

Letzte Änderung: 5.11.2013: Überprüfen der links zu den Internet-Nachweisen und kleinere Textkorrekturen und –erweiterungen

Vorletzte Änderungen: 2.12.2012: Überprüfen der links zu den Internet-Nachweisen und kleinere Textkorrekturen und –erweiterungen; 20.10.2011: Überprüfen der links zu den Internet-Nachweisen und kleinere Textkorrekturen und –erweiterungen;

Prof. Dr. Ulrich Naumann

Unterlagen für das Fernstudium

am Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin

Postgraduales Fernstudium Bibliothekswissenschaft

Modul 3.3: Methoden der Speicher- und Vervielfältigungstechnik

1. Einleitung
2. Notwendigkeit der Speicherung
3. Grundsätzliche (technische) Formen der Speicherung von Informationen
4. Wichtige Kenndaten der Speicherung und ihre Abhängigkeiten
5. Speicherhierarchien und Speicherarray
 - Literatur zu Speichertechniken
6. Vervielfältigungstechniken

1. Einleitung

Die Aufgabe, in einer Fernstudien-Kurseinheit von zwei Stunden „Methoden der Speicher- und Vervielfältigungstechnik“ zu vermitteln, zeigt erhebliche didaktische Probleme. Schon eine nur oberflächliche Analyse des Stoffes bringt bald die Erkenntnis, dass der Stoff kaum sinnvoll so zu komprimieren ist, dass es sich nicht nur um eine Aufzählung von Speichermedien und Vervielfältigungsverfahren handelt.

Ein gutes Beispiel für die Komplexität der Materie ist das genau zum Thema passende, 1996 erschienene Buch von Horst Völz¹: Informationsspeicher: Grundlagen - Funktionen - Geräte (s. Literaturangaben). Auf über 300 Seiten mit 390 Bildern legt der Autor eine nach seinen Ausführungen extrem komprimierte Zusammenfassung vor, wobei der gesamte Text in 6-Punkt-Schrift gesetzt wurde, um auf der wohl vom Verlag vorgegebenen Seitenzahl noch mehr Stofffülle unterzubringen. Inzwischen hat Horst Völz ein umfassenderes „Handbuch zur Speicherung von Information“ veröffentlicht: Band 1: Grundlagen und Anwendung in Natur, Leben und Gesellschaft ist 2003 erschienen, Band 2: Technik und Geschichte vorelektronischer Medien 2005 und Band 3: Geschichte und

¹ Zur Person s. http://de.wikipedia.org/wiki/Horst_V%C3%B6lz [URL geprüft 5.11.2013]

Zukunft elektronischer Medien 2007².

Wie soll man das auch nur ansatzweise in eineinhalb Zeitstunden vermitteln ?

Vor allem mit der Feststellung von Horst Völz, dass der Stoff auch nicht in einer zweistündigen Vorlesung in einem Semester zu schaffen ist und deswegen eine Auswahl zu treffen ist, ist das didaktische Problem für ein Fernstudien-Modul nicht gelöst, sondern erweitert:

- **Mengenproblem:** Stoff ist insgesamt nicht in 1,5 Stunden zu vermitteln
- **Inhaltsproblem:** die Materie, wie ein Vergleich zwischen den entsprechenden Kapiteln in den Büchern von Sebastian Dworatschek (Grundlagen der Datenverarbeitung, 8. Aufl. 1989, Online-Reprint 2011) und z. B. von Andreas Gadatsch: Grundkurs Geschäftsprozess-Management: Methoden und Werkzeuge für die IT-Praxis: Eine Einführung für Studenten und Praktiker, 6. Aufl. 2010) und die eigene und die Erfahrungen der Kursteilnehmerinnen und Kursteilnehmer zeigen, entwickelt sich so rasant, dass Einzelheiten über Kapazität, Zugriffszeit usw. einzelner Medien zu lehren und dann zu erwarten, dass dies später auch als Wissen anwendbar ist, eigentlich nutzlos vertane Zeit ist.

Dazu gibt es drittens das

- **Problem der didaktischen Reduktion:** die Darstellung des Stoffes gegenüber einem Zuhörerkreis, der in der Regel nicht aus Mathematikern, Physikern oder Informatikern und solchen, die es werden wollen, besteht, muss auf einem einsichtigen und verständlichen Niveau vorgenommen werden. Hierbei ist es „hilfreich“, dass auch der Dozent selbst nur „Anwender“ der Speichertechniken ist, so dass eine zu tiefgehende fachliche Darstellung vermieden wird.

Deshalb wird folgender **didaktischer Ansatz** gewählt:

Da die Kursteilnehmer neben dieser schriftlichen Unterlage die empfohlene Literatur selbst lesen können / sollen, erfolgt eine Beschränkung auf das Grundsätzliche und (beim Vortragen des Stoffes) die Demonstration von Objekten.

Was ist als **Grundsätzliches**, jenseits aller technischen Entwicklungen **Bleibendes** darstellbar?

- die Notwendigkeit der Speicherung
- Grundsätzliche (technische) Formen der Speicherung (hier auch mit Hardware-Demonstrationen)
- Wichtige Kenndaten der Speicherung und ihre Abhängigkeiten
- Speicherhierarchien und Speicherarray

Danach wird auch der folgende Stoff zu den Speichertechniken gegliedert.

Ungleich leichter ist das zweite Thema der Lehreinheit, die Vervielfältigungstechniken, darstellbar. Hier wird nach einer begrifflichen Abgrenzung auf die einzelnen Verfahren einzugehen sein, wobei

² S. <http://r-h-voelz.de> [URL geprüft 5.11.2013]. - Die drei Bände (einschließlich seines Buches „Wissen - Erkennen - Information: Allgemeine Grundlagen für Naturwissenschaft, Technik und Medizin. Shaker Verlag, Aachen 2001) gibt es als **CD** mit immerhin fast 6.000 Seiten Volltext unter dem Titel "Wissen – Erkennen – Information: Datenspeicher von der Steinzeit bis ins 21. Jahrhundert". (Digitale Bibliothek. Band 159, Berlin 2007).

auch hier betont werden muss, dass durch die zunehmende Digitalisierung auch in diesem Technikbereich ein „abgegrenztes Wissen“ nicht vermittelt werden kann. Aus Zeitgründen werde ich mich allerdings darauf beschränken, dieses Thema nur im Skript zum Modul 3.3 abzuhandeln.

2. Notwendigkeit der Speicherung (Speichergruppen und Speichermöglichkeiten)

Grundlage der folgenden Darstellung sind Überlegungen von Horst **Völz**.

Speichern ist ein Ur-Geschehen. Ohne Speichern gäbe es nicht die Welt, so wie sie ist. Ohne Speichern gäbe es uns nicht. Speichern ist ein fundamentales Bedürfnis des Lebens und ein Grundelement der gesamten Welt. Die gesamte Evolution ist ein Prozess unterschiedlichen Speicherns von Veränderungen und ihrer Wiedergabe.

Wenn wir heute etwas über die Vergangenheit wissen, so ist hierfür Gespeichertes unabdingbare Voraussetzung. Speichern hebt immer etwas von der Gegenwart für die Zukunft auf und wirkt so der Flüchtigkeit des Zeitablaufs und der damit zusammenhängenden Vergänglichkeit entgegen.

Dabei unterscheiden wir grundsätzlich drei große **Speichergruppen**:

- **Speicher für Stoffe** als Behältnisse, Räume oder Gebäude
- **Speicher für Energie** (Brennstoffe, Batterien, Akkumulatoren oder Kondensatoren)
- **Speicher für Information** (genauer für die stofflich-energetischen Informationsträger)

Nur die letzte Gruppe, die Speicher für stofflich-energetische Informationsträger, sind in unserem Zusammenhang von Interesse.³

Die **Speichermöglichkeiten**, die innerhalb der drei genannten Speichergruppen zum Einsatz kommen, sind prinzipiell als evolutionärer Prozess zu interpretieren:

- **physikalisch-chemische** Speicherung (Nachvollzug des „Urknalls“ anhand von Relikten; ein Relikt ist auch die Erde selbst)
- **genetische** Speicherung (genetischer Code der Lebewesen in den Chromosomen, jeweils vollständiger Satz in allen Zellen, eine gewaltige Verschwendung angesichts der Spezialisierung einzelner Zellen, zugleich aber auch ein Zeichen für die Wichtigkeit, die die Natur der Speicherung zumisst)
- **egotrope**[selbstbezogene] Speicherung (Unterscheidung von Ich und Fremd, Nahrung und Feind)
- **neuronal**e Speicherung (Entstehen des Gehirns, da die Mobilität der sich entwickelnden Tierwelt die Erweiterung der festen Speicherung des Codes in allen Zellen (-> genetische Speicherung) in Richtung individuelles Lernen erfordert.)
- **kollektive** Speicherung (staatenbildende Insekten, Mensch: Die Menschheit weiß mehr als der Einzelne)

³ S. auch den sehr informativen Überblick <http://de.wikipedia.org/wiki/Datenspeicher> [Letzter Zugriff: 5.11.2013]

- **technische** Speicher (hat nur der Mensch entwickelt, um sein mühsam erworbenes Wissen zu erhalten, in folgender Evolutionsfolge der Informationsträger:

Werkzeug als Wissensspeicher des Gebrauchs

Höhlenbilder als Wissensspeicher für Handlungsabläufe

Schrift (zunächst auf Steinen und Tontafeln)

Papyrus

Gutenbergs Druck mit beweglichen metallischen Lettern

Photographie

Film

Schallplatte

elektronische Speicher

Bionik⁴

???

Ein **Kennzeichen** dieses evolutionären Prozesses der Entwicklung von stofflich-energetischen Speichern sind **immer kürzere Abstände** zwischen den Evolutionsschritten.

3. Grundsätzliche (technische) Formen der Speicherung von Daten / Informationen

Zentrale Grundlage für das sinnvolle und zweckgerichtete Speichern in jeglicher Form sind drei Funktionalitäten, die durch die Formen der Speicherung realisiert werden müssen:

- der **Aufzeichnungsvorgang** (aus Daten oder Signalen wird durch Einwirken von Energie der Speicherzustand erzeugt, das Trägermaterial verändert)
- der **Speicherzustand** (Festhalten der Veränderung des Trägermaterials , also des Aufzeichnungsvorgangs in möglichst unveränderter und dauerhaft stabiler Form)
- der **Wiedergabevorgang** (aus den Speicherzuständen werden die ursprünglichen Daten/Signale möglichst originalgetreu wiedergewonnen)

Aufzeichnen, Bewahren und **Wiedergeben** sind also Kernfunktionen jedes Speicherns.

Einen weiteren Aspekt stellt noch die [inhaltlich unveränderte] **Übertragung** des Gespeicherten als Teil des Kommunikationsprozesses dar, auf den wir aber nicht eingehen werden.

Wie diese drei Kernfunktionen realisiert werden können, kann man wiederum aus drei verschiedenen Blickwinkeln betrachten, bei denen die für diese Funktionalitäten eingesetzten Träger der Spei-

⁴ S. <http://de.wikipedia.org/wiki/Bionik> [URL geprüft 5.11.2013]

cherung, die Form der Speicherung und die Methoden des Wiederfindens des Gespeicherten im Mittelpunkt stehen:

- das chemisch-physikalische Prinzip (Träger der Speicherung)
- das Speicherungsverfahren (Form der Speicherung)
- das Suchverfahren (Wiederfinden des Gespeicherten)

Wir wollen uns diese drei Sachverhalte näher anschauen.

Zum chemisch-physikalischen Prinzip

Bei Dworatschek (Grundlagen der Datenverarbeitung, 8. Aufl. 1989) findet sich eine von mir erweiterte tabellarische Übersicht, die nach dem chemisch-physikalischen Prinzip die einzelnen Speicherungsverfahren gliedert:

chemisch-physikalisches Prinzip	Speicherverfahren
mechanisch	Papierblatt, Katalogkarte gedrucktes Medium Lochstreifen Lochkarte
elektro-mechanisch	Relais, Beispiele: ZUSE-Rechner Z 3 (1941) ⁵ , Aiken-Rechner MARK II (1943) ⁶
magnetisch ⁷ „Die magnetische Speicherung von Informationen erfolgt auf magnetisierbarem Material. Dieses kann auf Bänder, Karten, Papier oder Platten aufgebracht werden. Magnetische Medien werden (außer Kernspeicher) mittels eines Lese-Schreib-Kopfes gelesen bzw. beschrieben. Wir unterscheiden hier zwischen rotierenden Platten(stapeln), die mittels eines beweglichen Kopfes gelesen bzw. beschrieben werden und nicht rotierenden Medien, die üblicherweise an einem feststehenden Kopf zum Lesen bzw. Beschreiben vorbeigeführt werden. Ein weiteres Unterscheidungsmerkmal ist, ob auf dem Medium Daten üblicherweise analog, digital oder in beiden Formen gespeichert werden.“ (Wikipedia)	<ul style="list-style-type: none"> • magneto-elektronisch <ul style="list-style-type: none"> ○ digitale Medien <ul style="list-style-type: none"> ▪ Kernspeicher • nicht rotierende Speichermedien <ul style="list-style-type: none"> ○ digitale Medien <ul style="list-style-type: none"> ▪ Magnetband, (z. B. DLT); DAT und Tonband (digital) ▪ Magnetkarte ▪ Magnetstreifen ▪ Compact Cassette (Datasette) ▪ Magnetblasenspeicher ○ analoge Medien <ul style="list-style-type: none"> ▪ Tonband (Musikkassette) ▪ Videoband (Videokassette) • rotierende Speichermedien <ul style="list-style-type: none"> ○ digitale Medien <ul style="list-style-type: none"> ▪ Trommelspeicher ▪ Festplatte (hard disk) ▪ Diskette (floppy disk) ▪ Wechsellplatte z. B. Zip-Diskette (von iomega)

