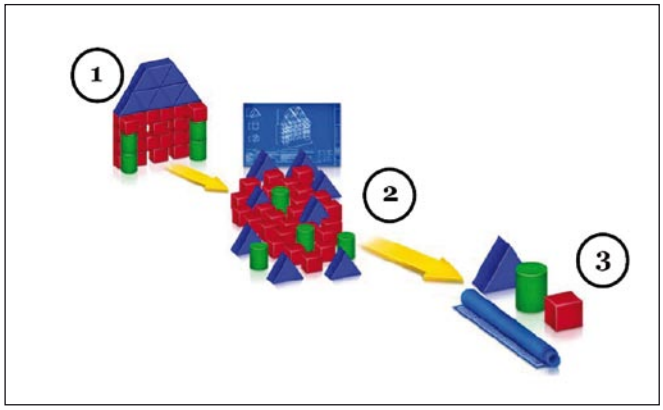
Data De-Duplication

Das Thema De-Duplication wurde im Jahr 2007 zu einem oft diskutierten Thema. De-Duplication ist **nichts Neues**. Es handelt sich um einen **mathematischen Algorithmus**, der Bit-Block-Vergleiche durchführt. Das einzige Ziel ist die Vermeidung von Duplikaten. Dabei gibt es die unterschiedlichsten Ansätze und Verfahrensmöglichkeiten. Man kann solche Verfahren im Betriebssystem etablieren. Ein Beispiel dafür ist das z/OS mit der Funktion Hyper PAV. Auch über Software Tools sind solche Verfahren möglich, wie z. B. Analyse Tools, ILM, TPC oder Common Store oder über Vergleichsalgorithmen in SW und HW. Heutige De-Duplication-Verfahren werden vor allem beim Backup auf Virtuelle Tape Libraries (VTLs) eingesetzt, weil die meisten Duplikate beim Backup und bei der Archivierung erzeugt werden.

Das erste professionelle De-Duplication-Verfahren führte IBM bereits im Jahr 2006 für das Produkt **IBM Common Store** ein. Common Store führt diese Vergleiche bei der E-Mail-Archivierung auf Mail-Basis und/oder Attachment-Basis durch und stellt sicher, dass eine E-Mail und/oder ein Attachment nur einmal archiviert wird. De-Duplication ist ein Feature von Common Store und ist nur in Verbindung mit dem IBM Content Manager als Backend Repository verfügbar.

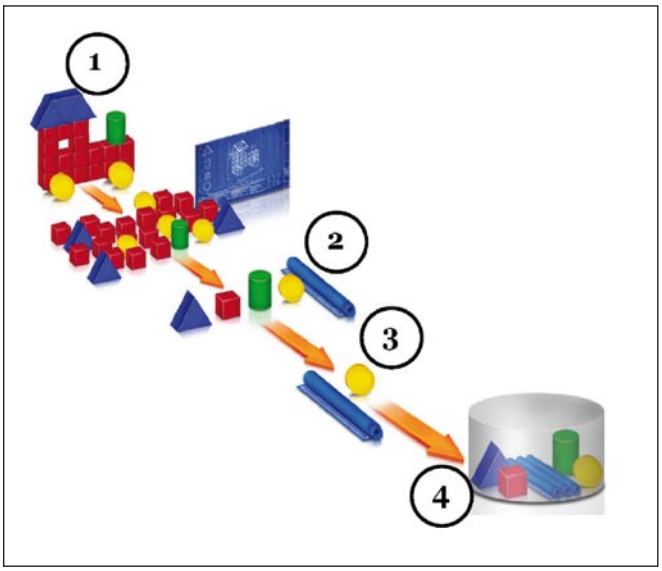
Der De-Duplication-Ansatz findet vor allem Freunde bei **Virtuellen Tape Libraries (VTLs)**. Der mathematische Vergleichsalgorithmus läuft dabei auf der VTL mit und führt Bit-Block-Vergleiche durch. Die Einsparungen an Plattenplatz können dabei durchaus den Faktor 10 – 20 erreichen. Es sind verschiedene Ansätze verfügbar: über Software, über Microcode mit Hardware oder eine Kombination von Software und Hardware. Man unterscheidet grundsätzlich zwei Verfahren. Beim **Online-Verfahren** werden die Vergleiche durchgeführt, bevor der Bit-Block auf Platte abgespeichert wird. Hier ergibt sich das Problem der Skalierbarkeit und Leistung. Werden zuviele TB mit De-Duplication bearbeitet, gehen die VTLs sehr schnell in der Leistungsfähigkeit zurück, weil der Rechner nur noch den De-Dup-Algorithmus durchführt. Die Empfehlung der Anbieter ist allgemein, nicht mehr als 15 – 20 TB mit De-Dup zu bearbeiten. Das andere Verfahren ist das **Post Processing-Verfahren**, wobei der Algorithmus nachgelagert stattfindet. Man schreibt die Blöcke zuerst ohne De-Dup auf Platte und führt die Vergleiche anschließend durch. Dafür wird dann aber zusätzlicher Speicherplatz benötigt.

Um De-Duplication besser zu verstehen, sehen wir uns folgendes einfaches Beispiel an:



Unsere abzuspeichernde File ist das hier abgebildete Häuschen (1). Es wird in seine Bauteile zerlegt (2). Es besteht aus roten Würfeln, grünen Zylindern und blauen Dächlein. De-Duplication speichert jetzt jedes Bauteil nur einmal ab. Den drei Bausteinen wird noch eine Aufbauanleitung beigefügt, wie das Haus wieder aufzubauen ist und wieviele Bausteine dafür von jedem Element benötigt werden (3).

Unsere zweite abzuspeichernde File ist in Form einer Lokomotive (1) dargestellt:



Hier finden wir dieselben Bauteile, die bei der ersten File enthalten waren: der rote Würfel, der grüne Zylinder und das blaue Dächlein. Ein weiteres zusätzliches Element kommt hinzu, das gelbe Rad. Allen vier Bauteilen wird wieder eine Aufbauanleitung beigefügt (2). Da der rote Würfel, der grüne Zylinder und das blaue Dächlein bereits abgespeichert sind, müssen wir nur noch das gelbe Rad und die Aufbauanleitung abspeichern (3). Für beide Files, das 'Haus' und die 'Lokomotive', wurden nur vier Bauelemente und zwei Aufbauanleitungen abgespeichert. Das Kritische an diesem Verfahren ist: Wenn eine Aufbauanleitung verloren geht, gibt es keine Möglichkeit der Rekonstruktion!

De-Duplication-Verfahren werden heute von den **Firmen Diligent, Data Domain, Quantum, FalconStore** und **Network Appliance** angeboten. Auch für die IBM Nseries steht De-Duplication in Form der Funktion A-SIS (Advanced Single Instance Storage) zur Verfügung. Leider sind die heutigen Verfahren noch nicht ausgereift und müssen so weiterentwickelt werden, dass sie hochperformant sind und unbegrenzt in der Kapazität skalieren können.

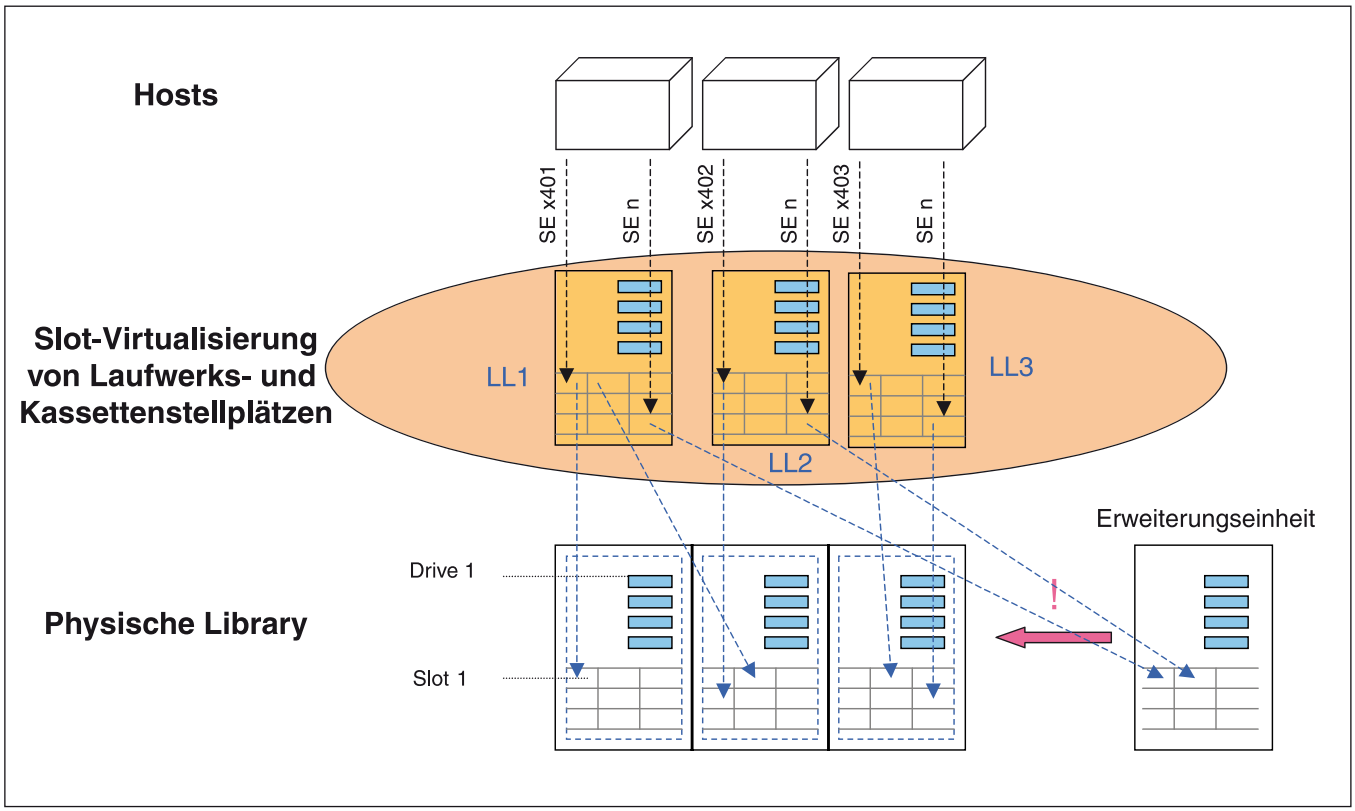
Die IBM arbeitet schon länger an einem eigenen De-Duplication-Verfahren mit dem Ziel, einen Algorithmus zur Verfügung zu stellen, der zum einen hochperformant ist und zum anderen unbegrenzt skaliert. Bereits im Januar 2004 wurde im IBM Labor Almaden in Kalifornien das Projekt **'Bit Block Mathematics'** ins Leben gerufen. Die Gruppe besteht aus Microcode-Spezialisten und Mathematikern. Seit Januar 2007 wird die Gruppe vom Lehrstuhl für diskrete Mathematik in Bonn unterstützt. Das erste Ziel ist, einen leistungsfähigen und hochskalierbaren Algorithmus zur Verfügung zu stellen, der sowohl in Hardware- als auch in Software-Lösungen integrierbar ist. IBM plant diesen Algorithmus im Laufe des Jahres 2008 in die Backup Software TSM (Tivoli Storage Manager) zu integrieren und zur Verfügung zu stellen.

Virtualisierung von Tape Libraries

TS3500 (3584) Library-Virtualisierung mit ALMS (Advanced Library Management System)

Eine einmalige Funktionalität bietet die TS3500 mit ALMS (Advanced Tape Library Management System). ALMS steht für die TS3500 schon seit April 2006 zur Verfügung, wurde aber erst richtig im Jahre 2007 genutzt. Mit ALMS wird die physische Library virtualisiert, d.h. die HW wird von den Hosts abgekoppelt. Zu den Hosts werden virtuelle Laufwerke und virtuelle Kassettenslots dargestellt. Dies ermöglicht eine noch nie dagewesene Flexibilität, weil logische Partitionen als logische Libraries im laufenden Betrieb dynamisch verändert, also verkleinert oder vergrößert, werden können.

Die Hosts, also die Backup-Applikationen, sehen völlig transparent virtuelle Laufwerke und virtuelle Kassetten-Slots so, als ob sie physisch vorhanden wären. ALMS zieht zwischen der Library Hardware und den Hosts eine virtuelle Schicht ein, die es erlaubt, im laufenden Betrieb Hardware-Erweiterungen vorzunehmen, logische Partitionen umzukonfigurieren, zu verkleinern und zu vergrössern, ob es nun die Laufwerke oder die Slots betrifft. ALMS ist eine einmalige Funktionalität, die es nur mit der TS3500 Library gibt. ALMS bietet vier wichtige Funktionalitäten für die TS3500 Library.



Storage Slot und Drive-Virtualisierung mit ALMS

1.) Dynamisches Partitioning
Mit ALMS können im laufenden Betrieb Partitionen erstellt oder verändert werden. Die physische Position in der Library ist dabei nicht mehr relevant. Dies bezieht sich auf Kassetten- und Laufwerksstellplätze.

2.) Laufwerkssharing
Mit ALMS können Laufwerke verschiedenen Backup-Servern zugeordnet werden und für den Backup-Server online gesetzt werden, der sie gerade benötigt.

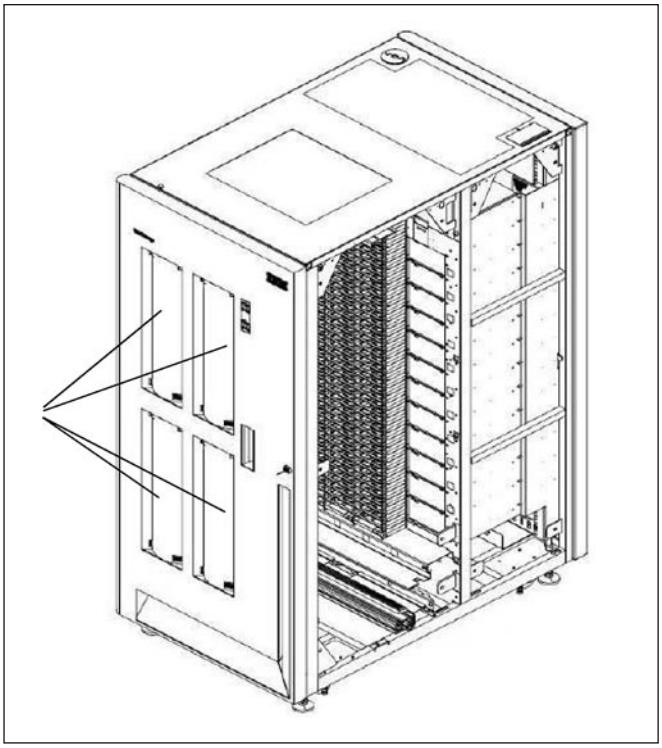
3.) Virtual I/O
ALMS bietet eine größere logische I/O-Station. Die Funktion 'Virtual I/O' ermöglicht es, dem Backup-Server eine virtuelle I/O-Station von bis zu 255 Slots darzustellen. Das vereinfacht das Kassettenmanagement beim Ein- und Auschecken.

4.) Overallocation/Underallocation:
Mit ALMS ist eine flexible Darstellung der Library-Größe möglich und somit Over- und Underallocation. Overallocation verringert den Administrationsaufwand bei einer Library-Erweiterung. Underallocation spart Lizenzkosten, z. B. im Legato-Umfeld.

TS3500 Library Hardware-Erweiterungen 2007
Neben der Virtual I/O-Option des ALMS besteht seit **Juni 2006** die Möglichkeit, D-Frames mit **zusätzlichen physischen I/O-Stations** zu versehen. Bis zu vier I/O-Stations mit jeweils bis zu 16 I/O-Slots können in eine D-Frame-Tür integriert werden.

Die TS3500 unterstützt drei solchermaßen ausgestattete D-Frames innerhalb der Library. Zusammen mit der I/O Station im L-Frame mit 32 I/O Slots, drei D-Frames mit bis zu 64 I/O Slots, bietet die TS3500 die Möglichkeit, **224 physische Kassetten** in die Library als 'Bulk' einzulagern oder als 'Bulk' aus der Library zu entnehmen.

Für die TS3500 stehen **flexible Modellkonvertierungen** zur Verfügung. So können alte Frames wie z. B. L22 und D22 mit TS1120 Laufwerken in die neuen L23 und D23 Frames umgebaut werden. Dasselbe gilt für die alten LTO Frames L52 und D52, die in die neuen LTO Frames L53 und D53 umgebaut werden können.



TS3500 Library 4 I/O Station D-Frame

Bei den neuen Frames können nun TS1120 Frames (L23, D23) in LTO Frames (L53, D53) umgebaut werden. Ebenso können die neuen LTO Frames (L53, D53) in neue TS1120 Frames (L23, D23) konvertiert werden. Damit bietet die TS3500 eine sehr flexible Möglichkeit, bestehende Frames auf neu gewünschte Ziel-Frames zu konvertieren.

Die TS3500 Library bietet mit der Funktionalität des **Data Gatherings** die Möglichkeit, Informationen über die unterschiedlichsten Dinge, die in der Library ablaufen, zu erhalten.

Die **Laufwerk-Statistik** enthält Informationen über die letzten Mounts jedes Laufwerks (nicht mit LTO1 möglich), die **Port Statistic** enthält FibreChannel Port-Informationen der letzten Mounts (nicht mit LTO1 möglich), die **Mount History** zeigt die Mount-Statistik der letzten 100 Kassetten, die 'demounted' wurden und die **Library Statistics** gibt Aufschluß über eine effektive Auslastung der logischen Libraries, d. h. es wird deutlich, ob manche logische Libraries überdurchschnittlich genutzt werden. Dadurch können die Partitionen geändert werden, um die Library Performance zu verbessern. Dies ist nur mit den neuen Frames möglich. Zusätzlich kann eine Vielzahl von Informationen wie **Residency Max Time** (maximale Mountzeit einer Kassette), **Residency Avg Time** (durchschnittliche Mountzeit einer Kassette), **Mounts Total** (Anzahl der gesamten Mounts), **Mounts Max Time**

(Maximalzeiten der Mounts), **Mounts Avg Time** (Durchschnittszeiten der Mounts), **Ejects Total** (Anzahl der Kassettenausgaben), **Ejects Max Time** (Maximalzeiten von Kassettenausgaben), **Ejects Avg Time** (Durchschnittszeiten von Kassettenausgaben) und **Total Inputs** (Anzahl der Kassetten-Inserts, die über die I/O Station der Library zugeführt werden) jeweils der letzten Stunde, abgerufen werden.

Tape Library-Virtualisierung mit dem Enterprise Removable Media Manager (eRMM, iRMM)

Der Enterprise Removable Media Manager bildet ein Virtualisierungslayer zwischen Datensicherungs-Anwendungen und der Tape Library Hardware. Durch die Entkoppelung von Datensicherungsanwendungen und Tape Library Hardware wird größtmögliche Flexibilität und Skalierbarkeit der Tape Library-Infrastruktur erreicht und der Funktionsumfang von Libraries mit SCSI Media Changer, wie z. B. IBM TS3500, durch erweitertes Media Management ergänzt. eRMM (seit **August 2007 als Programmpaket** iRMM – integrated Removable Media Manager – verfügbar) bietet die effizientere Nutzung von Drives und Libraries durch dynamisches Sharing und Pooling der Bandlaufwerke. Es erlaubt **zentrales Management von Bändern und Bandlaufwerken**, die von verschiedenen Anwendungen genutzt werden und bietet so größtmögliche Flexibilität und Skalierbarkeit der Tape Library-Infrastruktur. Das Ergebnis sind verkürzte Backup-Laufzeiten durch flexible Tape Device-Zuordnung und eine deutliche Ressourceneinsparung bei der Tape Library-Infrastruktur.

In modernen Umgebungen mit Speichernetzen sind Server und Tape Libraries über ein Speichernetz miteinander verbunden. Damit sind die technischen Voraussetzungen für das effiziente Sharing von Tape-Ressourcen zwischen verschiedenen Backup- und Archivierungsservern nur bedingt erfüllt. Für Tapes fehlt eine Abstraktionsschicht, die mit einem Volume Manager (SAN Volume Controller) oder einem Dateisystem (SAN File System) vergleichbar ist. Das führt dazu, dass Tape-Ressourcen nur sehr eingeschränkt von mehreren Anwendungen gemeinsam genutzt werden können. Die Datensicherungsanwendungen verlangen meist exklusiven Zugriff auf einen Teil der Ressourcen. Es fehlte bisher eine zentrale Verwaltung von Tape-Ressourcen, welche die Voraussetzung für ein anwendungsübergreifendes Sharing ermöglichen. Mit eRMM (iRMM) macht IBM eine Lösung verfügbar, die genau diese Anforderung erfüllt. Dieses Programmprodukt wurde in der IBM Lokation in Mainz entwickelt.

TS3400 Minilibrary mit TS1120 Laufwerken für Mainframe und Open Systems

TS3400 neue Minilibrary für High End Tape-Technologie

Im Februar 2007 kündigte IBM eine kleine Library **TS3400** an, die wie die große TS3500 mit den TS1120 High End-Laufwerken arbeitet. Die Library wurde speziell für die High End Tape-Laufwerke TS1120 (Jaguar-Technologie) als Einstiegslösung konzipiert. Diese kleine Library kann sowohl im Open Systems-Umfeld als auch im Mainframe-Umfeld (über den TS1120 FICON Controller seit August 2007) betrieben werden. Die TS3400 bietet eine native Kapazität von 12.6 TB an.

In die TS3400 können **1 bis 2 TS1120 FC-Bandlaufwerke** mit 4-GB/s-Dual-Port-FibreChannel-Anschlüssen integriert werden. Man kann alle Kassetten, die ein TS1120 Laufwerk verarbeitet, verwenden (100-, 500- und 700-GB-Kassetten). Die Library unterstützt sowohl überschreibbare als auch WORM-Kassetten und bietet direkte Encryption und Encryption Key Management-Unterstützung an, eine Verschlüsselungstechnik, die für die TS1120 Laufwerke im Herbst 2006 zur Verfügung gestellt wurde. Die Library hat 2 herausnehmbare austauschbare Kassettenmagazine. Jedes Magazin kann bis zu 9 Kassetten aufnehmen. Das untere Magazin kann so konfiguriert werden, dass 3 Slots im Magazin als I/O-Station dienen. Das obere Magazin ist mit zwei Reinigungskassettenslots konfigurierbar. Die Ausstattung mit Barcode-Leser ist Standard.

Die TS3400 bietet eine Partitionierung von bis zu 2 logischen Libraries. Jede logische Library enthält ein Laufwerk und ein Magazin und kann im Sequential (Autoloader) Mode oder Random (Library) Mode arbeiten.

Die TS3400 ist über das lokale Operator Panel oder remote über ein Web GUI administrier- und managebar. Die Library kann 'Stand Alone' oder 'Rack Mounted' (5U) betrieben werden. Die Speicherkapazität bei Verwendung der 700-GB-Kassetten geht bis 12.6 TB (bis 37.8 TB mit 3:1 Kompression).

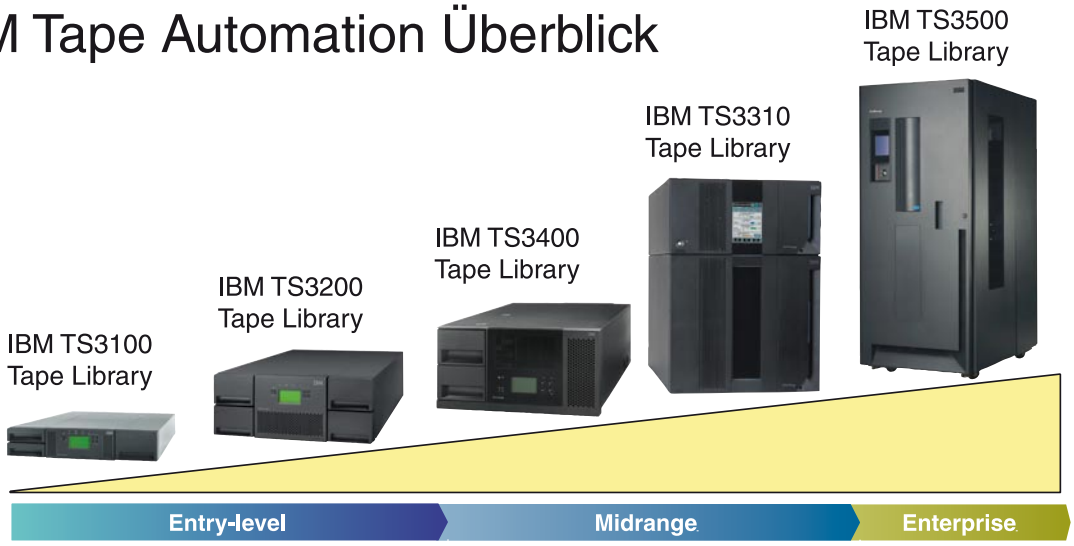
Die TS3400 ist eine hochredundante Library und mit redundanter Stromversorgung, 'Hot Swappable'-Laufwerken und Netzteilen, einem doppelten Stromanschluss sowie Control Path und Data Path Failover-Möglichkeit ausgestattet. Die Library kann an IBM Systeme i, p, x und zLinux sowie an z/OS via TS1120 Controller angeschlossen werden. Außer den IBM Systemen werden die Betriebssystem-Plattformen von HP, Sun und Microsoft Windows unterstützt.

