

IBM System Storage-Kompodium 50 Jahre Innovationen in Plattenspeichern

Die Ausarbeitung dieses IBM Storage-Kompandiums erforderte doch mehr Zeit und Aufwand und ich musste viel freie Zeit investieren. Deshalb möchte ich mich zuallererst bei meiner Ehefrau Gabriele auf das Herzlichste bedanken, dass sie die Geduld und das Verständnis für diese Arbeit aufbrachte und mir die Möglichkeit gab, immer wieder daran zu arbeiten. Ohne ihre Unterstützung hätte das Kompandium nicht in dieser kurzen Zeit fertiggestellt werden können. Ihre Geduld war bewundernswert.

Des Weiteren möchte ich mich ganz herzlich bei meinem ehemaligen Storage-Kollegen Peter Voss sowie bei Hans W. Spengler vom IBM Museum in Sindelfingen bedanken, die mir wichtige Informationen von der Anfangszeit bis hin zu den 3390-Platten liefern konnten, die ich aufgrund meines jungen Alters nicht mehr direkt erlebt habe.

Ein großer Dank gilt auch vielen IBM Kollegen, die für die unterschiedlichsten Themen Informationen lieferten, deren Beschaffung und Textgestaltung sonst wesentlich aufwendiger gewesen wäre. Ich möchte mich hier vor allem bei Niels Haustein und Hartmut Grund für ihren Input der DR550, Benedikt Braun für seine Mitarbeit an der DS4000, Herbert Vogel für seinen FASiT-Input, Ralf Colbus für seinen aktuellen Input der Nseries-Produkte, Volker Widder und Werner Bauer für ihren DS8000 Input, Peter Sauter für seinen Input in den unterschiedlichen Plattenbereichen, Stefan Neff für seine allgemeine Beschreibung der Verschlüsselungstechnik, Hartmut Lonzer für die vielen Internetlinks, die mir wesentlich die Recherchen erleichterten, Andreas Grottel für seine Informationen im Plattenumfeld und Roland Rauser, der mir viel Arbeit abnahm, sodass ich doch die eine oder andere Arbeitsstunde an der Broschüre arbeiten konnte. Allen anderen Kollegen, die mir dabei geholfen haben und jetzt nicht namentlich erwähnt sind, gilt ein ebenso großes Danke.

Mein Dank gilt auch dem IBM System Storage Marketing, vor allem Herrn Manfred Raiser und Herrn Jens Stolze, welche die Realisation und den Druck dieses Kompandiums ermöglichten.


Auch möchte ich mich bei meinen Vorgesetzten Klaus Wiethölter und Marcellus Scheefer bedanken, die mir bewusst das eine oder andere Mal den Rücken freihielten.

Der größte Dank allerdings gilt jemand ganz anderem, er gilt der Firma IBM und der IBM Storage Group mit ihren Labors. Ohne IBM als die treibende Kraft im Storage-Entwicklungsbereich würde es vielleicht diese Historie nicht geben. Meine Recherchen haben sehr klar ergeben, dass keine andere Firma außer IBM nahezu alles, was an Storage-Basistechnologien die letzten 55 Jahre entwickelt wurde, für sich verbuchen konnte. IBM leistete in diesem Bereich einmalige Arbeit mit nie da gewesenen Innovationen und ich bin überzeugt, dass sich dies in der Zukunft fortsetzen wird.

Ich hoffe und wünsche mir, dass Sie genauso viel Freude und Spannung beim Lesen dieses Kompendiums haben, wie ich es während meiner Recherchen im letzten halben Jahr hatte.

Ich hoffe und wünsche mir, dass dieses Kompendium dazu beiträgt, dass mehr Verständnis für die Entwicklungen im Speicherbereich entsteht und das Verständnis in den Entwicklungsbereichen der Speicherbasistechnologien vertieft wird. Ebenso möchte ich alle jungen Kollegen dazu auffordern, dieses Kompendium für eigene Schulungszwecke einzusetzen, da alle Storage-Produkte, die von der IBM im Jahr 2006 aktuell angeboten werden, ausführlich beschrieben sind.

In diesem Sinne möchte ich mich nochmals bei allen herzlichst bedanken!



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 50 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 'Online'-Speicherbasistechnologien (Platten) und Aufzeichnungsverfahren einschließlich anstehender Technologien näher beschreibt.

Entwicklung der Speichersysteme von 1952 bis 2006

1952 – 1961	die Anfangsepoche der elektromagnetischen Speicherung	Seite 7
1962 – 1974	die Epoche der Wechselplatten und die 'Winchester'-Zeit	Seite 13
1975 – 1993	die Epoche der fest eingebauten Platten mit externen Kontrolleinheiten	Seite 21
1994 – 1998	die Epoche der RAID-Systeme	Seite 31
1999 – 2005	die Epoche der Multiplattform-Systeme und des Fibre Channel SAN/NAS	Seite 43
2006	die Epoche der Server-basierenden Speichersysteme mit neuen Infrastrukturmöglichkeiten	Seite 73

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

Produktübersicht der IBM Speichersysteme

Vorwort	Seite	2	Die Epoche der Multiplattform-Systeme und des FC SAN und NAS	Seite	43
Speicherentwicklung, Gesamtüberblick	Seite	4	SAN Storage Area Network	Seite	44
Produktübersicht	Seite	5	NAS Network Attached Storage	Seite	45
Preis- und Kapazitätsentwicklung Plattenspeicher	Seite	6	iSCSI	Seite	46
Die Anfangsepoche der elektromagnetischen Speicherung	Seite	7	ESS Enterprise Storage Server (Shark)		
726/727 Magnetbandsystem	Seite	8	Plattensystem	Seite	46
RAMAC 305/350 Plattensystem	Seite	8	7133 + 7140 Hammer Shark SSA		
729 Magnetbandsystem	Seite	10	Plattenlösung für SAN	Seite	48
1405 Platteneinheit/1414 Steuereinheit	Seite	10	MSS Modular-Storage-Server-Plattensystem	Seite	48
1301 Plattensystem	Seite	10	7135 RAIDiant-Array-Plattensystem	Seite	49
7340 Hyper Tapes	Seite	11	FASTT-Plattensysteme	Seite	49
2321 Magnetstreifenspeicher	Seite	11	DS4000-Plattensysteme	Seite	52
Die Epoche der Wechselplatten und die 'Winchester'-Zeit	Seite	13	SVC SAN Volume Controller	Seite	54
1311 Wechselplattensystem	Seite	14	LTO 2/LTO 3 Linear-Tape-Open-Bandlaufwerke	Seite	54
1302 Plattensystem	Seite	15	3592/TS1120 Bandlaufwerk (Jaguar)	Seite	56
2401 Magnetbandsystem	Seite	15	3584/TS3500 Bandarchiv	Seite	59
2314 Wechselplattensystem	Seite	15	3581 Band Autoloader	Seite	61
2305 Plattenspeicher 'Zeus'	Seite	15	3582 Bandarchiv	Seite	61
3330 Wechselplattensystem/3830 Steuereinheit	Seite	16	3583 Bandarchiv	Seite	61
3340 Wechselplattensystem 'Winchester'-Platte	Seite	17	TS3310 Bandarchiv	Seite	61
3420 Magnetbandsystem	Seite	18	TS3100 Band Autoloader	Seite	62
3410 Magnetbandsystem	Seite	18	TS3200 Bandarchiv	Seite	62
3850 MSS-Maßenspeicher	Seite	19	TS7510 Tape-Virtualisierung für Open Systems	Seite	62
Die Epoche der fest eingebauten Platten mit externen Kontrolleinheiten	Seite	21	3995 Optisches Archivsystem	Seite	64
3350 Festplattensystem	Seite	22	DR 450 Data Retention 450 Archivlösung	Seite	66
3310 Magnetbandsystem	Seite	22	DR 550 Data Retention 550 Archivlösung	Seite	66
3370/3375 Plattensystem	Seite	22	3996 Optisches Archivsystem	Seite	68
3380 Plattensystem/3880 Steuereinheit	Seite	23	N3700 Nseries	Seite	69
3430 Magnetbandsystem	Seite	24	N5200/N5500 Nseries	Seite	70
3480 Magnetbandkassettensystem	Seite	24	N7000 Nseries	Seite	71
3390 Plattensystem/3990 Steuereinheit	Seite	25	Die Epoche der Server-basierenden Speichersysteme mit neuen Infrastrukturmöglichkeiten	Seite	73
9340/9345 Plattensystem	Seite	26	DS8000-Plattensystem	Seite	74
3490 Magnetbandeinheit	Seite	27	DS6000-Plattensystem	Seite	77
3495 Magnetbandarchiv	Seite	28	TS7700 Tape-Virtualisierung für z/OS (Mainframe)	Seite	80
3494 Magnetbandarchiv	Seite	29	Encryption TS1120 Bandlaufwerk	Seite	82
Die Epoche der RAID-Systeme	Seite	31	Infini-Band	Seite	86
RAID-Definitionen	Seite	32	Neue Basis-Technologien	Seite	88
RAMAC 1/2/3 RAID-Plattensystem	Seite	33	Technologie-Anhang: 'Online'-Technologien	Seite	89
RVA RAMAC Virtual Array	Seite	34	Induktive Aufzeichnung	Seite	90
RSA RAMAC Scalable Array	Seite	35	Magnetoresistive Aufzeichnung	Seite	90
7133 SSA-Plattensystem	Seite	35	PLMR Encoding	Seite	92
3590 Magstar-Magnetbandeinheit	Seite	36	Micro-Drives	Seite	92
3570 Magstar-MP-Magnetbandeinheit	Seite	37	AFC-Aufzeichnung	Seite	93
3494 B16, B18, B10, B20 Virtual-Tape-Server	Seite	38	Perpendicular Recording	Seite	94
LTO Linear-Tape-Open-Bandeinheiten	Seite	39	Millipede-Nanotechnologie	Seite	95

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	900	690	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	180.0
1999	ESS	720.0
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).

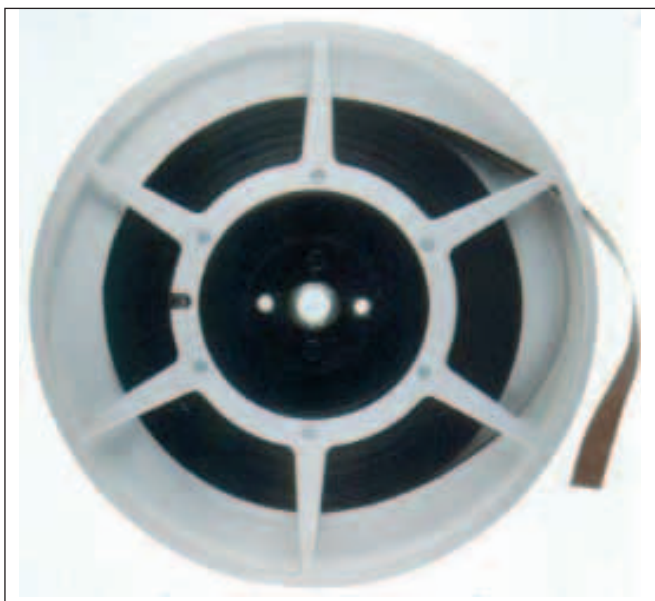
Die Anfangsepoche der elektromagnetischen Speicherung





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 Rechneinheit IBM 701 untergebracht. Erste dedizierte Bandsteuereinheiten gab es dann ab **1956**.