⁵ http://de.wikipedia.org/wiki/Zuse_Z3 [URL geprüft 5.11.2013]

⁶ <http://www-groups.dcs.st-andrews.ac.uk/~history/Mathematicians/Aiken.html> [URL geprüft 5.11.2013]

⁷ http://de.wikipedia.org/wiki/Datenspeicher#Magnetische_Speicherung [URL geprüft 5.11.2013]

elektronisch	<p><u>Halbleiter</u>⁸ Nur-Lesespeicher (nicht-flüchtig)</p> <p>ROM [Read Only Memory, herstellerprogrammiert] PROM [Programmable ROM, durch Anwender einmal programmierbar] EPROM⁹ [Erasable Programmable ROM, durch Anwender mehrfach lösch- und programmierbar] EAPROM¹⁰ [Electrically Alterable Programmable ROM, durch Anwender mehrfach lösch- und programmierbar]</p> <p>Schreib-/Lesespeicher (RAM) [Random Access Memory]</p> <p>SRAM [Static RAM ; laufend Ruhestrom erforderlich, flüchtig] DRAM [Dynamic RAM ; zyklisches Erneuern durch refreshing, flüchtig] MRAM [nicht flüchtig, kann gleichzeitig als Arbeits- und Massenspeicher auf dem gleichen Modul eingesetzt werden, was wiederum einfachere Architekturen, geringere Abmessungen und letztlich die Produktion preiswerterer Geräte ermöglicht]. F(e)RAM [nichtflüchtiger und gleichzeitig schneller Speicherchip-Typ auf der Basis von ferroelektrischen Materialien] FLASH¹¹ [Beschreibbar, nicht-flüchtig]</p>
<u>optisch</u> ¹²	<p><u>Laserprinzip:</u> CD <u>nur Wiedergabe</u> (durch Nutzer):</p> <p>CD-DA [Digital Audio-CD] CD-ROM [Read Only Memory] CD-ROM/XA [Extended Architecture, hohe Speicherdichte, max. 19 Std. Audio] CD-ROM mixed mode [1 Track Daten, Rest kann Audio sein] CD-I [Interactive, z.B. mit Filmsequenzen]</p> <p><u>beschreibbar</u> (durch Nutzer):</p> <p>CD-R [Recordable, einmal beschreibbar in CD-Brenner, <u>WORM</u>] Photo-CD [Entwicklung von Kodak] Video CD [Komprimiertes Video bis zu 74 Min.] CD-RW [Rewritable, bis zu 1000 x beschreibbar, ähnlich CD-R] CD Extra¹³ [Multisession von Audio- und komprimierten Bilddaten]</p> <p>DVD¹⁴ [Digital Versatile Disk, mit hoher Datendichte] DVD-ROM [entspricht CD-ROM]</p>

⁸ <http://www.uni-potsdam.de/u/physik/didaktik/projekt/halbleiter/index.html> [URL geprüft 5.11.2013]

⁹ <http://www.webopedia.com/TERM/E/EPROM.html> [URL geprüft 5.11.2013]

¹⁰ <http://www.webopedia.com/TERM/E/EEPROM.html> [URL geprüft 5.11.2013]

¹¹ <http://de.wikipedia.org/wiki/Flash-Speicher> [URL geprüft 5.11.2013]

¹² http://de.wikipedia.org/wiki/Optischer_Speicher [URL geprüft 5.11.2013]

¹³ <http://www.mediabit.de/lexikon/cd-extra.html> [URL geprüft 5.11.2013]

¹⁴ <http://de.wikipedia.org/wiki/DVD> [URL geprüft 5.11.2013]

	DVD-R [entspricht CD-R] DVD-RAM [ähnlich CD-RW] HV-DVD [blauer Laser] Blue-Ray-Disk ¹⁵ [blauer Laser] Transphasor ¹⁶ [„optischer Transistor“]
magneto-optisch	<u>Aufzeichnungsvorgang</u> : (elektromagnetischer) Curie-Effekt <u>Wiedergabevorgang</u> : (optischer) Kerr-Effekt ¹⁷
bio-chemisch/ -elektrisch organisch	<u>Biotechnik-Speicher</u> ¹⁸ <u>Protein-Speicher</u> ¹⁹ Wasser (?) ²⁰

An der Vielzahl der Verfahren und der sich daraus ergebenden Speichermedien wird die Stofffülle erkennbar, die sich bei dem Versuch einer auch nur halbwegs in jedem Sinn (also auch technologisch) erschöpfenden Darstellung der chemisch-physikalischen Prinzipien der Speicherung ergäbe.

Zum Speicherungsverfahren (Form der Speicherung) und zum Suchverfahren (Wiederfinden des Gespeicherten) (Datenorganisation²¹)

Speicherungsformen²²

Das chemisch-physikalische Prinzip des Speichers bedingt auch die Form des Speicherverfahrens und des Zugriffs auf das Gespeicherte. Zum Beispiel steht auf dem (physikalisch-chemischen) Träger gedrucktes Buch oder einem Magnetband nur die Speicherungsform „sequentielles Speichern“ zur Verfügung. Daher sollen am Beispiel des (physikalisch-chemischen) Trägers Magnetplatte bestimmte realisierbare Speicherungsformen erklärt werden.

- **Sequentielle Speicherung** (Sätze lückenlos hintereinander)
 - Sätze sind in lückenloser Reihenfolge hintereinander auf dem Datenträger abgelegt
 - Es wird kein Zusammenhang (funktionale Vorschrift) zwischen Ordnungsbegriff (Schlüssel) und physischer Satzadresse (Beginn-Adresse des Satzes) gebildet, daher ist kein frei wählbarer Zugriff möglich

¹⁵ http://de.wikipedia.org/wiki/Blu-ray_Disc [URL geprüft 5.11.2013]

¹⁶ <http://staff.science.uva.nl/~robbe/OC/node6.html> [URL geprüft 5.11.2013]

¹⁷ http://www.physik.uni-kl.de/w_fp/htmltexte/Moke.html [URL geprüft 5.11.2013]

¹⁸ http://www.at-mix.de/biologische_speicher.htm [URL geprüft 5.11.2013]

¹⁹ <http://de.wikipedia.org/wiki/Bacteriorhodopsin> [URL geprüft 5.11.2013]; für weitergehende Information s. die drei sehr informativen Überblicksartikel von Hermann Strass 2003 über zukünftige Speichertechnologien in der Online-Zeitschrift tecCHANNEL (<http://www.tecchannel.de/storage/grundlagen/402141/>; <http://www.tecchannel.de/storage/grundlagen/402142/>; <http://www.tecchannel.de/storage/grundlagen/402168/>) [URL geprüft 5.11.2013]

²⁰ S. z. B. <http://www.deam.de/allergie/wasser004.htm> [URL geprüft 5.11.2013]

²¹ Zum Begriff s. <http://www.wirtschaftslexikon24.net/d/datenorganisation/datenorganisation.htm> [URL geprüft 5.11.2013]

²² S. Stahlknecht, Peter; Hasenkamp, Ulrich: Einführung in die Wirtschaftsinformatik, 11. Aufl. – Berlin u.a.: Springer, 2005, S. 150-158.

- Bei der Suche eines Satzes muss jeder Satz angesprungen und ein Schlüsselwert-Vergleich zwischen momentanem Satzschlüssel und gesuchtem Schlüssel durchgeführt werden
- Beim Einfügen von Sätzen muss durch Aufbau einer Zwischendatei (scratch file), in die die neuen Sätze an der vorgesehenen Stelle eingefügt werden, ein neuer Datenträger erzeugt werden
- **Index-sequentielle Speicherung** (Index = Schlüsselbegriff + Speicheradresse; Indextabellen werden als separate Indexdateien mitabgespeichert) (dies entspricht der sequentiellen Speicherung, nur merkt man sich jetzt die Speicherstelle)

Zusätzlich zur Datendatei wird eine Indexdatei mit einer Indextabelle abgespeichert. In der **Indextabelle** wird zu jedem Primärschlüsselwert die zugehörige physische Satzadresse vermerkt:

- Die Speicherplätze der einzelnen Sätze müssen nicht zusammenhängend sein
- Bei der Suche nach einem Datensatz wird nicht im originären Datenbestand gesucht, sondern der Index wird in den Hauptspeicher geladen und durchsucht.
- Bei unsortierter Reihenfolge der Schlüsselwerte ist im Durchschnitt die halbe Index-Datei zu durchsuchen
- Sortierung nach den Primärschlüsselwerten (Nummer, die ein Datenobjekt eindeutig identifiziert) ermöglicht binäres Suchen
- **Index-verkettete Speicherung (gekettete Speicherung) (chaining)** (Index = Schlüssel + Speicheradresse + Adresse nachfolgender Satz (simple chain) und vorhergehender (double chain) Satz)

Vorteil: Nur Änderung der Verweise in der Indextabelle;

Nachteil: Zeitaufwendigeres Suchen;

In der Indextabelle werden bei diesem Speicherverfahren zusätzlich zu jedem Schlüsselwert vom nachfolgenden / vorhergehenden Satz

entweder der Schlüsselwert oder die Position in der Indextabelle oder die physische Speicheradresse als drittes Feld (Spalte) festgehalten.

Vorteile:

- Verteilung des Datenbestands über den gesamten Speicherbereich
- Neu hinzukommende Datensätze werden mit dem Verweis auf den Nachfolgersatz jeweils am Ende der Indextabelle angefügt, d.h. keine Sortierung und entsprechende Einordnung erforderlich)
- Einfache Änderung beim Einfügen und Löschen

Nachteil: hoher Aufwand für den Änderungsdienst

Neben diesen einfacheren Basis-Speicherformen gibt es weitere, auf mathematischen Prinzipien

beruhende Speicherverfahren wie etwa die **Index-Speicherung durch binäre Bäume** (B-Bäume, hierarchische Graphen vom Grad 2^{23}), die auch als gestreute oder gekettete Formen realisierbar sind. Diese Speicherformen sind vor allem von Bedeutung, wenn das schnelle Suchen in großen Datenmengen optimiert werden soll.

Der Anwender, sofern nicht selbst Mathematiker oder sonst Fachkundiger, wird bei den komplizierteren Speicherformen nicht mehr ohne weiteres darstellen können, in welcher Speicherungslogik die Speicherung vollzogen wird.

Er wird aber eine für seine Arbeit bedeutsame Feststellung machen können: Ob die gewählte Speicherform für seinen Zweck geeignet ist, und diese Eignung wird vornehmlich durch die folgenden Parameter bestimmt:

- die Benutzungshäufigkeit der Daten
- den Umfang des Datenbestandes
- die Beziehung zwischen Datenzugriff und Datenveränderung

Allerdings ist bei der Entscheidung für eine Speicherungsform zu bedenken, dass die Speicherform vom für notwendig erachteten Komfort beim Dateizugriff abhängt und dieser sich wiederum nach den Erfordernissen der Verarbeitung der Daten richtet. Die Speicherorganisation wird aufgrund dieser im Entscheidungszeitpunkt gesehenen Erfordernisse konzipiert, und ist nur in seltenen Fällen ohne großen Aufwand auf geänderte Bedürfnisse anzupassen.

Wer also glaubt, mit der einfachen Form der sequentiellen Speicherung auszukommen und erst bei einer Million gespeicherter Datensätze zu komfortableren index-orientierten Verfahren übergehen will, weil ihm nun die Suche in den Daten zu lange dauert, wird nicht umhin kommen, eine Million Indexsätze neu anzulegen (oder hoffentlich seine Daten so strukturiert erfasst haben, dass ein maschineller Indexaufbau möglich ist.)

Wiederfinden des Gespeicherten

Von der Form der Datenspeicherung hängt dann auch die Möglichkeit ab, wie auf die gespeicherten Daten wieder zugegriffen werden kann:

- **seriell** (starr fortlaufend) in der Reihenfolge, wie die Daten für die Speicherung angefallen sind (das kann am Beispiel dieses Textes verdeutlicht werden)
- **sequentiell** (fortlaufend gemäß einer festgelegten Reihenfolge) (das kann am Beispiel eines Lexikons verdeutlicht werden, wobei das Alphabet in mehreren Stufen die Reihenfolge festlegt)
- **direkt** (wahlfrei, Random Access) über die Nutzung des Index (das kann an einem Register eines Buches verdeutlicht werden)

Es ist von der zu erfüllenden Aufgabe abhängig, welche Form der Speicherung gewählt wird, weil jede „luxuriösere“ Form arbeitsaufwendiger und teurer ist.