Damit stellt IBM die Möglichkeit zur Verfügung, große TS3500 Libraries mit TS1120 Laufwerken in einem zentralen Rechenzentrum zu betreiben, während Außenstellen mit derselben Technologie in Form einer Minilibrary arbeiten können. Dies ermöglicht den Datenträgeraustausch in Form von 3592 Kassetten zwischen Rechenzentrum und Außenstellen. Um hier höchste Sicherheitsaspekte zu gewährleiten, bieten die TS1120 Laufwerke die Möglichkeit, die Kassetten in verschlüsselter Form zu beschreiben (siehe auch unter Tape Encryption).

Mit der Verfügbarkeit der TS3400 erweitert sich das IBM Library-Portfolio in der Weise, dass kleine bis ganz große Lösungen sowohl mit TS1120 Technologie als auch mit LTO-Technologie ermöglicht werden.

Untenstehende Grafik zeigt im Überblick das komplette Library-Angebot der IBM im Jahre 2007, vom Autoloader TS3100 bis zur High End Tape Library TS3500. Die Einstiegslösungen TS3100 und TS3200 kommen von der Firma BDT und werden per OEM als IBM Logo Produkte vertrieben. Beide Entry-Lösungen bieten LTO-Laufwerkstechnologie an. Eine Einstiegslösung für die High End-Laufwerke TS1120 bietet die TS3400 Library, die neben Open Systems auch an System z angeschlossen werden kann. Als Midrange Library für LTO-Laufwerke hat die IBM die TS3310 im Portfolio, eine Library die von Quantum/Adic gebaut wird und per OEM als IBM Logo Produkt vetrieben wird. Quantum selbst vertreibt diese Library als Scalar 500i auch selbst. Der Produktunterschied liegt darin, dass die IBM TS3310 partitionierfähig ist und 'End to End' Path Failover unterstützt. Im High End-Bereich verwendet die IBM die TS3500, die IBM selbst produziert und konstant weiterentwickelt. Die TS3500 ist derzeit die schnellste Library auf dem Markt und bietet viele Alleinstellungsmerkmale und Vorteile gegenüber vergleichbaren Libraries. Das angebotene Library-Portfolio ist für einen Hersteller/Anbieter einmalig und bietet Tape Automation für alle Anforderungen.

IBM Tape Automation Überblick



LTO3	8.8 TB	17.6 TB		up to 159 TB	up to 2.7 PB
LTO4	17.6 TB	35.2 TB		up to 318 TB	up to 5.5 PB
TS1120			12.6 TB		up to 4.3 PB

Kapazitäts-Angaben sind 'native' ohne Kompression

File Virtualisierung und File Area Networks (FAN)

Global Parallel File System (GPFS) und Scale Out File Services (SOFS)

Unter Verwendung von erprobten IBM Baugruppen wie GPFS, BladeCenter und Storage stellte IBM im November 2007 ein hochskalierbares und hochperformantes Dateisystem mit integrierten ILM-Funktionalitäten vor. Das System ist in der Lage, linear mit den Anforderungen an Speicherplatz und Durchsatz zu wachsen und bietet neben extremer Hochverfügbarkeit eine einfache Administrierbarkeit.

IBM Global Parallel File System (GPFS)

IBM GPFS (Global Parallel File System) ist ein Cluster-fähiges Dateisystem. Es läuft nativ in Linux- und AIX-Umgebungen und ist sowohl für den technisch-wissenschaftlichen als auch kommerziellen Bereich geeignet, beispielsweise für Oracle RAC. Aber auch andere Clients, wie Windows- und UNIX-Systeme, können hochverfügbar und skalierbar auf die Daten zugreifen.

Prinzipiell greifen sämtliche GPFS-Knoten auf einen gemeinsamen physischen Speicherpool (Shared Storage) zu. Die Knoten müssen dabei nicht immer im SAN-Verbund sein. Der Datentransfer erfolgt über so genannte Recoverable Virtual Shared Disk (RVSD) bei AIX bzw. bei Linux mit Network Shared Disk (NSD). Bei dieser Variante wird nur eine Teilmenge der Server als Storage-I/O-Server dediziert.

Die anderen Knoten greifen über ein Netzwerk auf die I/O-Server zu. Stark vereinfacht bedeutet dies, dass die RVSD bzw. NSD ein SAN über einen IP Locking-Mechanismus bilden.

Die Koordination des Zugriffs erfolgt über einen verteilten Sperrmechanismus (Distributed Locking Mechanismus). Dabei läuft ein globaler Lock-Manager auf einem GPFS-Knoten zusammen mit den jeweiligen lokalen Lock-Managern auf jedem der GPFS-Knoten. Ein solcher Ansatz ermöglicht ebenfalls ein Client-seitiges Caching und führt zu einer höheren Performance.

GPFS ist im Moment nur für AIX und Linux nativ verfügbar, IBM arbeitet aber bereits an Clients für andere Betriebssysteme. GPFS kann dabei weitgehend unabhängig von der Speichersystem-Hardware eingesetzt werden. So können unterschiedliche Speichersysteme von unterschiedlichen Herstellern benutzt werden.

IBM Scale Out File Services (SOFS)

Effizientes Datenmanagement ist ein ständiges Abwägen zwischen schnellem Zugriff und Skalierbarkeit. Die Bereitstellung von Daten auf Dateiebene wird immer schwieriger, je mehr Benutzer hinzukommen und je größer die Datenbestände werden.

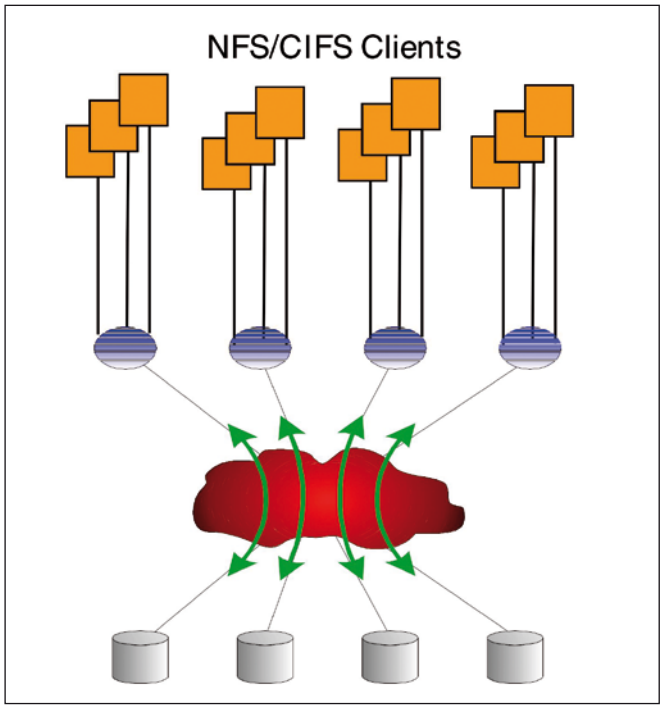
Mit IBM Storage Optimization and Integration Services – **Scale Out File Services** – werden die Daten in einer einzigen Speicherlösung zentralisiert und virtualisiert. Scale Out File Services (SOFS) konsolidieren isolierte Informationsbestände in einer globalen, beliebig erweiterbaren und heterogenen Speicherlösung.

SOFS wurde entwickelt, um mittels der Implementierung eines höchstskalierbaren, globalen, in einem Speicherverbund zusammengefassten NAS-Systems Probleme bei der gemeinsamen Nutzung von Informationen zu lösen.

Die Technologie, die Scale Out File Services zugrunde liegt, bietet einen hochmodernen Sperrmechanismus, der mehrere parallel stattfindende Lese-/Schreibzugriffe auf dieselbe Datei von verschiedenen Client-Plattformen aus ermöglicht. Der Aufbau in Clustern unterstützt eine nahezu unbeschränkte Anzahl von Knoten und quasi unbeschränkten Speicherplatz in einem einzelnen Cluster.

SOFS-Komponenten können vor Ort und/oder in einem IBM Delivery Center implementiert werden. Mit den IBM SOFS-Services übernimmt IBM den Betrieb einer Speicherlösung und bietet rund um die Uhr einen zuverlässigen Betrieb einer SOFS-Umgebung.

SOFS verwendet Speicherpools um die Dateien zu speichern. Diese Speicherpools können mit Speichern verschiedener Güte erstellt werden. Die Dateien werden dabei wie gewohnt im Dateibaum abgelegt. Regeln (Policies) legen fest, auf welchem Speichertyp die Dateien physisch abgelegt werden. Die Zuweisung erfolgt komplett transparent für den Anwender. Speicherpools können auf der Basis von Disk- und Tape-Systemen erstellt werden. Damit besteht die Möglichkeit, ein regelgesteuertes ILM von Disk bis Tape zu betreiben.



Schematische Darstellung GPFS/SOFS

Mit SOFS ist ein einheitlicher Zugriff auf alle Daten gewährleistet. Der Cluster erscheint als ein einzelnes System, das so skalierbar ist, dass alle Unternehmensanforderungen erfüllt werden können.

Dabei werden Metadaten und Nutzdaten parallel verarbeitet. Daher sind keine Engpässe existent und Failover-Aktivitäten notwendig. Dabei sind höchste Zuverlässigkeit und Fehlertoleranz sichergestellt. Anstelle von Failover verwendet SOFS 'bypassing' in der parallelen Architektur. Das bietet extremen Schutz vor Knoten- und Plattenausfällen.

Ein extrem hoher Systemdurchsatz wird durch 'wide striping' von großen Blöcken erreicht. Der Zugriff auf eine File erfolgt auf GB/s-Niveau (>100 GB/s aus einer Datei).

Encryption für Tape

Im August 2006 kündigte IBM die Verschlüsselungstechnik und ein dazugehöriges Schlüsselmanagement für die **TS1120 Bandlaufwerke (Jaguar 2)** an. Damit können 3592 Kassetten so beschrieben werden, dass nur die Instanz sie wieder auslesen kann, die den dazugehörigen Encryption Key kennt und hat. Missbrauch von Daten, die sich auf den Kassetten befinden, kann somit ausgeschlossen werden. Fallen Kassetten aus irgendwelchen Gründen in falsche Hände, können sie nicht ausgelesen werden.

Encryption-Möglichkeiten auf Tape gab es bisher in unterschiedlichen Ausprägungen. Mit speziell entwickelter Encryption-Software auf einem Server, mit speziell entwickelten Encryption-Anwendungen oder externen Verschlüsselungsgeräten konnte auf Tape verschlüsselt aufgezeichnet werden. Das Bandlaufwerk selbst hatte mit der eigentlichen Encryption nichts zu tun. Solche Lösungen konnten bisher immer nur in Teilbereichen zur Anwendung kommen. Eine Implementierung auf unternehmensweiter Basis, unabhängig von den unterschiedlichen Betriebssystem-Plattformen und auch über Unternehmensgrenzen hinaus, war bisher nicht sinnvoll gestaltbar. Hinzu kommen noch die vielen Verschlüsselungsarten wie zum Beispiel Mars, RC5, Serpent, Twofish, DES oder AES, um nur einige zu nennen, die das Ganze entsprechend komplex gestalten. Um Verschlüsselungen vorzunehmen, wird zum einen ein Verschlüsselungsalgorithmus benötigt und zum anderen ein sogenannter Datenschlüssel (Data Key), der die verschlüsselten Daten vor unautorisiertem Zugriff absichert. Nur über den Datenschlüssel können die verschlüsselten Daten wieder entschlüsselt werden. Zum besseren Verständnis werden im Folgenden die verschiedenen Verschlüsselungsmethoden beschrieben.

Symmetrische Verschlüsselung: Um das Ausspionieren von versendeten Daten durch eine dritte Partei zu verhindern, werden im Allgemeinen kryptografische Verfahren angewendet. Bei der symmetrischen Verschlüsselung werden die Daten mittels eines geheimen Schlüssels ver- bzw. entschlüsselt. Der Schlüssel muss dabei sowohl dem Sender als auch dem Empfänger bekannt sein und zu diesem Zweck vorher persönlich ausgetauscht werden.

Beispiele für bekannte symmetrische Verschlüsselungsalgorithmen sind der Data Encryption Standard (DES), der von IBM Anfang der siebziger Jahre entwickelt wurde und mit einer Schlüssellänge von 56 Bit arbeitet, sowie der International Data Encryption Algorithm (IDEA), der von den Schweizern Lai und Massey entwickelt und 1990 veröffentlicht wurde und mit einer Schlüssellänge von 128 Bit deutlich sicherer ist als der DES. Der generelle Nachteil dieses Algorithmus ist der direkte Austausch der geheimen Schlüssel, was seine Anwendung in einer Kunde-Händler-Beziehung erschwert. Der Vorteil besteht in der relativ geringen benötigten Rechenleistung.

Asymmetrische Verschlüsselung: Ein weiteres Verschlüsselungsverfahren ist die sogenannte asymmetrische Verschlüsselung. Sie basiert auf der Verwendung eines zusammengehörenden Schlüsselpaares, wobei ein Schlüssel zur Ver- und einer zur Entschlüsselung genutzt wird. Beim Public-Key-Verfahren wird nun einer der Schlüssel veröffentlicht und kann von jedem Sender dazu genutzt werden, eine Nachricht an den Empfänger zu verschlüsseln. Nur der Empfänger, der im Besitz des zweiten, privaten Schlüssels ist, kann die Nachricht dann entschlüsseln.

Ein Vertreter der asymmetrischen Verschlüsselungs-Algorithmen ist der RSA Algorithmus (RSA Data Security Inc.), benannt nach seinen Entwicklern Ron Rivest, Adi Shamir, Leonard Adleman, der in den USA 1977 entwickelt und patentiert wurde und für den Export in jedoch nur begrenzter Verschlüsselungstiefe (40 Bit) zur Verfügung steht.

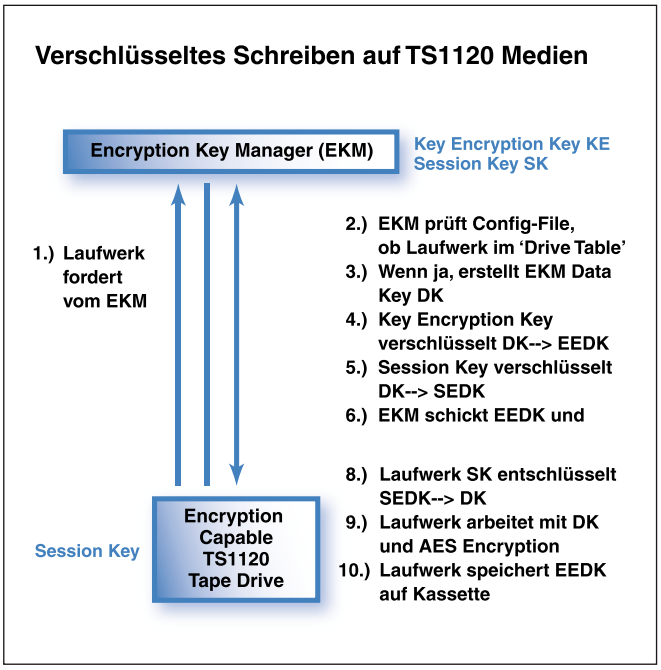
Die asymmetrische Verschlüsselung kann auch genutzt werden, um das Problem der Authentifizierung zu lösen. Zu diesem Zweck werden die öffentlichen Schlüssel von Sender und Empfänger gegenseitig bekannt gemacht. Der Sender verschlüsselt die Nachricht zunächst mit seinem eigenen, privaten und dann mit dem öffentlichen Schlüssel des Empfängers. Nach Erhalt der Nachricht entschlüsselt der Empfänger die Nachricht zunächst mit seinem privaten und dann mit dem öffentlichen Schlüssel des Senders. Dieser letzte Schritt ist jedoch nur dann erfolgreich, wenn die Nachricht wirklich von dem bezeichneten Sender kam, da andernfalls der verwendete öffentliche Schlüssel nicht passend ist.

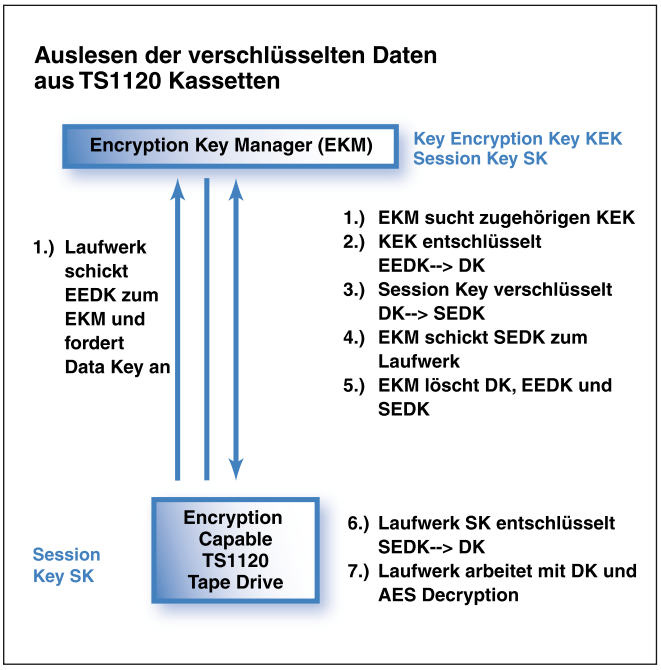
Hybride Verfahren: Der Nachteil der asymmetrischen Verschlüsselung ist der hohe Rechenaufwand. Deswegen wird oftmals eine Kombination aus symmetrischer und asymmetrischer Verschlüsselung genutzt. Dabei wird eine Nachricht durch den Sender zunächst mit einem speziellen geheimen Schlüssel (Session Key) symmetrisch verschlüsselt. Anschließend wird dieser Schlüssel mit dem öffentlichen Schlüssel des Empfängers asymmetrisch verschlüsselt und übertragen. Der Empfänger kann nun asymmetrisch mit seinem privaten Schlüssel den Session Key und somit die eigentliche Nachricht symmetrisch entschlüsseln. Da die asymmetrische Verschlüsselung nur für die Verschlüsselung des symmetrischen Schlüssels verwendet wird, bleibt der Rechenaufwand bei der asymmetrischen Verschlüsselung relativ gering.

IBM ist die erste Firma, die für Bandlaufwerke eine direkte Verschlüsselungstechnik anbietet. Um mit den TS1120 Laufwerken mit der Option Encryption arbeiten zu können, muss auf installierten TS1120 Laufwerken ein Hardware-Upgrade und ein Mikrocode-Update durchgeführt werden. TS1120 Laufwerke, die seit September 2006 ausgeliefert werden, sind standardmäßig mit der Encryptionfähigkeit ausgestattet. IBM entschied sich für die AES-Verschlüsselung (Advanced Encryption Standard), die mit 128, 192 und 256 Bit Schlüssellänge zur Verfügung steht und den Rijndael-Algorithmus als symmetrisches Kryptografieverfahren verwendet. AES ist der Nachfolger von DES (Data Encryption Standard mit 56 Bit Schlüssellänge). AES ist anerkannt von der US-Behörde NIST (National Institute of Standards and Technology) und vom US-Handelsministerium und soll anerkanntermaßen nicht 'knackbar' für die kommenden Jahre sein. Das Bandlaufwerk TS1120 arbeitet mit dem sicheren AES-256-Bit-Verschlüsselungsalgorithmus. Damit ist zum ersten Mal die Möglichkeit gegeben, dass das Laufwerk selbst die Verschlüsselung durchführt.

Die Verschlüsselungsmethode ist die eine Seite. Genauso wichtig sind die Datenschlüssel, die vor unautorisiertem Zugriff schützen und über die sich die verschlüsselten Daten wieder entschlüsseln lassen. Im Mainframe-Umfeld, wo schon seit Jahren Encryption mit entsprechenden Krypto-Prozessoren eingesetzt wird, bildet das Herzstück für die Datenschlüsselverwaltung ein **EKM Encryption Key Manager** im z/OS-Betriebssystem. Der im z/OS eingesetzte EKM bietet eine absolut nicht manipulierbare und hundertprozentig sichere zentrale Datenschlüsselverwaltung. Dieser EKM wurde nun auf die JAVA-Plattform portiert und steht so für alle Betriebssystemplattformen zur Verfügung. Der neue EKM liefert die Keys zum TS1120 Laufwerk, kann auf unterschiedlichsten Systemplattformen laufen und unterstützt dabei zentralisiertes Encryption Key Management, das unternehmensweit einsetzbar ist.

Der **EKM** steht für die Betriebssysteme z/OS 1.6 oder 1.7, AIX 5.2 oder später, I5/OS 5.2 oder später, HP-UX 11.0, 11i, und 11.23PI, Sun Solaris 8, 9 und 10, Linux – System z, System p und Intel, Red Hat Enterprise Linux 4 (REHL 4), SuSE Linux Enterprise Server 9 (SLES 9), Windows 2000 und Windows 2003 zur Verfügung.





Muss nun ein TS1120 Bandlaufwerk mit Encryption aufzeichnen, fordert das Laufwerk den Datenschlüssel vom EKM Encryption Key Manager an. Der EKM prüft dann, ob das Laufwerk zugelassen ist und im 'Drive Table' des EKM bekannt ist. Falls nicht, muss das Laufwerk über die Konfigurationsfile im Drive Table des EKM aufgenommen werden. Dann erstellt der EKM den angeforderten Datenschlüssel, der im Folgenden als 'Data Key' bezeichnet wird.

Da die Datenschlüssel aus Sicherheitsgründen nicht unverschlüsselt über ein Netzwerk zum Bandlaufwerk übertragen werden können, arbeitet der EKM intern mit zwei verschiedenen Verschlüsselungsmethoden für den Datenschlüssel. Der eine Algorithmus ist der KEK Key Encryption Key, der ausschließlich im EKM läuft und mit dem nur der EKM etwas anzufangen weiß. Der andere Algorithmus ist der SK Session Key, der sowohl im EKM als auch im Bandlaufwerk zur Anwendung kommt. Ist zwischen EKM und Bandlaufwerk eine Session aufgebaut und wird der Data Key mit dem Session Key SK verschlüsselt und zum Bandlaufwerk übertragen, entschlüsselt das Bandlaufwerk wieder den orginalen Data Key, um dann mit der AES-256-Bit-Verschlüsselung die Daten aufzuzeichnen. Bei jeder neu aufgebauten Session zwischen EKM und Bandlaufwerk kommt immer ein neuer SK Session Key zum Einsatz. Der mit dem SK verschlüsselte Data Key wird als SEDK (Session Encryption Data Key) bezeichnet. Neben dem SEDK überträgt der EKM aber auch den über den KEK (Key Encryption Key) verschlüßelten Data Key zum Laufwerk, der als EEDK (Externally Encryption Data Key) bezeichnet wird und mit dem das Laufwerk absolut nichts anfangen kann, weil ausschließlich der EKM den KEK kennt und anwenden kann. Das Laufwerk hat also keine Möglichkeit, den übertragenen EEDK zu entschlüsseln, um an den eigentlichen Data Key zu gelangen. Das Laufwerk kann den Data Key ausschließlich über den übertragenen SEDK entschlüsseln. Hat das Laufwerk nun die verschlüsselte Aufzeichnung abgeschlossen, wird am Bandende in die Tape Volume Control Region der EEDK hinterlegt, mit dem das Bandlaufwerk nichts anzufangen weiß. Zusätzlich wird der EEDK auch im Memory Chip der Kassette gespeichert. Sollte nun unautorisiert diese Kassette auf einem Encryption-fähigen TS1120 Laufwerk ausgelesen werden, ist dies nicht möglich, weil das Laufwerk mit dem mitaufgezeichneten EEDK nichts anfangen kann. Ein nicht autorisiertes Auslesen ist nicht möglich.

Zum Auslesen fordert das Laufwerk vom EKM wieder den Data Key an und überträgt den aufgezeichneten EEDK zum EKM. Der EKM sucht den zum EEDK zugehörigen KEK Key Encryption Key und entschlüsselt den EEDK. Der orginale Data Key steht also wieder zur Verfügung. Der Data Key wird nun wieder über einen neuen SK Session Key verschlüsselt und zum Laufwerk übertragen. Das Laufwerk erhält den neuen SEDK Session Encryption Data Key und entschlüsselt den SEDK, um den orginalen Data Key zu erhalten. Danach können die verschlüsselt aufgezeichneten Daten auf der Kassette entschlüsselt und ausgelesen werden.

Es besteht also keine Möglichkeit, die verschlüsselten Daten ohne den originalen Data Key auszulesen. Die Raffinesse dieser Implementierung liegt in der Tatsache, dass das Laufwerk eine symmetrische Encryption über einen Data Key durchführt, während die darübergelegte Key-Verwaltung asynchron implementiert ist. Dabei kann der auf dem Tape verschlüsselt abgespeicherte Data Key (EEDK) nicht von TS1120 Laufwerken sondern ausschließlich vom EKM (Encryption Key Manager) für die Wiederherstellung des originalen Data Keys verwendet werden. Das Laufwerk muss den abgespeicherten verschlüsselten Data Key zum EKM übertragen. Der EKM stellt den originalen Data Key wieder her und schickt ihn für das Laufwerk verständlich verschlüsselt (Session Key) zurück. Das Laufwerk ist nun in der Lage, den originalen Data Key herzustellen und damit die Daten entschlüsselt auszulesen. Die Kombination von synchroner und asynchroner Technik machen diese Verschlüsselung absolut 'wasserdicht'.