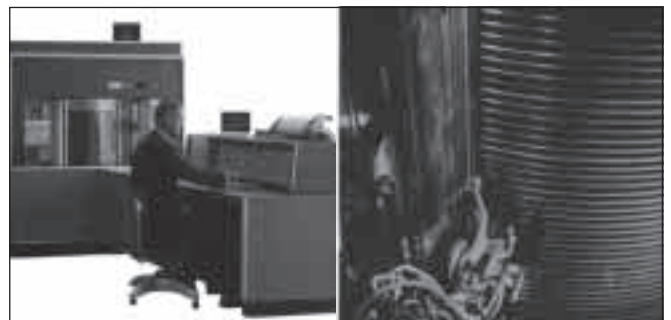
Medium-Rollenband für IBM 726



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.

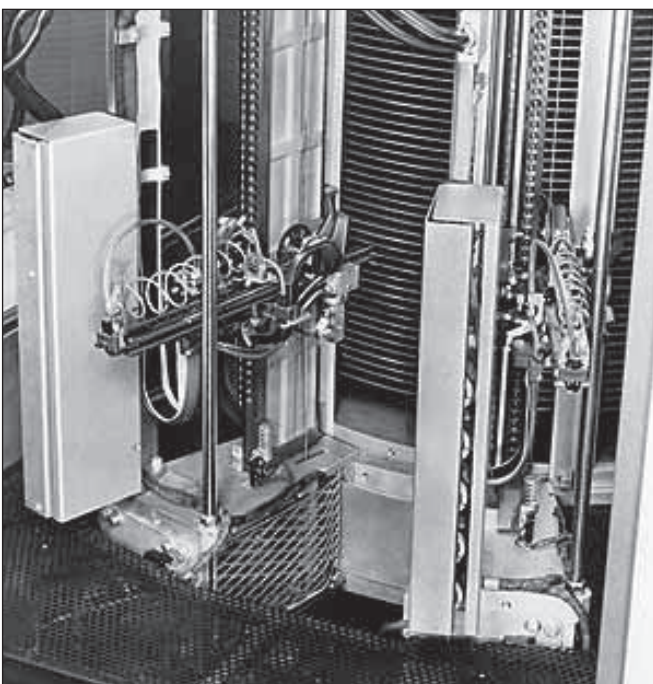


IBM 350 RAMAC Platte, 51 Platten im Stapel mit 24 Zoll Durchmesser. 1956: Modell mit 5 MB (5 Millionen 'Characters'), 1958: Modell mit 10 MB (10 Millionen 'Characters'), Zurückziehung vom Vertrieb im Jahr 1960



IBM RAMAC 350 Plattenstapel

Damals gab es noch keine Zugriffskämme, sondern nur einen Zugriffssarm, 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, so dass 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 entsprechen, 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



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.

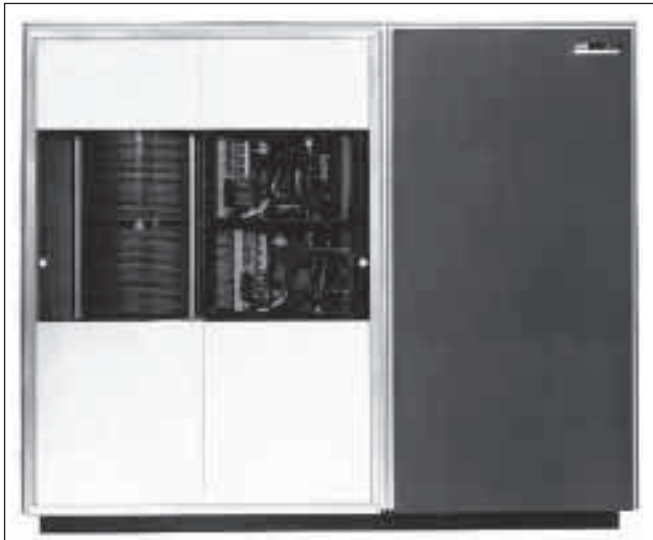
Im Oktober 1960 wurde der **Plattenspeicher IBM 1405** angekündigt, der an den Systemen IBM 1401, 1410 und 1460 (nicht an System 7000) betrieben wurde. Dieser Plattenspeicher vervierfachte die Kapazität eines Plattenstapels im Vergleich zur RAMAC von 1956. Sowohl die Anzahl der Spuren pro

'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 Platten-einheit** angekündigt. Sie wurde ein Jahr später, im 3. Quartal 1962, an Kunden ausgeliefert. Die Entwicklung dieser Platten-einheit 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

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.



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

Die **Hyper Tape Drives IBM 7340** reflektierten die neueste 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 Bandensystem, 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 aufwendige Konstruktion und der damit verbundene hohe Wartungsaufwand standen einer weiten Verbreitung entgegen.



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



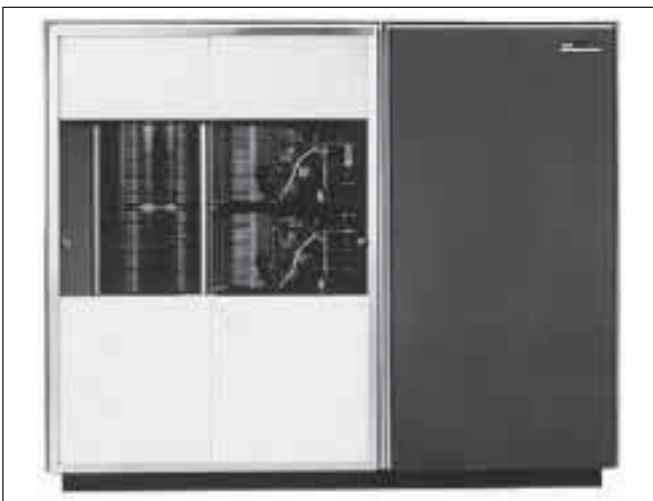
Die Epoche der Wechselplatten und die 'Winchester'-Zeit



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

Die **zwischen 1962 und 1964** eingeführten **Magnetplatten-speicher IBM 1311, IBM 2311 und IBM 2314** zum Anschluss an die 360er-Systeme arbeiteten mit auswechselbaren Magnetplattenstapeln. Die neue **Plattensteuereinheit IBM 2841** war **erstmalig 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

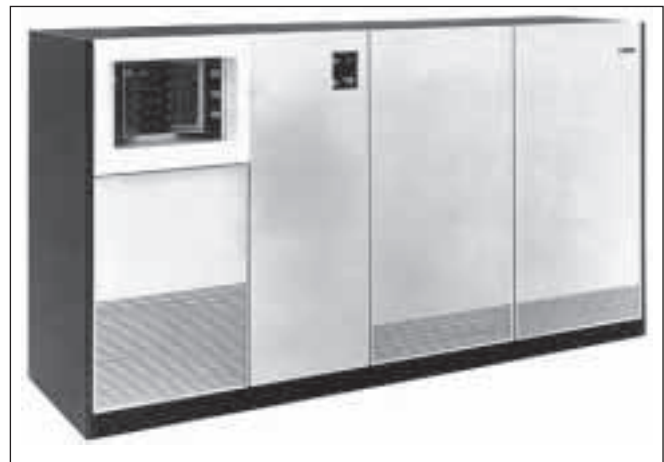


IBM 2401 Magnetbandeinheit

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%



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

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.

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

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 Mikroprogramm-speicher, 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 Antrieb 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 Kanal des Rechners, sobald der Anfang des im Kanalbefehl vorgege-



*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*

benen 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 Magnetplatten-speicher 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.



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

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 Magnetband-system** 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 3410 Magnetbandsystem mit horizontaler Bändeinlegemöglichkeit

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.



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





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 Micro-fiches, rechts hinten: Steuerkonsole des Systems/370 Modell 168

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 austauschbare Speichermedium für Datensicherung, Offline-



IBM 3310 Direct Access Storage, Plattensystem für IBM 4331 Rechner

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 meldete 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 Lauf-

werke 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.

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 Magnetbandsystemseite 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).

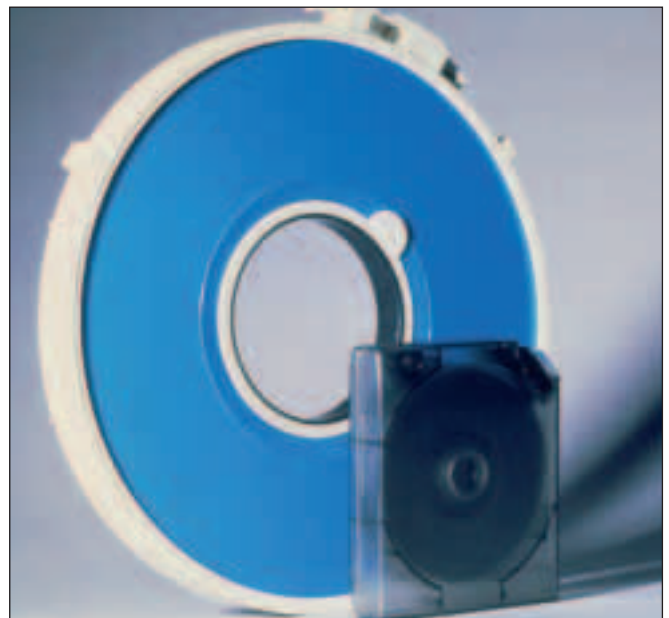
Aufgrund der ein Jahr später angekündigten neuen IBM 3480 Kassettentechnik als Ablösung der Rollenbänder fanden 3430 Einheiten nicht in großen Stückzahlen ihren Absatz, hielten sich aber dennoch ziemlich lange auf dem Markt, um der Datenträgeraustauschanforderung gerecht zu werden, die in den Folgejahren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



IBM 3430 Magnetbandsystem, 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 Aufzeichnungsichte 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



Größenvergleich Magnetbandsysteme IBM 3420 gegenüber IBM 3480

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.

Die neue Schreib-/Lesekopf-Technik der IBM 3390 benötigte eine neue Beschichtung in sogenannter Dünnschichttechnik. Dabei beschichtete man die Oberflächen von Glasplatten aus einem speziell gehärteten Glas mit einer Legierung und unterzog sie anschließend einem Einbrennprozess. Die magnetoresistive Schreib-/Lesetechnik in Verbindung mit Dünnschichtbeschichtungen auf den Platten kam innerhalb der 3390 Baureihe mit den Modellen 3 und 9 erstmals zum Einsatz. Die seit der Einführung von Magnetplatten verwendete 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 3390 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 ausreichen.

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 neuester 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 Datenbank-anwendungen 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 Kom-

primierungsverfahren 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



IBM 3495 Bandarchiv mit 5660 bis 18920 Kassetten (13.5 bis 45.5 Terabyte), 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 Terabyte 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 Kanalraten 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

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, leistungstärkere 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 Dünnschicht-Beschichtungen wurden eingeführt. Dies führte zu wesentlich höherer Zuverlässigkeit, weil mit Grain-Strukturen gearbeitet werden konnte und Eisenoxidpartikel (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 **'RAID Architecture with Multilevel Adaptive Cache'**.