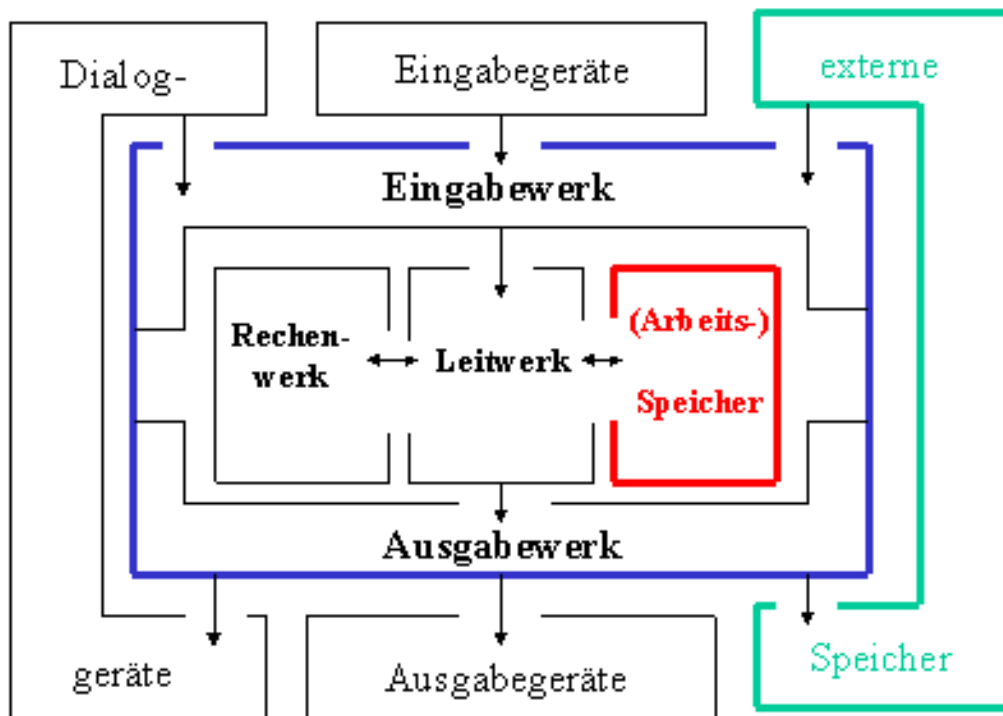
²³ S. http://de.wikipedia.org/wiki/Bin%C3%A4rer_Suchbaum [URL geprüft: 5.11.2013]

4. Wichtige Kenndaten der Speicherung und ihre Abhängigkeiten

Daher wollen wir uns im Folgenden mit wichtigen Kenndaten der Speicherung und den dabei bestehenden Abhängigkeiten beschäftigen, um ein Grundverständnis für die anzustellenden Wirtschaftlichkeitsbetrachtungen bei der Speicherung zu entwickeln.

Zunächst aber einige Hinweise zum **grundsätzlichen Aufbau von Rechenanlagen** und der dort vorhandenen Speicher.

Für den grundsätzlichen Aufbau einer Datenverarbeitungsanlage (Rechner) ist die prinzipielle Darstellung des Von-Neumann-Rechners, den John von Neumann²⁴ bereits kurz nach dem 2. Weltkrieg beschrieben hat, ein beliebtes Erklärungsmodell. Auch wenn der Von-Neumann-Rechner inzwischen für die „modernen“ EDV-Fachleute ein unzureichendes Erklärungsmodell darstellt (für Hochinteressierte dazu das Buch von Giloi²⁵), kann er für uns in diesem Zusammenhang als Erklärungsmodell genügen.



Funktionseinheiten eines Von-Neumann-Rechners

Rechner

Nach DIN-Definition ist ein **Rechner** eine Einheit zur Verarbeitung von Daten, nämlich zur Durchführung mathematischer, umformender, übertragender und speichernder Operationen. Als Synonyme werden in der DIN-Norm die Benennungen Rechensystem, Rechenanlage, Datenverarbeitungssystem und Datenverarbeitungsanlage sowie Computer und Data Processing System

²⁴ http://www-groups.dcs.st-andrews.ac.uk/~history/Mathematicians/Von_Neumann.html [URL geprüft 5.11.2013]

²⁵ S. Giloi, Wolfgang K.: *Rechnerarchitektur*. Springer, 1993.

genannt.

Damit die Einheit „Rechner“ funktionieren kann, bedarf sie bestimmter Funktionseinheiten.

Funktionseinheiten eines Rechners

Ein Rechner besteht aus Funktionseinheiten

- zur Aufnahme von Daten von außen (**Eingabewerk**),

Eingabewerk

Ein Eingabewerk ist eine Funktionseinheit eines Rechners, durch die Daten von aussen aufgenommen werden können. Die Funktionseinheit „Eingabewerk“ verbindet den Rechner mit verschiedenen Eingabeeinheiten wie

Tastatur, Maus, Lichtgriffel, Digitalisieretablett, Pen-based Computer, Schriftenleser, Scanner, *Speicher*

- zur Verarbeitung dieser Daten, die also interpretiert, verändert und aufbewahrt werden können (**Prozessor**, der Leitwerk, Rechenwerk und den internen (Arbeits-) Speicher steuert)

Prozessor

Ein **Prozessor** ist eine Funktionseinheit, die Leitwerk, Rechenwerk und Verbindungskomponenten, z.B. zum *Arbeitsspeicher*, umfasst.

- zur Abgabe der verarbeiteten Information nach außen (**Ausgabewerk**).

Ausgabewerk

Ein Ausgabewerk ist eine Funktionseinheit eines Rechners, durch die die verarbeiteten Daten wieder nach außen abgegeben werden können. Auch die Funktionseinheit „Ausgabewerk“ ist das Bindeglied zwischen verschiedenen Ausgabegeräten wie

Monitor, Kathodenstrahlröhre, Flüssigkristallbildschirm, Plasmabildschirm, Elektrolumineszenzbildschirm, Elektronenstrahlbildschirm, Drucker, Zeichendrucker, Zeilendrucker, Seitendrucker, Plotter, *Speicher*

Hier kann das bekannte „EVA“ (= Eingabe, Verarbeitung, Ausgabe) als Eselsbrücke dienen. Jeder der genannten Funktionseinheiten können in der Realität eine oder mehrere Baueinheiten entsprechen.

Wir sehen, dass in allen drei Bestandteilen eines Rechners nach dem Von-Neumann-Modell funktional Speichereinheiten eingebunden sein können. Daher wollen wir die Form dieser Speicher noch näher betrachten.

Speicher

Ein **Speicher** ist eine Funktionseinheit eines Rechners, in den Daten eingebracht werden können, der sie dann aufnimmt, aufbewahrt und abgibt.

Dabei ist zu unterscheiden in den internen und die externen Speicher.

- **Interner Speicher (Zentralspeicher)**

Ein Zentralspeicher ist ein Speicher innerhalb der Zentraleinheit. Typische Merkmale sind der unmittelbare Zugang durch den Prozessor, die vorübergehende Speicherung und der schnelle Zugriff. Weitere gängige Benennungen für Zentralspeicher sind: Speicherwerk und **interner Speicher**. Der größte interne Speicher ist der **Arbeitsspeicher**, der auch vielfach Hauptspeicher genannt wird. Der Arbeitsspeicher und die anderen internen Speicher arbeiten mit einem ausserordentlich schnellen Zugriff, der noch durch das Zwischenschalten eines Cache-Speichers ²⁶ beschleunigt werden kann. Das Fassungsvermögen dieser Speicher ist jedoch aus technischen und Kostengründen begrenzt. Sie dienen während der Programmausführung zur Speicherung von Programmcodes und Daten und werden im Allgemeinen nicht für eine dauerhafte Aufbewahrung von Daten herangezogen. Diese Funktion übernehmen die externen Speicher.

- **Externer Speicher**

Jeder Speicher, der nicht Zentralspeicher ist, wird als **externer** beziehungsweise **peripherer Speicher** bezeichnet. Externe Speicher sind langsamer (geringere Zugriffsgeschwindigkeit), aber dafür billiger als Zentralspeicher, und sie verfügen über sehr große Speicherkapazitäten. Nicht unmittelbar benötigte Daten und Programme, die aus Platzgründen nicht ständig im Zentralspeicher stehen, werden extern gespeichert und können bei Bedarf mit großer Geschwindigkeit in den internen Speicher übertragen werden. Wegen ihrer hohen Kapazität bezeichnet man sie auch als **Massenspeicher** (Magnetische Datenträger, Optische Datenträger, Elektronische Datenträger)

Periphere Einheit (Peripheriegerät)

Eine Funktionseinheit eines Rechners, die nicht zur Zentraleinheit gehört, wird **periphere Einheit** genannt. Dementsprechend werden externe Speicher auch als periphere Speicher bezeichnet. Der Verkehr mit den peripheren Geräten zur Eingabe (input) und Ausgabe (output) von Programmen und Daten wird in Rechnern meist durch selbstständige Funktionseinheiten gesteuert, die als Ein-Ausgabe-Prozessoren bezeichnet werden.

Intelligente Peripherie

Durch eingebaute Prozessorchips (und Speicherchips) werden die **Peripheriegeräte „intelligent“**. Das heißt im IT-Jargon, dass sie über Eigensteuerungsvermögen verfügen und damit im Betrieb nicht auf die ständige Fremdsteuerung durch eine Zentraleinheit angewiesen sind. Bei Bedarf werden sie vom Prozessor der Zentraleinheit „angestoßen“ und wickeln dann die gerätespezifischen Aufgaben weitgehend selbstständig ab. Die Zentraleinheit wird damit wesentlich entlastet, und die Parallelarbeit erhöht die Gesamtleistung des Rechners.

Konfiguration

Das Zusammenschalten von mindestens einer Zentraleinheit mit den an diese angeschlossenen peripheren Geräten wird **Konfiguration** genannt.

²⁶ <http://www.elektronik-kompodium.de/sites/com/0309291.htm> [URL geprüft 5.11.2013]

Von diesen **prinzipiellen Teilen eines so modellierten Von-Neumann-Rechners** wollen wir uns nur mit den Speichern beschäftigen. Für diese Speicher sind unter dem Gesichtspunkt des wirtschaftlichen Einsatzes bei der Investitionsentscheidung verschiedene Kenndaten zu berücksichtigen:

- die Zugriffszeit
- die Speicherkapazität
- der Datenfluss
- die Permanenz
- der Preis

Zugriffszeit

Als **Zugriffszeit** oder **Zugriffsgeschwindigkeit** bezeichnet man die Zeit für den Zugriff auf eine definierte Position des Speichermediums mit dem Ziel, Daten zu schreiben oder zu lesen. Sie hängt wesentlich von den chemisch-physikalischen Prinzipien der Speicher und der Organisation des Signalzugriffs ab.

Generell kann man sagen, dass der Zugriff auf interne Speicher wesentlich schneller erfolgt als auf externe Speicher.

Völz nennt für die verschiedenen Zugriffstechniken folgende Werte:

- mentaler Zugriff: Gedächtnis 0,1 sec, Suchen in Gedrucktem: Minuten und Stunden
- mechanisch bewegte Bauteile (Magnetband, Diskette, Festplatte): 10^{-1} Millisek.
- Halbleiter-Speicher: untere Grenze bei mehreren Nanosekunden (10^{-9} sec)
- optische oder supraleitende Speicher: Werte bis in den Picosekunden-Bereich (10^{-12} sec)

Die bisher bekannte Obergrenze in der Zugriffsgeschwindigkeit ist die **Lichtgeschwindigkeit**: 300.000 km/sec. Die Angabe **km/sec** macht zugleich deutlich, dass die Geschwindigkeit eines Speichers auch von der Länge des Weges abhängt, der innerhalb des Speichers zurückgelegt werden muss, um die eingespeicherten Daten wiederzufinden. Je kompakter solche Speicher gebaut werden können, desto schneller ist der Speicherinhalt aufzufinden und wieder auszugeben.

Speicherkapazität

Horst Völz weist darauf hin, dass seit Beginn der Rechentechnik, also etwa seit 70 Jahren, die Speicherkapazität um das 10^9 , also um das Milliardenfache, gestiegen ist. Dennoch, so Völz, gab es und gibt es zu jedem Zeitpunkt drei Aussagen:

1. Die gerade verfügbare Speicherkapazität ist zu klein.
2. Die Speicher müssten um das 10-fache größer sein, dann werden sie dauerhaft ausreichen. (Wenn das 10-fache erreicht ist, gilt wieder Aussage 1.)

3. Die Speicher sind groß genug, man geht nur nicht ökonomisch genug damit um.

Diese ironische Feststellung beschreibt das Problem der Speicherkapazität in der Wahrnehmung der Nutzer. Fest steht jedoch: wenn viel Speicher da ist, wird er auch gefüllt. Deutliches Beispiel sind die heutigen Software-Angebote, die ein Speicher-Volumen erfordern, das vor einigen Jahren nicht leistbar war.

Ein Grund ist auch darin zu sehen, dass die Software-Programmierer häufig mit der neuesten Technologie arbeiten und damit immer sorgloser mit den Ressourcen umgehen, wodurch der „Kunde“ ständig zum Nach- und Aufrüsten gezwungen wird. Unter diesem Gesichtspunkt sollte man sich einmal die Programme betrachten, die etwa von einer Software-CD auf den eigenen Rechner „gedankenlos“ automatisch heruntergeladen werden. Vorausgesetzt, man kann die Bedeutung einzelner Programme interpretieren, wird dabei schnell klar werden, wie wenig man davon für den eigenen Bedarf braucht. Wer braucht 450 Druckertreiber, die ständig auf der Festplatte präsent sind, wenn man nur über ein Modell verfügt?