Werden Kassetten, die mit Verschlüsselung aufgezeichnet wurden, in eine andere Lokation transportiert, zum Beispiel zu einem Geschäftspartner, müssen dort TS1120 Laufwerke mit Encryptionfähigkeit und ein EKM installiert sein. Will die Lokation nun die Daten auslesen, schickt ein TS1120 Laufwerk den auf der Kassette gespeicherten EEDK zum örtlichen EKM, der einen Private Key hat, um den auf der Kassette abgespeicherten EEDK zu entschlüsseln.

Es können auf der 3592 Kassette bis zu zwei verschlüsselte EEDKs mit zwei unterschiedlichen Private Keys abgespeichert werden. Der erste Key gehört zum EKM, der den Schlüssel erzeugt hat und der zweite Key kann einem anderen EKM gehören, der aufgrund des dort verfügbaren Private Keys den zweiten EEDK in den orginalen Schlüssel umwandeln kann. Mit zwei Keys auf der Kassette zu arbeiten ist dann sinnvoll, wenn in verschlüsselter Form Datenträgeraustausch zwischen zwei unterschiedlichen Lokationen durchgeführt wird. Das Laufwerk schickt immer den ersten EEDK zum EKM. Kann der EKM damit nichts anfangen, wird automatisch der zweite EEDK zu dem vorhandenen EKM übertragen.

Mit der Ankündigung der LTO4-Laufwerke im **April 2007** wurde für **LTO4** auch **Encryption** angekündigt. LTO4-Laufwerke arbeiten wie TS1120 Laufwerke auch mit der AES 256-Bit-Verschlüsselungstechnik. Der Unterschied zu TS1120 liegt darin, dass keine verschlüsselten Schlüssel (EEDK) auf der LTO4-Kassette abgespeichert werden, sondern nur der Key Identifier, also der Schlüsselanhänger. Die Key-Verwaltung und Key-Abspeicherung erfolgt über den EKM. Soll eine verschlüsselte Kassette ausgelesen werden, überträgt das LTO4-Laufwerk den Identifier zum EKM, der dann in der Lage ist, den richtigen Originalschlüssel dem Identifier zuzuordnen.

Der EKM wurde im Mai 2007 mit dem EKM Release 2 so erweitert, dass er sowohl das Key Handling für TS1120 als auch für LTO4 durchführen kann. Damit können beide Technologien mit demselben EKM gemanaged werden.

IBM bietet drei Encryption-Implementierungs-Möglichkeiten auf Tape an:

- *Die Anwendung selektiert die Daten, die mit Encryption auf Band geschrieben werden müssen, und stellt die Keys dem TSM zur Verfügung. Der TSM Tivoli Storage Manager verwaltet dann als Backup-Server die Encryption-Key-Datenbank.*
- *Im Mainframe-Umfeld kann über System Utilities des DFSMS im z/OS-Betriebssystem über neue Data-Class-Parameter selektiert werden, was mit Encryption verarbeitet werden soll. Ähnlich können Policies für Encryption auch im 'AIX Tape Device Driver' eingerichtet werden. In beiden Fällen verwaltet der neue Enterprise Key Manager (EKM) die Encryption-Key-Datenbank.*
- *Die dritte Option geht über die Library selbst. Sowohl die 3494 als auch die TS3500 Library etabliert entsprechende Policy-Regeln, welche 'VolSers' oder welche logischen Libraries mit Encryption verarbeitet werden sollen. Die Verwaltung der Encryption-Key-Datenbank wird durch den neuen Enterprise Key Manager (EKM) durchgeführt.*

Alle drei Optionen lassen sehr flexible, unternehmensweite und über Unternehmensgrenzen hinausgehende Verschlüsselungslösungen auf Tape mit einem zentralen Key Management zu und sind einmalig in ihrer Lösungsform.

Encryption auf Tape sichert die auf Bandkassetten geschriebenen Daten vor Missbrauch und bietet einen **neuen Sicherheitsstandard auf Tape**, der die geschriebenen Daten vor unautorisiertem Zugriff schützt.

SAN

Brocade hatte im **Januar 2007** McData übernommen und eine Roadmap für die nächsten 18–24 Monate angekündigt. Die Architektur der übernommenen McData SAN Direktoren IBM 2027-140/256, sowie die Architektur des Brocade 2109-M48 SAN Direktors, wird in einer gemeinsamen neuen Architektur in einem neuen Produkt, dem 'Converged' SAN Direktor, mit 8 Gbps abgebildet. Die Ankündigung des **IBM SAN768B** erfolgte im **Februar 2008**. Dabei können dann die 2027-140/256 native an den neuen 8-Gbps-Direktor angeschlossen und weiterverwendet werden.

Der IBM Direktor SAN768B wird von Brocade (Brocade DCX) gebaut und stellt eine Plattform bereit, die sich durch hochleistungsfähige IT- und zukünftige FCoE-(FibreChannel over Ethernet) Technologien kommenden IT-Entwicklungsstufen anpasst. Es ist das erste 'Direktor-artige' Produkt, das über FibreChannel-Verbindungen Transferraten von 8 Gigabit pro Sekunde (Gbps) unterstützt, was die Geschwindigkeit beim Datentransfer in Zukunft nahezu verdoppeln kann. Der Direktor bietet die Unterstützung von FibreChannel, FibreChannel over Ethernet (FCoE), Data Center Ethernet (DCE), Gigabit Ethernet und das iSCSI-Protokoll.

Der SAN768B ist das erste Produkt der Branche, das wirkliche Interoperabilität zwischen Direktoren des b-Typs (Brocade) und solchen des m-Typs (McDATA) in einer 8- oder 10-Gbps-FC-SAN-Infrastuktur unterstützt. Für Nutzer des m-Typs und Großkunden, die ihre Mainframes über FICON-Kanäle betreiben, bietet diese Interoperabilität eine Roadmap für die Erweiterung der bestehenden FICON-Struktur. Die neue Plattform reduziert Komplexität, Ausfallrisiko und Energieverbrauch mit einer speziellen Skalierbarkeit von bis zu 768 FibreChannel Ports (externe User Ports) auf Basis von 8 Gbps über zwei Domänen.

IBM Direktor SAN768B (Brocade DCX) mit bis zu 896 Ports

In einer großen Konfiguration mit zwei Domänen (Bild oben, Abbildung rechts) bietet der Direktor bis zu 896 Ports. Davon werden 768 Ports als 'User Ports für externen Anschluss' verwendet. Die restlichen 128 Ports werden über sogenannte ICL Inter Chassis Links an die Backplane der beiden Direktoren geführt und als ICL-Ports bezeichnet. Die Übertragungsbandbreite zwischen den beiden Domänen (SAN868B Chassis) kann auf bis zu 1 Tbit/s gesteigert werden.

Die adaptiven Networking Services Features im SAN768B (Brocade DCX) erlauben in der Fabric eine dynamische Zuweisung verteilter Ressourcen in dem Moment, in dem die Anforderungen der virtuellen Server oder des vernetzten Speichers auftreten. Sollten Engpässe auftreten (oder vorhersehbar sein), können die Bandbreiten in der Fabric automatisch und entsprechend der vordefinierten Service Levels angepasst werden. Dies hilft dabei, Operationen mit einer höheren Priorität automatisch die benötigten Ressourcen zur Verfügung zu stellen.

Das Jahr 2008 ist das Jahr der Einführung von 8-Gbps-Technologie in Storage Area Networks und im Laufe des Jahres können 8-Gbps-'End-to-End'-Lösungen in diesem Umfeld implementiert werden.

Green IT

Das Thema Energieeffizienz rückt immer mehr in den Mittelpunkt. Manche Rechenzentren können schon heute keine Server oder Speichereinheiten mehr aufnehmen, weil es klimatechnisch unmöglich ist.

Laut einer IDC-Umfrage von 2007 sind Themen rund um Energieverbrauch und Energieversorgung von zunehmender Bedeutung. Steigende Strompreise belasten die verfügbaren Budgets. In vielen Rechenzentren wird das Thema Energieeffizienz inzwischen zum Top-Thema. 'Stromlose' Datenträger wie Bandkassetten oder optische Platten werden in den nächsten Jahren eine neue Renaissance erleben. Nach einer Gartner-Studie aus dem Jahr 2006 werden 70 % aller Daten auf Band abgespeichert. Wenn alle Daten, die heute im produktiven Einsatz auf Bandkassetten abgespeichert sind, auf Plattensysteme (Disk) verlagert würden, müßten ca. 40 neue Kernkraftwerke für die Stromversorgung gebaut werden. Das Internet und die steigende Nutzung des Internets erhöhen die Energieproblematik zudem immens: Vier Internet-Suchanfragen benötigen soviel Strom wie eine 40-Watt-Glühbirne mit einer Stunde Brenndauer.

Im **Mai 2007** kündigte IBM das Projekt **'Big Green'** an. IBM investiert seit dieser Ankündigung **1 Milliarde Dollar** pro Jahr in das Projekt zur Verbesserung der 'green' Technologien und Serivces und arbeitet an einer Roadmap, um die eingetretene IT-Energie-Krise für Rechenzentren zu bewältigen. Dazu hat IBM ein Team aus mehr als 850 Energie-Spezialisten aufgebaut. Der Mehrwert für IT-Betriebe ergibt sich oft aus der Tatsache, dass Energiekosten für ein typisches Rechenzentrum halbiert werden können. Das entspricht im Vergleich einer Emmissionsreduzierung, die ca. 1300 Autos erzeugen. IBM bietet dazu neben der Hardware und Software eine Vielzahl entsprechender Services für Rechenzentrumsbetriebe an. Teil dieser Services beinhalten Energie-Effizienz-Analysen und/oder 'Energy-Self-Assessments'.

IBMs beispielhafte Aktivitäten im Rahmen des Projekts 'Big Green' und die damit verbundenen langfristigen Verpflichtungen führten dazu, dass IBM im Februar 2008 als die 'Top Green IT Company' des Jahres 2008 sowohl von der IDG (International Data Group) als auch von der Computerworld gewählt und ausgezeichnet wurde.

Virtualisierung allgemein

Virtualisierung in jeglicher Form wird für die Epoche der Server-basierenden Speicherarchitekturen prägend sein. Neben Tape-Virtualisierungen und Plattenvirtualisierungslösungen im SAN wird sich die Virtualisierung auf weit mehr Bereiche ausdehnen. Die **IBM** vertritt hier weltweit die **Strategie 'Virtualize Everything'**. Nur so wird es möglich werden, heterogene Server- und Speichereinheiten und unterschiedlichste Infrastrukturen optimal zu nutzen und ein flexibles, nach Regelwerken betriebenes Datenmanagement zu ermöglichen. Nur durch Virtualisierung in allen Bereichen werden Automationsprozesse in den heute unterschiedlichen Infrastrukturen möglich werden. Server-basierende Speicherarchitekturen erleichtern den Virtualisierungs-Gesamtansatz erheblich.

Mit effektiven Virtualisierungslösungen erzielt man eine möglichst maximale Nutzung der vorhandenen physikalischen Ressource, d. h. vorhandene Speicherkapazitäten werden wesentlich höher ausgenutzt. Somit die liefert Virtualisierung einen maßgeblichen positiven Beitrag zur Bewältigung der Energieproblematik im IT-Umfeld.

Neue Infrastrukturen und Bandbreiten

Die heutigen FibreChannel-Netze und IP-Infrastrukturen werden in den nächsten Jahren mit Sicherheit weiterentwickelt werden. Von heutigen 4-Gbit-SANs geht es in Richtung **8-Gbit- und 12-Gbit-**SAN-Technologie weiter. IP-Netze heute mit bis zu 2-Gbit- und 10-Gbit-Ethernet könnten in den nächsten Jahren durch **100-Gbit-Ethernet-Technologie** abgelöst werden. Der IP-Infrastruktur, speziell dem Ethernet, wird in den nächsten Jahren noch eine besondere Bedeutung zukommen. Es ist geplant, innerhalb der nächsten 5 Jahre jedem Privathaushalt in Deutschland einen 100-Mbit-Ethernet-Anschluss zur Verfügung zu stellen. Damit bekommen die Privathaushalte neue, riesige Übertragungsbandbreiten. Ethernet-Infrastrukturen werden also zu einem maßgeblichen Faktor auch im privaten Umfeld.

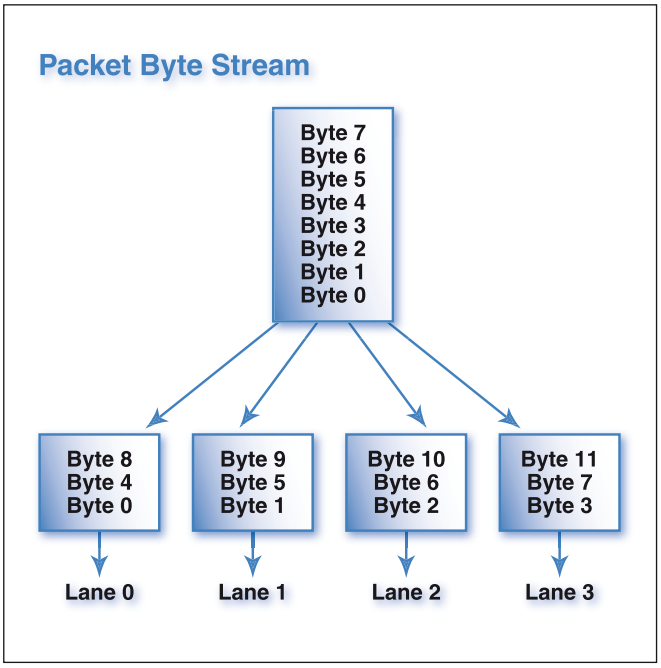
So wie die ersten SAN-Infrastrukturen zwei Jahre früher in den Hochleistungsrechenzentren der Universitäten ihre Anwendung fanden, bevor sie im kommerziellen Rechenzentrumsumfeld zum Einsatz kamen, entwickeln sich heute ähnliche Tendenzen für eine neue Übertragungstechnik, die als **InfiniBand** bezeichnet wird. **InfiniBand** steht für **'Infinite Bandwidth'**. InfiniBand ist ein neues Transport-Protokoll, das Servereinheiten, Speichereinheiten und ganze Netz-

werke von Rechenzentren verbinden kann. Seit 1999 wird der Standard von der InfiniBand Trade Association vorangetrieben, nachdem damals Compaq, Dell, HP, IBM, Intel, Microsoft und SUN angekündigt hatten, ihre zukünftigen I/O-Entwicklungen der InfiniBand-Technologie anzupassen. Seit Herbst 2005 ist der Standard verfügbar.

InfiniBand schafft die Möglichkeit, heutige LAN-, SAN-, NAS- und IPC(Interprocessor Communication)-Infrastrukturen in einem einzigen **universellen Netzwerk** zu vereinigen. Ebenso könnten separate InfiniBand-Netze neben vorhandenen SAN- und IP-Infrastrukturen über iSCSI betrieben werden.

Für das Transport-Protokoll können Kupferkabel (Entfernungslimitierung bei 17 m) und Fibre-Optic-Kabel (bis 10 km) verwendet werden. Die Fibre-Optic-Kabel für InfiniBand werden im Vergleich zu heutigen FibreChannel-Verbindungen anders hergestellt und sind lange nicht so empfindlich. Ein Knick im Kabel verursacht fast keine Dämpfung und damit keinen Verlust. Würde man das mit dem heutigen FibreChannel machen, müssten die Kabel komplett ausgetauscht werden. Durch diese Eigenschaft ist InfiniBand Fibre Optic ideal geeignet, um für interne Verkabelungen von Maschinen zur Anwendung zu kommen.

Für den externen Einsatz ist es wichtig, die Fibre-Optic-Lösungen sowohl in der Unterstützung der seriellen Schnittstellen als auch in den Entfernungsmöglichkeiten voranzutreiben.



4, 8 oder 12 'Physical Lanes' können auf einen InfiniBand-'Physischen Link' geschaltet werden, die Paketübertragung ist 'byte-multiplexed' über die Physical Lanes.

InfiniBand repräsentiert ein extrem hohes skalierbares Link Interface SLI. Es ist möglich, die Anzahl der Physical Lanes dynamisch – abhängig von Durchsatzanforderungen und Datenverkehr – zu verändern und dynamisch anzupassen.

InfiniBand bietet eine hoch skalierbare Verbindungstechnik mit 'nX' Multiple Lane Pairs mit einer Datenübertragungsrate von 2.5 Gbit/s (Single Data Rate SDR), 5 Gbit/s (Double Data Rate DDR) oder 10 Gbit/s (Quad Data Rate QDR) in jede Richtung. Die untenstehende Tabelle stellt die maximale Bandbreite für jede der Möglichkeiten dar:

	SDR	DDR	QDR
1X	2.5 Gb/s	5.0 Gb/s	10.0 Gb/s
4X	10.0 Gb/s	20.0 Gb/s	40.0 Gb/s
8X	20.0 Gb/s	40.0 Gb/s	80.0 Gb/s
12X	30.0 Gb/s	60.0 Gb/s	120.0 Gb/s

InfiniBand hat einen weiteren wesentlichen Vorteil. Die InfiniBand-Upper-Level-Protokolle wie SDP, SRP und iSER benützen die Methode des Remote Direct Memory Access RDMA, um direkt Daten in anwendungszugeordnete Pufferbereiche des Rechners zu schreiben und gleichzeitig auszulesen.

Dies reduziert den Overhead-Datentransfer zu einem Drittel und setzt CPU-Kapazitäten frei, die anders genutzt werden können.

Nachdem die ersten InfiniBand-Netze in einigen Hochleistungsrechenzentren der Universitäten etabliert sind, kann man – analog zum FibreChannel – davon ausgehen, dass die ersten InfiniBand-Netze und -Infrastrukturen innerhalb der nächsten 2–3 Jahre in 'kommerziellen' Rechenzentren implementiert werden. Beschleunigend kommt noch hinzu, dass InfiniBand im Vergleich zum FibreChannel kostengünstiger ist.

Die Epoche der Server-basierenden Speichersysteme mit neuen Infrastrukturmöglichkeiten

Neue Basis-Technologien

Die Fachwelt im Storagebereich ist sich darin einig, dass **Kunststoff das Speichermedium der Zukunft** werden kann. Mit organischen Polymeren wird schon seit einigen Jahren experimentiert. Organische Polymere sind geeignet, durch optische Technologien für die Speicherung von Daten genutzt zu werden. Mit der neuen **blauen Laser-Technik** lassen sich Interferenz-Felder auf dem Kunstoff erzeugen, die eine Vielzahl von Bits abbilden. Diese Technologie wird auch als holografische Speicherung bezeichnet und erlaubt viel größere Kapazitäten als mit herkömmlichen optischen Verfahren. Hierzu wird ein blauer Laser als 'Daten-Strahl' eingesetzt. Ein Teil des Lasers wird über Spiegelverfahren durch einen Lichtmodulator geschickt, damit er eine andere Lichtbeugung bekommt. Anschließend kommen beide Strahlen auf der Polymeroberfläche wieder zusammen. Durch die unterschiedliche Lichtbeugung wird ein Interferenzfeld im Polymer erzeugt, das man heute als Hologramm bezeichnet.

Kunststoffe können aber auch anders bearbeitet werden. Hierzu bietet sich die **Nanotechnologie** an, die auf mechanische Weise ein Kunststoff-Chip beschreiben kann. Diese Technologie wurde von IBM im Grundlagenforschungslabor in Rüschlikon in der Schweiz nahe Zürich entwickelt und wird als **Millipede-Technologie** bezeichnet. Bereits im Jahr 2005 zeigte IBM die ersten Millipede-Chips 'live' auf der CeBit in Hannover (detaillierte Informationen sind im Technologie-Anhang dieses Kompendiums zu finden).

Da heute noch niemand abschätzen kann, wie sich Kunststoffe und veränderte Kunststoffe im Laufe ihrer Alterung verhalten, arbeitet IBM an sinnvollen Alterungsverfahren für Kunststoffe, um genauere Aussagen zu bekommen. Heute geht man davon aus, dass ein veränderter Kunststoff einen Lebenszyklus von cirka 100 Jahren haben müsste. Diese Aussage muss aber noch durch entsprechende Alterungsverfahren belegt und bestätigt werden.

Kommentar zur Epoche der Server-basierenden Speicherarchitekturen mit neuen Infrastrukturmöglichkeiten

Server-basierende Speicherarchitekturen wie DS8000, SVC, DR550, die neuen Tape Virtualisierungseinheiten TS7520 und TS7700, TS3500 Tape Library Virtualisierung und File Virtualisierung über File Area Networks, bieten eine noch nie dagewesene Flexibilität und Skalierung in Leistung, Kapazität und Funktionalität. Sie werden die vor uns liegende Epoche maßgeblich bestimmen, weil Konsolidierungen auf einer ganz neuen Basis möglich werden. Diese neuen Architekturen schaffen die Möglichkeit, viele Server über Cluster zusammenzuschließen. Damit sind die ersten Anfänge von Grid Computing im kommerziellen Umfeld gelegt.

Durch neue Bandbreiten der Datenübertragung und neue Vernetzungstechniken ergeben sich neue Ansätze für eine optimale Gestaltung der Rechenzentren und deren Infrastruktur. Virtualisierungslösungen in den unterschiedlichsten RZ-Bereichen werden zu einer wesentlichen Vereinfachung der RZ-Abläufe führen und die Administration stark vereinfachen.

Durch Standards werden die IT-Anbieter gezwungen, näher zusammenzurücken, und der Endbenutzer bekommt endlich mehr Kompatibilität zwischen den unterschiedlichen Systemen, wie es bisher nicht der Fall war. Virtualisierung trägt hier einen ganz entscheidenden Anteil dazu bei.

Neue Speichertechnologien werden rechtzeitig auf den Markt kommen, um den Anforderungen an ständig wachsende Kapazitäten Rechnung zu tragen. Kunststoffe als Speichermedium der Zukunft werden maßgeblich diese Epoche prägen. Ebenso können Stromunabhängige RAM Speicher (Random Access Memory) wie MRAM und PCRAM, die unendlich oft beschrieben werden können, die heutigen Flash-Speicher wie auch teure FC-Platten ersetzen, weil sie zukünftig um ein Vielfaches billiger produziert werden können.

Das Tempo in der Entwicklung und in den Produktzyklen wird sich noch weiter verstärken. Es wird immer schwieriger werden, einen Gesamtüberblick über alle Möglichkeiten im Storage-Umfeld zu behalten, die man aktiv nutzen kann, und Beratung wird eine größere Bedeutung als bisher haben. Hardware allein spielt dabei nur noch eine geringe Rolle.

Die Epoche wird sicherlich die innovativste werden und vielleicht Dinge hervorbringen, an die wir heute noch gar nicht zu denken wagen. Hätte man in den neunziger Jahren jemandem erzählt, was heute alles verfügbar ist, wäre man sicher zum Fantasten gestempelt worden.

Was wird in 10 Jahren sein? Es wird eine spannende Epoche werden und der Autor fragt sich selbst, ob seine Aussagen zutreffen werden.

Vielleicht liegt unsere Vorstellungskraft noch weit hinter den Dingen, die vor uns liegen ...

Anhang

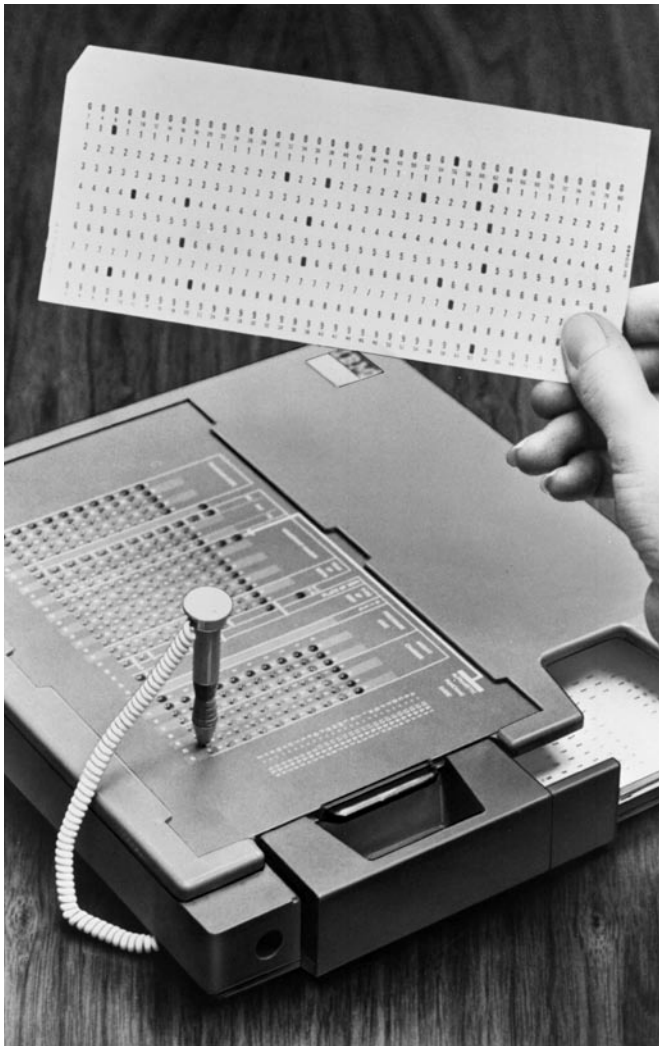
Technologie-Anhang

Die Vorgänger der heutigen Massenspeicher

Lochkarten

Lochkarten als Massenspeicher existierten lange vor den Computern. Bereits Holleriths Volkszählungsmaschine setzte 1891 Lochkarten ein. Lochkarten stellen Binärdaten dar, in dem eine bestimmte Stelle auf der Karte gelocht (oder eben nicht gelocht) war. Beim Aufkommen der Computer in den 50er-Jahren beherrschte man diese Technik inzwischen sehr gut: Die Karten konnten sortiert, gezählt und in jeder erdenklichen Form ausgewertet werden.