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 Übertragungsraten 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 2006) 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 Hardware- 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.

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 aufwendig 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, Minimal Konfiguration: 2 Arrays, Maximal Konfiguration: 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 Terabyte 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.

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



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

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 Laufwerksträngen 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.

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 Fibre-Channel-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 Fibre-Channel-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 Mal in der Bandtechnologie magnetoresistive Schreib-/Lese-



Blick in die IBM 3590 Magstar Laufwerke

köpfe mit dem 'Track Following Servo'-Prinzip eingesetzt, wo die Leselemente 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 Terabyte (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 Kassetten speicherte (native ohne Komprimierung). Die Daten-



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

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

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 Kassetten-Design 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.



IBM Magstar MP 3570 Bandsystem

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.

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 Platten-cache 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.



IBM Virtual Tape Server VTS, 1997: Modell B16, 1998: Modell B18, 2002: Modell B10 und B20

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.

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- $\frac{1}{2}$ 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 und 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 Fertigungslizenzen 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, 2006, die Marktanteile betrachtet, ging diese Rechnung auf. DLT ist heute nahezu vom Markt verschwunden.

Als die LTO Group **1997** anfang, 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 (LTO 1) und 35 MB/s mit 512 Spuren (LTO 2) beschrieben wird. Dabei wurden immer 8 Spuren gleichzeitig geschrieben. Die Kassetten erreichten bei diesen Spurdichten native Kapazitäten von 100 GB (LTO 1) und 200 GB (LTO 2). Bei dieser hohen Spurdichte 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 sieben 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). 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.

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



LTO 1 Laufwerk mit LTO 1 Ultrium Kassette



Kassettenladevorgang beim IBM LTO 1 Laufwerk



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

Fibre-Channel-basierende Steuereinheit integrierte, um den Libraries über die dadurch geschaffene Multipfadarchitektur die Möglichkeit zu geben, logische Partitionen als Sub-Libraries anzulegen und um direkt Fibre-Channel-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.

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 Wettbewerb. 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 LTO-1-Technologie fertig. Danach begann ein Prestige-Wettrennen zwischen IBM, HP und Seagate. Jeder wollte der Erste sein, der mit LTO-1-Laufwerken auf den Markt kommt.

Seagate kündigte als erste Firma, bereits im Frühjahr 2000, ihre LTO-1-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 Fibre Channel 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 Fibre Channel Switches auf 8-Port-Basis mit 1-Gbit-Ports auf den Markt. Eine **neue Infrastruktur, das SAN (Storage Area Network)** auf Fibre-Channel-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 Rechenzentrumsambiente 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 Fibre-Channel- 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 Fibre Channel SANs entwickelten sich rasch weiter. **2002** kamen die ersten **Switches mit 2-Gbit-Ports** auf den Markt. Heute, im **Jahr 2006**, sind alle Hersteller im Begriff, von 2-Gbit- auf 4-Gbit-Technologie umzustellen und es ist davon auszugehen, dass bis Ende 2006 'End to End'-SAN-Implementierungen auf 4-Gbit-Basis realisiert werden, da sowohl die Host-Adapter aller Server-Hersteller als auch die Host-Ports aller Storage-Anbieter auf 4-Gbit-Technologie umgestellt worden sind.

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 Fibre-Channel-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 Fibre Channel 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 Terabyte. 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 Fibre Channel 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 Fibre Channel 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(Transaktion 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 Infrastrukturverbesserungen 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-Hostanbindungstechnologie, also 2 Gbit FICON und 2 Gbit Fibre Channel, ausgelegt. Daneben konnte die Maschine immer auch noch mit ESCON- und Ultra-SCSI-Adaptoren 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 Fibre-Channel-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 Fibre Channel 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.

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 Architektur-entwicklung 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 Fibre Channel 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 Einstiegsconfigurationen einer Shark für den Mittelstandsbereich zu teuer waren. Deshalb wurde bereits im **Jahr 2000** mit der Firma **Compaq** ein **OEM-Vettrriebsvertrag** unterzeichnet, der IBM erlaubte, Compaqs aktuelles Plattensystem mit Fibre-Channel-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 FASt500 und die IBM 3542 FASt200 und im Jahr 2001 die IBM 1742 FASt700.

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 **FASt200**, damals auch als IBM 3542 bekannt, das mit bis zu zwei Fibre Channel 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 '**FASt**' 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 FASt mit der Intention, dass das kleine 't' nicht ausgesprochen wird. Trotz allem setzte sich die Aussprache als 'fast_T' in ganzer Breite durch.

Die Reihe der FASt-Plattensysteme wurde im **Oktober 2001** um die leistungsfähigeren Produkte **FASt500 und FASt700** erweitert, die mit bis zu 4 Fibre-Channel-Anschlüssen höhere Kapazitäten boten. Im Hochleistungsbereich kam im **Februar 2003** die **FASt900** dazu, um die Baureihe auch im oberen Leistungsbereich abzurunden. Im **April 2003** wurde die FASt500 durch die **FASt600 und FASt600 Turbo** ersetzt. Neben den Plattenerweiterungseinheiten EXP500 und EXP700 (EXP steht für Expansion), die mit Fibre-Channel-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 **FASt100** angekündigt. **SATA-Platten** (Serial Advanced Technology Attached) sind günstige IDE-Platten mit serielltem Interface, die bisher ausschließlich im Heim-PC-Bereich ihre Anwendung fanden. Mit Ausnahme der schon länger verfügbaren FASt200 waren alle Systeme jetzt mit 2 Gbit Fibre Channel ausgestattet, und das sowohl Host-seitig als auch in der Anbindung der Plattenloops an

	FASt900	FASt700	FASt600 Turbo	FASt600	FASt200	FASt200
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 FASt-Plattensubsystemfamilie, Spezifikationsübersicht

den jeweiligen FASTT Controller. Über 90% des Vertriebs dieser LSI-Plattensubsystem-Serie lief über IBM oder IBM Geschäftspartner und war im Fibre-Channel-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, das 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.

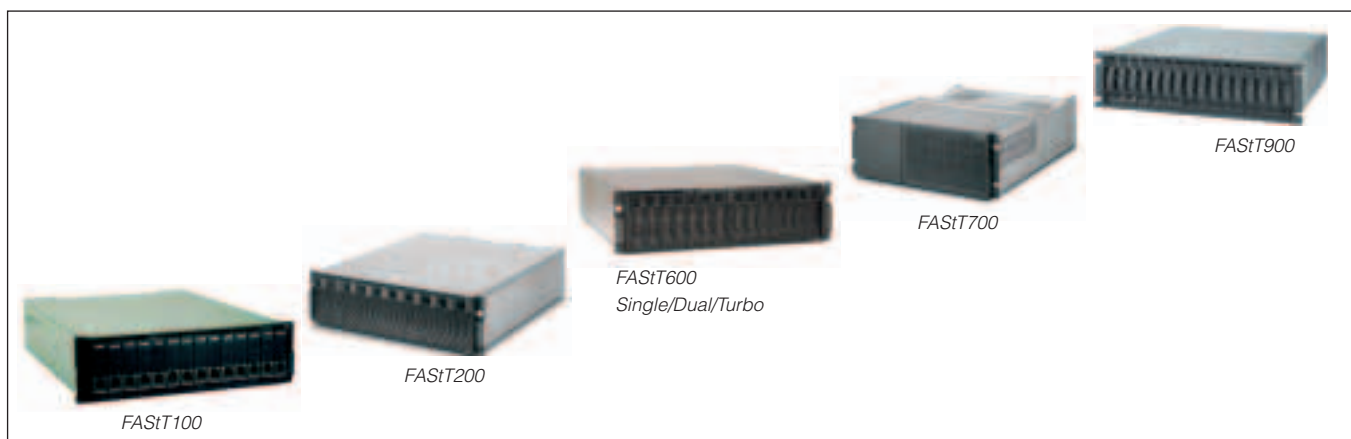
DAE: Dynamic Array Expansion. Sollte innerhalb einer FASTT weitere physische Kapazität benötigt werden, so gibt es die Möglichkeit, im laufenden Betrieb die angeschlossenen Erweiterungseinheiten mit weiteren FC-Laufwerken in verschiedenen Kapazitätsgrößen und mit unterschiedlichen RPMs zu bestücken (bis 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 FAST-Systemen?*

Über dem FAST 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 Repositories ist abhängig von der Anzahl der Änderungen am Originalvolume. Die Erfahrung zeigt, dass eine Repositorygröße von 20% des Originalvolumens 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 (Block A), 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 FAST-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 FAST-Systems zur Verfügung stellen, erstellt Remote Mirroring eine Kopie von Daten über eine SAN-Infrastruktur von einer FAST auf eine weitere. Diese Funktion ist für die FAST-Systeme 700 und 900 verfügbar. Remote Mirroring ist eine bidirektionale, synchrone Kopie. Über dedizierte Fibre Channel Ports können bis zu 32 Volumes auf eine zweite FAST gespiegelt werden. Remote Mirroring wird für Katastrophenvorsorge und High-Performance Copys eingesetzt.*
- *Partitioning: Um unterschiedliche Speicherbereiche auf der FAST 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.*



FAST RAID Kontroller Familie



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 FAST-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 FAST-Plattenfamilien wurden im Einzelnen folgende Rebrandings durchgeführt:

Naming Prior to Sept 7, 2004	New naming as of Sept 7, 2004
IBM TotalStorage FAST Storage Server	IBM TotalStorage DS4000
FAST	DS4000
FAST Family	DS4000 series
FAST Storage Manager vX.Y (example FSM v9.10)	DS4000 Storage Manager vX.y (example 9.10)
FAST100	DS4100
FAST600	DS4300
FAST600 with Turbo Feature	DS4300 Turbo
FAST700	DS4400
FAST900	DS4500
EXP700	DS4000EXP700
EXP100	DS4000EXP100
FAST FlashCopy	FlashCopy for DS4000
FAST VolumeCopy	VolumeCopy for DS4000
FAST Remote Volume Mirror (RVM)	Enhanced Remote Mirroring for DS4000
FAST Synchronous Mirroring	Metro Mirroring for DS4000
FAST Asynchronous Mirroring (New Feature) w/o Consistency Group	Global Copy for DS4000
FAST 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-Richtlinie 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 Fibre-Channel-Platten durch die EXP810 mit FC-Platten auf 4-Gbit-Basis ersetzt.

Die DS4500, ehemals FAST900, 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.

Sowohl bei Kapazität und Leistung als auch beim Preis bietet die neue Modellfamilie der DS4000 für jeden Endbenutzer die maßgeschneiderte Plattenlösung im Fibre-Channel-Umfeld an.

DS4000 Series aktuell im Jahr 2006

- DS4200**
4 Gbit
2 GB Cache
RoHS compliant
SATA Controller
- DS4700 Modell 70**
4 Gbit
2 GB Cache
RoHS compliant
- DS4700 Modell 72**
4 Gbit
4 GB Cache
RoHS compliant
- EXP420**
4 Gbit
RoHS compliant
SATA-Platten 500 GB
- EXP810**
4 Gbit
RoHS compliant
- DS4800 Modell 80**
4 Gbit
4 GB Cache
RoHS compliant
- DS4800**
4 Gbit
4/8/16 GB Cache
RoHS compliant

Aktuelle Modelle der IBM DS4000-Plattenfamilie im Jahre 2006

SAN Virtualisierung

Die **Epoche der Multiplattform-Systeme** war stark durch die Einführung von Fibre-Channel-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 Entkopplung 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**.