Hinzu kommt, dass nach vorsichtigen Schätzungen in immer kürzeren Zeiträumen immer mehr Daten produziert werden. Matthias Jehn weist 2007 darauf hin, dass „in den nächsten fünf Jahren (...) voraussichtlich mehr Daten produziert [werden], als in der gesamten Geschichte der Menschheit gesammelt wurden“.²⁷

Der Datenfluss

Der Datenfluss in Bytes/sec besitzt besonders bei großen Dateien Bedeutung. Hier kommt es wesentlich darauf an, wie schnell z. B. von einem externen Speichermedium die Daten an den Arbeitsspeicher übertragen werden können. Sind hierbei die Kanäle zu gering ausgelegt, kommt es zu einem bottle-neck-Effekt. Daher ist man inzwischen bei 32-Bit- bzw. auch 64-Bit-Kanälen angelangt, mit denen 4 bzw. 8 Zeichen (8 bit = 1 Byte) gleichzeitig auf der Leitung übertragen werden können. 64-Bit-Kanäle können auch gekoppelt werden, um den Datenstrom von 128 Bit zu ermöglichen, was für umfangreiche graphische Anwendungen angezeigt ist.

Um den Datenfluss zwischen Prozessor und Speicher zu organisieren, werden Schnittstellen eingesetzt. Hierbei ist die **IDE-Schnittstelle** (IDE bedeutet Integrated Device Electronics oder auch Intelligent Drive Electronics, später als „ATA“ ANSI Standard genormt) weit verbreitet. Es handelt sich dabei um eine **Festplattenschnittstelle**, bei der die Steuerungselektronik bzw. der Controller in das Festplattengehäuse integriert ist. Sie ist im Gegensatz zur SCSI-Schnittstelle billiger und etwas langsamer. Sie bringt aber trotzdem eine beachtliche Datenübertragungsrate (10 MBits/sec) zustande. Für leistungsfähigere Anwendungen (z. B. auf Servern) wird die **SCSI-Schnittstelle** (SCSI = Small Computer Systems Interface) eingesetzt, an die auch externe Geräte (wie etwa Massenspeicher) angeschlossen werden können. An Wide SCSI Schnittstellen können bis zu 15 Geräte, an normalen SCSI und Ultra-SCSI 7 Geräte angeschlossen werden. Die verschiedenen Geräte besitzen eigene Identifikationsnummern (Adressen) und können somit vom Controller unterschieden werden. Die höhere Performance im Gegensatz zu IDE wird durch zwei Funktionen der SCSI-Schnittstelle erreicht: Bei IDE-Geräten wird das Datenbus²⁸ solange blockiert, bis eine Anforderung eines Gerätes bedient wurde. Der SCSI-Controller kann dagegen das Bus freigeben und in der Zwi-

²⁷ S. Jehn, Matthias: Das digitale Gedächtnis der Informationsgesellschaft: Überblick über Fragen und Strategien der Langzeitarchivierung. In: BuB, 59 (2007), S. 534-537, hier S. 534.

²⁸ Ein Bus (*Binary Unit System*) ist ein System zur Datenübertragung zwischen mehreren Teilnehmern über einen gemeinsamen Übertragungsweg, bei dem die Teilnehmer nicht an der Datenübertragung zwischen anderen Teilnehmern beteiligt sind. [http://de.wikipedia.org/wiki/Bus_\(Datenverarbeitung\)](http://de.wikipedia.org/wiki/Bus_(Datenverarbeitung)) [URL geprüft 5.11.2013]

schenzeit von anderen Geräten nutzen lassen. Der andere Vorteil des SCSI ist das Command Queueing. Damit werden Zugriffe, die über die Festplatte verteilt sind, so optimiert, daß die Schreib-/Leseköpfe weniger Bewegungen ausführen müssen. In der größten Ausbaustufe (Ultra 320) sind dabei Datenübertragungsgeschwindigkeiten von 320 MBits/sec. erreichbar. SCSI wird in neuen Servern zunehmend von schnelleren und kostengünstigeren [Serial ATA \(SATA\)-Schnittstellen](#)²⁹ abgelöst.

Die Permanenz

Für jeden Speichertyp ist die Frage zu stellen, ob zur Speichererhaltung eine dauernde, eine zyklische oder gar keine Energiezufuhr erforderlich ist und ob durch etwaigen Energieausfall, in der Regel also Stromausfall, die Speicherinhalte zerstört werden.

Die überwiegende Zahl der hier behandelten Speicher sind als permanente Speicher definiert (z. B. die Magnetspeicher oder die CD). Halbleiterspeicher sind temporäre Speicher (SRAM und DRAM) oder nicht-flüchtige Speicher (ROM., Flash).

Aber auch diese Aussage ist nicht allgemeingültig. So können sich Magnetspeicher im Laufe der Zeit verändern, weil sich die Magnetisierungsrichtung der Informationsträger durch die permanente Einwirkung des [Erdmagnetismus](#)³⁰ ändert. Von den CDs wird immer noch behauptet, dass sie nur 10 Jahre halten, Optimisten sprechen aber von 30 bis 50 Jahren. Und wie lange hält inzwischen der Informationsträger GUTENBERG-BIBEL?

Manuel Masiero kommt in dem sehr lesenswerten Report „Nur Steintafeln halten ewig: Digitale Amnesie (PC Professionell, 5/2007³¹) zur Feststellung, dass allen digitalen Speichermedien trotz aller Versprechen um ihre Haltbarkeit nur eines gemeinsam ist: bei keinem lässt sich hundertprozentig sicherstellen, dass es sich noch nach mehr als fünf Jahren wieder auslesen lässt.

Eine weitere Problematik liegt in der technischen Zugreifmöglichkeit auf permanente Speicher. Wer kann heute noch einen Stapel mit Lochkarten verarbeiten? In einer Ausgabe der Zeitschrift Chip wurde 1998 vom amerikanischen Wetterdienst berichtet, dass dort Milliarden von Daten aus den fünfziger Jahren auf Magnetbändern gespeichert sind, damals angelegt, um langfristige Klima- und Wetteranalysen zu ermöglichen. Heute ist man technisch nicht mehr in der Lage, diese Magnetbänder zu lesen. Und wer schon länger im EDV-Geschäft ist und wichtige Erkenntnisse auf 5 1/4 Zoll-Disketten gespeichert hat, ist arm dran, wenn er die Daten direkt von diesem Speichermedium lesen will. Welcher Rechner hat noch ein 5 1/4-Zoll-Disketten-Laufwerk? Und wer hat in seiner Software noch ein lauffähiges WORD3-Textverarbeitungs-Programm, mit dem diese Daten erzeugt wurden?

Die digitale Speicherung bietet deshalb schon mittelfristig keine Sicherheit, einmal gespeicherte Daten auch zu jeder Zeit wieder nutzen zu können. Vielmehr muss man die Speicher laufend mit den noch und schon vorhandenen technischen Möglichkeiten abgleichen, im Sinne einer sequentiellen Permanenzsicherung. Hier entsteht zum Beispiel für die Deutsche Nationalbibliothek als Archivbibliothek auch des elektronisch Publizierten ein massives Problem der Langzeitarchivierung³² unter dem Gesichtspunkt der jederzeitigen Wiedergabemöglichkeit. Und muss es nicht nachdenklich stimmen, dass die britische Atomenergiebehörde beschlossen hat, Daten über abgebaute For-

²⁹ <http://www.itwissen.info/definition/lexikon/serial-ATA-SATA-SATA-Schnittstelle.html> [URL geprüft 5.11.2013]

³⁰ <http://www.raumfahrer.net/astronomie/planeterde/erdmagnet.shtml> [URL geprüft 5.11.2013]

³¹ <http://www.itespresso.de/2007/06/25/nur-steintafeln-halten-ewig-pcpro/> [URL geprüft 5.11.2013]

³² <http://www.langzeitarchivierung.de/> [URL geprüft 5.11.2013]. – S. auch Jehn, Matthias: Das digitale Gedächtnis der Informationsgesellschaft: Überblick über Fragen und Strategien der Langzeitarchivierung. In: BuB, 59 (2007), S. 534-537.

schungsreaktoren, deren strahlende Überreste noch in Jahrtausenden eine Gefahr für die Menschheit sein werden, auf papyrusartiges Trägermaterial zu kopieren und so der Nachwelt zu erhalten?³³

Für denjenigen, der Dokumente für das Internet aufbereitet, entsteht das Problem, in welcher „Beschreibungssprache“ er seine Dokumente abfassen soll. Hier ist abzusehen, dass, kaum erlernt, HTML als Dokumentbeschreibungssprache³⁴ bald ausgedient haben wird. Aber das ist ein weites Feld, das in anderen Lerneinheiten des Fernstudiums behandelt wird. In unserem Zusammenhang, dem Speichern, soll hier nur wieder einmal auf die Problematik des elektronischen Speicherns und des inhaltsgetreuen Wiedergebens hingewiesen werden.

Der Preis

Als letztes Kriterium wurde der Preis genannt, und zwar bezogen auf die Kosten pro gespeichertem Bit. Unabhängig von der allgemeinen Trendaussage, dass elektronische Speicher immer billiger werden, führt uns dies unmittelbar zum folgenden Punkt, der sich mit Speicherhierarchien beschäftigt, die vor allem aus Kostengründen entstehen, während die ebenfalls in diesem Zusammenhang behandelten Speicherarrays sowohl Kosten- als auch Sicherheitsgründe miteinander zu verbinden versuchen.

5. Speicherhierarchien und Speicherarray

5.1 Speicherhierarchien

Der ideale Speicher sollte besitzen ³⁵

- nahezu unbegrenzte Speicherkapazität
- kurze Zugriffszeit bei wahlfreiem Zugriff
- hohe Zugriffsraten
- geringe Speicherkosten
- Nichtflüchtigkeit

Da diese Bedingung „**größtmögliche Speicherkapazität im direkten Zugriff bei geringsten Kosten**“ als „Minimax-Prinzip“ nicht erreichbar ist, ist das Prinzip der Speicherhierarchie entwickelt worden. Ich will mich hier auf den Grundgedanken beschränken.

Die Speicherhierarchie versucht eine Annäherung an den oben skizzierten idealen Speicher durch die Kombination realer Speichermedien unter Nutzung von Eigenschaften der jeweiligen Speicherformen zu erreichen

- eine Pufferung / Bereitstellung von Daten mit hoher Zugriffswahrscheinlichkeit in schnellen (relativ kleinen) Speichern
- die Mehrheit der Daten verbleibt auf langsameren, kostengünstigeren Speichern

³³ S. Rötzer, Florian: Papier anstatt digitale Archivierung. Online verfügbar unter der URL <http://www.heise.de/tp/r4/artikel/20/20696/1.html> [URL geprüft 5.11.2013]

³⁴ <http://www.swe.uni-linz.ac.at/teaching/lva/ws99-00/seminar/WebClients.pdf> [URL geprüft 5.11.2013]

³⁵ Vgl. <http://www.hardwaregrundlagen.de/oben08.htm> [URL geprüft 5.11.2013]

Auf jeder Ebene der Speicherhierarchie stellen sich ähnliche Verwaltungsaufgaben:

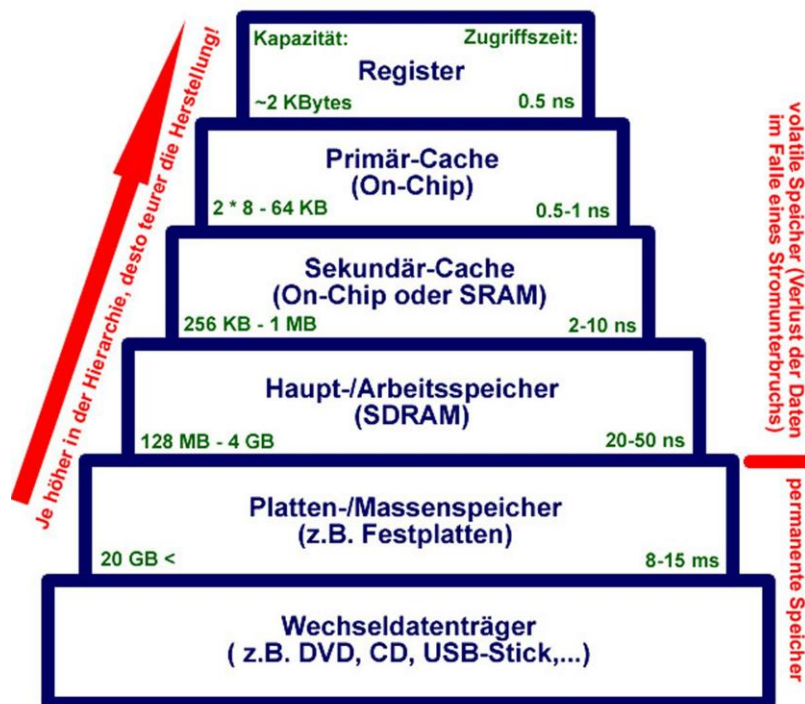
- Lokalisieren von Datenobjekten
- Bereitstellen von Speicherplatz
- Ersetzung von Inhalten
- Schreibstrategie (Write-through (direktes) vs. Write-back (gepuffertes) Schreiben)

Um diese Verwaltungsaufgaben optimal zu lösen, wird das Prinzip der Zwischenspeicherung angewendet:

Kleinere, schnellere und teurere **Cache-Speicher**³⁶ werden benutzt, um Daten und Programme für einen laufenden Verarbeitungsprozess zwischenzuspeichern, die sich sonst in größeren, langsameren und billigeren Speichern befinden

Die Auswahl des jeweiligen Speichers als dauerhafter Aufbewahrungsort für die Daten richtet sich nach dem Wirtschaftlichkeitsprinzip:

Ein schneller Speicher ist teuer und deshalb relativ klein;
Ein Speicher hoher Kapazität ist typischerweise billiger und langsamer



Eigene Darstellung, Grafikinhalte übernommen aus dem Buch:
"Mikrocontroller und Mikroprozessoren" von Birkschulte/Ungerer, S. 352
Zugriffszeit/Kapazität: Stand 2002

© by Andreas B. G. Baumann (2009)

37

Das **Ziel einer Speicherhierarchie** ist die Geschwindigkeitsanpassung des jeweils größeren und langsameren Speichers an den schnelleren und kleineren Speicher zu vertretbaren Kosten (Kosteneffektivität).