Hergestellt wurden die Karten mit sogenannten Lochkartenstanzern. Mit den fertig codierten Karten fütterte man über einen Leser den Computer und führte ihm auf diese Weise die gewünschten Programm- und Verarbeitungsdaten zu. Statt einzelner Karten wurden später auch Lochstreifen verwendet, die eine höhere mechanische Verarbeitungsgeschwindigkeit gestatteten.



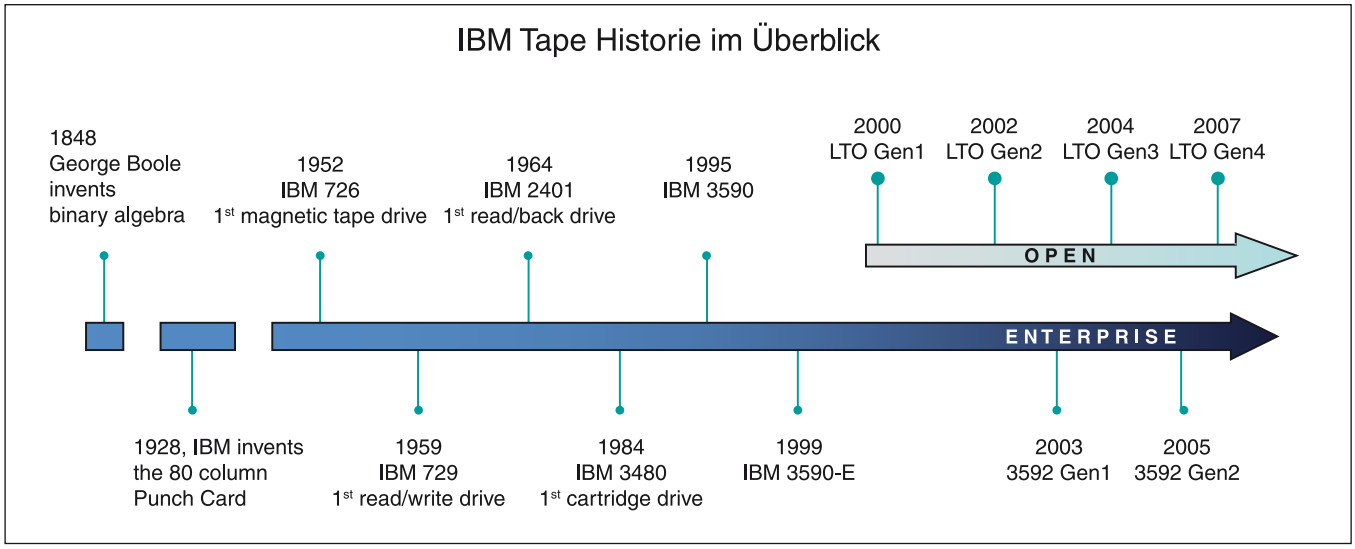
Lochkarten

Trommelspeicher

Die ersten elektrischen Massenspeicher, die im Zusammenhang mit Computern eingesetzt wurden, waren 1947/48 die sogenannten Trommelspeicher. Auf einer massiven Trommel wurde eine Beschichtung aus einem magnetischen Material aufgebracht. Dieses konnte induktiv durch einen Strom magnetisiert werden oder erzeugte beim Auslesen einen induzierten Strom. Durch die kleineren Ausmaße der Trommelspeicher verringerte sich auch die Größe der damaligen Computer erheblich. Bereits ab 1950 wurden vereinzelt Trommelspeicher eingesetzt, um Programme oder Daten zu laden und zu speichern.

Magnetband

Ab 1952 verdrängte das Magnetband (IBM 726, IBM 727) die Lochkarten als Massenspeicher. Dabei handelte es sich um ein 12,4 mm breites Plastikband, bei dem eine Seite mit einer magnetisierbaren Beschichtung überzogen war. Die Datenaufzeichnung bestand darin, dass die auf dem Band enthaltenen Ferritstreifen magnetisiert oder nicht magnetisiert waren und so die Informationen in dem für den Computer benötigten Binärsystem darstellten. Auf einer zusätzlichen Bandspur konnten später sogenannte Paritätsbits untergebracht werden, wodurch eine hohe Datensicherheit gewährleistet wurde. Der Vorteil der Magnetbänder lag neben einer hohen Verarbeitungsgeschwindigkeit vor allem in der hohen Speicherkapazität. Allerdings konnten Daten auf einem Band immer nur sequentiell (also hintereinander) abgelegt und wieder gelesen werden. Trotzdem findet das Magnetband bei Großrechner-Systemen mit hohem Archivierungsbedarf und als billiger Massenspeicher (heute in Form von Bandkassetten) bis heute Verwendung. Zukünftig dürfte das Magnetband aufgrund der Energie-Effizienz-Problematik und der ‘Green IT’-Diskussionen eine neue Renaissance erfahren. Magnetbänder brauchen keinen Strom. Tape ist ‘cool’!



Magnetplatten

Magnetplatten sind die Nachfolger der voluminösen Trommelspeicher und gelten als die direkten Vorläufer der heutigen Festplatten. Am 13. September 1956 stellte IBM die erste Magnetplatte mit der Bezeichnung RAMAC 350 (Random Access Method of Accounting and Control) als Teil des RAMAC 305 Systems und mit einer Kapazität von 5 MB der Öffentlichkeit vor. Diese Kapazität verteilte sich auf 51 Scheiben mit je 24 Zoll (60 cm) Durchmesser. Die Abtastung der Informationen erfolgte durch einen Schwebe-Schreib/Lesekopf (siehe unter RAMAC 305)

Noch immer dominieren Magnetplattenspeicher die Informationstechnologie, obwohl auch die optischen Speicher in Form der wiederbeschreibbaren Compact Discs sich einen gewichtigen Anteil am Massenspeichermarkt erobert haben. Magnetplattenspeicher funktionieren alle nach dem gleichen Grundprinzip: Als Speichermedium dient eine runde Scheibe, auf der sich eine Schicht aus hartmagnetischem Material (früher verschiedene Ferrite, heute Dünnfilm) befindet. Die Platte ist in konzentrische Spuren unterteilt. Ein beweglicher Magnetkopf wird radial über diese Platte bewegt, um die nadelförmigen Ferrite auf den einzelnen Spuren so zu magnetisieren, dass Binärdaten abgebildet werden. Er ist auch in der Lage, durch Verschiebung des Laufwerksarms schnell von der einen in die andere Spur zu wechseln. Die Spuren wiederum sind in Sektoren unterteilt. Die Lage und Ausdehnung der einzelnen Sektoren werden durch die sogenannte Formatierung festgelegt.

Prinzipiell sind Magnetplattenspeicher auf wahlfreien Zugriff ausgelegt. Dies bedeutet, dass das Medium nicht – wie z. B. bei Bandlaufwerken – von Beginn an sequentiell durchsucht werden muss, um eine bestimmte Stelle (Datei) zu finden. Die Köpfe der Magnetplatten können – vergleichbar mit dem Tonarm eines Plattenspielers und dem Anwählen eines bestimmten Musikstücks – direkt zu der Stelle springen, an der die gewünschte Datei angelegt ist.

Disketten

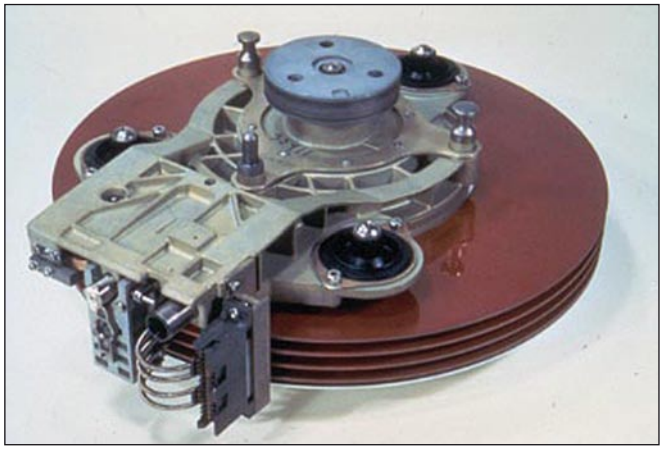
Alan Shugart, der in den späten 60er-Jahren für IBM arbeitete, wird die Erfindung der 8“-Diskette im Jahre 1971 zugeschrieben. Die Diskette besteht aus einer Kunstofffolie, die mit einer nichtorientierten Magnetschicht versehen ist. Die Datenaufzeichnung erfolgt entweder einseitig (SS) oder doppelseitig (DS). Zum Schutz und zur besseren Handhabung befindet sich die Scheibe in einer rechteckigen Kunststoffhülle, die mit einem Gleit- und Reinigungsvlies ausgekleidet ist. Die Hülle besitzt Öffnungen für den Arbeitskonus (über den die Scheibe angetrieben wird), das Indexloch und den Schreib-/Lesekopf. Zusätzlich besitzt die Hülle noch eine Aussparung für das Setzen eines Schreibschutzes. Je nach System wird der Schreibschutz durch Abdecken oder Freilassen dieser Aussparung gesetzt. Der Schreib-/Lese-Kopf berührt beim Schreiben und Lesen die Diskettenoberfläche, ansonsten ist der Kopf angehoben. 1971 stellte IBM das erste 8“-Diskettenlaufwerk der Öffentlichkeit vor. Damit wird zum ersten Mal bei einem beweglichen Datenträger der wahlfreie Zugriff realisiert, denn der Schreib-/Lesekopf kann auf jeder Stelle des Datenträgers positioniert werden.

Seine eigene Firma, Shugart Associates, gründete Alan Shugart dann 1973, zur Entwicklung und Herstellung von Diskettenlaufwerken. Schon lange wollte er Prozessoren und Diskettenlaufwerke als Teile in komplette Computersysteme integrieren. 1976 ist es Alan Shugarts Firma, die das erste 5.25“-Diskettenlaufwerk (Auftraggeber war die Fa. Wang Laboratories) auf den Markt bringt. 1978 stellten bereits 10 Firmen 5.25“-Laufwerke her. 1980 stellt Sony die 3.5“-Diskette der Öffentlichkeit vor. Dieses Diskettenformat mit der gesteigerten Kapazität von 1.44 MB ist heute noch in sehr alten PCs gebräuchlich. Die Miniaturisierung bei den Diskettenlaufwerken brachte es u. a. mit sich, dass eine Kombination von 5.25“- und 3.5“-Floppy in einem halbhohen Gehäuse untergebracht werden konnte. Verschiedene 3“- (Amdisk) und 2“-Formate wie auch die LS-120-Diskette (Panasonic, sogar abwärts kompatibel zur 1.44 MB Diskette) und die Zip-Diskette (Iomega, 1994, 100 MB – 750 MB) konnten sich nicht mehr als Standard durchsetzen.

Festplatten

Die ersten Ansätze für mobile Festplatten sah man 1973. IBM stellte 1973 die Festplatte ‘Winchester 3340’ vor. Kapazität: 35 MB. Der Name Winchester stammt daher, da das produzierende Werk in Winchester (England) stand. Die IBM 3340 benutzte als auswechselbaren Datenträger das Plattenmodul IBM 3348. Dieses Plattenmodul war leicht zu transportieren, da es im Vergleich zum Vorgänger IBM 3330 kleiner war und nur etwa die Hälfte wog (siehe auch unter IBM 3340).

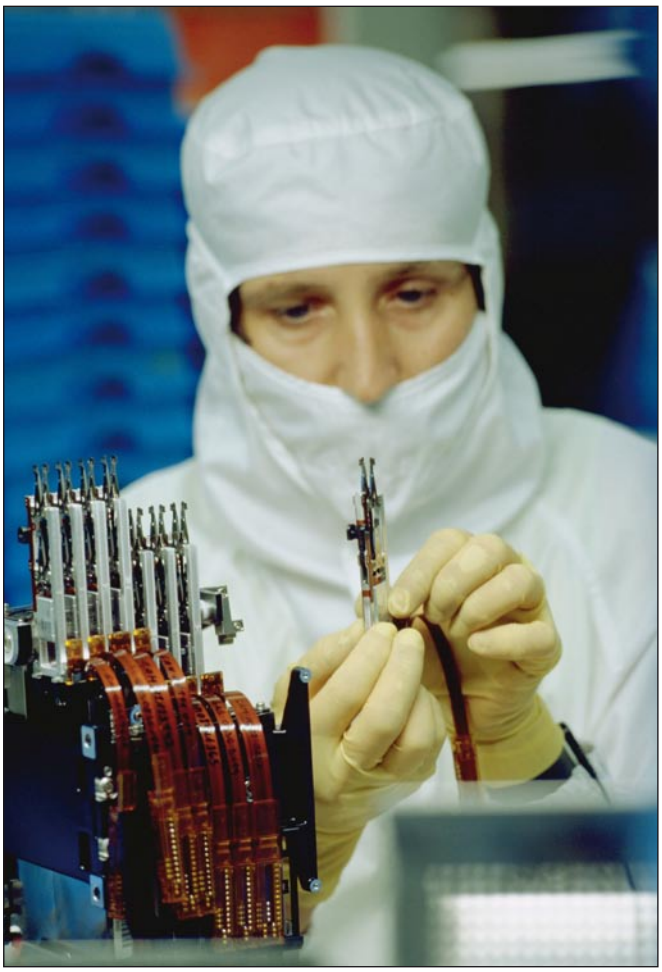
Wie funktionieren nun Festplatten? Eigentlich bis heute nach dem gleichen Prinzip. Im einem luftdicht verschlossenen Gehäuse (beinahe luftdicht, denn ein gewisser Luftaustausch findet statt) sind mehrere übereinander rotierenden Magnetplatten montiert. Bei neueren Festplatten sind das – zur Reduzierung der Bauhöhe – nur noch wenige Magnetplatten.



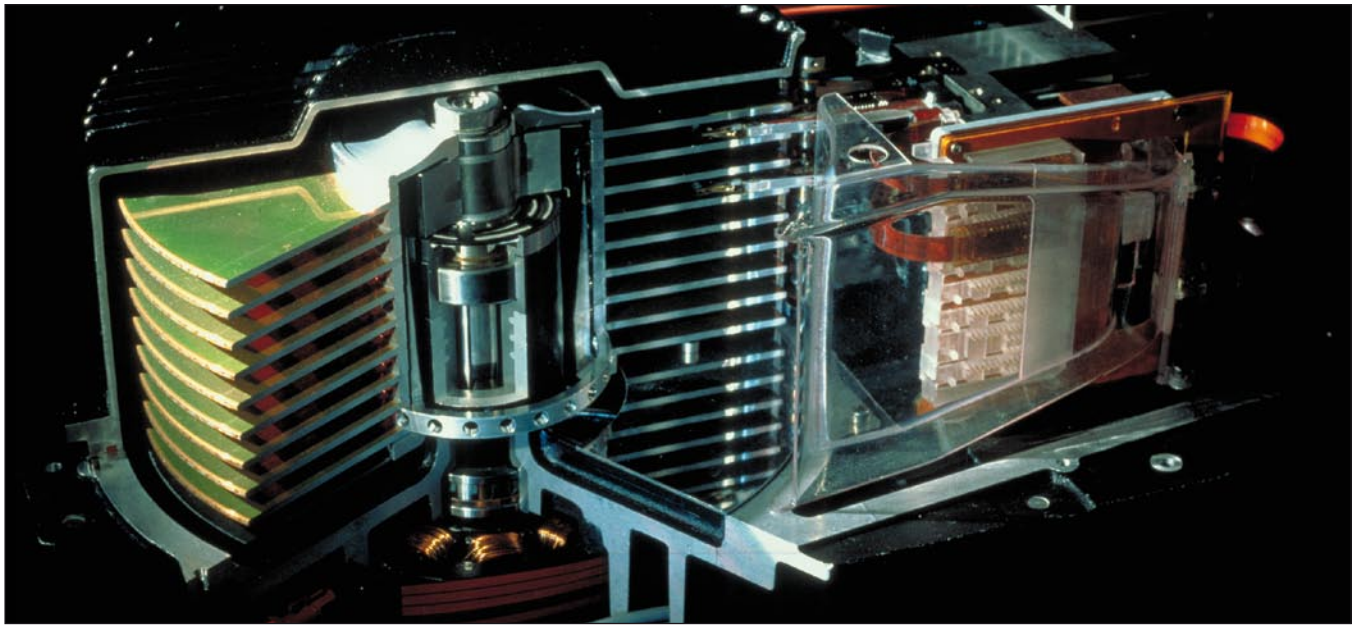
1973: IBM 3340 Winchester Wechselplatte

1977 brachte Shugart das erste preiswerte Laufwerk auf den Markt (14“, 30 MB). Die weitere Entwicklung führte zu kleineren Platten. Seagate baute 1979 die erste Festplatte im 5.25-Zoll-Format. 1981 kam SCSI und 1982 gab es die ST506-Schnittstelle von Seagate, aus der sich IDE, E-IDE, ATA und ATAPI entwickelt haben. Das Seagate-ST506-Laufwerk, nach dem die Schnittstelle benannt wurde, verfügte (wie das RAMAC-Laufwerk aus dem Jahre 1956) über eine Kapazität von 5 MB.

1996 hatte Seagate mit der Cheetah-Serie erste Festplatten mit 10000 U/min präsentiert. 1998 bot die Seagate-Barracuda-Serie eine Maximalkapazität von 50 GB. Nur zwei Jahre später waren es schon fast 200 GB. Dies übertraf die bislang übliche Steigerung von 60 Prozent in einem Jahr oder die Verdoppelung in 18 Monaten bei weitem. Zwischen 1957 und 1990 lag die Steigerungsrate noch bei etwa 25 Prozent im Jahr. Die Flächendichte auf den Festplattenscheiben stieg von 2000 Bit/inch^2 im Jahr 1957 auf über 1 Gbit/inch^2 in den Jahren 1995 bis 1997.

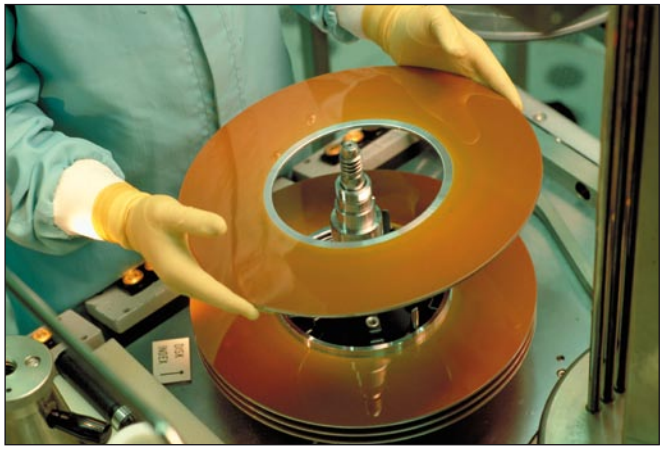


1990: IBM 3390 Montage der Zugriffsarme



Im Schnitt: IBM 3390 Laufwerk

Die Datentransferraten der Festplatten waren mindestens das Zehnfache höher als bei Disketten, weil sich die Platten mit bis zu 20facher Geschwindigkeit einer Diskette drehten, je nach Typ zwischen 3500 und 15000 U/min. Jede Platte besitzt mindestens zwei Lese- und Schreibköpfe, die die Platten auf der Ober- oder Unterseite abtasten. Es gibt allerdings auch Festplatten, die über mehrere Sets an Schreib-/Leseköpfen verfügen, z. B. Festplatten in Hochleistungsrechnern oder SCSI-Festplatten mit 4 R/W-Köpfen, wodurch die Zugriffszeit verringert wird. Durch das Rotieren der Magnetscheiben entsteht ein Luftpolster, auf dem sich dann die Schreib-/Leseköpfe bewegen. Diese sind mechanisch über den Schreib-/Lesearm (Kamm) auch mit den anderen Köpfen verbunden, so dass ein Spurwechsel für alle Platten gleichzeitig vollzogen wird. Um die Bewegung von Kopf zu Kopf zu beschreiben, hat man den Begriff des Zylinders eingeführt. Dieser umfasst alle Spuren, die die gleiche Spurnummer tragen. Jede Platte ist also in Spuren aufgeteilt, die in konzentrischen Kreisen um den Mittelpunkt angeordnet sind (ähnlich einer Schallplatte). Spuren werden mit den Nummern von 0 bis N bezeichnet, wobei N die Gesamtzahl der Spuren minus 1 darstellt. Die äußere Spur trägt die Nummer 0, die darauffolgende die Nummer 1. Jede Spur nimmt dabei eine konstante Anzahl von Sektoren auf, die die Spur in einzelne Abschnitte gleicher Größe unterteilen. Der Aufbau gleicht also dem einer Diskette. Jeder Sektor beinhaltet 512 Bytes an Daten (Fixed Block Architecture FBA) und stellt zugleich die kleinste Zugriffseinheit dar. Jede Festplatte verfügt über wesentlich mehr konzentrische Spuren als eine Diskette, die Positionierungsgenauigkeit ist höher und somit wird auch die mögliche Speicherdichte größer.



1987: IBM 3380 Fertigung im IBM Werk Berlin

1990: IBM 3390 Montage

MFM (Modified Frequenz Modulation)

MFM setzte sich als neues Verfahren zur Übertragung von Daten über den Schreib-/Lesekopf auf die Oberfläche der Festplatte durch. MFM–Festplatten waren bis Mitte der 80er-Jahre der Standard bei den PC-Festplatten. Sehr verbreitet waren zuerst Festplatten mit einer Kapazität von 20, später von 40 MB. Man bedenke: Zur gleichen Zeit passte ein komplett lauffähiges MS-Word 2.0 auch noch auf eine 360-K-Diskette. Aktuelle Motherboards haben keine Anschlüsse mehr für diese Art von Festplatten, entsprechende 8- und 16-Bit-ISA-Controller können aber noch auf Flohmärkten erworben werden.

RLL (Run Length Limited)

Prinzipiell ist der Aufbau mit dem der MFM–Festplatte identisch, nur die Speicherkapazitäten waren größer. Das ergab sich aus der verbesserten Oberfläche der Platten. Auch die Steuerung der Laufwerke verbesserte sich sehr stark. Dadurch wurden pro Spur 26 Sektoren möglich, was eine erhebliche Erhöhung der Speicherdichte bedeutete. Von RLL-Platten ist allerdings heute nur noch die Art und Weise des Aufzeichnungsverfahrens übriggeblieben. Ansonsten sind sie – wie die MFM-Festplatten – veraltet.