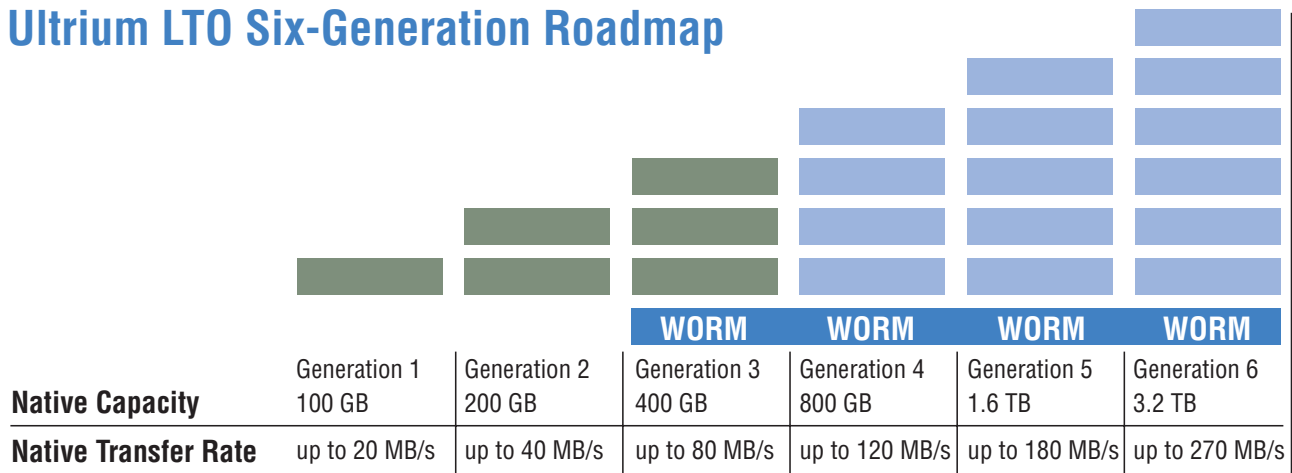
Bandsysteme

Neben der Plattensubsystem-Weiterentwicklung und den Fortschritten 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. LTO-2-Laufwerke arbeiteten bereits mit 35 MB/s mit 8-Spur-Technik. Bei **LTO 3** wurde auf 16-Spur-Technik und eine Datenrate von 80 MB/s umgestellt.

Wie vorgesehen wurde die Schreib-/Lese-Rückwärtskompatibilität eingehalten. So können LTO-3-Laufwerke immer noch Kassetten der Generation 1 lesefähig verarbeiten und Kassetten der Generation 2 sowohl schreiben als auch lesen.

Mit der Einführung der LTO-3-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



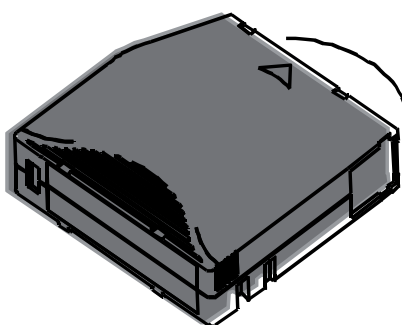
Standardisierte LTO-Roadmap bis Generation 6

Kassette und in der Tape Volume Control Region am Bandanfang. Um absolut manipuliertsicher zu sein, wird bei der Herstellung dieser WORM-Kassetten die Kennzeichnung noch auf die vorgeschriebenen Servobänder gebracht: eine Vierfach-Kennzeichnung, die sicherstellt, dass LTO-3-WORM-Kassetten nicht mehr überschrieben werden können. LTO-WORM-Kassetten sind zweifarbig gekennzeichnet. Die eine Hälfte ist schwarz, und die Andere in weißer Farbe.

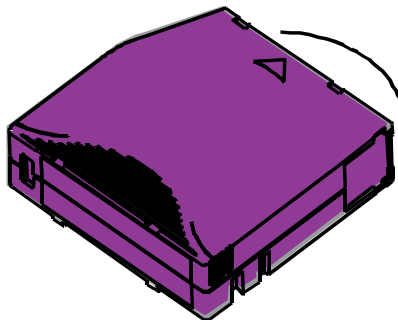
Lag der Marktanteil 1999 mit 90% noch auf Quantums DLT-Technologie, hat sich das Bild seither um 180 Grad gedreht und die LTO-Technologie dominiert heute mit über 80%. Der LTO-3-Streamer von IBM gilt heute als das leistungsstärkste LTO-Bandlaufwerk auf dem Markt.

Mit der Verfügbarkeit von LTO 3 wurde die offizielle Roadmap um zwei weitere Generationen mit LTO 5 und LTO 6 erweitert. Es ist davon auszugehen, dass alle 2 bis 2 1/2 Jahre eine neue LTO-Generation gemäß dieser Roadmap auf dem Markt verfügbar wird.

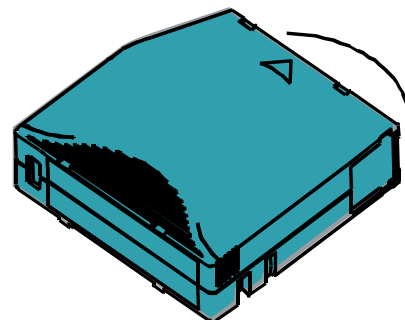
Die etwas eigenwillige Farbwahl der standardisierten LTO-Kassetten kommt aufgrund einer wichtigen Anforderung zustande. Farbenblinde müssen in der Lage sein, die Kassetten optisch zu unterscheiden.



LTO-1-Kassette, 100 GB Kapazität (native)



LTO-2-Kassette, 200 GB Kapazität (native)



LTO-3-Kassette, 400 GB Kapazität (native)

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 LTO-2- und LTO-3-Laufwerk integriert) **erstmalig PRML Encoding** (Partial Response Maximum Likelihood – 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, solche hochqualitativen 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 Kopf-technologie kam erstmals in vollem Umfang bei LTO 2 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 LTO 2 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 Tape-mark, 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.

Im **November 2005** wurde die zweite Generation der Jaguar-Laufwerke verfügbar. Die **3592 Generation 2**, die Marketingseitig als TS1120-Laufwerk (TS steht für Tape System) bezeichnet wird, war und ist das erste Bandlaufwerk, das mit **zwei 4-Gbit-Fibre-Channel-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 LTO-3-WORM-Kassetten, bieten aber neben allen WORM-Sicherheiten den Vorteil, dass das **Fort-schreiben 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. LTO-3-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.

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.



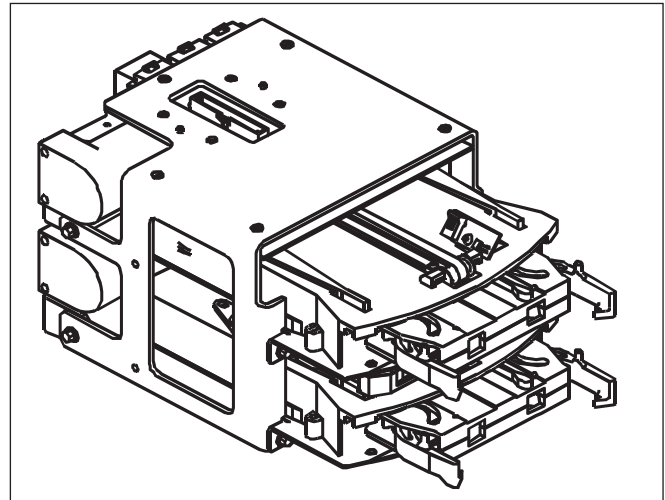
IBM TS1120(3592 Generation 2)-Bandlaufwerk der Spitzenklasse



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



IBM 3584 Grundeinheit mit Ein-/Ausgabestation



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.

Auf der CeBIT in Hannover **sorgte die 3584 immer wieder für viele Mensentrauben**. 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.

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 **Multi-pfad-Architektur** mit direkter Anbindung von LTO- und 3592 Fibre-Channel-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 Laufwerktausch, 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-Fibre-Channel-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 LTO-1-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 Fibre-Channel-Netz anschließen zu können, wurden über ein eingebautes FC-AL Gateway die bis zu sechs SCSI-Laufwerke 'daisy chained' angeschlossen.

Diese Fibre-Channel-Anschlussweise der IBM 3583 verursachte bei bestimmten Host-Fibre-Channel-Adaptoren 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 LTO-3-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 LTO-3-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 LTO-3-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 3582 wurde durch die **IBM TS3100** ersetzt, ein neuer Autoloader mit einem 4-Gbit-Fibre-LTO-3-Laufwerk und bis zu 22 Kassettenstellplätzen. Die Mini-Library 3582 wurde durch die **IBM TS3200** ersetzt, die bis zu zwei **4-Gbit-LTO-3-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.



IBM TS3100

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 LTO-3-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 TS3200

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 Rechneinheit emuliert. Bei einer Dual-Server-Konfiguration werden bis zu 1024 virtuelle Laufwerke emuliert. Abgebildet werden LTO-2-, LTO-3- und 3592(Jaguar 1)-Laufwerke. Es stehen bis zu 128 virtuelle Libraries und bis zu 8192 virtuelle Volumes zur Verfügung, also pro Rechneinheit 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 Minimal-konfiguration 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.

In Kürze wird ein Nachfolgesystem zur Verfügung stehen, das neben höheren Plattenpufferkapazitäten, höherer Durchsatzleistung und einer wesentlich höheren Anzahl an emulierten Laufwerken, Libraries und virtuellen Volumes neue Funktionalitäten bietet, die den Einsatz im Open-Systems-Bereich optimieren.