³⁶ <http://www.elektronik-kompodium.de/sites/com/0309291.htm> [URL geprüft 5.11.2013]

³⁷ Abbildung entnommen aus: <http://n.ethz.ch/~abaumann/uebung02.html> [URL geprüft 5.11.2013]

5.2 Speicherarray

Neben den Überlegungen zum Aufbau einer Speicherhierarchie, die innerhalb einer Rechnerkonfiguration nach dem Von-Neumann-Modell (mit externen Speichern) zum Tragen kommen, soll zum Schluss der Speicherbetrachtungen, auch im Zusammenhang mit der Kostenproblematik, noch ein Begriff angesprochen werden, der bei Investitionsentscheidungen zum Aufbau großer Massenspeicher eine Rolle spielt: die Frage nämlich, ob ein großer oder viele kleine Speicher eingesetzt werden sollen. Diese Frage ist vor allem sinnvoll, wenn es um häufigen parallelen Zugriff auf einen Datenbestand geht, was etwa bei dem großen Katalogdatenbestand einer Bibliothek der Fall ist.

Stellt man mehrere oder viele kleine Speicher nebeneinander, wird ein „array“ gebildet, so dass sich hierfür der Begriff „RAID“ eingebürgert hat (RAID = **R**edundant **A**rray of **I**nexpensive **D**isks, ab 1995: **R**edundant **A**rray of **I**ndependent **D**isks).

Redundante Datenhaltung ist eines der Hauptmotive für RAID. Hier ahmt der Mensch die milliardenfache Speicherung der Erbinformationen in jeder einzelnen Zelle nach. Durch die dynamische Bereitstellung der Redundanz ergibt sich jedoch zunächst eine Leistungseinbuße, vor allem beim Überschreiben/Ergänzen/Ändern von bereits gespeicherten Daten. Das ist das am intensivsten bekämpfte Problem der RAID-Technik.

Zur RAID-Technologie gibt es im Internet mehrere gute Einführungen. Ich benutze hier Texte und Abbildungen aus einer Darstellung der Fa. EDV-Systemberatung Klopff (Wissenswertes über RAID), die online leider nicht mehr verfügbar ist.

Eine weitere, sehr übersichtliche Darstellung zur RAID-Technologie findet sich in dem Aufsatz von Jörg Luther, Christian Vilsbeck und Bernhard Haluschak (2001): Raid im Überblick ³⁸.

Grund für die Aktivitäten rund um RAID sind die bereits oben erwähnten exponentiell steigenden Datenmengen. Im längerfristigen Durchschnitt rechnet die Industrie derzeit jährlich mit etwa 50 Prozent Zuwachs an Speicherbedarf für Festplatten.

Wer größere Datenmengen speichern muss, wird eher früher als später mehr als ein Plattenlaufwerk benötigen. (Das gilt allerdings weniger für reine PC-Anwendungen, wo die Speicherkapazität der internen Festplatten inzwischen hohe Gigabyte-Werte erreicht hat. Inzwischen werden Speicherkapazitäten auch als billige externe Festplattenspeicher (2000 GB = 2 Terabyte für ca. 80 €) angeboten.) Aus statistischen Gründen steigt in diesem Fall die Fehlerwahrscheinlichkeit sehr schnell, abhängig von der Qualität des einzelnen Laufwerks. Die Qualität der Einzelkomponenten anzuheben ist entweder nicht machbar oder unerschwinglich teuer. Also müssen andere Maßnahmen greifen, wenn eine bestimmte Qualität 'mission critical' ist. Den Ausweg bieten die Methoden Fehlererkennung und Fehlerkorrektur. Diese Methoden sind zum Teil seit Jahrzehnten bekannt; manche werden aber erst jetzt angewendet, weil der rechentechnische Aufwand dazu bisher zu hoch oder die daraus resultierenden Leistungseinbußen unverträglich groß waren. „Heilende“ Korrekturmethode sind verständlicherweise aufwendiger als „diagnostische“ Fehlererkennungsmethoden.

Viele billige statt weniger teurer Platten

Zur Zeit der ‚Erfindung‘ von RAID (1987) gab es noch ein zusätzliches Problem. Damals gab es nur extrem teure, aber qualitativ gute Laufwerke für die hohen Kapazitäten. Die gerade neu eingeführten ‚kleinen‘ Laufwerke mit einem Formatfaktor von 5,25 Zoll waren noch nicht so zuverlässig,

38

http://www.tecchannel.de/storage/extra/401665/raid_sicherheit_level_server_storage_performance_festplatten_controll_er/ [URL geprüft 5.11.2013].

aber in Relation zur Kapazität sehr viel billiger als die in Rechenzentren gebräuchlichen ‚großen‘ Platten. Der Kerngedanke von RAID war, durch die Koppelung einer Anzahl preisgünstiger (‚inexpensive‘) Laufwerke die benötigte Kapazität zu schaffen und durch geeignete Maßnahmen (beispielsweise Spiegelung oder Fehlerkorrektur) für das entstandene Disk-Array die gleiche Zuverlässigkeit wie bei einem großen Plattenlaufwerk zu erreichen.

Braucht man aus Kapazitätsgründen in jedem Fall mehrere Platten, wird der Betrieb sehr viel einfacher, wenn der Administrator nicht jede Platte einzeln verwalten muss (Namen vergeben, Dateien gezielt speichern und wiederfinden). Werden die Platten zu einem RAID-System zusammengefasst, dann sehen Betriebssystem und Anwender nur noch ein großes logisches Laufwerk. Diese Vereinfachung in der Handhabung war von Anfang an ein erklärtes Ziel der RAID-„Erfinder“.

Parität als Methode zur Fehlererkennung

Um die vorgenannten Ziele zu erreichen, mussten die entsprechenden Verfahren geschaffen werden. Ein Grundgedanke war die Zusammenfassung von Laufwerken zu einer Matrix (Array) mit einem oder später zwei Laufwerken zur Speicherung der Redundanz, um mögliche Fehler korrigieren zu können. Eine der einfachsten Fehlererkennungsmethoden ist die Parität (manchmal Paritätsprüfung genannt). Sie wird bei alten Magnetpurbändern oder bei der Übertragung von Bits auf einer RS-232-Leitung (V.24) angewendet. Diese Methode vermag einzelne Bitfehler zu erkennen, aber nicht zu korrigieren.

Bei allen RAID-Verfahren mit Fehlerkorrektur wird daher die logische XOR-Operation (eXclusive OR) angewendet. XOR kann man auch als Parität ohne das Hinzufügen eines Paritätsbits verstehen. Wenn man die Daten vorübergehend in Form einer Matrix anordnet (zum Beispiel Bits nebeneinander, Bytes untereinander), dann kann man Parität oder XOR horizontal und vertikal bilden. Tritt nun ein einzelner Bitfehler auf, dann muss man nur am Kreuzungspunkt von Reihe und Spalte mit dem fehlerhaften Bit dieses invertieren (von 0 nach 1 und umgekehrt, eine andere Möglichkeit gibt es ja nicht), um diesen Bitfehler innerhalb der Matrix zu korrigieren. Diese Methode wird bei den RAID-Anordnungen mit Fehlerkorrektur angewendet. Man spart sich jedoch den doppelten Satz XOR-Werte und ersetzt die daraus resultierende Fehlerinformation durch die Fehlermeldung des Laufwerks. Erscheint eine Laufwerksfehlermeldung, dann wird mit den Daten der fehlerfreien Laufwerke der Fehler korrigiert.

Wird ein zu bearbeitender Datenstrom in gleichbleibend kleine Teilstücke zerlegt, die ‚nebeneinander‘ auf die verfügbaren Laufwerke verteilt werden, dann ist Lesen oder Schreiben schneller, weil ja eine Art von Parallelverarbeitung stattfindet. Als Bremse wirkt jedoch der unvermeidliche Zeitverlust, bedingt durch Errechnen und Speichern der XOR-Werte.

RAID-Level

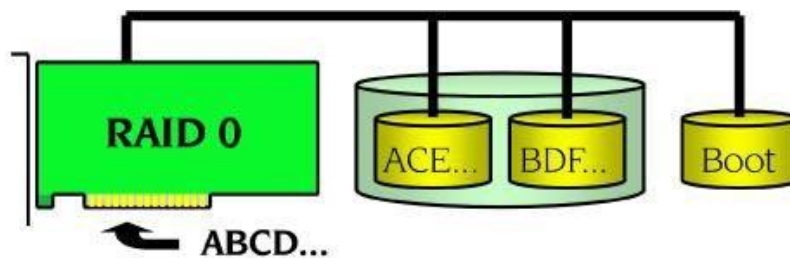
Bei der Namensfindung für die diversen RAID-Konfigurationen haben sich die „Erfinder“ (drei Professoren der Universität Berkeley, Ca.) unglücklicherweise für den Begriff „Level“ mit einer Zahl zwischen 1 und 5 entschieden. Dabei sind die Level nur Nummern zur Unterscheidung der Verfahren, weil man keine einprägsamen Kurzbezeichnungen für jedes einzelne Verfahren fand. Die Zahlen drücken keinerlei Wertung oder Rangfolge aus, die RAID-Level-Zahl ist also kein Qualitätsmerkmal. Zu einem späteren Zeitpunkt kamen zu den Leveln 1 bis 5 noch Level 0 (strenggenommen kein RAID, weil die Redundanz fehlt) sowie 6 und 7 (seit Frühjahr 1995) hinzu.

Zu den Leveln, die im Folgenden noch detaillierter behandelt werden, ist generell anzumerken: RAID 0 implementiert kein Sicherheitsverfahren. Fällt eine Platte aus, dann sind alle Informationen, die sich in Kombination mit den Daten auf den nicht defekten Platten zusammensetzen, verloren. Bei den RAID-Varianten 1, 2, 3, 4, 5 und 7 darf nur eine einzige Platte zu einer Zeit ausfallen,

weil dann die XOR-Funktion die Information aus den verbleibenden Daten restaurieren kann. Bei RAID 6 dürfen zwei Platten gleichzeitig ausfallen. Jeder weitere Plattenausfall bei den verschiedenen Varianten führt zum Totalverlust der Daten. Auch bei RAID-Anwendungen ist daher eine regelmässige Datensicherung unerlässlich. Im Gegensatz zu früher, als die Hersteller sich mehr oder weniger an die Berkeley-Definitionen von RAID hielten oder neue Level hinzuerfanden, legt heute ausschliesslich das RAB (RAID Advisory Board) neue RAID-Definitionen fest.

Die einzelnen RAID-Levels ³⁹

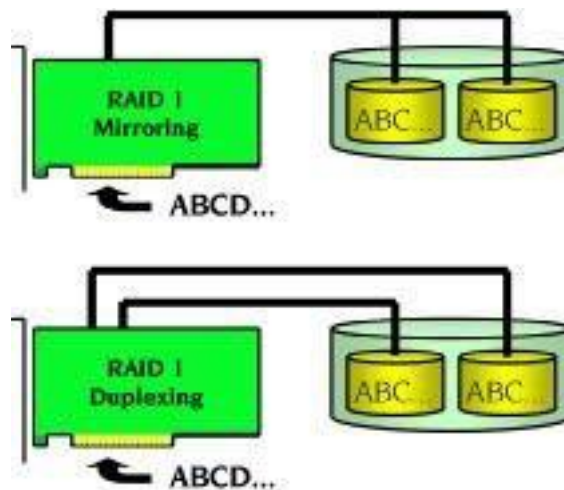
RAID 0 („disk striping“) ist keine definierte RAID-Variante; es dient ausschliesslich zur Beschleunigung des Zugriffs und sollte daher nur gemeinsam mit einer der anderen Varianten eingesetzt werden. Der Datenstrom wird in kleine Teilstücke (zum Beispiel Blöcke) zerlegt und parallel auf die vorhandenen Laufwerke geschrieben. Sind also drei Laufwerke vorhanden, dann werden die Blöcke 1 bis 3 auf die Laufwerke 1 bis 3, danach die Blöcke 4 bis 6 ebenfalls auf die Laufwerke 1 bis 3, und so weiter in numerischer Reihenfolge geschrieben oder davon gelesen. Redundanz und damit Fehlersicherheit gibt es bei diesem Verfahren nicht. Fällt ein Laufwerk aus, dann sind alle Daten, die Blöcke von diesem Laufwerk zur Vervollständigung benötigen, auch auf den anderen Laufwerken verloren.



RAID 1 („mirroring“) betreibt Datenspiegelung zur doppelten Speicherung. Nutzbar ist nur die Hälfte der vorhandenen Plattenkapazität. Fällt ein Laufwerk aus, dann wird mit dem anderen aus dem Paar weitergearbeitet.

Die Daten werden jeweils auf zwei Festplatten gespeichert. Beim Ausfall einer Platte sind die Daten identisch auf der zweiten Festplatte vorhanden. Beim Spiegeln von Festplatten an einem Kanal spricht man von **Disk Mirroring**, beim Spiegeln an unabhängigen Kanälen von **Disk Duplexing** (zusätzliche Sicherheit). RAID 1 ist eine einfache und schnelle Lösung zur Datensicherheit und Datenverfügbarkeit, besonders geeignet für kleinere Nutzkapazitäten.