Entwicklung Speicherdichte pro Produkt von 1956 bis 2000

Jahr	Produkt	Bits/Inch (bpi)	Tracks/inch (tpi)	Areal Density (MBit/inch²)	Areal Density (MBit/cm²)	Kopftechnik
1956	350	100	20	0,002	0,00031	Schwebekopf
1957	355	100	20	0,002	0,00031	
1961	1405	220	40	0,009	0,00140	
1962	1301	520	50	0,026	0,00400	Flugkopf
1963	1311	1025	50	0,051	0,00790	
1964	2311	1100	100	0,110	0,01700	
1966	2314	2200	100	0,220	0,03400	
1971	3330	4400	192	0,780	0,12100	
1973	3340	5636	300	1,690	0,26200	
1976	3350	6425	478	3,070	0,47600	
1979	3370	12134	635	7,800	1,20000	Dünnfilmkopf
1981	3380-D	15200	840	12,000	1,86000	
1985	3380-E		1400	19,290	2,99000	
1987	3380-K	15240	2083	36,000	5,47000	
1989	3390-1/2	27483	2235	60,510	9,38000	
1991	3390-3	29718		90,000	13,50000	DFpl + MR
1991	9340/9345	24130	4450	111,000	16,00000	
1993	3390-9		2921	270,000	41,85000	
1994	RAMAC 1			260,000	40,30000	
1995	RAMAC 2			544,000	84,32000	
1996	RAMAC 3			829,000	129,20000	
1999	2105 ESS			2758,000	427,50000	GMR-Lesekopf
2000	2105 ESS			55160,000	855,00000	

DFpl + MR = Dünnfilmplatte und Magnetoresistiver Schreib-/Lesekopf, GMR = Giant Magnetoresistiver Lesekopf

IBM Travelstar 60-GB-Serial-ATA-Laufwerk mit 7200 RPM

IBM Ultrastar 146-GB-Laufwerk mit 10000 RPM

Festplatten in den ersten IBM PCs

Der erste Schnittstellen-Standard für Festplatten wurde 1980 von Seagate entwickelt: ST-506. Dieser wurde in einer 5 MB großen Festplatte mit eben dieser Modellbezeichnung verwendet. Bereits ein Jahr später entstand – ebenfalls von Seagate – der ST-412 Standard. Diese 10 MB große Festplatte wurde u.a. im populären IBM XT (5160) verbaut. Diese beiden frühen Standards unterschieden sich in einem wesentlichen Aspekt von ihren Nachfolgern IDE/ATA und SCSI: Sie hatten keine eigene Intelligenz. Die komplette Steuerung wurde von einem komplex aufgebauten Controller erledigt, an den sie mit zwei Kabeln angeschlossen waren. Für die Steuerung der Plattenmechanik wurde der Controller über ein 34-adriges Kabel mit der Platte verbunden, für den Datentransport mit einem 20-adrigen Kabel. Die ganze Sache war noch sehr fehlerträchtig und langsam, denn die von den Schreib-/Leseköpfen verarbeiteten Daten mussten erst auf den Controller transportiert werden, bevor ein Fehler festgestellt werden konnte. Zudem waren diese frühen Festplattenstandards aufwändig in der Installation. Viele komplizierte Parameter (z.B. der 'Interleave-Faktor') und die gesamte Geometrie der Festplatte mussten von Hand eingestellt werden, denn es gab noch kein BIOS, in dem diese Werte voreingestellt gewesen wären. Und Plug&Play (also die automatische Erkennung der Festplattengeometrie), so wie wir es heute kennen, kam erst Mitte der 90er-Jahre.

Der Nachfolger der MFM/RLL-Festplatten war die IDE-Festplatte (Integrated Drive Electronics) von IBM. Aber auch diese werden heute eigentlich nicht mehr verwendet, denn die Kapazität war auf netto 504 MB begrenzt. Die heute verwendeten Festplattenarten sind EIDE (Enhanced Integrated Drive Electronics), SCSI (Small Computer Systems Interface) und aktuell SATA (Serial Advanced Technology Attachment), eine erst im Jahre 2001 definierte Weiterentwicklung von ATA (Advanced Technology Attachment).

Der Unterschied lag bei diesen neueren Standards hauptsächlich im Controllerbereich. SCSI-Festplatten waren zu der Zeit noch etwas schneller, aber auch wesentlich teurer. Der Vorteil von SCSI lag in seiner Flexibilität, denn so ziemlich alle Arten von Geräten (Floppy, Streamer, CD-ROM, Festplatte, Scanner etc.) können an einen SCSI Controller angeschlossen werden, und zwar bis zu sieben Stück pro Controller (Weiterentwicklung dann 16 pro Kanal). Ein EIDE Controller konnte dagegen nur 2 Geräte verwalten, denn diese Schnittstelle war eigentlich ausschließlich auf den Betrieb mit Festplatten ausgelegt. Trotzdem gibt es auch für EIDE mittlerweile CD-Laufwerke oder Streamer. Moderne Motherboards enthielten oft bereits zwei integrierte EIDE Controller. Dadurch konnten 4 entsprechende Geräte betrieben werden, was für die meisten Nutzer vollständig ausreichte. SATA überträgt die Daten im Gegensatz zu allen anderen Systemen seriell, d.h. Bit für Bit. Trotzdem werden Übertragungsraten von bis zu 600 MB je Sekunde erreicht. SATA ist kompatibel zum ATAPI-Standard und den verschiedenen ATA-Vorgängern. Ein weiterer großer Vorteil: Eine Konfiguration entfällt fast vollständig und Geräte können bei eingeschaltetem Rechner ein- und ausgesteckt werden (Hot-Plugging).

In den 80er-Jahren bis 1991 wurden **groß dimensionierte braune Platten gebaut**. Die IBM 3380 Technologie mit 14-Zoll-Platten, die später von 10.5-Zoll-Platten, der 3390 Technologie, abgelöst wurden, beschäftigte sich konstant mit der Frage, wie viel Kapazität unter einem Actuator, also unter einem Schreib-/Lesekopf, verwaltet werden kann, ohne dass es zu Warteschlangen kommt und die Leistungsfähigkeit des Laufwerks beeinträchtigt wird. Aufgrund dieser Problematik wurden auf dem Zugriff in der Regel 2 Schreib-/Leseköpfe aufgebracht. Dabei war ein Kopf für den Innenbereich und der andere für den Außenbereich der Platte zuständig.

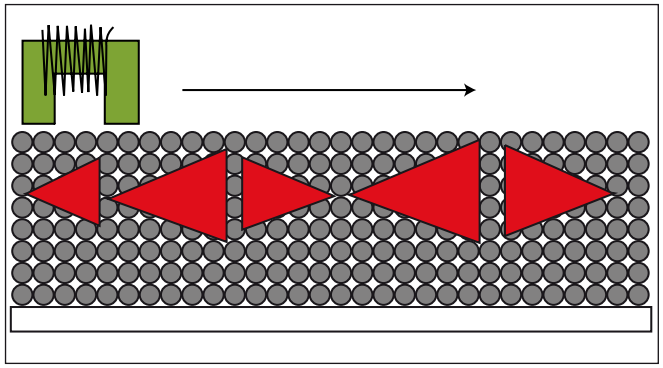
Die **induktive Aufzeichnung** wurde **bis 1991** verfolgt und eingesetzt. Stellen wir uns einen solchen Aufzeichnungs-Schreib-/Lesekopf als aufgeschnittenen Transformator mit einer entsprechenden Kupferspule vor. Der Schreibvorgang verlangt nun eigentlich wenige Kupferwindungen bei Verwendung eines möglichst dicken Drahtes, damit möglichst viel Strom transportiert werden kann und ein gut durchmagnetisiertes Bit auf der Magnetschicht erzeugt wird. Das Lesen eines Bits verlangt aber genau das Gegenteil, nämlich möglichst viele Kupferwindungen bei einem möglichst dünnen Kupferdraht, damit das Streufeld einer Informationseinheit eine möglichst hohe induktive Spannung hervorruft, die dann als eindeutiges Lesesignal abgegriffen werden kann. Der Kopf wurde aufgrund der gegensätzlichen Anforderungen bei der Anzahl der Windungen und der Dicke des Kupferdrahtes so gestaltet, dass beide Vorgänge möglich waren. Dieser **Kompromiss** schaffte aber eine direkte Abhängigkeit bezüglich der Geschwindigkeit zwischen Kopf und Platte. Drehte man zu schnell, wurde der Schreibvorgang schwierig, drehte man zu langsam, reichte die Drehgeschwindigkeit nicht aus, um eine entsprechend hohe induktive Spannung zu generieren, die dann als sauberes Lesesignal abgegriffen werden konnte. Der komplexeste Fertigungsprozess in dieser Zeit war die **Herstellung der Platte** selbst.

Als Magnetisierungsträger wurden Eisenoxydpartikel, also Rostpartikel, verwendet und die Kunst war es, diese Rostpartikel möglichst homogen in einer Kunstharzmasse zu verteilen, um dann einigermaßen gleichmäßig magnetisieren zu können. Der Herstellprozess der Köpfe war im Vergleich dazu einfacher.

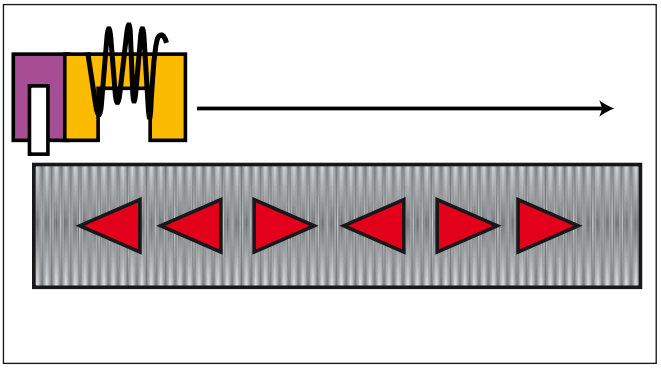
Diese braunen Platten bereiteten Anfang der 90er-Jahre (1990 – 1993) enorme Probleme, speziell bei den Modellen 1, 2 und 3 der 3390 Familie, weil sich nach langjähriger Benutzung die Eisenoxydpartikel durch die andauernden Drehkräfte aus der Kunstharzmasse herausarbeiteten und sich an den Köpfen anlagerten. Die Anlagerung dieser Rostpartikel erzeugte dann den berühmten Flugeffekt, das heißt, der Abstand zwischen den Köpfen und der Plattenoberfläche wurde immer größer, bis eine Höhe eintrat, die es nicht mehr erlaubte, das geschrieben oder gelesen werden konnte. In einer einzigartigen Aktion, die über zwei Jahre in Anspruch nahm, tauschte die IBM damals alle weltweit installierten Plattenstapel aus.

Bei der induktiven Aufzeichnungstechnik wurden **lineare Zugriffsmechanismen** eingesetzt, die von der Mechanik sehr aufwändig gestaltet waren, um den Kopf, der ja sowohl für das Schreiben als auch das Lesen zuständig war, exakt über einer Spur zu positionieren.

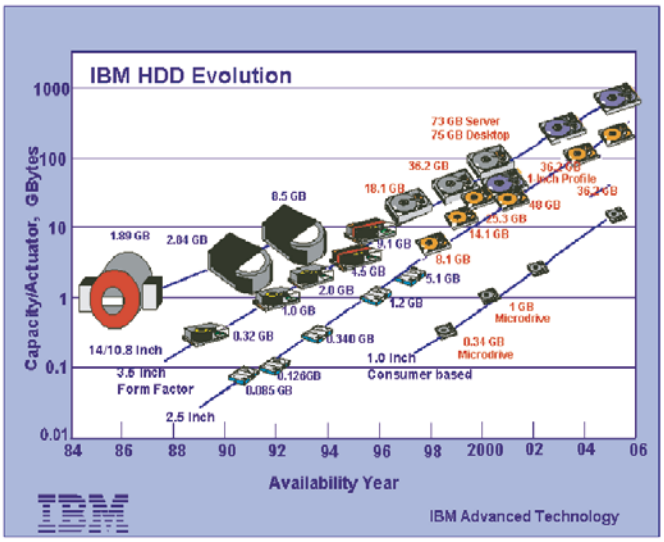
1991 verabschiedete sich IBM von der damals eingesetzten induktiven Aufzeichnung und führte als erstes Unternehmen **MR-Technologie** für die Kopftechnik (**magnetoresistive Aufzeichnung**) bei den Produkten 3390 Modell 3 und 9 sowie bei dem 9340 System mit seinen 5.25-Zoll-Platten ein. Heute arbeitet jeder Hersteller mit dieser Technologie, die IBM entwickelt hat. Die MR-Kopftechnik hatte so viel Potenzial für die Weiterentwicklung, dass uns diese Technologie bis in die heutige Zeit begleitet (2008). Der Unterschied zur induktiven Aufzeichnung liegt darin, dass man nicht mehr mit einem



Prinzip der induktiven Aufzeichnungstechnik



Prinzip der magneto-resistiven Aufzeichnungstechnik



1991:
RAID-Architekturen
Dünnfilmbeschichtung
MR-Aufzeichnung

1995 – 1997:
PRML Encoding

1999:
Micro Drives
Paramagnetischer Effekt

2000:
USA National Medal of Technology for Innovations in Storage

Schreib-/Lesekopf arbeitet, bei dem die Anzahl und die Dicke der Kupferspulen so ausgelegt sein muss, dass sowohl das Schreiben als auch das Lesen möglich wird. Der MR-Kopf verwendet nach wie vor für den Schreibvorgang einen induktiven Kopf, bei dem die Anzahl und Dicke der Kupferspulen jetzt aber so ausgelegt ist, dass ein optimaler Schreibvorgang stattfinden kann. Mit dem Lesen hat dieser Kopf nichts mehr zu tun. Damit werden wesentlich schnellere Drehgeschwindigkeiten möglich. Für den Lesevorgang macht man sich den 'magnetoresistiven Effekt aus der Physik' zunutze, dass Metalle oder Legierungen ihren Widerstand verändern können, wenn sie in den Einfluss eines Magnetfelds geraten. Gearbeitet wird in der Regel mit einem Nickel-Eisen-Film, der an Schwachstrom angeschlossen ist. Gerät dieser Slider in den Einflussbereich eines Magnetfelds, ändert sich die Ausrichtung der Elektronen. Wenn man das entsprechend verstärkt, bekommt man ein klares, eindeutiges Lesesignal. Mit diesem Verfahren können selbst kleinste Streufelder einzelner Informationseinheiten abgegriffen werden. Auf diese Weise schafft man einen **Spezialisten für den Schreibvorgang und einen Spezialisten für den Lesevorgang**.

Parallel dazu, als **zweites wichtiges Element**, änderte sich die Plattenherstellung. Die Laminierverfahren und Vakuumprozesse waren inzwischen so weit, dass auf der Platte mit Dünnfilmbeschichtung (Legierungen) gearbeitet werden konnte und anstelle von Eisenoxyd nun ein Metallfilm zum Einsatz kam. Der Metallfilm wird unter Vakuum auf eine speziell gehärtete Glasplatte gesprüht (dieser Vorgang wird als Kathodenzerstäubung bezeichnet). Anschließend läuft die Platte direkt vom Vakuum in einen Ofen, um 'gebacken' zu werden (dieser Vorgang wird als Sputtern bezeichnet). Durch den Erhitzungsprozess (Sputtern) entstehen im Dünnfilm 'Grains', also Körnerstrukturen, und die einzelnen Grains stellen die Magnetisierungsträger der Platte dar. Die Herstellung solcher Dünnfilmplatten ist vom Prinzip bis in die heutige Zeit beibehalten worden und wird etwas später im Teil der AFC -Technik näher beschrieben. Mit der Kombination von MR-Kopftechnologie und Dünnfilmplatten war es möglich, die Aufzeichnungsdichte von damals – 1991, bei der Einführung – 16 Millionen Bits auf dem Quadratzentimeter auf über 5 Milliarden Bits auf dem Quadratzentimeter zu steigern. Das entspricht einer Steigerung von über Faktor 300 und stellt eine einmalige

technologische Evolution dar, wie sie in keinem anderen technologischen Segment vom Steigerungsfaktor her jemals erzielt wurde.

Der **dritte treibende Faktor 1991** war die Einführung neuer Subsystemarchitekturen, der RAID-basierenden Systeme. **Die RAID-Architekturen** leiteten den Schritt ein, von großen Plattengrößen in kleinere Formfaktoren zu wechseln. Die kleinen Platten hatten zwar nicht den vergleichbaren Zuverlässigkeitsfaktor wie die großen braunen Platten, die sogenannten SLEDs (Single Large Expensive Disk), wie sie am Schluss ihrer Lebensperiode bezeichnet wurden, benötigten diesen aber auch nicht, weil Laufwerkausfälle in einem Array durch RAID abgesichert waren (siehe auch unter RAID).

Im Jahr **1990** integrierte IBM in die damals gefertigten 5,25- und 3,5-Zoll-Laufwerke ein neues Encoding-Verfahren, das **als PRML Encoding** bezeichnet wird und heute noch im Einsatz ist. **Partial Response/Maximum Likelyhood (PRML)** ist ein Schreibverfahren, um Daten möglichst platzsparend auf magnetische Datenträger zu schreiben. Dazu werden diese Daten vorher kodiert. Durch diesen Trick erhält man eine höhere Datendichte auf der Oberfläche einer Festplatte. Dies geschieht vom Nutzer völlig unbemerkt durch den Controller der Festplatte. Im Gegensatz zu älteren Verfahren wie MFM (Modified Frequence Modulation) werden beim Lesen des Datenträgers keine einzelnen Magnetfeld-Umkehrungen im analogen Signalstrom identifiziert und deren Abfolge als Sequenz von Daten- und Synchronisierungsbits interpretiert. Stattdessen wird das analoge Signal in Abschnitte zerlegt, diese aufbereitet ('Partial Response') und das Ergebnis mit vorgegebenen Mustern verglichen, um das ähnlichste zu finden ('Maximum Likelyhood'). Jedes der vorgegebenen Muster steht für eine bestimmte Bitfolge. PRML erlaubt eine 30 – 40 % höhere Datendichte als MFM. PRML wird heute noch bei modernen Festplatten eingesetzt, aber mehr und mehr von EPRML (Enhanced) verdrängt, das wegen seiner verbesserten Algorithmen und Signalverarbeitung 20 bis 70 Prozent höhere Datendichten ermöglicht.

PRML wurde 1990 erstmals für Platten der Systeme IBM RS/6000 und AS/400 eingesetzt, weiterentwickelt und danach auch bei damals vorrangig gefertigten 3.5-Zoll-Laufwerken aller IBM Produktreihen genutzt, ebenso in vielen Folgeprodukten. Entwickelt wurde das Verfahren im IBM Labor in Rüschlikon bei Zürich.

Ein Beispiel: Wurden herkömmlich etwa 800 – 900 durchmagnetisierte Grains benötigt, um eine stabile Informationseinheit abzubilden, konnte durch das PRML-Verfahren im Laufe der Jahre die Anzahl der Grains auf 200 – 250 reduziert werden, d. h. vierfach höhere Kapazitäten.

Die kapazitiven Auswirkungen von PRML sah man vor allem von 1995 bis 1997, wo sich die Kapazitäten der damals produzierten 3.5-Zoll-Laufwerke innerhalb von 12 bis 15 Monaten verdreifachten.

1997 wurde als Weiterentwicklung der MR-Köpfe die **GMR-Technologie**, die 'giant'-magnetoresistance Lesekopftechnik, eingeführt, die vor allem das Auslesen von kleinsten Streufeldern erlaubte. Den **GMR-Effekt** entdeckten Ende der 1980er-Jahre unabhängig voneinander **Peter Grünberg** an der KFA in Jülich und **Albert Fert** an der Universität Paris-Süd. Bei beiden Entdeckungen wurden extrem niedrige

Einführung von PRML-Kanälen in IBM Platten

Platte	Kapazität GB	Durchmesser Zoll	Datenrate MB/S	Aufzeichnung Dichte Mb/sq"	Kopftype Plattenart	PRML
Redwing 0681 Feb. 1990	0.857	5¼	3	63	MIG Dünnfilm	1. Gen.
Corsair 0663 Sep. 1991	1.0/1.2	3½	3	132	MR Dünnfilm	1. Gen.
Allicat 0664 Nov. 1992	2.0	3½	5.2	264	MR Dünnfilm	2. Gen
Spitfire 0662 März 1993	1.0	3½	5.2	354	MR Dünnfilm	2. Gen
Starfire-Reihe 0667/0668 Nov. 1993	1.0 – 5.2	3½	12.2	564	MR Dünnfilm	3. Gen

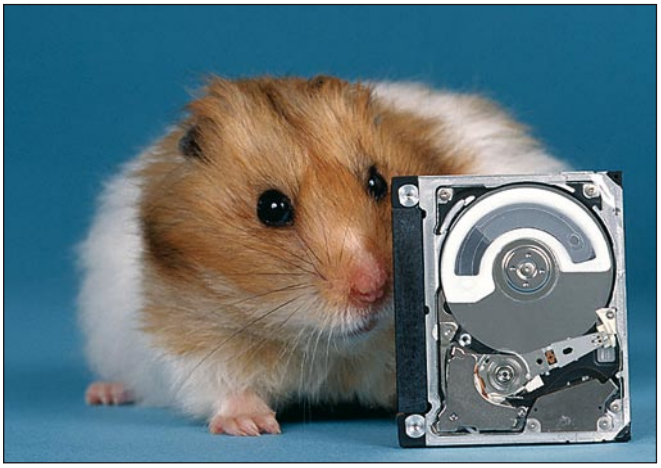
Temperaturen und starke magnetische Felder zur Erzeugung des Effekts benötigt. Einige Zeit nach diesen Arbeiten begannen **Stuart Parkin** und andere Forscher im **Almaden-Forschungszentrum von IBM** in San Jose, Kalifornien, mit dem Versuch, den Effekt bei Normaltemperatur und mit weniger exotischen Materialzusammensetzungen für die Massenfertigung weiterzuentwickeln. Sie benötigten dazu viele Jahre und erforschten mehr als 30 000 unterschiedliche Multilayer-Materialkombinationen. 1997 erhielten die drei genannten Forscher gemeinsam den **Europhysik-Preis** für die GMR-Entwicklung. Peter Grünberg und Albert Fert wurden zudem **2007** mit dem **Nobelpreis für Physik** ausgezeichnet.

Der GMR-Effekt (engl. giant Magnetoresistance, dt. 'Riesenmagnetwiderstand') wird in Strukturen beobachtet, die aus sich abwechselnden magnetischen und nichtmagnetischen dünnen Schichten mit einigen Nanometern Schichtdicke bestehen. Der Effekt bewirkt, dass der elektrische Widerstand der Struktur von der gegenseitigen Orientierung der Magnetisierung der magnetischen Schichten abhängt, und zwar ist er bei Magnetisierung in entgegengesetzte Richtungen deutlich höher als bei Magnetisierung in die gleiche Richtung.

Dabei handelt es sich um einen **quantenmechanischen Effekt**, der durch die Spinabhängigkeit der Streuung von Elektronen an Grenzflächen erklärt werden kann. Elektronen, die sich in einer der beiden ferromagnetischen Schichten gut ausbreiten können, weil ihr Spin günstig orientiert ist, werden in der zweiten ferromagnetischen Schicht stark gestreut, wenn diese entgegengesetzt magnetisiert ist. Sie durchlaufen die zweite Schicht aber wesentlich leichter, wenn die Magnetisierung dieselbe Richtung aufweist wie in der ersten Schicht.

Werden zwei Schichten eines ferromagnetischen Materials durch eine dünne nichtmagnetische Schicht getrennt, so richten sich die Magnetisierungen bei bestimmten Dicken der Zwischenschicht in entgegengesetzten Richtungen aus. Schon kleine äußere magnetische Felder reichen aber aus, um diese antiferromagnetische Ordnung wieder in die ferromagnetische Ordnung zurückzuführen.

In Verbindung mit dem GMR-Effekt bewirken Variationen des äußeren Magnetfeldes in geeigneten Strukturen daher große Änderungen des elektrischen Widerstandes der Struktur.



IBM Micro-Drive

Die Möglichkeiten, den Effekt in einem Sensor für ein magnetisches Feld einzusetzen (und damit als einen neuen Typ von Lesekopf in einer Festplatte), wurden schnell durch ein von Stuart Parkin geleitetes IBM Forschungsteam entdeckt, indem er zeigte, dass der Effekt auch in polykristallinen Schichten auftritt.

IBM stellte im **Dezember 1997** das erste kommerzielle Laufwerk her, das diesen Effekt nutzte. GMR ist bis heute im Einsatz. Neben der Anwendung in Festplatten wird der GMR-Effekt auch in Magnetfeldsensoren der Automobilindustrie und Automatisierungsindustrie ausgenutzt.

1999 konnten die ersten **Micro-Drives** – damals mit einer Kapazität von 340 MB – auf dem kommerziellen Markt eingeführt werden. Die Kapazitäten der Micro-Drives entwickelten sich rasch weiter und heute, im Jahr 2008, werden bereits Kapazitäten von 6 GB und 8 GB angeboten.

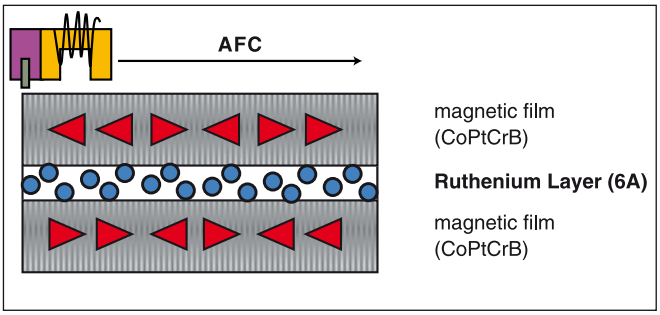
Durch einen glücklichen Zufall entdeckten die IBM Entwickler in der 2. Hälfte des Jahres **1999** die Möglichkeit einer neuen Aufzeichnungstechnik. Man sprach und spricht heute noch von der **Entdeckung des paramagnetischen Effekts**. Schon am 4. Oktober 1999 konnte das IBM Labor in Almaden eine Hard-Disk-Laborversion in dieser neuen Technik demonstrieren. In einer Presseveröffentlichung am 4. Oktober 1999 wurde von einer Aufzeichnungsdichte von 35.3 Gbits/inch^2 gesprochen – das sind **über 5 Milliarden Bits auf den Quadratzentimeter**. Damit war die Sprengung der bis dahin für MR-Technik gesehenen physikalischen Grenzen eingeleitet. Mit dieser Technik lassen sich vierfach höhere Kapazitäten auf demselben Raum realisieren!