Virtuelle Tape Library IBM TS7510

Archivierungslösungen

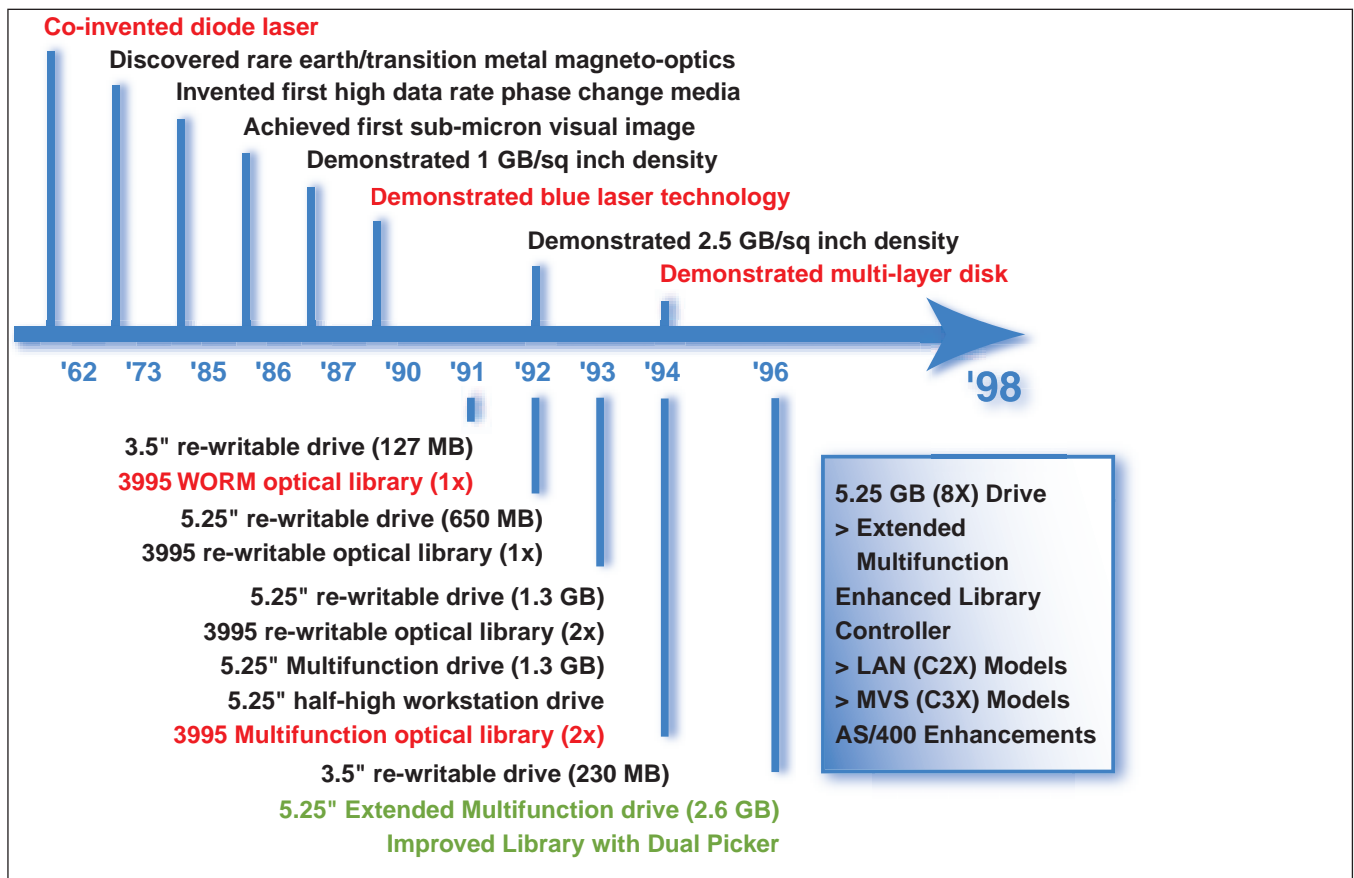
Archivierung von Informationen ist schon ein sehr altes Thema, denkt man an die Höhlenmalereien der Grotte Chauvet im Vallon Pont d'Arc in Südfrankreich, die auf ca. 31.000 Jahre geschätzt werden – die wohl erste Überlieferung von Informationen aus Menschenhand.

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.

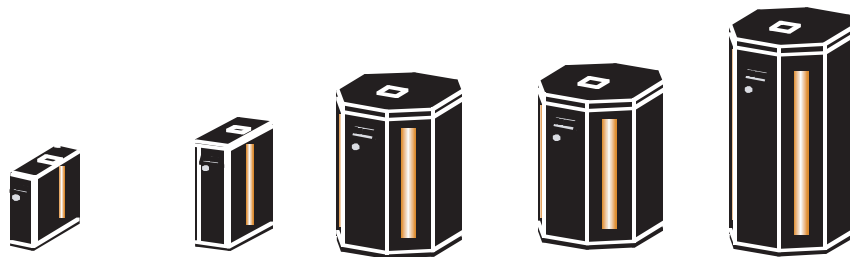
Mit der Erfindung des Buchdruckes durch Johannes Gutenberg aus Mainz, Mitte des 15. Jahrhunderts, war es dann auch möglich, Informationen einfach zu vervielfältigen, was zu vielen interessanten Überlieferungen beigetragen hat und natürlich auch zum Wachstum an Informationen.

Bereits **1982** gründeten IBM und Sony eine Entwicklungsalianz mit dem Ziel, **optische Technologien** gemeinsam weiterzuentwickeln. So entstanden 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	SCSI	SCSI	SCSI	SCSI	SCSI
X = 2	LAN	LAN	LAN	LAN	LAN
X = 4	AS/400	AS/400	AS/400	AS/400	AS/400
X = 3	--	S/390	S/390	S/390	S/390
X = 1	--	S/390	--	S/390	S/390
(Expansion)					
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.

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. Revisions-sicherheit 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 pharmazeuti-schen 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 über-führt 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-Betriebssystem) 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 FAST600 mit EXP100-Erweiterungseinheiten abgespeichert. Kapazitäten von 3.5 TB bis 56 TB konnten konfiguriert werden. Das Magnetplattensystem war über ein redundant ausgelegtes, Fibre-Channel-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 Einbauschränk kann optional mitbestellt werden.
- Das DR550-Single-Node-System besteht aus einem Einbauschränk, 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 Einbauschränk 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 Zugriffszeiten. 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

Die Epoche der Multiplattform-Systeme und des Fibre Channel SAN und NAS

Die physikalische Anbindung der DR550 an die Serversysteme erfolgt über Ethernet-Schnittstellen und basiert auf dem TCP/IP-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 optisch Archive – 960 GB–5.2 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, boten 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. Die kapazitiven Möglichkeiten liegen also deutlich unter den Möglichkeiten einer DR550.

Es bleibt abzuwarten, ob basierend auf der Blaue-Laser-Technik ein klassischer optischer 2 x Standard verabschiedet wird, da andere Lösungsoptionen für die Langzeitarchivierung inzwischen 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' 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 Fibre Channel 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 'löschen/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 Systemen 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 Hochleistungssysteme Nseries 7000 bestehen aus der **N7600**, die kapazitiv bis 420 TB (bis zu 672 LUNs) skaliert, und der **N7800**, die 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.

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 zur Zeit 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 von statten! 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 Packet für den Systemadministrator. Statt aufwendiger 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, Attachements) für das MS Exchange Umfeld nach einem Fehlerfall einfach wieder herzustellen.

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 Seite. 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 der nSeries auf eine andere (oder von vielen auf eine). Durch den Einsatz von günstigen SATA Disks kann auf das Backup-system schnell und kostengünstig gesichert werden. Da ein Restoring von Disk erfolgt, sind die Daten auch schnell wieder herstellbar. 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.

Kommentar zur Epoche Multiplattform-Systeme und Fibre Channel 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 Fibre-Channel-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 Fibre-Channel-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 Fibre Channel 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. Heute, im Jahr 2006, sind 4-Gbit-Fibre-Channel-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 zu wenig 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 LTO 2 und LTO 3 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



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-Kompodium 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 kön-

nen. 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.



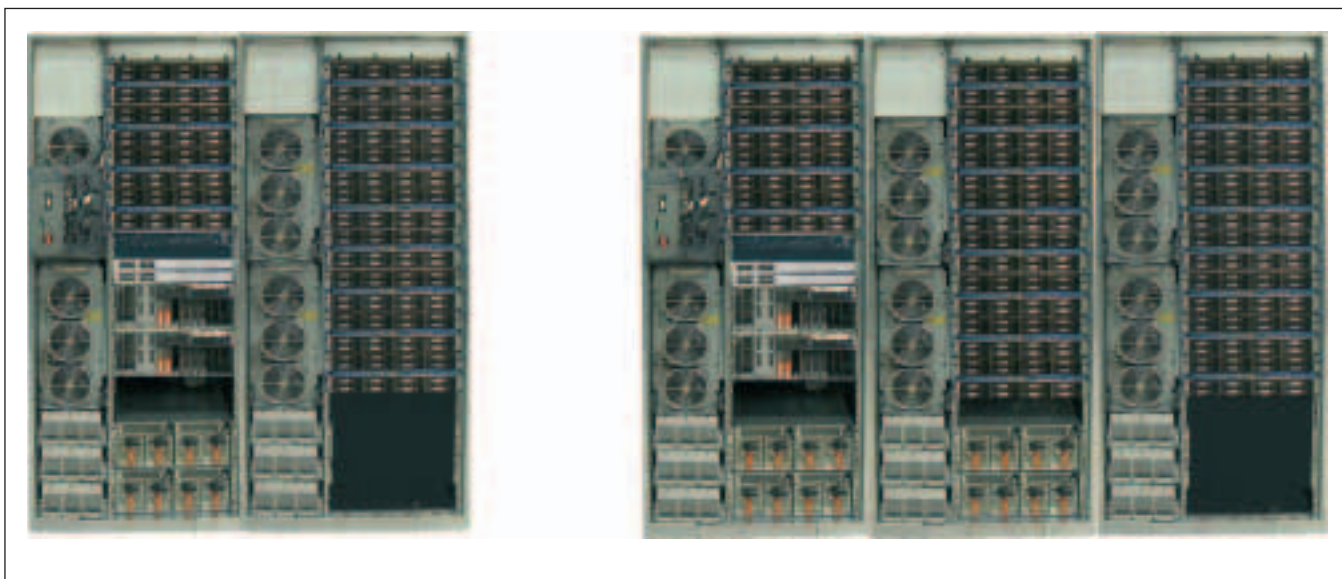
IBM DS8100 2-Wege-System

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. Das 8-Wege-System ist geplant für Anfang 2007, das 12-Wege-System wird realisiert, wenn sich eine Nachfrage nach solchen großen Systemen 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
Fibre Channel (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



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.

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.

Die Jahre 2007 und 2008 werden die Applikationsjahre für die DS8000-Architektur werden. Es ist geplant, viele Applikationsformen wie z. B. TSM, SVC, Dokumentenmanagementsysteme bis hin zu Datenbankapplikationen auf die Maschine zu bringen. Ein Umdenken wird im Markt erforderlich. Eine DS8000 ist nicht nur zum Abspeichern von Daten da, sondern kann auch als Server genutzt werden.

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 Designanforderungen entspricht das **Speichersystem DS6000**. Das System ist voll und ganz für den Maßmarkt 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 Terabyte.

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 Hilfsfunktionen vereinfachen die Reparatur. Zur einfachen Administration und schnellen Implementierung verfügt das Speichersystem über einen Express Configuration Wizard, der das erstmalige Konfigurieren oder Umkonfigurieren stark vereinfacht.