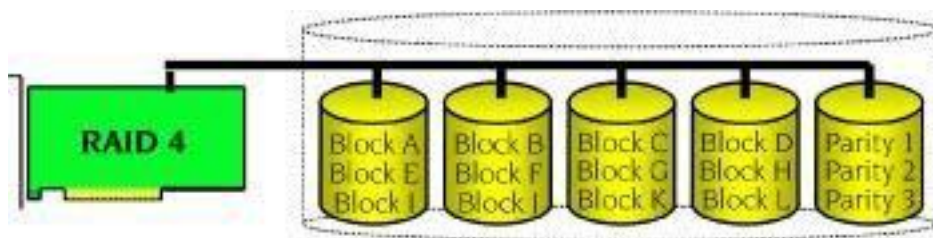
³⁹ S. hierzu auch das Online-Tutorial der Fa. Advanced Computer & Network Corporation AC&NC, <http://www.acnc.com/raid.php> [URL geprüft 5.11.2013]



RAID 2 arbeitet meistens mit einem Hamming-Code⁴⁰ zur Fehlerkorrektur, wie bei Hauptspeicherblöcken im Rechner. Die bei allen Plattenlaufwerken verfügbare interne Fehlererkennung wird nicht genutzt. RAID 2 wird kommerziell nicht eingesetzt, da es zu aufwendig ist, ohne dafür irgendwelche Vorteile zu bieten.

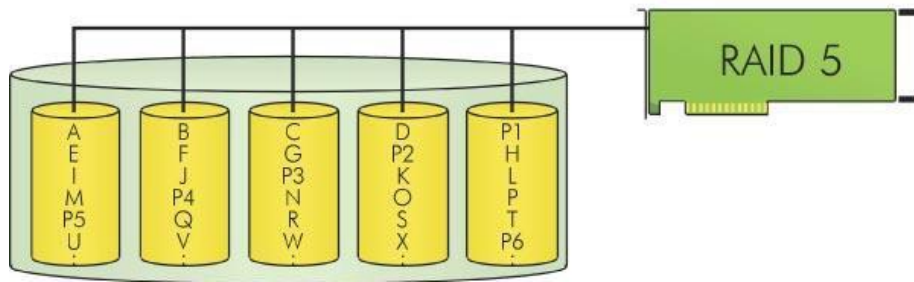
RAID 3 ist günstig für die Speicherung großer, zusammenhängender Datenmengen, wie bei der Bildverarbeitung oder im CAD-Bereich. Wie bei RAID 0 wird parallel geschrieben oder gelesen. Zusätzlich benötigt die Methode jedoch ein weiteres Laufwerk für die Speicherung der Paritätsinformation, die sie mittels XOR-Verfahren bildet. Bei einer Konfiguration mit 4 + 1 Platten stehen 80 Prozent der Bruttokapazität für Nutzdaten zur Verfügung. Es kann nur eine Aktion gleichzeitig ablaufen, weil alle Laufwerke dazu benötigt werden. Daher werden diese oft rotationssynchron eingesetzt. Bei wahlfreiem Zugriff und kleinen Datenmengen ist RAID 3 sehr ineffizient.

RAID 4 ist fast identisch mit RAID 3. RAID 4 muss aber nicht auf alle Laufwerke gleichzeitig zugreifen, solange die Daten auf unterschiedlichen Laufwerken gespeichert sind. Bei Schreibvorgängen ist RAID 4 besonders ineffizient. Ausserdem gelten alle anderen bei RAID 3 genannten Nachteile. RAID 4 wird nur sehr selten kommerziell genutzt.



RAID 5 ist logisch identisch mit RAID 4. Allerdings wird die Paritätsinformation (in der folgenden Abbildung P1 - P5) nicht auf einem bestimmten Laufwerk gespeichert, sondern über alle Laufwerke verteilt. Somit kann sehr häufig mehr als eine Aktion gleichzeitig stattfinden. RAID 5 ist besonders für viele Zugriffe mit kleinen Datenmengen geeignet, daher wird diese Variante besonders gerne in transaktionsorientierten Umfeldern wie Datenbanken eingesetzt.

⁴⁰ <http://eprog.sourceforge.net/eprog/1150/hamming.html> [URL geprüft 5.11.2013]



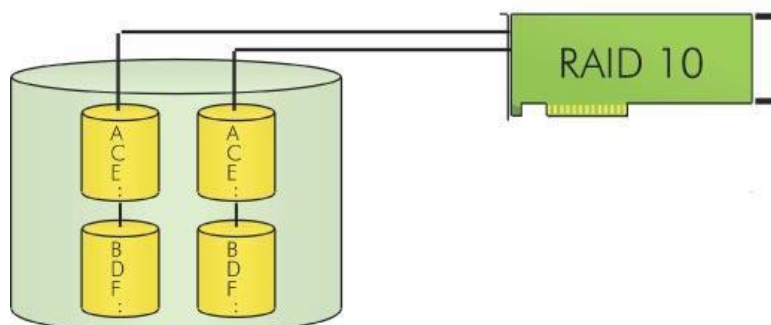
RAID 6 bietet von den standardisierten Verfahren die höchste Datensicherheit. Man fügt einfach zu RAID 5 eine weitere unabhängige Paritätsinformation auf einem zusätzlichen Laufwerk hinzu. Damit werden allerdings Schreibzugriffe noch langsamer als bei den anderen Paritätsverfahren (RAID 3 bis RAID 5).

„**RAID 7** ist eine Erweiterung von RAID 6. Hier geht man etwas weiter und übergibt alle eingehenden und ausgehenden Zugriffe einem zentralen Cache. Die Daten werden per Sector Striping auf mehrere Festplatten gespeichert. Mehrere Reserve-Festplatten übernehmen beim Ausfall einer Festplatte die gespeicherten Daten. Statt nur einer, können gleich mehrere Paritätslaufwerke definiert werden. Die Verarbeitungsgeschwindigkeit der Zugriffe ist hoch. Doch auch hier benötigt der zentrale Cache mehrere redundante Netzteile und eine unterbrechungsfreie Stromversorgung (USV). RAID 7 ist besonders bei sehr großen Speichermengen ein flexibles System, da am laufenden System Festplatten getauscht und Erweiterungen vorgenommen werden können.“⁴¹

Mischformen sind ebenfalls gebräuchlich

Von diesen Grundformen abweichende Verfahren sind herstellerspezifisch oder bestehen aus Mischungen der hier beschriebenen Verfahren. Beliebte sind Kombinationen von Variante 0 und Variante 1 ('0&1' oder '10'), Variante 0 und Variante 3 ('53') sowie von Variante 3 und Variante 5 ('8'). Variante 53 verhält sich bei den Kosten wie eine Kombination aus 5 und 3, daher die Bezeichnung 53. Das von Hewlett-Packard entwickelte AutoRAID⁴² hat eine Verwaltungssoftware zur automatischen Anpassung der RAID-Varianten bei laufendem Betrieb. Daher ist keine RAID-Variante fest vorgegeben.

In der Freien Universität Berlin wird für den dort implementierten ALEPH-Server zum Beispiel die Variante 0&1, also RAID 10, eingesetzt. Dabei dient RAID 0 (disk striping) dem schnelleren Datenzugriff und RAID 1 kennzeichnet den Einsatz der Datenspiegelung auf einer weiteren Festplatte.



⁴¹ Zitat aus: Elektronik-Kompendium <http://www.elektronik-kompendium.de/sites/com/1001031.htm> [URL geprüft 5.11.2013]

⁴² Wilkes, J., Golding, R., Staelin, C. and Sullivan, T.: The HP AutoRAID Hierarchical Storage System. In Proceedings of the 15th Symposium. on Operating Systems Principles, December 1995. Online verfügbar unter: <http://portal.acm.org/citation.cfm?doid=225535.225539> [URL geprüft 5.11.2013]

RAID ist häufig von aussen nicht erkennbar

Erstaunlicherweise ist selbst für Experten nicht immer erkennbar, wo RAID ‚drin‘ ist. Das Marktforschungsinstitut Disk/Trend, Inc. in den USA hat herausgefunden, dass die meisten EDV-Leiter nicht wussten, dass sie fast immer RAID Level 1 (Mirroring = Plattenspiegelung) bei ihren NetWare-Installationen im Einsatz hatten (weil die Fa. Novell, die das Produkt NetWare vertreibt, das nicht RAID 1 nennt).

Die TecChannel-Autoren schließen ihren Aufsatz mit dem

„Fazit: Redundanz mit Grenzen

Zwar lässt sich durch den Einsatz von Diskarrays die Verfügbarkeit von Rechnersystemen deutlich erhöhen. Ein Allheilmittel gegen Datenverluste stellt RAID allerdings nicht dar. Um eine Ausfallsicherheit nahe 100 Prozent zu erreichen, müssen alle Komponenten des Speichersubsystems inklusive Controller, Netzteil und Lüftern redundant ausgelegt werden. Solche Lösungen bietet die Storage-Industrie zwar an, preiswert fallen sie aber nicht gerade aus.

Zudem ereignen sich Ausfälle von Laufwerken und anderen Komponenten nicht immer unabhängig voneinander. In der Praxis treten gelegentlich Situationen ein, durch die sich die Ausfallwahrscheinlichkeit des gesamten Arrays schlagartig erhöht. Dazu zählen etwa durch Blitzschlag verursachte Überspannungen, Überschwemmungen oder Brände. Auch Viren und Würmer befallen RAID-Systeme ebenso gern wie Einzelaufwerke.

Schließlich kann selbst das zuverlässigste Array den Risikofaktor Nummer 1 nicht ausschalten - den Menschen. Den weitaus größten Teil irreparabler Datenverluste verursacht nicht etwa versagende Technik, sondern Fehlbedienung durch den Benutzer. Gelöschte Dateien sind auch auf RAID-Systemen verloren. Selbst für das ausgefeilteste RAID-System gilt deshalb: Den einzig wirklich zuverlässigen Schutz gegen Datenverluste bietet ein konsequent geplantes und vorgenommenes Backup.

Sollte auch kein *Backup* vorhanden sein, lassen sich immer noch professionelle Datenrettungslabore konsultieren. Selbst bei komplexen RAIDs können die Labore die Daten mit Erfolgsquoten von 70 bis 80 Prozent rekonstruieren. (jlu/cvi/hal)⁴³

Und Pete Perlegos fragt sich in bezug auf die RAID-Technologie bereits 2009:

Has this system [er bezieht sich zwar auf das AUTO-Raid von HP, ich denke aber, die Überlegung ist generell auf RAID-Systeme zutreffend] become obsolete with the trend towards clusters of computers networked together? The cost of storage has gotten so cheap that RAID is not needed in such systems. Redundancy can be achieved by replicating data across multiple computers.⁴⁴

Kostenaspekt der unterschiedlichen RAID-Varianten

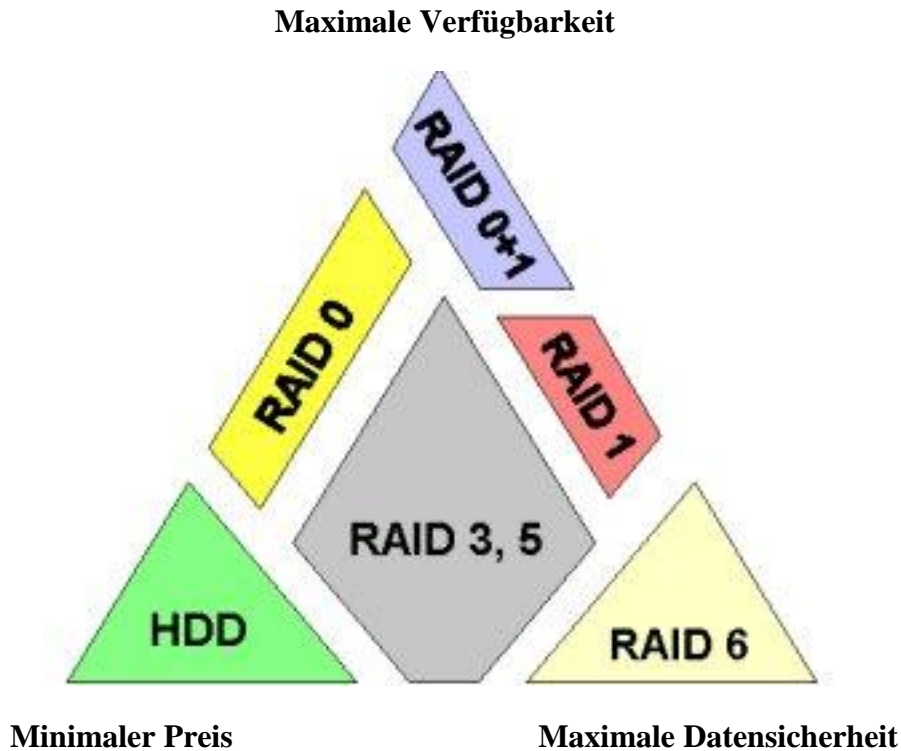
In unserem Zusammenhang muss auch noch das sogenannte RAID-Dreieck erwähnt werden, das die Entscheidungsfaktoren **Preis**, **Verfügbarkeit** (Geschwindigkeit) und **Datensicherheit** in einen

⁴³

http://www.tecchannel.de/storage/extra/401665/raid_sicherheit_level_server_storage_performance_festplatten_controll_er/, S. 18 [URL geprüft 5.11.2013]

⁴⁴ <http://infinitepolygon.blogspot.de/2009/09/hp-auraid-hierarchical-storage-system.html> [URL geprüft 5.11.2013]

Zusammenhang stellt.