Um die magnetische Beschichtung einer Festplatte heute zu fertigen, wird unter Vakuum eine komplexe Legierung auf eine Glasscheiben-Oberfläche gesprüht (Kathodenzerstäubung), die anschließend 'gebacken/gebrannt' wird (Sputtern). Dabei entstehen ca. **80 – 100 Nanometer große magnetische Körner** (MR-Technik). Für das Speichern von einem Bit richtet ein induktiver Schreibkopf die magnetische Orientierung von einigen Hundert solcher Partikel aus. Um Speicherdichten zu vergrößern, versuchte man, die Zahl der Körner zu reduzieren, die ein Bit bilden. Das hatte jedoch den gravierenden Nachteil, dass sich das Signal-Rausch-Verhältnis verschlechterte. Deshalb versuchte man, die Körner zu verkleinern und die Anzahl beizubehalten, die ein Bit bilden. Kommt allerdings die Korngröße unter ein bestimmtes Limit, verlieren die Partikel ihre ferromagnetischen Eigenschaften (paramagnetisches Limit). Diesen Effekt kann man teilweise ausgleichen, indem man 'härtere' magnetische Materialien mit einer höheren Koerzitivfeldstärke verwendet. Problem hier ist die Tatsache, dass 'härtere' Materialien wiederum schwieriger zu magnetisieren sind und damit wesentlich mehr Energie für die Magnetisierung benötigen.

Mit der Entdeckung des **paramagnetischen Effekts** und der paramagnetischen Aufzeichnungsmöglichkeit wurde eine **Korngröße von ca. 25 – 30 Nanometer** möglich, ohne dass auf härtere Materialien ausgewichen werden musste. Das bedeutete für die Aufzeichnungsdichte eine Vervierfachung der Kapazitäten, die auf demselben Raum abgebildet werden konnten.

Dem MR-Kopf wird ein Spezialkopf vorgeschaltet, der ein vertikales Bit in der Metallkörnerstruktur erzeugt. Gleich danach wird das horizontale Datenbit geschrieben. **Beide Bits stabilisieren** sich bei bestimmten Legierungen und lassen es zu, das horizontale Datenbit wesentlich kleiner zu gestalten und mit wesentlich höheren Drehgeschwindigkeiten und Datenraten auf der Platte zu arbeiten.

Die Maßenproduktion dieser Technik wäre sehr komplex und kostenintensiv geworden. Deshalb beschritten die Entwickler **andere Wege**. Dies führte dazu, dass IBM die Maßenproduktion der sogenannten **Pixie-Dust-Technologie** im Mai 2001 einleitete, was damals die gesamte Fachpresse aufrüttelte, weil noch wesentlich höhere Kapazitäten bei noch schnelleren Drehgeschwindigkeiten realisiert werden konnten.



Prinzip der AFC(AntiFerromagnetically Coupled)-Aufzeichnungstechnik

Im **Jahr 2000** erhielt die IBM als erste Firma die **'National Medal of Technology for Innovations in Storage'** von der amerikanischen Regierung, eine Auszeichnung, die bisher ausschließlich an Einzelpersonen verliehen wurde. Grund für die Auszeichnung die Maßenproduktion von Micro-Drives und die Entdeckung des paramagnetischen Effekts. Sie ist eine der höchsten technologischen Ehrungen.

Die AFC-Aufzeichnungstechnik (Pixie Dust) erlaubt es, bei **mittleren Korngrößen von 4 – 8 Nanometer** im Dünnfilm eine wesentlich höhere thermische Langzeitstabilisierung zu erzielen, ohne dafür eine höhere Koerzitivfeldstärke des Materials in Kauf nehmen zu müssen. Das heißt: Verwendung von wesentlich kleineren Körnern als Magnetisierungsträger bei weicheren Materialien. Dazu wird dem bisher verwendeten, klassischen Metallfilm aus Kobalt, Platin und Chrom noch Bor zugesetzt.

Diese Technologie nennt sich **AFC (AntiFerromagnetically Coupled)** oder, von den IBM Forschern in Almaden geprägt, **'Pixie Dust'**. AFC-Medien bestehen aus drei Schichten. Zwischen zwei magnetischen Lagen, die aus einer komplexen Kobalt-Platin-Chrom-Bor-Legierung bestehen, befindet sich eine nur **drei Atomlagen (6 Angström) dicke Schicht aus dem nichtmagnetischen Metall Ruthenium**. Bei genau dieser Dicke sorgt die Ruthenium-Schicht dafür, dass die Magnetisierungsausrichtungen der beiden anderen magnetischen Schichten **antiparallel gekoppelt** sind. Das heißt, die untere magnetische Schicht ist stets andersherum magnetisiert als die obere Schicht. Dies führt zu einer **enormen Stabilisierung der beiden Bits** und lässt erheblich höhere Schreibdichten (bis zu 100 Gigabits pro Quadratzoll und mehr – ca.15 – 20 Milliarden Bits auf dem Qudratzentimeter) bei wesentlich höheren Drehgeschwindigkeiten und Datenraten zu. Wegen der wundersamen Wir-

kung der Ruthenium-Schicht tauften die IBM Forscher in Almaden diese neue Technik 'Pixie Dust', also Feenstaub, angelehnt an Peter Pan's Märchen.

Ein weiterer **Vorteil** liegt in der Tatsache, dass mit der heutigen MR-Schreib-/Lesekopftechnologie weitergearbeitet werden kann. Die Miniaturisierung dieser Köpfe ist heute bereits so weit, dass Schreib-/Lesespaltbreiten von 1 – 2 Nanometer als Laborversionen ermöglicht werden. Dies stellt ein riesiges Potenzial dar, die Aufzeichnungsdichte in noch nicht absehbare Dimensionen voranzutreiben.

Die Plattenaufzeichnungstechnologien werden sicherlich mit der neuen AFC-Technik in kapazitive Bereiche von über 1 TB in den Jahren 2008 – 2009 gehen können.

Wegen der Hitzentwicklung in den Laufwerken muss man sich immer mehr über das **Packaging**, im Speziellen die Kühlung dieser Platten Gedanken machen. **Wassergekühlte Racks** stellen eine neue Alternative dar oder die von der IBM in der Entwicklung befindliche IceCube-Subsystem-Architektur, die die Hitzeproblematik der Laufwerke entsprechend löst, um noch höhere Kapazitäten bei noch höheren Drehgeschwindigkeiten abbilden zu können. Bei IceCube werden Laufwerke in einem kleinen Kühlschrank betrieben. In der Entwicklung befinden sich auch Möglichkeiten einer 'Alttechnologie', wie sie bei den 3380 Platten eingesetzt wurde, allerdings in entsprechend miniaturisierter Form. Dabei wird mit zwei Köpfen auf dem Arm gearbeitet, der eine kontrolliert den Innenbereich der Platte, der andere den Außenbereich. Dies ist eine Möglichkeit, die Mehrkapazität unter dem Aktuator leistungsmäßig auszugleichen, ohne die Drehgeschwindigkeit des Plattenstapels massiv erhöhen zu müssen. In diesen Richtungen bestehen Möglichkeiten genug, die Kapazitäten in den nächsten Jahren voranzutreiben.

Im Jahr **2003** gründeten **IBM und Hitachi eine gemeinsame Entwicklungs- und Fertigungs-Allianz** für Plattenlaufwerke. In dieser Allianz ist IBM für die Weiterentwicklung der Plattentechnologie zuständig und Hitachi für die Maßenproduktion. Diese Allianz hat sich bis heute bewährt und es ist davon auszugehen, dass sie noch viele Jahre Bestand haben wird und auf andere technologische Bereiche ausgedehnt wird. Inzwischen ist Hitachi der drittgrösste Plattenlieferant weltweit. Das zweitgrößte Produktionsvolumen wird durch Western Digital abgedeckt. Der größte ist nach wie vor die Firma Seagate, die etwa 40 % des weltweiten Gesamtvolumens produziert. Heute, im Jahr 2008, werden zwei Formfaktoren produziert: 2½-Zoll- und 3½-Zoll-Laufwerke.

Spätestens innerhalb der nächsten zwei Jahre – da sind sich alle Plattenhersteller einig – wird generell nur noch mit einem Formfaktor gearbeitet werden, weil sich kein Hersteller bei Preisverfällen von etwa 40 % pro Jahr die Fertigung von zwei unterschiedlichen Formfaktoren leisten kann. Die 3½-Zoll-Laufwerke, die bisher vor allem in Plattensubsystemarchitekturen eingebaut sind, werden durch 2½-Zoll-Laufwerke ersetzt.

Voraussetzung für diese Strategie ist die Produktion von leistungsstarken 2½-Zoll-Laufwerken. Es kristallisiert sich heraus, dass dann innerhalb der 2½-Zoll-Bauweise drei unterschiedliche Techniken eingesetzt werden: die **AFC-Technik**, eine neue Technik mit **'Perpendicular' Recording** und **IDE**-Platten, die als ATA-, SATA- oder FATA-Platten bekannt sind. Diese Vorgehensweise macht die Welt einfacher.

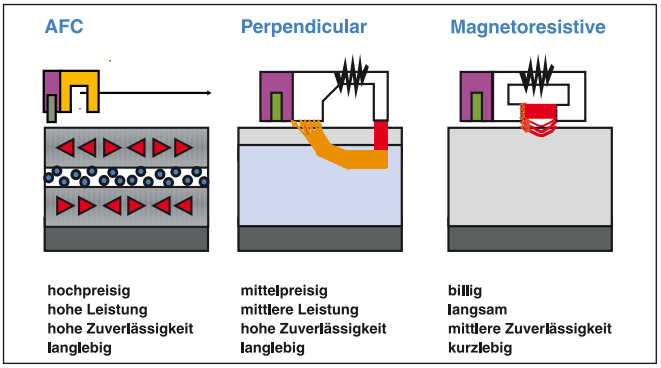
Nur noch ein Formfaktor, das bedeutet:
AFC als hochpreisiges Hochleistungslaufwerk mit extremer Zuverlässigkeit, Laufwerke mit Perpendicular Recording als mittelpreisiges Laufwerk mit guter Leistung und vergleichbarer Zuverlässigkeit wie AFC und die billigen SATA-/FATA-Platten, die vom Aufbau her IDE-Platten entsprechen.

Perpendicular Recording

Bei Perpendicular Recording stehen die magnetischen Momente, die zusammen mit den verwendeten logischen Schreibverfahren wie PRML jeweils ein logisches Bit repräsentieren, nicht parallel zur Rotationsrichtung des Datenträgers, sondern senkrecht dazu. Dies führt zu einer potenziell wesentlich höheren Datendichte, womit auf der gleichen Oberfläche mehr Daten gespeichert werden können. Festgestellt wurde dies erstmals bereits 1976 von Professor Shun-ichi Iwasaki am Tohoku Institute of Technology in Japan.

Nachteilig ist, dass die kleineren Weiss-Bezirke auch einen kürzeren Abstand zwischen Schreib-/Lesekopf und der Magnetoberfläche bedingen, um die Daten noch schreiben und lesen zu können. Daher ist diese Aufzeichnungstechnik technisch schwieriger zu realisieren. Das Gegenstück zum Perpendicular Recording ist das bisher eingesetzte Longitudinal Recording (englisch für Längsaufnahme). Hier tritt jedoch bei zu enger Aneinanderreihung der Magnetpartikel (zu hohe Datendichte) der sogenannte superparamagnetische Effekt auf, d. h., die einzelnen Bits verlieren ihre magnetische Richtung und beginnen zu 'kippen', was zu Datenverlust führt. 3.5"-Festplatten erreichen somit eine maximale Kapazität von derzeit 1 TB, 2.5"-Festplatten kommen auf derzeit 320 GB. Das entspricht einer durchschnittlichen Datendichte von 132 Gbit pro Quadratzoll bzw. 20.46 Gbit pro Quadratzentimeter.

Die Festplattenkapazitäten können durch diese neue Aufzeichnungstechnik um das maximal Zehnfache gesteigert werden. Zudem wird durch die deutlich höhere Datendichte ein Zuwachs der Lese-/Schreibgeschwindigkeit erreicht, da der Lesekopf pro Umdrehung mehr Daten liest und damit bei gleicher Umdrehungszahl die Datenrate steigt. Seagate liefert bereits seit 2006 Platten mit Perpendicular Recording aus.



Aufzeichnungsverfahren im Vergleich

HAMR (Heat Assisted Magnetic Recording)

HAMR (Heat Assisted Magnetic Recording) bezeichnet ein Verfahren, das noch größere Datendichten bei Festplatten erlaubt als das neue Verfahren mit Perpendicular Recording, welches bereits die Marktreife erreicht hat. Somit sollen künftig Festplatten mit Volumina im mittleren einstelligen Terabytebereich möglich sein. Mit Hilfe von HAMR gelingt das Schreiben von Daten auch auf Materialien mit einem schwach ausgeprägten superparamagnetischen Effekt. Der sogenannte Effekt tritt bei heutigen Festplatten auf, wenn man die Datendichte zu sehr vergrößert. Dies bedeutet, dass die Domänen, in die Daten geschrieben werden, ihre ferromagnetischen Eigenschaften verlieren und superparamagnetisch werden, wodurch die Bits zerstört werden.

Mit HAMR wird die Domäne, in welche Daten geschrieben werden sollen, zunächst lokal durch einen Laser erhitzt, um das für einen Schreibvorgang nötige Magnetfeld möglichst klein zu halten und das Schreiben trotz eines schwach ausgeprägten superparamagnetischen Effekts zu ermöglichen. Da beim Erhitzen aber das schützende Schmiermittel der Festplatte verdampft, will der Festplattenhersteller Seagate Nachschub in nanometerdünnen Kohlenstoffröhrchen speichern, welchen letztere bei Bedarf auf die betroffenen Domänen sprühen sollen.

Es wird bereits spekuliert, welche Speicherkapazität mittels dieser Technologie erreicht werden kann. So berichtet itwire.com am 4. Januar 2007, dass es vielleicht bereits im Jahre 2010 Kapazitäten von 37.5 TB von Seagate geben könnte. Das wäre 37.5-mal so viel wie die zur Zeit größte Festplatte von Hitachi mit rund 1 TB. Allerdings wird auch berichtet, dass sich mittels dieser Technologie die Anzahl der Head-Crashes erhöhen könnte und somit die Qualität der Festplatte darunter leiden würde. Diese Spekulation steht jedoch im Widerspruch zu der auf Techworld zitierten Aussage von Seagate, wonach 2010 mit 3-TB-Festplatten zu rechnen sei.

Optische Speichertechnologien

CD-ROM (Compact Disc Read Only Memory)

1983 stellten Philips und Sony die CD-ROM vor, quasi ein Ableger der Musik-CD. Das Medium ist auch ähnlich aufgebaut wie die Musik-CD. Die Datenspeicherung erfolgt während der Herstellung der Platte und die Daten können nur gelesen werden (Analogie: ROM). Im Gegensatz zu Magnetplatten erfolgt die Aufzeichnung – wie bei einer Schallplatte – in einer einzigen, spiralförmigen Spur. In diese vorgeprägte, reflektierende Schicht werden bei der Herstellung der Masterplatte mit einem Laser Löcher (Pits) eingebrannt. Von der Masterplatte lassen sich dann beliebig viele Kopien herstellen.

Die Kopie wird vom Laserstrahl abgetastet, der durch die unterschiedliche Struktur der Speicherfläche mit einer digitalen Information moduliert wird. Die Spurdichte beträgt bis zu 16000 Spuren/Zoll. Als Aufzeichnungsstandard hat sich das Format ISO 9660 durchgesetzt (Transferrate: 1.2 MBit/s, Kapazität: ca. 600 MB). Die CD-ROM dient hauptsächlich der Verbreitung größerer Datenmengen, als Photo-CD und jüngst auch als Videoträger. Die erste CD-ROM-Applikation, die auf CD-ROM ausgeliefert wird, ist 1987 Microsoft Bookshelf. 1990 kommt der Commodore CDTV auf den Markt, ein auf dem Amiga 500 basierender Computer mit integriertem CD-ROM-Laufwerk. 1993 und 1994 gibt NEC den Takt mit Dreifach- (Triple Speed) bzw. Vierfach- (Quad Speed) CD-ROM-Laufwerken an.

Bei der beschreibbaren CD-R ist der Aufbau komplexer als bei der CD-ROM. Unten liegt die Trägerschicht aus Polycarbonat, darauf folgt eine lichtempfindliche organische Substanz, die durchscheinend ist. Dann kommt eine reflektierende Goldschicht und schließlich eine Lack-Schutzschicht. Mit erhöhter Laserenergie kann das organische Material verfärbt bzw. verschmolzen werden und es erhält so eine andere Reflexionseigenschaft. Die Platte kann danach wie eine CD-ROM gelesen werden.

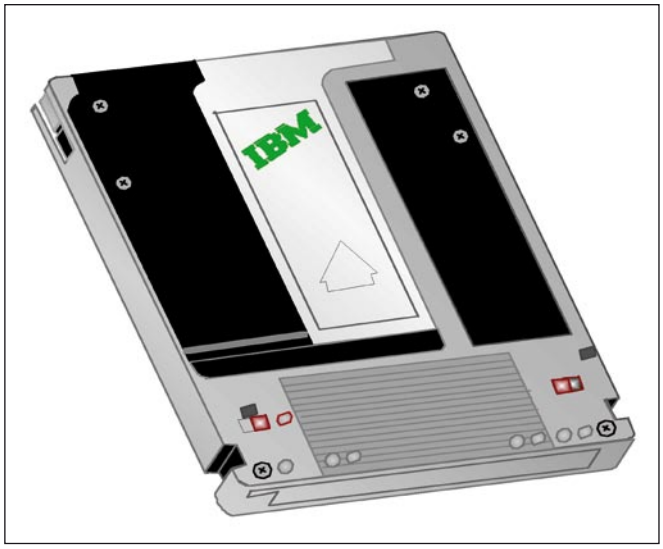
WORM (Write Once Read Many)

WORM-Platten lassen sich vom Anwender beschreiben, jedoch nur einmal (Analogie: PROM). Bei 5,25-Zoll-Platten sind Speicherkapazitäten von weit über 1 GB (Plasmon bietet heute im IT-Umfeld basierend auf blauer Laser-Technik 60-GB-Platten an) möglich. WORM kann zur Archivierung von Daten aller Art verwendet werden (Backup-Medium). Die Platte arbeitet wie ein Magnetplattenlaufwerk und kann genauso angesprochen werden, die Treibersoftware sorgt dafür, dass bei mehrfacher Speicherung einer Datei immer die jüngste Version angesprochen wird (ältere Versionen lassen sich über spezielle Programme lesen) --> Speicherung einer Dateichronologie. Beim Schreiben wird durch hohe Laserenergie die Plattenstruktur dauerhaft verändert. Beim Lesen wird diese Veränderung mit niedriger Laserenergie abgetastet und detektiert. Man unterscheidet zwei Speichertechniken:

Bei der Blasenerzeugung wird durch den Laserstrahl eine Polymerschicht erhitzt, die unter einem dünnen Metallfilm liegt. Es kommt zur Bildung einer Blase, die den Metallfilm dauerhaft verformt. Bei der Abtastung mit geringer Laserenergie kann die geänderte Streuung ausgewertet werden.

Bei der Pit-Erzeugung durchbrennt der Laserstrahl eine lichtundurchlässige Schicht, die über einer Reflexionsschicht liegt (Pit entsteht). Beim Lesen werden die so entstandenen Hell-Dunkel-Zonen ausgewertet.

Ende der 90er-Jahre wurde die Haltbarkeit von Daten auf WORM-Platten auf ca. 300 Jahre geschätzt.

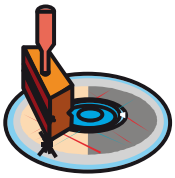


120 mm Optical Disk im IT-Umfeld für Langzeitarchivierung (WORM, MO und CCW)

MOD (Magneto Optical Disc)

Mit WORM verwandt ist die MOD, ein modernes Speichermedium für Computer auf magnetisch-optischer Basis. Sie erlaubt das digitale Aufnehmen und Löschen über Laser. Die MOD enthält in Aufzeichnungsspuren geordnet bis zu einer Milliarde Mikromagnete. Diese sind in einer bestimmten Richtung vormagnetisiert. An den mit einem Laserstrahl erwärmten Stellen lassen sich diese Mikromagnete durch ein angelegtes Magnetfeld umpolen. Beim Abtasten wird der Laserstrahl durch die nun verschieden gepolten Magnete zirkular entweder rechts- oder linksdrehend zurückgeworfen (magneto-optischer Kerr-Effekt). Die Änderung der Polarisation kann über eine entsprechende Optik gelesen werden. Die MOD ist weniger lichtempfindlich als die CD-R, kann schneller beschrieben werden, ist unempfindlich gegen Nässe, hat eine hohe Lebenserwartung und bietet hohe Datensicherheit.

Optische Technologien im IT-Umfeld für Juke-Boxen



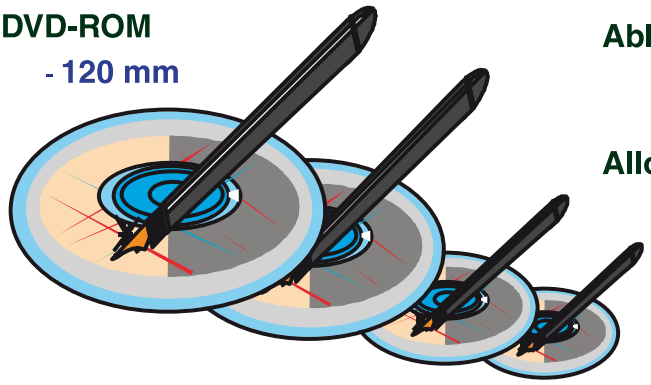
Read Only Media (ROM)
- CR-ROM
  - 120 mm
- DVD-ROM
  - 120 mm

Permanent WORM

Phase Change
- 14 inch

Ablative
- 12 inch
- 5.25 inch

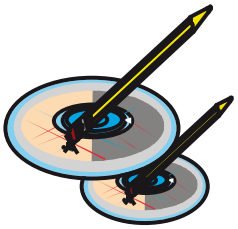
Alloy Melt
- 5.25 inch (8X)



CCW WORM
- Magneto Optic
  - 5.25 inch

Magneto Optisch

Re-writable (Erasable)
- MO (Magneto Optic)
  - 5.25 inch
  - 3.5 inch
- CD-RW
  - 120 mm
- DVD
  - 120 mm



Recordable
- CD-R
  - Phase Change
  - 120 mm
- DVD
  - 120 mm

CCW (Continuous Composite WORM) basiert auf der MO-Technik. Die Platte erhält bei der Produktion eine Kennzeichnung, damit sie nicht überschrieben werden kann. Zusätzlich schützt eine Software im Laufwerk die Platte vor dem Überschreiben. CCW ist also eine 'Schein'-WORM-Platte, die auf der MOD-Technologie basiert.

Im IT-Umfeld wurden Anfang der 90er-Jahre die Technologien auf 120-mm-Basis (5.25 Zoll) als MO-, CCW- und WORM-Platten standardisiert. Die 120-mm-Technologie findet bis heute ihren Einsatz im IT-Umfeld für Archivierungszwecke. Optische Platten auf 90-mm-Basis (3.25 Zoll) werden vorrangig im Consumer Umfeld eingesetzt.

UDO (Ultra Density Optical Disk)

In den Jahren 2003 und 2004 wurde die 120-mm-Technologie, die bis dahin mit einem roten Laser arbeitete, durch einen neuen Standard ersetzt, die UDO-Technologie (Ultra Density Optical). Die UDO-Technologie ist nicht rückwärtskompatibel zur alten 120-mm-Technologie und arbeitet mit einem blau-violetten Laser mit 405-nm-Wellenlänge. Die Plattenformate WORM, MO und CCW wurden beibehalten. UDO bietet ein 'Phase Change' WORM und 'Phase Change' Rewritable-Medium zum Einsatz an. Die ersten Platten auf dieser neuen Laser-Basis als neuer 1-x-Standard boten eine Kapazität von 30 GB. Die Zweitgeneration als 2-x-Standard, der 2006 verfügbar wurde, bot das Doppelte an Kapazität, also 60 GB. Die UDO-Technologie als neuer 120-mm-Standard wird gemäß einer Roadmap weiterentwickelt, die im UDO-Standard verankert wurde.

UDO Roadmap

	Generation 1	Generation 2	Generation 3
Capacity	30 GB	60 GB	120 GB
Transfer Rate	8 MB/s	12 MB/s	18 MB/s
RPM	2000	3000	3600
Avg Seek Time	25 ms	25 ms	25 ms
Numerical Aperture	0.7	0.7	0.85
Media Layers	1	2	2
Encoding	1,7	1,7	ML
Sector Size	8 KB	8 KB	8 KB
SCSI Transfer Rate	80 MB/s	80 MB/s	80 MB/s
Load Time	5 s	5 s	5 s
Unload Time	3 s	3 s	3 s
MSBF	750000	750000	750000

DVD (Digital Versatile Disk)

DVD steht für 'Digital Versatile Disk' (ehemals 'Digital Video Disk'). Das Medium ist so groß wie eine normale CD-ROM, jedoch wird mit einer wesentlich höheren Speicherdichte gearbeitet. Dabei unterscheidet man vier verschiedene Medien. Die einfache einseitige DVD kann 4.7 GB auf einer Schicht speichern. Es gibt aber auch zweischichtige DVDs. Dabei wird die Information auf zwei übereinanderliegenden Schichten gespeichert, eine davon ist durchsichtig.