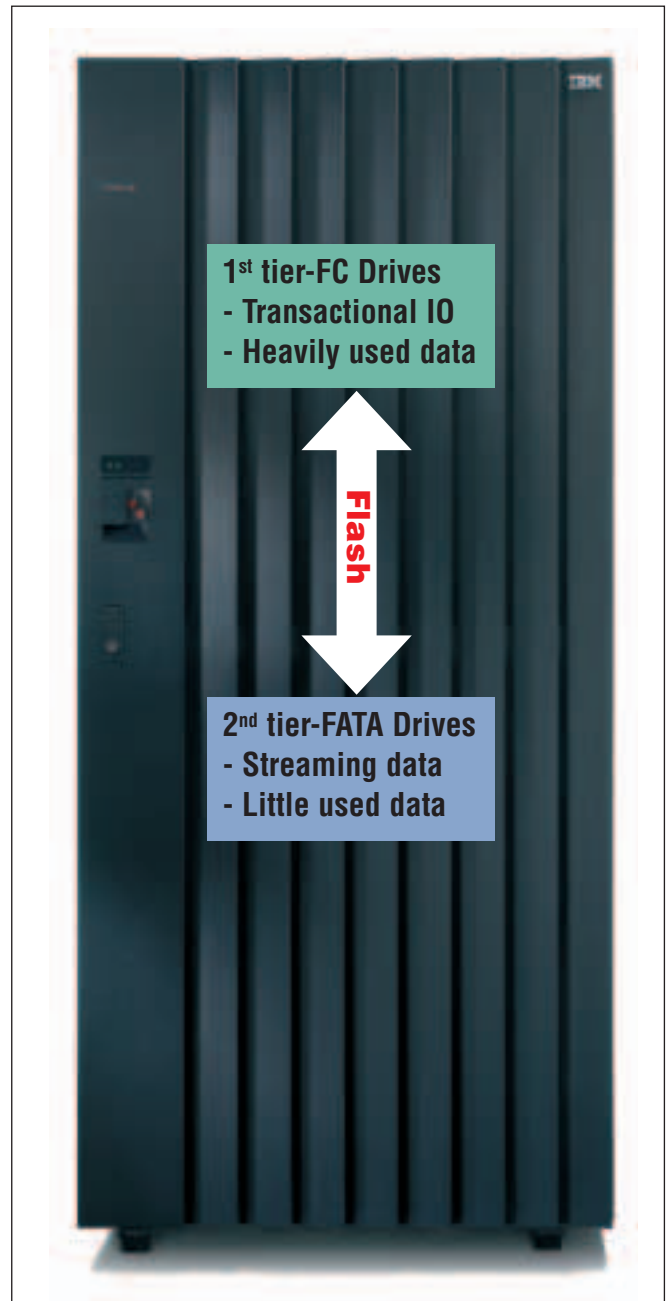
Die **DS6000 und DS8000** sind die ersten Plattensysteme von IBM, die **mit vier Jahren Gewährleistung** angekündigt wurden. Aufgrund dieser langen Gewährleistung erkennt man bereits, dass diese neue Architektur sehr lange Bestand haben wird.

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 Fibre-Channel-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-Fibre-Channel- 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.

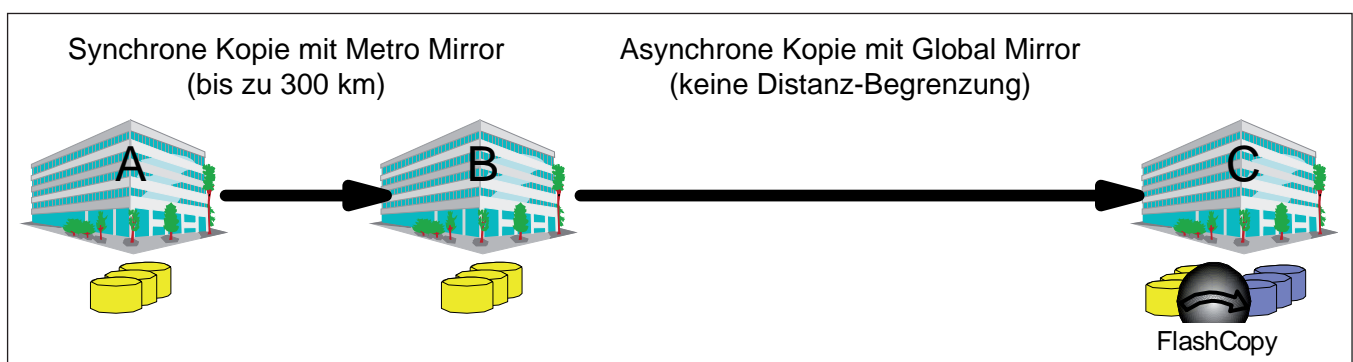
Das **dreifache Remote-Spiegelverfahren** setzt sich aus den Komponenten 'Metro Mirror' und 'Global Mirror' (ehemals PPRC XD – Peer to Peer Remote Copy eXtended Distance – und FlashCopy) zusammen. Von der Lokation A wird mit der DS8000-Turbo synchron zur Lokation B gespiegelt. Die Lokation B spiegelt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

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.



Neue Server-basierende Tape-Virtualisierung für zSeries

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 Hierarchical Storage Management) ist für die Verwaltung der logischen Volumes im Plattenpuffer zuständig. Es managt 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 Kapa-

zitäten 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

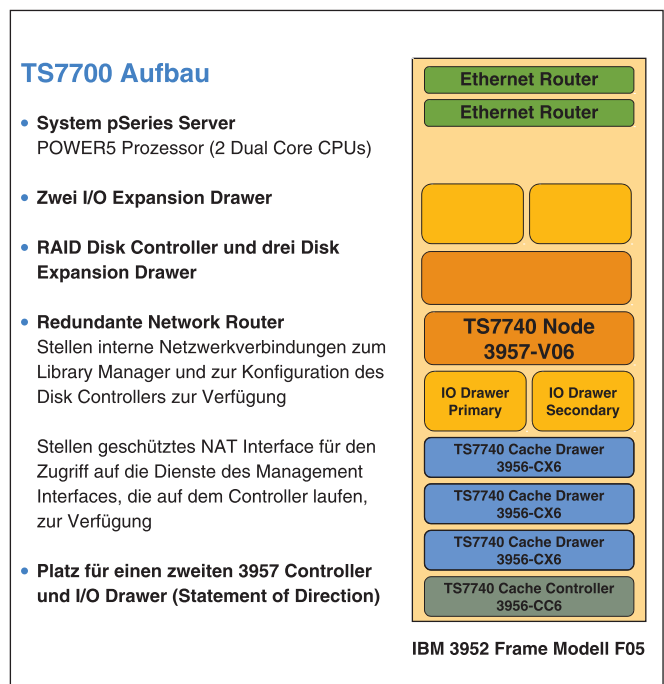
Zwei Cluster können heute in einem **Dual Cluster Grid** (oder auch Multi Cluster Grid) zusammengeschlossen werden und dienen dazu, entsprechende Disaster-Recovery-Lösungen aufzubauen. Sie sind in der Funktionsweise dem Peer to Peer VTS mit den Modellen B10 und B20 vergleichbar (synchrone und asynchrone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

Im mitangekündigten Statement of Direction wird eine zukünftige Erweiterung auf drei Cluster in einem Grid avisiert (3-Site-Spiegelung). 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 Basis-einheit zur Verfügung. Unterstützt ist die TS3500 Library (3584 Library) über die Anbindung des IBM 3953 Library Managers Modell L05. Die Einheit unterstützt im Backend TS1120 Laufwerke im Emulation-Mode der 3592 J1A. Später ist auch die 'Native'-Unterstützung vorgesehen. 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 zum Betriebssystem, zum Bandverwaltungssystem und zum Katalogeintrag. Die emulierten Laufwerke bezeichnet man als virtuelle Laufwerke. Die TS7700 unterstützt bis zu 128 virtuelle Laufwerke sowie bis zu 500000 virtuelle Volumes und dies erhöht sich später durch die Bildung von Multi Node Clusters entsprechend. In einem Dual Cluster Grid (Spiegelung) stehen 256 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 'premigriert' 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 System auf dem Markt.

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

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 unterscheidet. Der Nachfolger der TS7510 mit größeren Plattenpufferkapazitäten und einer weit höheren Leistungsfähigkeit ist in Arbeit und wird die neue Einheit durch funktionale Erweiterungen zum Tape-Virtualisierungsspezialisten im Open-Systems-Umfeld machen. Neue intelligente Algorithmen werden eine neue Basis für die Migration vom Plattenpuffer auf Tape ermöglichen, sodass Recalls wesentlich schneller durchgeführt werden können, wenn das virtuelle Volume nicht mehr im Plattenpuffer verfügbar ist (siehe auch Beschreibung TS7510).

Die Weiterentwicklung beider Systeme, TS7700 für den Mainframe und TS7510 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. Es ist geplant, im nächsten Jahr denselben Support auch für die 3494 Library zur Verfügung zu stellen.

Encryption für Tape

Ein weiterer wichtiger Teil der Tape-Ankündigung vom 29. August 2006 war die Encryption-Möglichkeit mit den neuen **TS1120 Laufwerken (Jaguar 2)**. 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 Betriebssystemplattformen 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 Verschlüsselung zu machen, 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 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 ab 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.

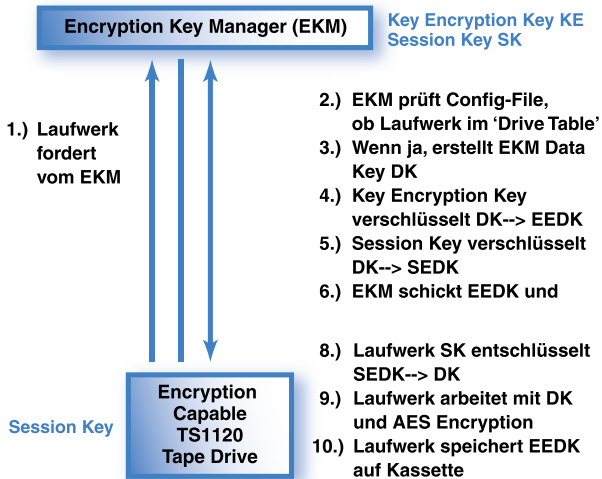
Die Verschlüsselungsmethode ist die eine Seite. Genauso wichtig sind die Datenschlüssel, die vor unautorisiertem Zugriff schützen und über die sich die verschlüsselten Daten wieder entschlüsseln lassen. Im Mainframe-Umfeld, wo schon seit Jahren Encryption mit entsprechenden Krypto-Prozessoren eingesetzt 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.

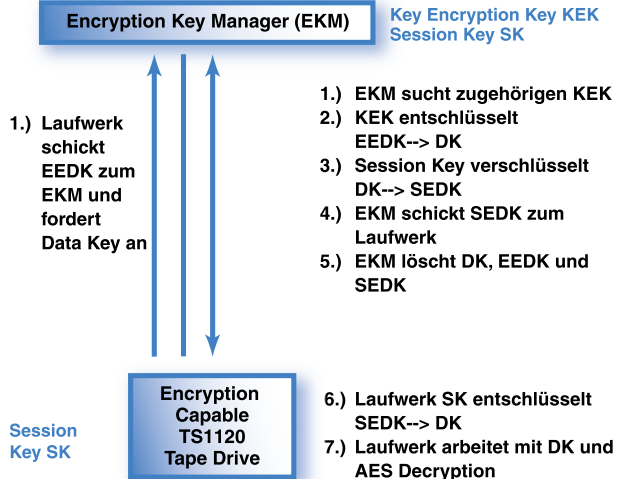
Muss nun ein TS1120 Bandlaufwerk mit Encryption aufzeichnen, fordert das Laufwerk den Datenschlüssel vom EKM Encryption Key Manager an. Der EKM prüft dann, ob das Laufwerk zugelassen ist und im 'Drive Table' des EKM bekannt ist. Falls nicht, muss das Laufwerk über die Konfigurationsfile im Drive Table 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 originalen 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.