(Abb. entnommen aus dem Online-Text von Markus Fischer: Hardwaregrundlagen ⁴⁵)

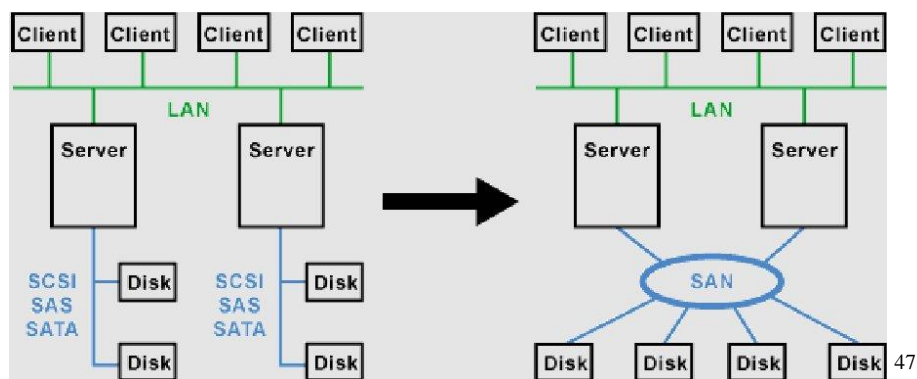
Je nach Zweck werden Systeme zwar schneller, diese geht dann aber zu Lasten der Datensicherheit (RAID 0). Umgekehrt sind Systeme mit einer hohen Datensicherheit in den seltensten Fällen die preisgünstigen.

SAN-Technik ⁴⁶

Während die RAID-Technologie die kostengünstige fehlerfreie Abspeicherung und Bereitstellung von Daten aus einem verteilten Speichersystem sichern soll, ist ein Storage Area Network (SAN) ein sekundäres Netz parallel zu einem primären LAN (Local Area Network). Gespeicherte Daten und Informationen sind das zentrale Element in den vernetzten Systemen. Dies wird durch die logische Zusammenfassung verteilter Datenspeicher in ein Gesamtsystem, das SAN, erreicht. In einem SAN ist der Massenspeicher für Daten nicht mehr nur ein Peripheriegerät an einem bestimmten Rechner oder Server. Die Konsolidierung aller gespeicherten Daten wird so vereinfacht, da diese im SAN immer direkt verfügbar sind. Es entsteht eine „Parallelwelt“: Das SAN ist ein zusätzliches Netz, das für die schnelle Datenübertragung zwischen Speichern und Servern optimiert ist.

⁴⁵ <http://www.hardwaregrundlagen.de/> [URL geprüft 5.11.2013]

⁴⁶ S. http://de.wikipedia.org/wiki/Storage_Area_Network [URL geprüft 5.11.2013]; s. Strass, Hermann: SANs – Standards und Lösungen. In tecCHANNEL vom 4.4.2001, online verfügbar unter der URL: <http://www.tecchannel.de/storage/netzwerk/401636/> [URL geprüft 5.11.2013]; s. auch Schreier, Markus: SAN-Technik, online verfügbar unter der URL: http://www.ordix.de/ORDIXNews/1_2002/sm_3.html [URL geprüft 5.11.2013]



„Rechnen in der Wolke“⁴⁸

AUS DER WIKIPEDIA:

„2009 veröffentlichte das National Institute for Standards and Technology (NIST) eine Definition, die auf weitgehende Akzeptanz stieß und verschiedene Definitionsansätze bündelt:

Cloud-Computing enthält die drei verschiedenen Servicemodelle:

- IaaS – **Infrastructure as a Service** – Rechnerwolken bieten Nutzungszugang von virtualisierter Computerhardware Ressourcen, wie Rechnern, Netzwerken und Speicher. Mit IaaS gestalten sich Nutzer frei ihre eigenen virtuellen Computer-Cluster und sind daher für die Auswahl, die Installation, den Betrieb und das Funktionieren ihrer Software selbst verantwortlich.
- PaaS – **Platform as a Service** – Rechnerwolken bieten Nutzungszugang von Programmierungs- oder Laufzeitumgebungen mit flexiblen, dynamisch anpassbaren Rechen- und Datenkapazitäten. Mit PaaS entwickeln Nutzer ihre eigenen Software-Anwendungen oder lassen diese hier ausführen, innerhalb einer Softwareumgebung, die vom Diensteanbieter (Service Provider) bereitgestellt und unterhalten wird.
- SaaS – **Software as a Service**⁴⁹ – Rechnerwolken bieten Nutzungszugang von Software-Sammlungen und Anwendungsprogrammen. SaaS-Diensteanbieter offerieren spezielle Auswahlen von Software, die auf ihrer Infrastruktur läuft. SaaS wird auch als *Software on demand* (Software bei Bedarf) bezeichnet.¹

Zudem enthält die Definition des National Institute for Standards and Technology (NIST) vier Liefermodelle:

- Public Cloud – die öffentliche Rechnerwolke – bietet Zugang zu abstrahierten IT-Infrastrukturen für die breite Öffentlichkeit über das Internet. Public-Cloud-Diensteanbieter erlauben ihren Kunden IT-Infrastruktur zu mieten auf einer flexiblen Basis des Bezahls für den tatsächlichen Nutzungsgrad bzw. Verbrauch (pay-as-you-go), ohne Kapital in Rechner- und Datenzentrumsinfrastruktur investieren zu müssen.

⁴⁷ Abbildung entnommen <http://www.elektronik-kompodium.de/sites/net/0906071.htm> [URL geprüft 5.11.2013]

⁴⁸ Zum Begriff s. http://de.wikipedia.org/wiki/Cloud_Computing [URL geprüft 5.11.2013]

⁴⁹ Gunten, Andreas von: Software as a Service – Ein Paradigmenwechsel, online unter:

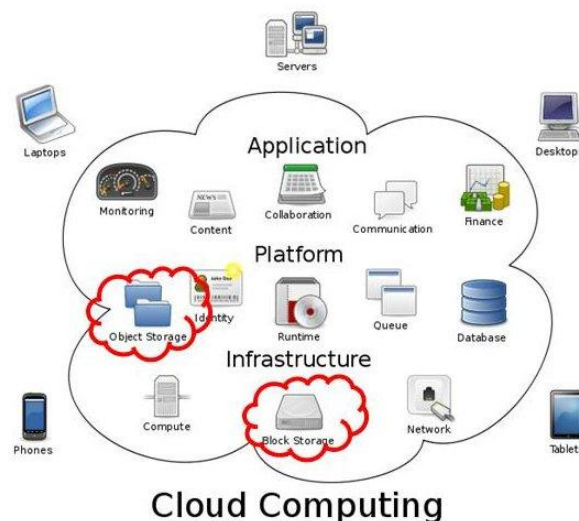
http://www.topsoft.ch/de/dyn_output.html?navigation.void=1546&content.cidid=1546&content.vcname=container_news&navivoid=401&comeFromOverview=true&collectionpageid=181 [URL geprüft 5.11.2013]

- Private Cloud – die private Rechnerwolke – bietet Zugang zu abstrahierten IT-Infrastrukturen innerhalb der eigenen Organisation (Behörde, Firma, Start-Up, Verein), z. B. Ubuntu One.
- Hybrid Cloud – die hybride Rechnerwolke – bietet kombinierten Zugang zu abstrahierter IT-Infrastrukturen aus den Bereichen von Public Clouds und Private Clouds, nach den Bedürfnissen ihrer Nutzer.
- Community Cloud – die gemeinschaftliche Rechnerwolke – bietet Zugang zu abstrahierten IT-Infrastrukturen wie bei der Public Cloud – jedoch für einen kleineren Nutzerkreis, der sich, meist örtlich verteilt, die Kosten teilt (z. B. mehrere städtische Behörden, Universitäten, Betriebe/Firmen mit ähnlichen Interessen, Forschungsgemeinschaften).

Das NIST listet fünf essenzielle Charakteristika für Cloud-Computing:

- Selbstzuweisung von Leistungen aus der Cloud durch den oder die Nutzer, die bei Bedarf bereitstehen soll (Self-service provisioning / As-needed availability).
- Skalierbarkeit bietet die Entkopplung von Nutzungsschwankungen und Infrastrukturbeschränkungen (Scalability).
- Zuverlässigkeit und Fehlertoleranz garantiert permanent definierte Qualitätsstandards der IT-Infrastruktur für den Nutzer (Reliability and fault-tolerance).
- Optimierung und Konsolidierung bietet Effizienz und Ökonomie in Anpassung an fortlaufende Umweltschutzstandards, die sukzessive vom Cloud-Diensteanbieter optimiert werden können (Optimization/Consolidation).
- Qualitätssicherung und -kontrolle kann fortlaufend durch den Diensteanbieter überwacht und sichergestellt werden, ohne dass die Nutzer belastet werden müssten (QoS – Quality of Service).

Demzufolge geht „Cloud-Computing“ über andere gegenwärtig diskutierte Ansätze („Organic Computing“) (Virtualisierung) hinaus. Unter der Bedingung einer öffentlichen Verfügbarkeit, ähnlich beispielsweise dem öffentlichen Telefonnetz, kann man „Cloud-Computing“ je nach Architektur auch als Summe von SaaS und „Utility Computing“ ansehen.



50

⁵⁰ Quelle: http://de.wikipedia.org/w/index.php?title=Datei:Cloud_computing.svg&filetimestamp=20110724162425
[URL geprüft 5.11.2013]

In der bibliotheksbezogenen Anwendung dieser Technologie⁵¹ werden Programme (Software) und Daten nicht mehr lokal oder (bei bibliothekarischen Verbundsystemen) regional gehalten, sondern in einer entfernten Umgebung bearbeitet, verwaltet und **gespeichert**. Deshalb wird das Cloud Computing in unserem Zusammenhang behandelt. Der oben stehenden Abbildung ist zu entnehmen, dass in die „Wolke“ auch zwei Speichermodule integriert sind: Object Storage und Block Storage. Neben den Auswirkungen auf die Arbeit mit den Programmen hat dies auch Rückwirkungen auf die eingesetzte Speichertechnologie, die nunmehr nicht mehr vom Anwender gestaltet, sondern vom Anbieter angeboten wird. Damit bleiben zwar die oben geschilderten Fragen zu den Zugriffszeiten, der Speicherkapazität, dem Datenfluss, der Permanenz, dem Preis und der Sicherung der Daten gegen Verlust und Verfälschung entscheidungsrelevant, werden aber von der Überlegung dominiert, ob man die EDV-Verarbeitung selbst organisieren will oder von einem Anbieter als Gesamtpaket kaufen oder mieten soll. Es ist offensichtlich, dass hiermit auch erhebliche betriebsorganisatorische Folgewirkungen bedacht werden müssen, die z. B. die Notwendigkeit einer eigenen großen EDV-Abteilung in Frage stellen.

Für Bibliotheken wird dabei auch zu bedenken sein, dass sie in Teilbereichen der Benutzung mit hochsensiblen und absolut vor fremdem Zugriff zu schützenden Benutzerdaten arbeiten, die sie ohne erhebliche Absicherungen⁵² nicht einer von ihnen nicht kontrollierten „Wolke“ anvertrauen dürfen.

Gegenwärtig wird mit dem DFG-geförderten Projekt „CIB“ des hessischen Bibliotheksinformationssystem (HEBIS), des Bibliotheksverbunds Bayern (BVB) und des Kooperativen Bibliotheksverbunds Berlin-Brandenburg (KOBV) die Entwicklung einer cloudbasierten Infrastruktur für Bibliotheksdaten mit der verstärkten Anpassung und Einbindung bestehender Strukturen in internationale Nachweissysteme vorsieht.⁵³ Alle andere Verbünde und die Deutsche Nationalbibliothek sind zur Teilnahme eingeladen. Allerdings herrscht dort große Skepsis wegen der dauerhaften Anbindung an eine herstellerabhängige Paketlösung.

6. Vervielfältigungstechniken

In der Abgrenzung des Stoffes „Methoden der Vervielfältigungsverfahren“ fällt bei der Materialsuche das Begriffspaar „**Druck- und Vervielfältigungsverfahren**“ auf. Da hier nur über Vervielfältigungsverfahren gesprochen werden soll, fallen somit alle Verfahren weg, die eine Vervielfältigung von Daten durch „konventionelle“ Druckverfahren ermöglichen, also Tiefdruck, Offsetdruck, Umdruckverfahren usw.

Im Englischen wird „Vervielfältigungsverfahren“ mit „reprography“ wiedergegeben. Um daher eine Eingrenzung vorzunehmen, sollen zunächst Reprographie-Verfahren besprochen werden. Später behandle ich dann als zweites großes Gebiet die Mikrofilmverfahren.

Bei den Reprographieverfahren soll zunächst eine Definition an den Anfang gestellt werden, die

⁵¹ Der Ausdruck „Cloud“ ist eigentlich wenig zutreffend, weil man zwar die Rechner nicht mehr selbst hat, sie aber nach wie vor irgendwo auf der Erde stehen, „Rechnen im Nebel“ wäre deshalb vom Bild her richtiger.