Durch unterschiedliche Fokussierung des Lasers wird die richtige Schicht angesteuert. Damit sind 8.5 GB möglich. Und dann gibt es das Ganze noch zweiseitig. Damit sind 17 GB Daten auf einer einzigen DVD möglich. Die ersten Laufwerke kamen in den 90er-Jahren auf den Markt und können einschichtige, einseitige DVDs lesen. Leider gab es zu diesem Zeitpunkt noch wenig DVD-Titel mit Videos. Die Videos wurden in MPEG-2 kodiert, was eine sehr gute Qualität bei der Wiedergabe ergibt. Auch die ersten Brenngeräte für einseitige, einschichtige DVDs waren in den 90er-Jahren bereits vorgestellt worden, der Brenner von Pioneer wurde im Herbst 1997 für etwas über 10000 DM angeboten. Aufgenommen wurde mit ca. 1–2 MB/s, und speichern konnte er maximal 3.9 GB.

Die Lesegeräte können auch normale CDs lesen, jedoch meist keine CD-Rs, also die beschreibbaren CDs. Dies kommt daher, dass ein Laser mit einer kürzeren Wellenlänge verwendet wird, der die selbstgebrannten CDs nicht mehr richtig lesen kann. Sony hatte dazu ein Laufwerk mit zwei Laser-Dioden entwickelt, mit dem man dann auch die CD-Rs wieder lesen kann.

HD DVD

Die HD DVD (High Density Digital Versatile Disk, zuvor: Advanced Optical Disk, kurz: AOD) ist ein Datenträgerformat und wurde von 2005 bis Februar 2008 neben Blu-ray Disc und VMD als ein mögliches Nachfolgeformat der DVD gehandelt.

Die HD DVD wurde durch das Advanced Optical Disc-Konsortium (AOD) spezifiziert, dem u. a. NEC, Microsoft, Toshiba, Intel, IBM und Hewlett Packard angehörten. Danach haben sich diese Firmen zur HD DVD Promotion Group zusammengeschlossen, um die HD DVD effizienter bekannt zu machen. Das DVD Forum hatte im November 2003 die HD DVD als HD-Nachfolger der DVD nach der HD-DVD-Spezifikationen für Read-Only-Discs verabschiedet. Für die HD DVD wurde der Kopierschutz Advanced Access Content System (AACS) aus dem Bereich des Digital Rights Management (DRM) vorgesehen.

Im März 2006 kam mit dem HD-XA1 von Toshiba der erste HD-DVD-Player in Japan auf den Markt. Im August 2006 erschien mit Elephants Dream die erste HD DVD in deutscher Sprache.

Im Februar 2008 erklärte Toshiba, dass man die Entwicklung, Herstellung und Vermarktung der Technologie Ende März 2008 einstellen werde.

Die HD DVD basierte auf einem blau-violetten Laser mit 405 nm Wellenlänge. Die Dicke der Trägerschicht ist mit 0,6 mm identisch mit der der DVD. Die numerische Apertur (NA) beträgt dagegen 0,65 im Vergleich zu 0,6 bei der DVD und 0,85 bei der Blu-ray Disc.

Die HD DVD hatte mit einer Schicht ausgestattet eine Speicherkapazität von:

- *15 GB (bei HD DVD-ROMs – gepresste Medien) bzw.*
- *15 GB (bei HD DVD-R/RWs – einmal und wiederbeschreibbare Medien) bzw.*
- *20 GB (bei HD DVD-RAM – wiederbeschreibbare Medien mit wahlfreiem Sektorzugriff)*

und bei zwei Schichten eine Speicherkapazität von

- *30 GB (bei HD DVD-ROMs – gepresste Medien).*

Zusätzlich wurde im August 2007 eine dreilagige Variante durch das DVD-Forum verabschiedet, bei der pro Schicht 17 GB Platz fanden und die somit eine Gesamtkapazität von 51 GB hatte. Der endgültige Standard für Filme auf HD DVD umfasste zunächst die 15-GB- und die 30-GB-Variante, wobei für Filme fast immer die zweischichtige 30-GB-Variante verwendet wurde.

Im Januar 2008 gab Time Warner bekannt, dass seine Studios Warner Bros. und New Line Cinema zukünftig keine weiteren Filme für die HD DVD veröffentlichen werden, sondern ausschließlich auf die Blu-ray Disc setzen. Dies wurde von einigen Medien als Entscheidung des Formatkrieges bezeichnet. In den darauffolgenden nächsten Tagen folgten weitere Anbieter, so am 10. Januar der große europäische Filmverleih Constantin Film, am 12. Januar der US-Pornoanbieter Digital Playground und die deutsche Senator Film. Die US-Anbieter Universal Studios und Paramount Pictures dementierten hingegen Gerüchte, ihre Verträge mit HD DVD aufzugeben, sie gelten als die letzten verbliebenen großen Anbieter. Nachdem Toshiba, von denen ein Großteil der HD-DVD-Player stammt, am 15. Januar die Preise der Geräte in den USA teilweise drastisch reduziert hatte, sprachen einige Medien von einem ‘Ausverkauf’. Auf der anderen Seite nahm der Druck auf die HD DVD weiter zu: Anfang 2008 nahmen Märkte der Media-Saturn-Holding beim Kauf eines Blu-ray-Abspielgerätes den HD-DVD-Player in Zahlung. Weitere Rückschläge musste das Format im Februar 2008 hinnehmen. Die US-Elektronikkette Best Buy teilte am 12. Februar 2008 mit, sich zukünftig nur noch auf das Blu-ray-Format zu konzentrieren. Einen Tag zuvor kündigte schon der größte US-Online-Videoverleih Netflix an, dass er bis zum Jahresende die HD DVD aus dem Sortiment nehmen wird. Am 15. Februar 2008 kündigte zudem Wal-Mart, größter Einzelhändler der USA, an, HD-DVD-Bestände auszuverkaufen und somit in Zukunft nur noch auf Blu-ray setzen zu wollen. Am 19. Februar 2008 gab Toshiba eine Pressemitteilung heraus, in der bekannt gegeben wurde, dass die Entwicklung, Herstellung und der Vertrieb der HD DVD sowie entsprechender Geräte nicht weiter vorangetrieben wird und Ende März 2008 endgültig eingestellt wird. Daraufhin gaben die Universal Studios noch am selben Tag bekannt, von der HD DVD auf das Blu-ray-Format zu wechseln.

Blu-ray Disc

Die Blu-ray Disc (abgekürzt BD oder seltener BRD) ist ein digitales optisches Speichermedium. Sie wurde neben HD DVD und VMD als ein möglicher Nachfolger der DVD beworben. Nachdem Toshiba im Februar 2008 die Einstellung von Produktion und Weiterentwicklung der HD-DVD-Technik – einschließlich der Geräte – für März 2008 angekündigt hatte, gilt die Blu-ray Disc als Sieger im Formatkrieg. Der Name Blu-ray ist englischen Ursprungs. Blue ray bedeutet wörtlich soviel wie 'Blauer Lichtstrahl', was sich auf den verwendeten blauen Laser bezieht. Die bewusste Abweichung von der orthografisch korrekten Schreibweise 'blue ray disc' zielt darauf ab, eine Registrierung des Ausdrucks als Marke zu begünstigen.

Die Spezifikationen für die Blu-ray Disc wurden am 19. Februar 2002 durch die neun Unternehmen der Blu-ray Group, Panasonic, Pioneer, Philips, Sony, Thomson, LG Electronics, Hitachi, Sharp und Samsung, beschlossen; dieser Gruppierung schlossen sich Ende Januar 2004 zusätzlich noch Dell und Hewlett-Packard sowie Mitte März 2005 Apple an. Hewlett-Packard trat allerdings 2005 wieder aus dem Blu-ray-Konsortium aus, nachdem einige Verbesserungsvorschläge abgewiesen worden waren, und wechselte in das HD-DVD-Lager.

Die Blu-ray Disc gibt es in drei Varianten: als nur lesbare BD-ROM (vergleichbar mit DVD-ROM), als einmal beschreibbare Variante BD-R (vergleichbar mit DVD±R) und als wiederbeschreibbare BD-RE (vergleichbar mit DVD±RW). Durch den Einsatz eines blau-violetten Lasers und dessen kurzer Wellenlänge (405 Nanometer) lassen sich höhere Datendichten erreichen als bei einer DVD. Pro Datenschicht lassen sich 25 GB Daten auf einer BD ROM speichern, so dass sich bei zweischichtigen Medien eine Kapazität von bis zu 50 GB ergibt.

Eine vierlagige Version der Blu-ray Disc, die auf einer Seite um 100 GB fassen soll, wurde von TDK vorgestellt. TDK ist es gelungen, auf einer sechslagigen Scheibe 200 GB unterzubringen. Dabei wurde die Kapazität einer Lage auf 33,3 GB erhöht. Dies sind jedoch in erster Linie Machbarkeitsstudien. Von vornherein sind auch beschreibbare Medien vorgesehen, soll die Blu-ray Disc doch vor allem in der Unterhaltungselektronik ihre Heimat finden und als Speichermedium für hochauflösende Filme dienen. Die wiederbeschreibbare Blu-ray Disc arbeitet mit der Phase-Change-Technik.

Die neue Phase-Change-Technik soll eine doppelt so hohe Übertragungsrate (Datentransferrate) von 9.0 MB/s (72 Mbit/s) beim Beschreiben anstatt der ursprünglich maximal spezifizierten einfachen 4.5 MB/s (36 Mbit/s) ermöglichen. Ein wichtiger Bestandteil der Spezifikation ist auch ein Kopierschutz in Form einer eindeutigen Identifikationsnummer. Außerdem eignen sich Blu-ray Discs besonders gut für Full-HD-Videos, die dank der hohen Auflösung eine bessere Qualität als die gängigen Systeme wie PAL und NTSC bieten, aber auch dementsprechend mehr Speicherplatz benötigen. BD-Spieler unterstützen in der Regel das zuerst von Panasonic und Sony eingeführte hochauflösende AVCHD-Format.

Eine weitere Neuerung gegenüber der DVD ist der verkleinerte Abstand des Lasers zum Datenträger sowie die geringere Wellenlänge (und daher andere Farbe) des Laserstrahls. Weiterhin ist die Schutzschicht auf der Laser-Seite mit 0,1 mm im Vergleich zu 0,6 mm der DVD deutlich kleiner. Aufgrund der daraus resultierenden größeren Anfälligkeit gegen Kratzer und Staub war zunächst geplant, die Blu-ray Disc nur in einer Cartridge anzubieten. Stattdessen wurde jedoch eine Beschichtung namens 'Durabis' entwickelt, die den Gebrauch einer Cartridge nicht mehr notwendig macht.

Wegen des geringeren Abstands zwischen Medium und Laseroptik sowie der dünneren Schutzschicht kann ein Objektiv mit günstigerer numerischer Apertur eingesetzt werden, das den Strahl effizienter bündeln kann. Somit werden Schreibfehler und stärkere Streuungen verringert und es ist möglich, eine Blu-ray Disc zum Beispiel aus Metall oder anderen stabilen, undurchsichtigen Materialien kombiniert

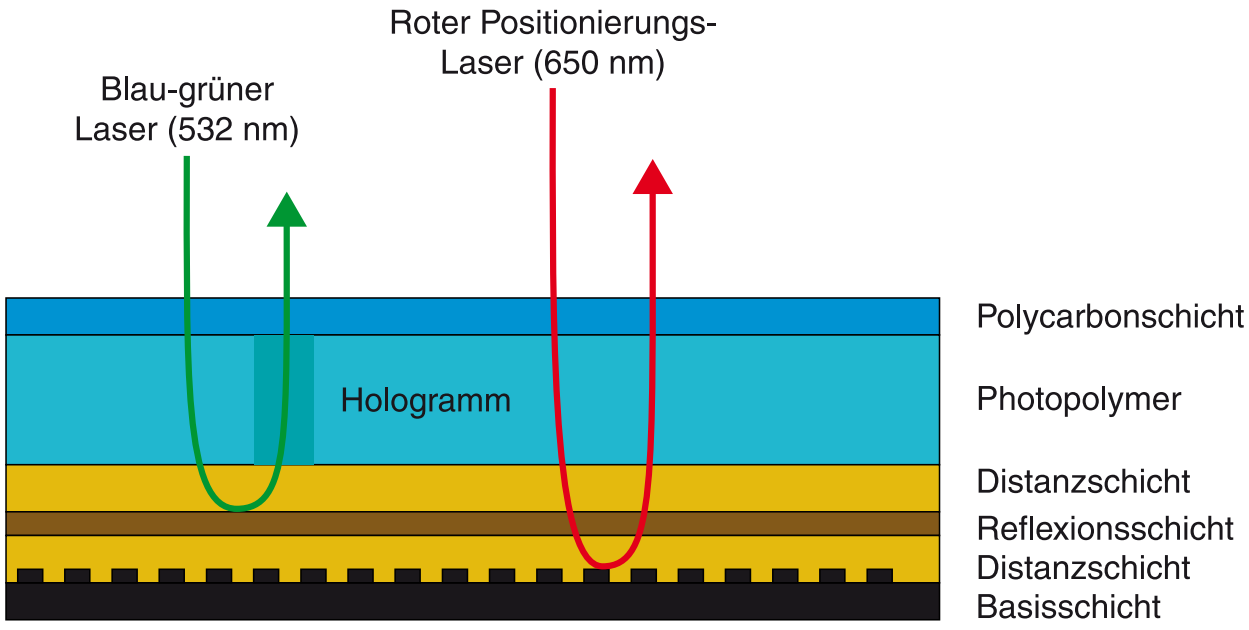
mit einer dünnen durchsichtigen Trägerschicht zu bauen, die mit erheblich höheren Drehzahlen als eine Scheibe aus Polycarbonat betrieben werden können, woraus dann höhere Übertragungsraten resultieren. Zudem erlaubt die gegenüber der DVD kleinere Wellenlänge des Laserstrahls eine wesentlich höhere Datendichte und damit eine erhöhte Speicherkapazität.

Die Firma Verbatim stellte auf der IFA 2005 eine 'Blu-ray Disc Rewriteable' vor. Dieses Medium soll laut Hersteller eine Speicherkapazität von 25 Gigabyte erlauben (also ca. 135 Minuten Video in MPEG2 HD-Qualität). Außerdem soll die BD-RE mit einer besonders stoß- und kratzfesten Schutzschicht versehen sein. Neben der unmittelbaren Verfügbarkeit von 25-GB-Rohlingen hat TDK für Oktober 2006 auch 50-GB-Discs in Aussicht gestellt. TDK ist eines der Gründungsmitglieder der Blu-ray Disc Association (BDA) und war maßgeblich an der Entwicklung beteiligt.

Holographic Versatile Disc (HVD)

Die Holographic Versatile Disc, auch HVD genannt, wird bereits heute in Fachkreisen als Nachfolger der heutigen HD DVD- und Blu-ray-Technologie gehandelt. HVDs haben eine Kapazität von bis zu 3.9 TB und damit um ein Vielfaches mehr als die größte heute zur Verfügung stehende Blu-ray-Platte mit 200 GB. Auch die Transferrate ist um ein Vielfaches höher und erreicht 1 Gbit/s im Vergleich zu 36 bzw. 72 Mbit/s bei der Blu-ray-Platte. Zudem ist die HVD dabei noch lange nicht ausgereizt und es sind neue Laufwerke mit noch höherer Rotationsgeschwindigkeit denkbar. Die Spezifikation für die HVD wurde im Dezember 2004 durch das TC44 Committee der Ecma International beschlossen. Bereits im Februar 2005 wurde die 'HVD Alliance' gegründet, um die Entwicklung der HVD voranzutreiben. Inzwischen gehören der Alliance 19 Firmen an: Alps Electric Corporation, CMC Magnetics, Dainippon Ink and Chemicals, EMTEC International, Fuji Photo Film, Konica Minolta Holdings, LiteOn Technology, Mitsubishi Kagaku, Nippon Kayaku, Nippon Paint, Optware Corporation, Pulstec Industrial, Shibaura Mechatronics, Software Architects, Suruga Seiki, Targray Technology, Teijin Chemicals, Toagosei Company und die Tokiwa Optical Corporation.

Zwei überlagerte Laser mit unterschiedlicher Wellenlänge, ein blau-grüner Schreib-/Leselaser mit 532 nm und ein roter Positionierungslaser mit 650 nm, werden auf zwei unterschiedlichen Lagen in der Platte reflektiert und erzeugen in der photopolymeren Schicht ein Interferenzfeld, ein sogenanntes Hologramm. In der schematischen Zeichnung sind die beiden Laser zur Vereinfachung nicht überlagert dargestellt, um das Prinzip der unterschiedlichen Reflektion besser darzustellen.



Prinzip der HVD (Holographic Versatile Disc)

Beim Auslesen liest der blau-grüne Laser das Hologramm der als Laser-Interferenz-Muster codierten Daten in der Polymerschicht im oberen Bereich des Mediums aus, während der rote Laser dazu dient, Hilfsinformationen von einer CD-vergleichbaren Aluminiumschicht im unteren Bereich auszulesen. Die Hilfsinformationen dienen zur exakten Positionierung, wo man sich gerade auf der Platte befindet (vergleichbar mit Sektor-, Kopf- und Segmentinformationen einer Festplatte). Eine dichroitische Spiegelschicht (Reflexionsschicht), die sich zwischen der Polymerschicht und der Aluminiumschicht (Basisschicht) befindet, lässt den blau-grünen Laser reflektieren, während der rote Laser hindurch geht. Durch diesen Trick wird die Interferenz durch Refraktion des blau-grünen Lasers von den Hilfsdaten-'Pits' verhindert und beide Informationen (Hologramm und Hilfsdaten) können unabhängig voneinander sauber angesteuert werden.

Polymere, also Kunststoffe, sind lange Kettenmoleküle, die aus den immer gleichen Bausteinen bestehen. Sie haben gegenüber Kristallen den Vorteil, nahezu unbegrenzt modifizierbar zu sein. Polymere sind extrem lichtempfindlich, hochtransparent und unempfindlich gegen Temperaturschwankungen. Sie verändern auch nach häufigem Auslesen ihre Leistungsfähigkeit nicht und sind ideal für die Laserbearbeitung. Nach der Bestrahlung mit Laserlicht verändern die lichtempfindlichen Moleküle im Polymer ihre Ausrichtung. An den belichteten Stellen lenkt das Material Licht stärker ab als an den unbelichteten Stellen. In dieser als Interferenzfeld gezielten Veränderung der molekularen Ordnung steckt die 'holografische' Information. Da die Daten nicht als einzelne Bits sondern als ganze Bitmuster aufgezeichnet und gelesen werden, lassen sich mit einem einzigen Laser-'Blitz' extrem viele Bits gleichzeitig abrufen. Je nach Polymerart und Polymerdicke können das mehrere hunderttausend Bits sein. Die Übertragungsraten werden gigantisch. Ein Spielfilm, der heute auf eine DVD passt, könnte in etwa 30 Sekunden ausgelesen werden. Die große Speicherkapazität kommt daher, dass die Hologramme nicht nur auf die Oberfläche eines Speichermaterials geschrieben werden, sondern in das gesamte Volumen.

Besonders sensible Daten könnten zusätzlich durch eine spezielle Maske verschlüsselt werden, die zwischen das Speichermaterial gesetzt wird. Im Gegensatz zu Informationen (z. B. auf Chips), die über eine Software verschlüsselt sind, ließe sich die Hardware-Verschlüsselung von Hologrammen prinzipiell nicht knacken. Aus denselben Gründen ist ein Hologramm auch schwer zu fälschen. Die Haltbarkeit von Informationen in holografischen Speichern wird heute auf etwa 50 Jahre geschätzt.

Millipede-Nano-Technologie

Massiv arbeitet die IBM an der Einführung der neuen **Millipede**-Nanospeicherarchitektur für Subsysteme. Es ist davon auszugehen, dass die Millipede-Technik im langfristigen Bereich die heutigen Magnetplatten ersetzen wird, zumal Millipede-Substrate parallel adressierbar sind und durch die Erzeugung von echten Pits in einer Polymerschicht eine wesentlich höhere Datensicherheit abbilden. Mit Millipede-Systemen kann ab 2009 gerechnet werden.

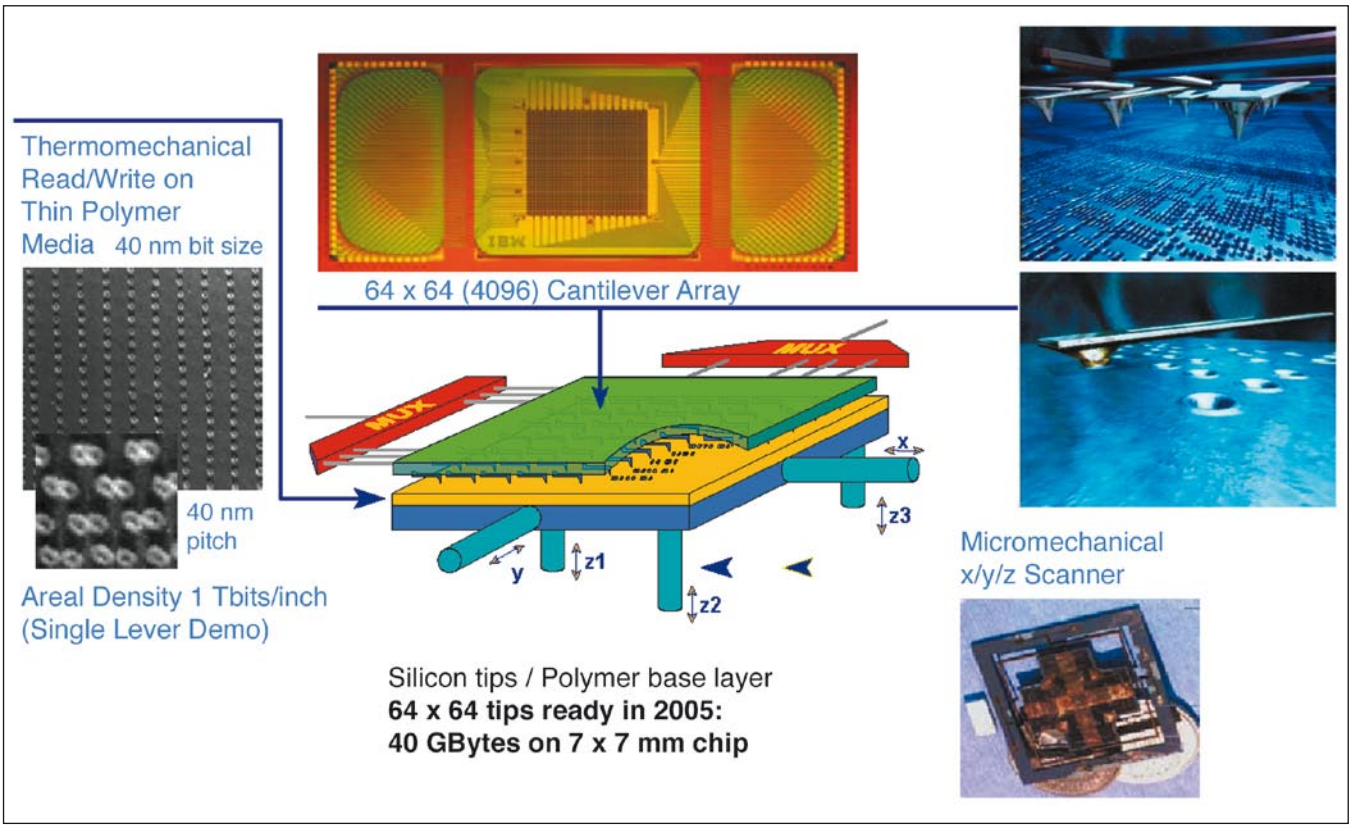
Die extreme Dichte von einem **Terabit pro Quadratzoll** – genug, um 25 Millionen Buchseiten oder den Inhalt von 25 DVDs auf der Fläche einer Briefmarke zu speichern – wurde im Rahmen des Projektes 'Millipede' (Tausendfüßler) am IBM Forschungslabor Zürich erzielt. In dieser Entwicklung spielt Nanomechanik anstelle von herkömmlicher magnetischer oder elektronischer Speichertechnologie die zentrale Rolle: Mit Tausenden von feinsten Spitzen 'schreibt' Millipede winzige Vertiefungen, die einzelne Bits repräsentieren, in einen dünnen Film aus Kunststoff. Das Resultat ist mit einer althergebrachten Lochkarte vergleichbar, allerdings auf der Skala

von Nanometern (Millionstelmillimetern), und die Bits lassen sich auch löschen und überschreiben. „Da Nanometer-scharfe Spitzen einzelne Atome adressieren können, sind Verbesserungen weit über den Terabit-Meilenstein hinaus möglich“, sagt **Nobelpreisträger Gerd Binnig**, eine treibende Kraft hinter dem Millipede-Projekt. „Während die heute eingesetzten Speichertechnologien allmählich an fundamentale Grenzen stoßen, steht unser **nanomechanischer Ansatz erst am Anfang** und hat ein Entwicklungspotenzial für tausendfach höhere Speicherdichte.“ Die Funktionstüchtigkeit des Konzepts hat das Team im IBM Forschungslabor Zürich mit einem experimentellen Speicherchip mit mehr als tausend Spitzen erprobt. Derzeit arbeiten sie am Prototyp eines kompletten Speichersystems, das im nächsten Jahr betriebsbereit sein und nachweisen soll, dass die vielversprechende neue Technologie die praktischen Anforderungen an ein marktfähiges Produkt erfüllen kann. Dieser Prototyp wird über mehr als 4000 Spitzen verfügen, die in einem kleinen Quadrat von 7 Millimetern Seitenlänge angeordnet sind. Die Dimensionen würden es ermöglichen, ein komplettes Speichersystem hoher Kapazität in das kleinste standardisierte Format für Flash-Memory zu packen. **Millipede** könnte Flash Memories, mobilen Geräten wie PDAs, Mobiltelefonen und multifunktionalen Armbanduhren enorme Speicherkapazität verleihen“, sagt Millipede-Projektleiter Peter Vettiger. „Wir denken aber auch bereits an die Nutzung unseres Konzeptes über die Datenspeicherung hinaus. Andere mögliche Anwendungen sind beispielsweise Lithografie im Nanometerbereich, mikroskopische Abbildungen von relativ großen Bereichen oder atomare und molekulare Manipulation.“

Der Speicherhunger mobiler Geräte nimmt beständig zu. Flash-Karten haben den Nachteil, dass bei ihnen die Kosten pro Speichereinheit relativ konstant sind – jedenfalls bis zum Erreichen einer bestimmten Grenze. Festplatten und ihre kleineren Verwandten – Micro-Drives, die vereinzelt bereits in Compact-Flash-Gehäusen verbaut werden – , sind andererseits vergleichsweise massiv, schwer und empfindlich.