Verschlüsseltes Schreiben auf TS1120 Medien



Auslesen der verschlüsselten Daten aus TS1120 Kassetten



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 originale 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 originalen 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 den EEDK zum EKM der originalen Lokation weiterschickt, wo der Key erzeugt wurde. Der originale EKM stellt den originalen Daten Key wieder her und schickt ihn mit SK verschlüsselt (SEDK) zum EKM der anderen Lokation zurück. Dieser gibt ihn dem Laufwerk weiter. So wird hundertprozentig gewährleistet, dass nur autorisierter Zugriff auf die Daten möglich ist.

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

Virtualisierung allgemein

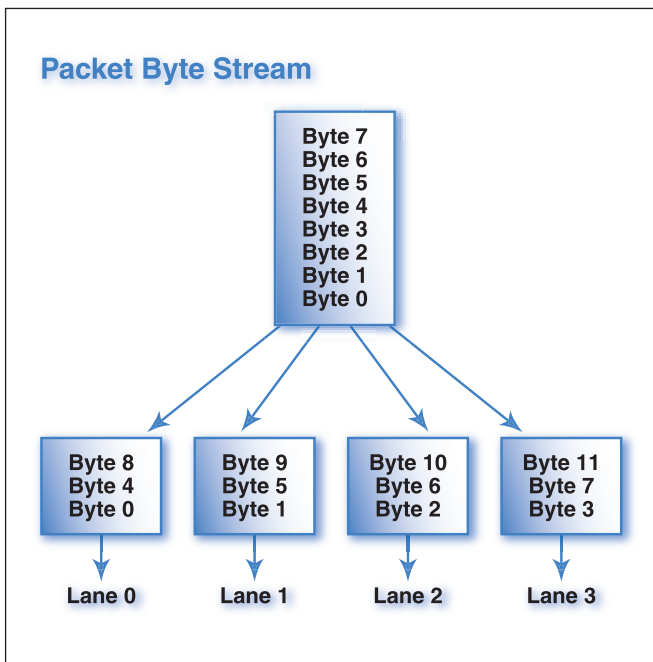
Virtualisierung in jeglicher Form wird für die Epoche der Server-basierenden Speicherarchitekturen prägend sein. Neben Tape-Virtualisierungen und Plattenvirtualisierungslösungen im SAN wird sich die Virtualisierung auf weit mehr Bereiche ausdehnen. Die **IBM** vertritt hier weltweit die **Strategie 'Virtualize 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.

Neue Infrastrukturen und Bandbreiten

Die heutigen Fibre-Channel-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 ihren Einsatz fanden, bevor sie im kommerziellen Rechenzentrums-umfeld 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 Netzwerke 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 Fibre-Channel-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 Fibre Channel 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 benötigen 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 Fibre Channel – 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 Fibre Channel kostengünstiger ist.

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 **BlueRay(blauer Laser)-Technik** lassen sich Interferenz-Felder auf dem Kunststoff 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 circa 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 bei dem Plattensystem DS8000 und der neuen Tape-Virtualisierungseinheit TS7700, bieten eine noch nie dagewesene 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.

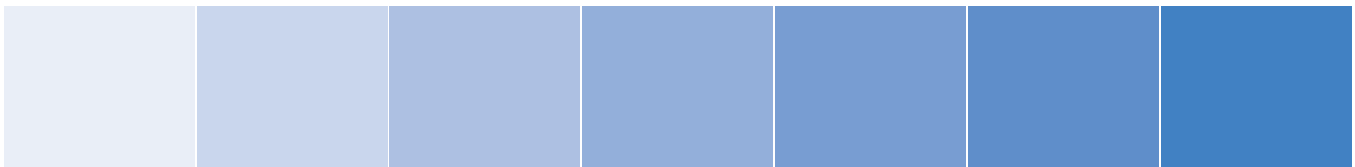
Das Tempo in der Entwicklung und in den Produktzyklen wird sich noch weiter verstärken. Es wird immer schwieriger werden, einen Gesamtüberblick über alle Möglichkeiten im Storage-Umfeld zu behalten, die man aktiv nutzen kann, und Beratung wird eine größere Bedeutung als bisher haben. Hardware allein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

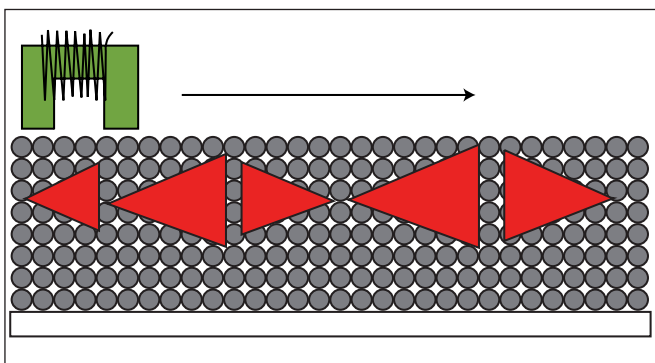
Technologie-Anhang: 'Online'-Technologien



Entwicklung der Plattentechnologie

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.



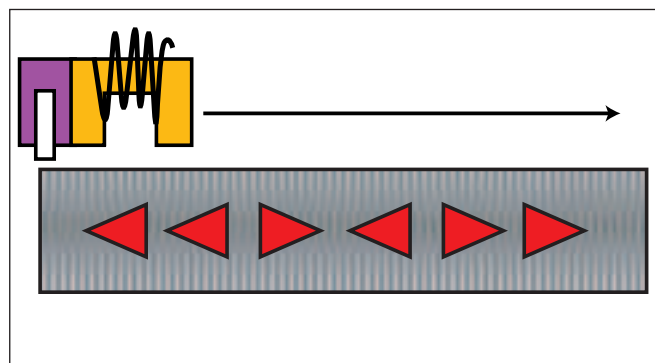
Prinzip der induktiven Aufzeichnungstechnik

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.

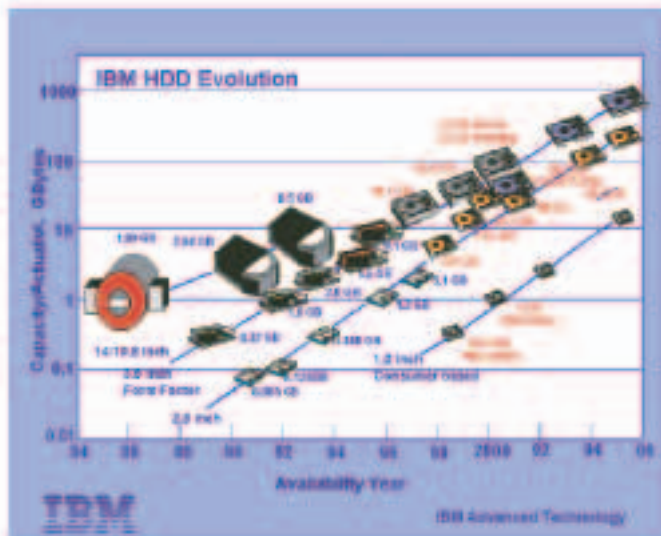
Diese braunen Platten bereiteten Anfang der 90er-Jahre (1990 – 1993) enorme Probleme, speziell bei den Modellen 1, 2 und 3 der 3390 Familie, weil sich nach langjähriger Benutzung die Eisenoxydpartikel durch die andauernden Drehkräfte aus der Kunstharzmasse herausarbeiteten und sich an den Köpfen anlagerten. 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 Geschriebene oder Gelesene zu lesen. 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 aufwendig 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 (2006). Der Unterschied zur induktiven Aufzeichnung liegt darin, dass man nicht mehr mit einem



Prinzip der magneto-resistiven Aufzeichnungstechnik



1991:
RAID-Architekturen
Dünnschichtbeschichtung
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 Laminiervorgänge und Vakuumprozesse waren inzwischen so weit, dass auf der Platte mit Dünnschichtbeschichtung (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 Dünnschichtfilm 'Grains', also Körnerstrukturen, und die einzelnen Grains stellen die Magnetisierungsträger der Platte dar. Die Herstellung solcher Dünnschichtplatten 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 Dünnschichtplatten 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 Laufwerk ausfälle in einem Array durch RAID abgesichert waren.

1995 integrierte IBM in die damals vorrangig gefertigten 3½-Zoll-Laufwerke ein neues Encoding-Verfahren, das als **PLMR Encoding** bezeichnet wird und heute noch im Einsatz ist (Partial Response Maximum Likelihood Encoding). Bei diesem Verfahren wird nicht mehr das Gesamtsignal aufgezeichnet, sondern nur noch der Bereich der Signalspitze. Grundvoraussetzung für ein solches Verfahren sind extrem gut durchmagnetisierte Grainstrukturen, weil nicht mehr alles aufgezeichnet wird. Das Gesamtsignal selbst wird direkt beim Auslesen einer Informationseinheit durch den PLMR-Encoding-Algorithmus rekonstruiert. Dadurch spart man jede Menge Datenträger ein und bekommt wesentlich höhere Kapazitäten auf die Platte. 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 3½-Zoll-Laufwerke innerhalb von 12 bis 15 Monaten verdreifachten.

1997 wurde als Weiterentwicklung der MR-Köpfe die **GMR-Technologie**, die 'Giant'-magnetoresistive Kopftechnik, eingeführt, die vor allem das Auslesen von kleinsten Streufeldern erlaubte. Der Lesekopf bekam eine Doppelfunktion und wurde vom Aufbau her neu gestaltet. Zum einen wurde wie bisher ein Streufeld durch die Widerstandsänderung des Nickel-Eisen-Sliders abgegriffen, zum anderen wurden aber auch zusätzlich die Spineffekte abgegriffen, die bei der Magnetisierung der Grains entstehen. Diese neue GMR-Technik ließ



IBM Micro-Drive

noch viel größere Miniaturisierungsmöglichkeiten zu und man beherrschte die Kopftechnik bereits 1999 so gut, dass Köpfe mit Spaltbreiten von wenigen Angström Dicke gebaut werden konnten. GMR ist bis heute im Einsatz.

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 2006, 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. Okt. 1999 konnte das IBM Labor in Almaden eine Hard-Disk-Laborversion in dieser neuen Technik demonstrieren. In einer Presseveröffentlichung am 4. Okt. 99 wurde von einer Aufzeichnungsdichte von 35.3 Gbits pro Square-Inch gesprochen – das sind **über 5 Milliarden Bits auf dem Quadratzentimeter**. Damit war die Sprengung der bis dahin mit 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

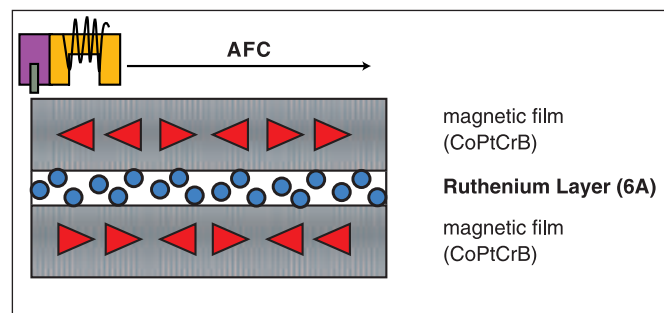
zu vergrößern, versuchte man, die Zahl der Körner zu reduzieren, die ein Bit bilden. Das hatte jedoch den gravierenden Nachteil, dass sich das Signal-Rausch-Verhältnis verschlechterte. Deshalb versuchte man, die Körner zu verkleinern und die Anzahl beizubehalten, die ein Bit bilden. Kommt allerdings 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 Vervielfachung 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.