Grundprinzip

Das Grundprinzip ist simpel und ist mit dem der früheren Lochkarte vergleichbar, nun aber mit Strukturgrößen im Bereich von Nanometern. Ein weiterer entscheidender Unterschied: mithilfe der eingesetzten Technologie lassen sich Bits löschen und überschreiben. Winzige Hebelchen mit einer feinen Spitze aus Silizium schmelzen ebenso winzige Vertiefungen in ein Polymer-Medium, um Bits zu schreiben. Dieselben Spitzen kann man auch verwenden, um diese Vertiefungen nachzuweisen, also die Bits wieder auszulesen. Dazu bringt man die Spitze in die Nähe des Polymerfilms und erwärmt



sie. Taucht die Spitze in einen Bit-Krater, erhöht sich der Wärmeaustausch zwischen ihr und dem Speichermedium, wodurch der elektrische Widerstand des Hebelchens, auch Kantilever genannt, abnimmt. Um ein Bit wieder zu überschreiben, erzeugt man mit der Spitze auf dem Kraterrand neue Vertiefungen, deren Ränder die alte Vertiefung überlappen und so das Polymer-Material in Richtung Krater drängen.

Weil die Löcher so enorm klein sind, kann man sie sehr dicht nebeneinander setzen und so fantastische Datendichten erreichen: Mit revolutionärer Nanotechnologie ist es IBM Wissenschaftlern im Forschungslabor Rüschlikon gelungen, in den Bereich der Millionstelmillimeter vorzudringen. So konnte bei der Speicherung von Daten eine Aufzeichnungsdichte von 1 Terabit pro Quadratzoll überschritten werden – was etwa dem Inhalt von 25 DVDs auf der Fläche einer Briefmarke entspricht. Die Terabit-Dichte wurde mit einer einzelnen Silizium-Spitze erreicht, die Vertiefungen mit einem Durchmesser von circa 10 Nanometern erzeugt. Um die Datenrate, also die Schreib- und Lesegeschwindigkeit, zu erhöhen, wird nicht nur eine Spitze verwendet, sondern eine ganze Matrix von Hebelchen, die parallel arbeiten. Die neue Version verfügt über mehr als 4000 solcher Spitzen, die in einem kleinen Quadrat von 6,4 Millimetern Seitenlänge angeordnet sind. Diese Dimensionen ermöglichen es, ein komplettes Speichersystem hoher Kapazität in das kleinste standardisierte Format für Flash-Speicher zu packen.

Ein Kantilever im **Millipede** schreibt und liest aus einer ihm zugeordneten rund 100 mal 100 Mikrometer kleinen Zelle. Während sich beispielsweise bei Festplatten der Schreib- und Lesekopf und auch das Speichermedium bewegen, wird beim Millipede-Speicher nur das Medium bewegt. Zwei winzige Spulen, die zwischen Magneten platziert sind, treiben die Bewegung des Plättchens an: Der Mikroscanner kann mit einer Genauigkeit von bis zu zwei Nanometern positioniert werden. Aus der überlappenden Fläche von streifenförmigen Sensoren kann man die Position jederzeit präzise bestimmen.

Kantilever-Array

Kern der Millipede-Technologie ist eine zweidimensionale Anordnung von v-förmigen Silizium-Federzungen (Kantilever), die 70 Mikrometer (Tausendstelmillimeter) lang sind. Am Ende jeder ‘Zunge‘ befinden sich ein Sensor zum Lesen und ein Widerstand oberhalb der Spitze zum Schreiben. Die Spitze ist knapp 1 Mikrometer lang und der Radius beträgt nur wenige Nanometer. Die Zellen mit den Kantilevern sind auf einem Kantilever-Array-Chip angeordnet. Der Chip ist 7 x 14 mm groß. Im Zentrum sitzt das Array, das momentan aus insgesamt 4096 (64 x 64) Kantilevern besteht, die aus dem Silizium herausgeätzt sind. Der eigentliche Datenträger besteht aus einem nur wenige Nanometer dünnen Polymerfilm auf einem Silizium-Substrat. Über Multiplexer einzeln angesteuert lesen, schreiben oder löschen die Köpfe das gewünschte Bit. Bis zu 100 000 Schreib- und Überschreib-Zyklen sind bisher erfolgreich getestet worden. Obwohl in der Konstruktion Mechanik eingesetzt wird, kann die Übertragungsgeschwindigkeit von bis zu 20 – 30 Megabit pro Sekunde erreicht werden (beim Chip der 2. Generation mit 64 x 64 Tips).

Mikroscanner

Die Bewegung des Speichermediums relativ zu dem Kantilever-Array wird mithilfe des siliziumbasierten x/y-Mikroscanners realisiert. Der Scanner besteht aus ca. 6.8 x 6.8 mm^2 Scan Table, der das Polymer-Medium und zwei elektromagnetische Auslöser trägt. Der Scanner-Chip wird auf der Siliziumplatte montiert, die als mechanischer Grund des Systems dient. Der Abstand zwischen seiner Oberfläche und der Oberfläche von beweglichen Teilen des Scanners beträgt ca. 20 µm. Die Scan Table kann durch Auslöser in x-und y-Richtungen um 120 µm bewegt werden. Jeder Auslöser besteht aus zwei Permanentmagneten, die in die Siliziumplatte eingebaut sind, und einer kleinen Spule, die sich zwischen den Magneten befindet. Um die Vibrationen von außen zu löschen, wird ein sogenanntes Pivot verwendet, das mit den Auslösern gekoppelt ist.

Positionsbestimmung

Die Informationen über die Positionierung werden durch vier thermische Sensoren zur Verfügung gestellt. Diese Sensoren befinden sich direkt über der Scan Table, auf dem Kantilever-Array. Die Sensoren haben thermischisolierte Erhitzer. Jeder Sensor wird über einen Rand der Scan-Tabelle in Position gebracht und wird durch Strom erhitzt. Ein Teil dieser Wärme wird durch die Luft auf die Scan Table geleitet, die nun als Kühler dient. Eine Versetzung der Scan Table verursacht eine Änderung der Effizienz dieses Kühlsystems, was zur Änderung der Temperatur des elektrischen Widerstands vom Erhitzer führt.

Ein raffiniertes Design gewährleistet die exakte Nivellierung der Spitzen über dem Speichermedium und dämpft Vibrationen und Stöße von außen. Zeitmultiplexing-Elektronik, wie sie in ähnlicher Art in Speicherchips (DRAM) verwendet wird, ermöglicht die Adressierung jeder einzelnen Spitze im Parallelbetrieb. Elektromagnetische Aktuation bewegt das Substrat mit dem Speichermedium auf dessen Oberfläche sehr präzise in x- und y-Richtung, sodass jede Spitze in ihrem Speicherfeld von 100 Mikrometern Seitenlänge lesen und schreiben kann. Die kurzen Distanzen tragen wesentlich zu einem geringen Energieverbrauch bei.

Für die Funktionen des Gerätes, das heißt Lesen, Schreiben, Löschen und Überschreiben, werden die Spitzen mit dem Polymerfilm auf dem Siliziumsubstrat in Kontakt gebracht.

Schreibtechnologie

Das Schreiben von Bits erfolgt durch Aufheizen des in den Kantilever integrierten Widerstands auf typischerweise 400 Grad Celsius. Die dadurch ebenfalls aufgeheizte Spitze weicht das Polymer auf, sinkt ein und hinterlässt eine Vertiefung von wenigen Nanometern. Zum Lesen wird der Lese-Sensor des Kantilevers erhitzt, ohne den Polymerfilm aufzuweichen. 'Fällt' nun die Spitze in eine Vertiefung, kühlt sich der Lese-Sensor wegen des geringeren Abstands zum Substrat geringfügig ab, was aber zu einer messbaren Veränderung des Widerstands führt. Um Daten zu überschreiben, setzt die Spitze Vertiefungen in die Oberfläche. Deren äußere Ränder überlappen die alten Vertiefungen und löschen so die alten Daten. Mehr als 100000 Schreib- und Überschreib-Zyklen haben den Nachweis erbracht, dass sich das Konzept für einen wiederbeschreibbaren Speichertyp eignet.

Um die Daten schneller in den Speicher und wieder hinauszubekommen, bearbeitet eine komplette Matrix-Anordnung an Hebelchen das Medium gleichzeitig. Es stellte sich allerdings heraus, dass es enorm schwierig ist, Mechanik und Elektronik in einem Stück auf einem Chip zu fertigen. Die Wissenschaftler entschlossen sich also, den Aufbau in zwei Stücken zu realisieren:

1. *Die Matrix der Mikrospitzen wird mit winzigen Kontaktstiften versehen, die unter dem Elektronenmikroskop aussehen wie die Noppen von Lego-Steinchen.*
2. *Diese Studs werden dann mit den Gegenstücken auf der elektronischen Platine kontaktiert.*

Mit den gezeigten Prototypen wurde unlängst die technische Machbarkeit eines Produktes etwa hinsichtlich Speicherdichte, Leistung und Verlässlichkeit demonstriert. Während die heute eingesetzten Speichertechnologien allmählich an fundamentale Grenzen stoßen, hat der nanomechanische Ansatz ein enormes Entwicklungspotenzial für eine tausendfach höhere Speicherdichte. Dieser nanomechanische Datenträger entwickelt nahezu keine Wärme, nimmt nur wenig Strom auf und ist völlig schockresistent.

In der jetzigen Epoche 2006 – 2008 kann in den Folgejahren mit den ersten Millipede Chips in neueren Generationen gerechnet werden. Der Chip der dritten Generation ist in der Fertigstellung und weist eine Speicherkapazität von 160 Gbyte bei einer durchschnittlichen Datenrate von 120 Mbit/s aus.

Revolutionäre Speichertechnologie aus der IBM Forschung rückt näher an die Realisierung

'Racetrack-Memory' eröffnet neue Dimensionen in Speicherdichte, Geschwindigkeit und Zuverlässigkeit

11. April 2008: IBM Forscher berichten im renommierten Wissenschaftsjournal Science über einen Durchbruch in der Erforschung einer neuen Speichertechnologie, die unter dem Namen 'Racetrack' (deutsch: Rennstrecke) bekannt geworden ist. Die Technologie könnte eine neue Klasse von Speichern hervorbringen, die die hohe Leistungsfähigkeit von Flash-Speichern mit der großen Kapazität und den niedrigen Kosten von Festplattenspeichern auf sich vereint.

In der aktuellen Ausgabe von Science beschreiben IBM Fellow Stuart Parkin und seine Kollegen vom IBM Almaden Research Center, USA, einen bedeutenden Schritt in der Entwicklung von Speicherbausteinen auf Basis des 'Racetrack-Verfahrens'. Innerhalb der nächsten 10 Jahre könnte die Technologie zum Einsatz kommen und neue Größenordnungen von Speicherkapazitäten eröffnen, ein extrem schnelles Einschalten von elektronischen Geräten ermöglichen und dabei wesentlich kostengünstiger, energieeffizienter und robuster als heutige Systeme sein.

Stuart Parkin war es zu verdanken, dass der GMR-Effekt (Plattentechnologie, siehe unter GMR) zu einem Produkt in Form eines GMR-Lesekopfes umgesetzt wurde und er benötigte sieben Jahre dazu. Ihm und seinem Team ist es zu verdanken, dass die hohen Festplattenkapazitäten heute möglich wurden. Der Autor findet es nicht richtig, dass Stuart Parkin bei der Vergabe des Nobelpreis für Physik im Jahr 2007 unberücksichtigt blieb, zumal der Nobelpreis an drei Personen verliehen hätte werden können. Mit Racetrack-Memories wird er wirkliche Geschichte im Speicherbereich schreiben! Davon ist der Autor überzeugt!

Universalspeicher der Zukunft

Das revolutionäre Konzept beruht auf der Speicherung von Informationen in Form von winzigen, gegensätzlich magnetisierten Regionen (Domänen) in einem Nanodraht. Bei der herkömmlichen als Speichermedium verwendeten Festplatte werden das Medium und ein Schreib-/Lesekopf bewegt, um Daten zu lesen, zu schreiben oder zu löschen. Anders beim 'Racetrack-Verfahren': Hier werden die magnetischen Domänen zu den zentralen Lese- und Schreibeinheiten, die in der Mitte des Nanodrahtes angebracht sind, hin verschoben – und dies mit extrem hoher Geschwindigkeit. Die gespeicherten Datenbits scheinen durch den Datenleiter zu 'rasen', daher der Name 'Racetrack'.

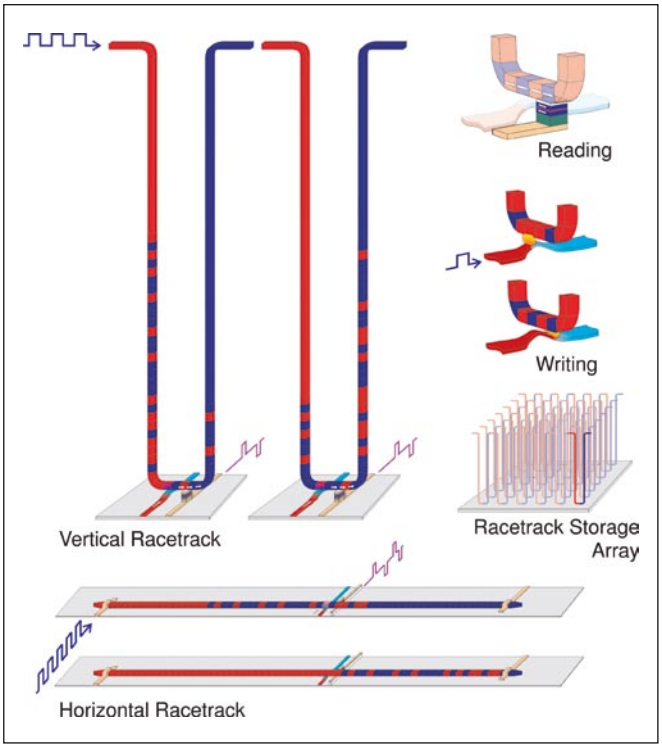
Stuart Parkin, IBM Fellow und einer der größten Speicher-Forscher und Entwickler bei IBM

Da ein einzelner ‘Racetrack’ nur wenige Nanometer groß ist und zwischen 10 und 100 Bits speichern kann, erlaubt die Technologie extrem hohe Speicherdichten. Im Vergleich zu Flash-Speichern könnte ein ‘Racetrack-Speicher’ eine 100-mal größere Datenmenge auf derselben Fläche aufzeichnen. Das entspräche rund 500 000 Musiktiteln oder 3 500 Filmen. Durch den minimalen Stromverbrauch eines ‘Racetrack-Speichers’ könnte ein MP3-Gerät zudem wochenlang mit einer einzigen Batterie betrieben werden.

Darüber hinaus benötigt das Verfahren keine beweglichen Teile, wodurch nahezu keine Abnutzungs- oder Verschleißerscheinungen auftreten. Das macht den ‘Racetrack-Speicher’ widerstandsfähiger als alle existierenden Speichertechnologien und verleiht ihm eine quasi unbegrenzte Lebensdauer.

„Die Verbindung zwischen der Erforschung von neuen physikalischen Phänomenen, die auf der molekularen und atomaren Ebene auftreten, und dem Nano-Engineering – der Fähigkeit, gezielt Materialstrukturen aus einzelnen Atomen und Molekülen zu erschaffen – birgt immer neue, spannende Herausforderungen“, erklärt Stuart Parkin. Er führt weiter aus: „Ein Ausschöpfen des Potenzials, das in unserer ‘Racetrack-Technologie’ liegt, könnte zu faszinierenden neuen Anwendungen führen – Anwendungen, die heute noch unvorstellbar sind.“

Bereits mehrmals in der Geschichte der IBM wurden durch radikale technische Innovationen, die in der Grundlagenforschung ihren Anfang nahmen, neue Märkte und Anwendungsfelder erschlossen. Hierzu zählen etwa der Memory Chip, die relationale Datenbank oder die Festplatte.



Schematische Darstellung der Racetrack Memory-Speicherung

Die ‘Racetrack-Methode’

Seit rund 50 Jahren versuchen Wissenschaftler digitale Informationen durch sogenannte magnetische Domänenwände – die Grenzflächen zwischen zwei gegensätzlich magnetisierten Regionen in einem magnetischen Material – zu speichern. Bis jetzt war die Manipulation von Domänenwänden nur durch kostspielige, komplexe und energieintensive Verfahren möglich.

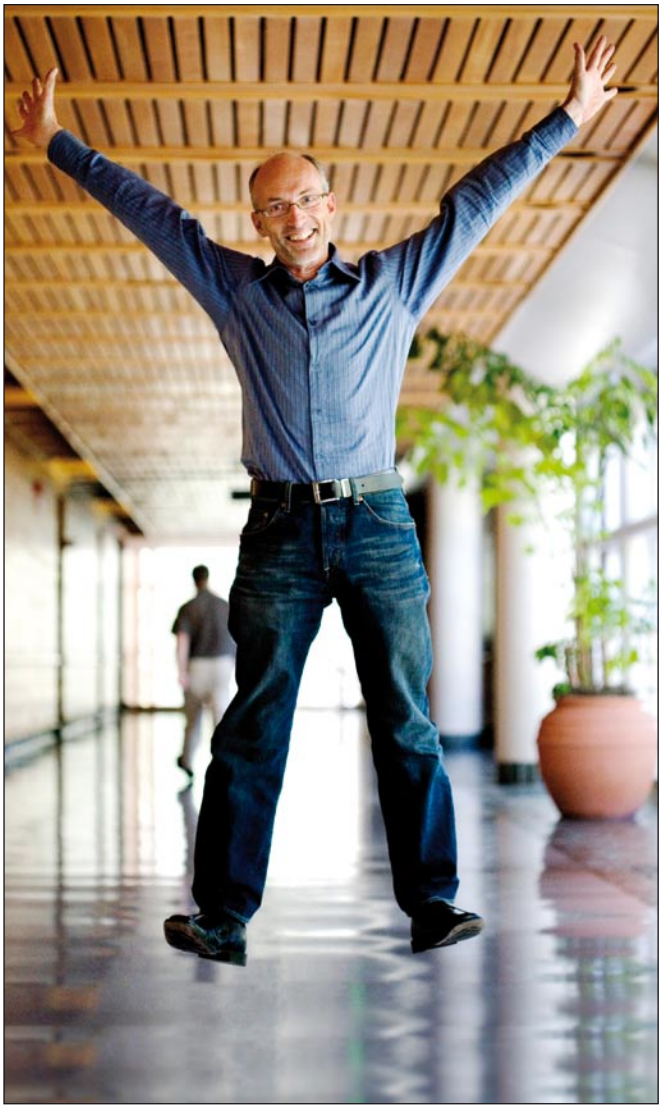
In ihrer nun veröffentlichten Arbeit ‘Current-Controlled Magnetic Domain-Wall Nanowire Shift Register’ demonstrieren Parkin und seine Teamkollegen eine neue, sehr effiziente Methode. Hierbei werden die Domänenwände durch einen spinpolarisierten Strom bewegt, der die Magnetisierung in der Wand drehen kann. Ursache hierfür ist der Spintransfer-Effekt, der die Speichertechnologie erheblich vereinfacht, weil der Strom direkt durch die Domänenwände fließt und keine zusätzlichen Magnetfelder generiert werden müssen.

In der zweiten Arbeit ‘Magnetic Domain-Wall Racetrack Memory’ fassen die Forscher die Grundprinzipien der ‘Racetrack-Technologie’ zusammen. Diese beruht auf der Speicherung von Daten in winzigen magnetischen Regionen innerhalb eines Datenleiters. Als Datenleiter verwenden die Forscher kleine Nanodrähte, die magnetisch mit Informationen beschrieben werden. Diese werden der Länge nach waagerecht oder in Form einer Schlaufe senkrecht auf einer Siliziumfläche angebracht. Wenn eine Spannung angelegt wird, bewegen sich dabei weder die Drähte noch das Silizium, sondern nur die magnetischen Domänenwände, in denen die Information gespeichert wird.

Jede magnetische Domäne, die ein Bit repräsentiert, verfügt über einen ‘Head’, einen magnetischen Nordpol, und einen ‘Tail’, einen magnetischen Südpol. Die Domänenwände, die sich an den Grenzflächen formen, wechseln sich so im Datenleiter zwischen ‘Head-to-Head’- und ‘Tail-to-Tail’-Konfigurationen ab. Die Abstände zwischen zwei hintereinander folgenden Domänenwänden gleichen der Bit-Länge und werden durch Markierungen entlang des Nanodrahtes, so genannte Pinningzentren, bestimmt.

In ihrem Experiment verwendeten die Forscher Nanodrähte aus Permalloy (einer FeNi-Legierung) und zeigten, wie sie Domänenwände gezielt kreieren, verschieben und auslesen, indem sie Stromimpulse von präzise kontrollierter Dauer im Nanosekundenbereich verwenden. Der Schreib- und Verschiebezyklus benötigt gerade einmal einige 10 Nanosekunden.

Ziel ist es, viele Tausende dieser Nanodraht-Speicher, die zwischen 10 und 100 Bits speichern können, dicht auf einer Fläche anzuordnen. Die senkrechte Anordnung von Nanodrähten eröffnet zudem neuartige dreidimensionale Architekturen – ein Paradigmenwechsel im Vergleich zu derzeitigen zweidimensionalen Chip- und Speichertechnologien. Die Ausnutzung der dritten Dimension schafft neue Spielräume für Leistungssteigerungen und Kostenreduktionen, die nicht mehr nur – wie bis jetzt – durch Miniaturisierung erzielt werden.



Stuart Parkin darf sich zu Recht freuen, er hat es geschafft und Racetrack kann die Welt verändern

Ein ganz herzlicher Dank gilt auch den Sponsoren dieses IBM System Storage Kompendiums, welche die Druckauflage unterstützt haben:

Accelerating Your Success™

Ihr kompetenter Partner für
IBM Storage Lösungen

Avnet Technology Solutions GmbH
Gutenbergstraße 15
D-70771 Leinfelden-Echterdingen
Telefon: 0711/51881–0
Fax: 0711/51881–506
ats.de@avnet.com
www.ts.avnet.com/de

A Trademark of ***Tech Data***

Tech Data GmbH & Co. OHG
Geschäftsbereich Azlan,
Ihr IBM Team bei Azlan
Kistlerhofstrasse 75
D-81379 München
Telefon: 089/4700–3003
Fax 089/4700–3007
IBMSupport@techdata.de
www.azlan.de

Ein ganz herzlicher Dank gilt auch den Sponsoren dieses IBM System Storage Kompendiums, welche die Druckauflage unterstützt haben:

more than distribution

Ihr persönlicher Distributor

ALSO Deutschland GmbH
IBM Division
Ernst-Heinkel-Straße 4 • 94315 Straubing
Tel. +49 94 21 93 30 10 • Fax +49 94 21 93 38 14
ibm-de@also.com • www.also.de

Technologie-Integration für Datacenter
Zertifiziertes IBM Business Partner Innovation Center

SVA GmbH
Borsigstraße 14
65205 Wiesbaden
Telefon: 06122/536–0
Fax: 06122/536–399
mail@sva.de
http://www.sva.de

IBM Deutschland GmbH
70548 Stuttgart
ibm.com/de

IBM Österreich
Obere Donaustraße 95
1020 Wien
ibm.com/at

IBM Schweiz
Vulkanstrasse 106
8010 Zürich
ibm.com/ch

Die IBM Homepage finden Sie unter:
ibm.com

Gedruckt in Deutschland.

IBM Form GM12-6777-03 (04/2008)