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.



Prinzip der AFC(AntiFerromagnetically Coupled)-Aufzeichnungstechnik

Die AFC-Aufzeichnungstechnik (Pixie Dust) erlaubt es, bei **mittleren Korngrößen von 4 – 8 Nanometer** im Dünnsfilm 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 Quadratzentimeter) bei wesentlich höheren Drehgeschwindigkeiten und Datenraten zu. Wegen der wundersamen Wirkung der Ruthenium-Schicht taufen 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 2006-2007 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 Hitze-problematik 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 'Alt-technologie', 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. 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 2006, werden zwei Formfaktoren produziert: 2½-Zoll- und 3½-Zoll-Laufwerke.

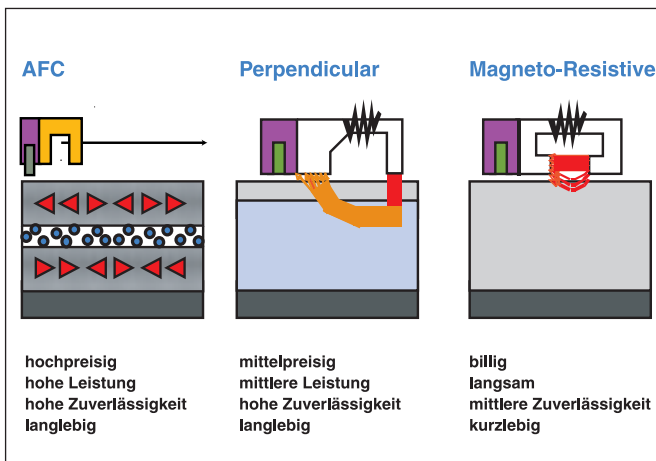
Spätestens im Jahr 2008 – da sind sich alle Plattenhersteller einig – wird generell nur noch mit einem Formfaktor gearbeitet werden, weil sich kein Hersteller bei Preisverfällen von etwa 40 % pro Jahr die Fertigung von zwei unterschiedlichen Formfaktoren leisten kann. Die 3½-Zoll-Laufwerke, die bisher vor allem in Plattensubsystemarchitekturen eingebaut sind, werden durch 2½-Zoll-Laufwerke ersetzt.

Damit werden generell nur noch 2½-Zoll-Laufwerke gebaut werden. Es kristallisiert sich heraus, dass innerhalb der 2½-Zoll-Bauweise drei unterschiedliche Techniken produziert 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.

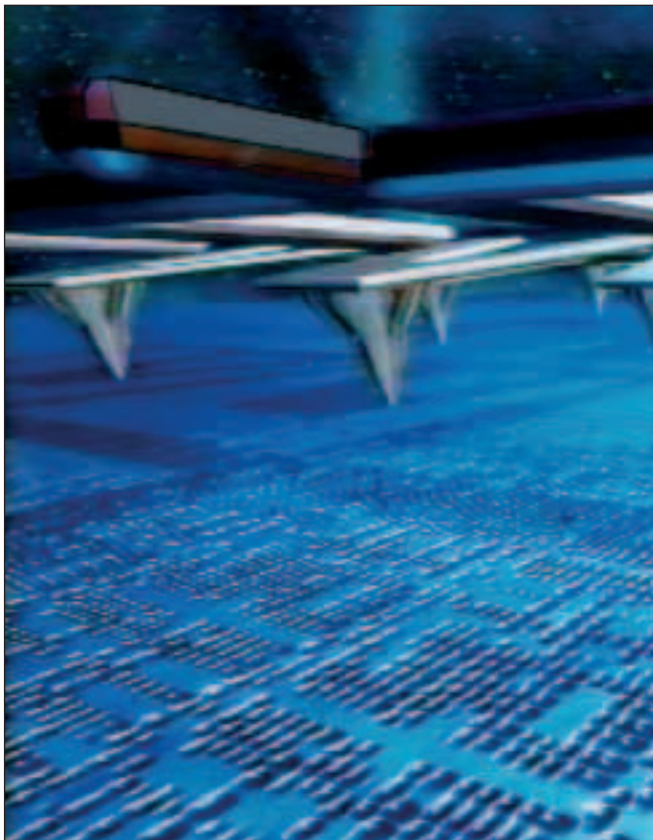
Perpendicular Recording verwendet einen speziell gebauten induktiven Schreibkopf, der ähnlich gut durchmagnetisierte Bits erzeugt wie die AFC-Technologie. Im Gegensatz zu AFC, wo über die Reflexion der Ruthenium-Schicht eine sofortige Polrückkopplung eingeleitet wird, erfolgt die Rückführung der Elektronen über einen sogenannten Softlayer, eine Legierung, die einen größeren Anteil Platin als nicht magnetisierbares Material enthält und die Induktionsstärke entsprechend abschwächt. Um die Polrückführung zu gewährleisten, ist der aufzunehmende Pol wesentlich breiter gestaltet. Der Vorteil liegt darin, dass die gut geschriebenen Bits durch die Abschwächung nicht wieder zerstört werden. Von der Kostenseite lassen sich Platten mit dem Perpendicular-Aufzeichnungsverfahren wesentlich günstiger produzieren (im Vergleich zu AFC-Platten), kommen allerdings nicht in die Leistungsbereiche von AFC-Platten, wo mit sehr hohen Datenraten und Drehgeschwindigkeiten gearbeitet wird.



Aufzeichnungsverfahren im Vergleich

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 2008 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.“

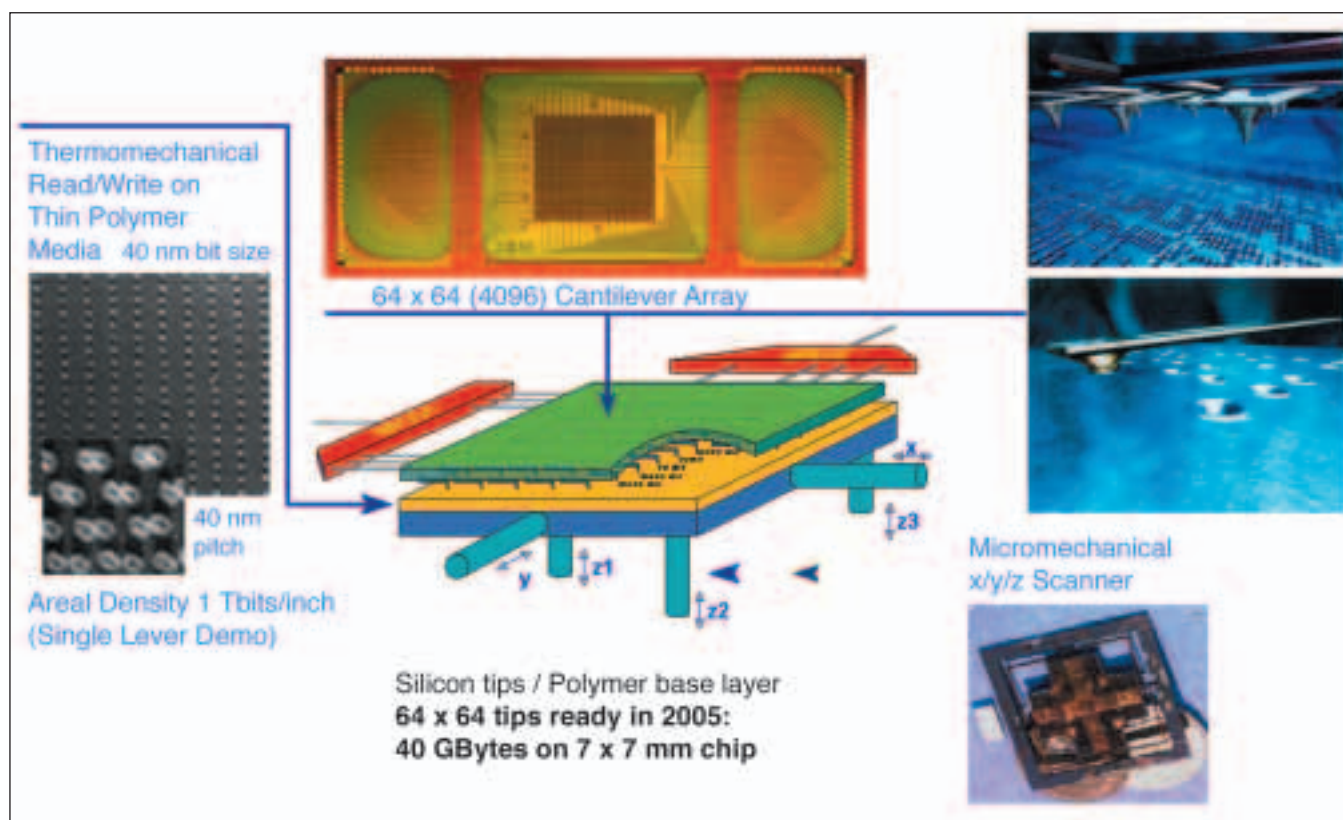
Der Speicherhunger mobiler Geräte nimmt beständig zu. Flash-Karten haben den Nachteil, dass bei ihnen die Kosten pro Speichereinheit relativ konstant sind – jedenfalls bis zum Erreichen einer bestimmten Grenze. Festplatten und ihre kleineren Verwandten – Micro-Drives, die vereinzelt bereits in Compact-Flash-Gehäusen 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 Krater

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 Millionstelmmillimeter vorzudringen. So konnte bei der Speicherung von Daten eine Aufzeichnungsdichte von 1 Terabit pro Quadrat Zoll 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 Mikroskanner 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.

Mikroskanner

Die Bewegung des Speichermediums relativ zu dem Kantilever-Array wird mithilfe des siliziumbasierten x/y-Mikroskanners realisiert. Der Scanner besteht aus ca. 6.8 x 6.8 mm² 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 und den Folgejahren 2007 und 2008 sind die ersten Millipede-Chips, die nächsten Chip-Generationen und die ersten SD-Karten mit deren Folgegenerationen zu erwarten.



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

IBM, das IBM Logo und ibm.com sind eingetragene Marken der IBM Corporation. On Demand Business und das On Demand Business Logo sind Marken der IBM Corporation in den USA und/oder anderen Ländern.

Weitere Unternehmens-, Produkt- oder Service-namen können Marken anderer Hersteller sein.

Java und alle Java-basierenden Marken und Logos sind Marken von Sun Microsystems, Inc. in den USA und/oder anderen Ländern.

Microsoft, Windows, Windows NT und das Windows-Logo sind Marken der Microsoft Corporation in den USA und/oder anderen Ländern.

Intel, Intel Inside (Logo), MMX und Pentium sind Marken der Intel Corporation in den USA und/oder anderen Ländern.

UNIX ist eine eingetragene Marke von The Open Group in den USA und anderen Ländern.

Linux ist eine Marke von Linus Torvalds in den USA und/oder anderen Ländern.

© Copyright IBM Corporation 2006
Alle Rechte vorbehalten.