⁵² Zur Datenschutzproblematik s. z. B. <http://itadministrator.de/themen/sicherheit/fachartikel/99227.html> [URL geprüft 5.11.2013]

⁵³ S. Albrecht, Rita, Ceynowa, Klaus, Degkwitz, Andreas, Kende, Jiri, Koch, Thorsten, Meßmer, Gabriele, Risch, Uwe, Rusch, Beate, Scheuerl, Robert, Voß, Michael: Cloudbasierte Infrastruktur für Bibliotheksdaten - auf dem Weg zu einer Neuordnung der deutschen Verbundlandschaft. http://www.b2i.de/fileadmin/dokumente/BFP_Preprints_2013/Preprint-Artikel-2013-AR-2898-Cloudbasierte_Infrastruktur.pdf [URL geprüft 5.11.2013]

auch Georg Thiele in seinem Beitrag im Sammelband „Zur Theorie und Praxis des Bibliothekswesens“ verwendet:

Erzeugung materieller und dauerhafter Bilder von flächigen Objekten durch Strahlung in ein- oder mehrfacher Ausfertigung, insbesondere zum Zweck der direkten Information.⁵⁴

Es handelt sich also im Wesentlichen um **Kopierverfahren** und um **fotografische Verfahren**.

Bei den Kopierverfahren hat bereits 1966 William R. Hawken in seinem umfangreichen Buch „Copying Methods Manual“ (Hawken, William R.: Copying Methods Manual. Chicago 1966.- 375 S.. Ill. (LPT publications ; 11) **21 Verfahren** nach **15 Merkmalsausprägungen** beschrieben. Die Übersicht zeigt jedenfalls auch auf diesem Gebiet den Erfindungsreichtum des Menschen.

Wenn wir den Blickwinkel auf die Bürokommunikation einengen, gibt es verschiedene Möglichkeiten, ein Bild auf optischem Weg auf ein strahlungsempfindliches Material zu übertragen.

Als **optische Wege** sind dabei feststellbar:

- * die **Kontaktkopie**⁵⁵ in Form der direkten Berührung des strahlungsempfindlichen Materials mit der Vorlage während der Belichtung: hierbei entstehen zur Vorlage negative Abbildungen nur in gleicher Größe
- * die **indirekte Übertragung** der Vorlage durch ein Objektiv auf das strahlungsempfindliche Material: hierbei können auch Vergrößerungen oder Verkleinerungen vorgenommen werden.

Beim **strahlungsempfindlichen Material** ist zu unterscheiden, ob es

- * **direkter Träger** der Information oder
- * **Zwischenträger** der Information

ist.

Ist das strahlungsempfindliche Material **direkter Träger** der Information, sind in der Regel weitere Arbeitsprozesse notwendig, um die Information wieder sichtbar zu machen. Wird zum Beispiel im Fotolabor ein mit Silbersalzemulsion beschichtetes Papier benutzt, sind aufwendige Entwicklungsprozesse notwendig, um das latente Bild in ein dauerhaftes Bild umzuwandeln (entwickeln, fixieren, trocknen). Werden bei diesem Prozess keine Fotopapiere verwendet, die den Einsatz einer Dunkelkammer erfordern, sondern Diazomaterialien, kann die in den siebziger Jahren des 20. Jahrhunderts weit verbreitete „Nasskopie“ hergestellt werden, die genau genommen in einem sogenannten „Halbfeuchtverfahren“ hergestellt wird.

Ist das strahlungsempfindliche Material nur **Zwischenträger** der Information, muss sie von dem Zwischenträger abgenommen und auf den dauerhaften Träger übertragen werden. Dies geschieht zum Beispiel bei den **Xerokopien**, wo die Information zunächst auf eine mit einem **Halbleiter** beschichtete Trommel übertragen wird und damit unterschiedliche Oberflächenspannungen auf der Trommel erzeugt werden. An den aufgeladenen Stellen bleibt ein Tonerpulver haften, das an-

⁵⁴ S. Thiele, Georg: Reprographie und Drucktechnik: Verfahren und Geräte. In: Zur Theorie und Praxis des Bibliothekswesens, Bd. 2, München 1976, S. 354-382, hier: S. 355.

⁵⁵ <http://de.wikipedia.org/wiki/Kontaktkopie> [URL geprüft 5.11.2013]

schliessend auf Normalpapier übertragen und durch Erhitzung fixiert wird.⁵⁶

Wie schon Thiele 1976 feststellt, geht die „Tendenz ... heute zur unbeschichteten Papierkopie“. Nach 30 Jahren ist zu bestätigen, dass die unbeschichtete Papierkopie in der Bürokommunikation alle anderen Verfahren verdrängt hat.

Und nicht nur die unbeschichtete Papierkopie! Vergleicht man zum Beispiel eine vor etwa 20 Jahren im Direktkopieverfahren hergestellte Overheadfolie mit den heutigen Vortragsfolien, die direkt mit dem Drucker oder dem Trockenkopierer produziert werden, kann man leicht erkennen, dass diese alte Direktkopie-Folie in der langen Zeit auch ohne ständige Lichtexposition doch erheblich gelitten hat.

Der technologisch neueste Stand im Bereich der Vervielfältigungsverfahren wird sicherlich durch die digitalen **Laserkopiersysteme** repräsentiert. Hier wird die Vorlage in einem Scanner durch Sensoren zeilenweise abgetastet und über einen Analog-Digital-Wandler in digitale Informationen umgewandelt.⁵⁷ Von der Abtasteinheit wird die (vielseitig bearbeitbare) Information über eine Datenleitung zu einem Drucker weitergeleitet, von einem Laserstrahl auf einen zuvor elektrostatisch aufgeladenen Zwischenträger (etwa eine Selentrommel) zeilenweise aufgetragen und anschliessend auf unbeschichtetes Papier oder auch eine Overhead-Folie übertragen.

Betrachtet man diesen Produktionsprozess genauer, wird man sofort die Verbindung zu unserem anderen großen Thema, den Speichermedien sehen. Denn die digitalisierte Information kann über die Datenleitung ja nicht nur zu einem Drucker, sondern auch in ein Speichermedium weitergeleitet oder als Datei per E-Mail versendet werden. Dies ermöglicht dann zum Beispiel eine Form des „publishing on demand“ von ehemals in Papierform vorliegenden Materialien oder eine direkte Weiterverarbeitung in eigenen Rechnersystemen. Eine andere Form des „publishing on demand“ wäre die ausschließlich digitale Produktion und Speicherung (wie beim vorliegenden Text).

Mit in diese Verfahrensfamilie gehört auch das **Faxgerät**, bei dem allerdings schon fraglich ist, ob diese Kommunikationsmöglichkeit zu den Vervielfältigungsverfahren gehört. Wenn wir uns an die oben genannte Definition halten (Erzeugung materieller und dauerhafter Bilder von flächigen Objekten durch Strahlung in ein- oder mehrfacher Ausfertigung, insbesondere zum Zweck der direkten Information) und den Abtastvorgang durch einen Scanner als Strahlung interpretieren, ist auch der output des Faxgerätes, die **Fernkopie** (so der eingedeutschte Ausdruck) in den Bereich der Kopierverfahren einzuordnen.

⁵⁶ „Die Wurzeln der Xerokopie gehen auf das Jahr 1778 zurück, als Georg Christoph Lichtenberg an der Universität Göttingen mit geladenen Harzschichten auf Metall experimentierte. Er machte erstmals die elektrostatischen Coulombkräfte sichtbar, die zur Abbildung von Schrift und Bildern ausgenutzt werden können. Der eigentliche Startschuss der Xerokopie fiel aber erst am 27. Oktober 1937, als der US-Amerikaner Chester Carlson ein Xeroxverfahren zum Patent anmeldete, das erstmals Kopien auf normalem unbeschichtetem Papier ermöglichte. Schlüssel zum Erfolg war eine auf einer Trommel befindliche Schicht des Photohalbleiters Selen, der mit Hilfe einer Hochspannungsanlage positiv aufgeladen und dadurch lichtempfindlich gemacht wurde. Carlsons erster erfolgreicher Kopierversuch fand in seinem Laboratorium im New Yorker Vorort Astoria am 22. Oktober 1938 statt. Unter eine Glasplatte, auf die er mit Tinte die Zeichen „10-22-38 Astoria“ geschrieben hatte, positionierte Carlson eine Metallplatte mit Schwefelbelag, die er durch Reiben mit einem Tuch elektrisch auflud. Nach einer kurzen Belichtung nahm er die Metallplatte heraus und puderte sie mit staubfeinem Bärlappsamen ein, so dass die Schrift als Negativ zu erkennen war. Nach Abzug eines auf die Platte gedrückten Wachspapiers war der Schriftzug auf ihm positiv zu lesen. Das war die erste xerographisch hergestellte Kopie der Welt. Carlson, ein ehemaliger (und später in der Rezession entlassener) Mitarbeiter der Firma Bell Laboratories, wirkte als freier Erfinder und war damals auf der Suche nach einem Partner. Schließlich entwickelte er das Verfahren gemeinsam mit dem Battelle-Institut zur Marktreife und 1947 kamen die ersten Photokopierer der Welt in den Handel.“ Haarer, Dietrich: Chemie und Informationstechnik – Potenziale konsequent nutzen; aus einer Firmenschrift der ehem. Fa. Degussa, die nicht mehr im Netz verfügbar ist)

⁵⁷ S. <http://de.wikipedia.org/wiki/Elektrofotografie> [URL geprüft 5.11.2013]

Auch hier haben wir übrigens beim Druckwerk solcher Geräte die technologische Wandlung vom Thermodruckverfahren zum unbeschichteten Papier zu beobachten. Während bei mit Hilfe des Thermodruckverfahrens (hierfür wird ein besonderes beschichtetes Papier verwendet, das sich durch Wärmeeinwirkung verändert) die Sichtbarkeit der Information bei Lichtexposition rasch verschwindet, werden heute Druckwerke eingesetzt, die unbeschichtetes Papier ohne Probleme verwenden können. Bei den Thermodruckern ist man nach Erhalt des Faxes sofort zum nächsten Kopierer gerannt, um die Information dauerhaft zu erhalten. Insofern könnte man Fernkopien mit Thermodruckern auch **nicht** zu den Reprographieverfahren zählen, weil keine „dauerhaften“ Abbildungen erzeugt werden.

Um zur Fernkopie nochmals die Meinung von Georg Thiele aus 1976 zu zitieren:

„Eine andere, teilweise in die Tat umgesetzte Weiterentwicklung ist die Fernkopie, deren Anwendung nicht unbestritten ist. Über das bereits bestehende große Fernschreibnetz können Mitteilungen irgendwelcher Art in kürzester Zeit und oft billiger übermittelt werden. Lediglich bei der Übermittlung von Tabellen, graphischen Darstellungen und Zeichnungen ist die reprographische Übertragung von Vorteil. Dieses wäre auf dem Fernschreibwege überhaupt nicht möglich.“

(s. Thiele, Georg, a.a.O., S. 377.)

Unter Personalkostengesichtspunkten wäre die Feststellung „oft billiger“ kritisch zu hinterfragen, da immer eine aufwendige manuelle Eingabe erforderlich ist bzw. war, denn welche Bedeutung spielt heute noch das Fernschreibnetz⁵⁸ ?

Bevor ich mich nun dem zweiten großen Bereich der Vervielfältigungsverfahren, den Mikrofilmverfahren, zuwende, will ich noch auf eine Spielart eingehen, die ein Zwischenglied darstellt, das heute jedoch weitgehend unbekannt ist: die **Mikrokarte**. Das Verfahren ist in dem Buch von Hawken ausführlich beschrieben (s. Hawken, William R., a.a.O, S. 77-82.) Es handelt sich um einen etwa DIN-A-4 große Karte aus Fotopapier, auf der Mikrokopien reihenweise angeordnet sind.

Wenn man es allerdings sehr genau nimmt, kann man im engeren Zusammenhang der Vervielfältigungstechnik diese Produktionsweise nicht behandeln, denn die Mikrokarten werden als feinste Drucke hergestellt, zählen also genau genommen zu den Druckverfahren.

Ein Hindernis für die breite Verwendung der Mikrokarte ist sicherlich die Tatsache, dass diese Karte nur mit Hilfe eines Gerätes gelesen werden kann, das eine Art **Epidiaskop** darstellt und davon keine unmittelbaren Kopien gezogen werden können, weil diese bei der geringen Auflösung der Kopierer nicht mehr lesbar wären.

Der zweite große Bereich, der im Rahmen der Methoden zur Vervielfältigungstechnik zu behandeln ist, ist die **Mikrofilmtechnik**. Auf sie legt Georg Thiele in seinem erwähnten Beitrag „Reprographie und Drucktechnik: Verfahren und Geräte“ das positive Schwergewicht, was nicht weiter verwundert, zählte er doch zu den Experten in Deutschland, die sich sehr gründlich damit im Rahmen der Dokumentation auseinandergesetzt haben und er auf diesem Gebiet selbst entwickelnd tätig gewesen ist. Deshalb kann das, was in seinem Beitrag zu diesem Verfahren dargestellt wird, uneingeschränkt zum Selbststudium empfohlen werden.

⁵⁸ S. <http://de.academic.ru/dic.nsf/dewiki/437510> [URL geprüft 5.11.2013]

Wie in unserem Zusammenhang üblich soll uns bei den Mikroformen nicht die Herstellungstechnik im Fotolabor oder mit anderen Maschinen interessieren, sondern nur das Ergebnis.

Von den Formen der Mikrofilme unterscheidet man zunächst **Rollfilme** und **Planfilme** bzw. etwas unschärfer **Mikrofilme** und **Mikrofiches**. Der Mikrofilm hat üblicherweise das 35-mm-Format, seltener das 16-mm-Format. Der Mikrofiche hat als Standard die Größe DIN A 6 (14,8 x 10,5 cm).

Aber auch davon abweichende Formen waren und sind produzierbar und in Gebrauch, vor allem in Apotheken und bei Autoteile-Listen. Ein Beispiel aus dem Buchbereich ist der LIBRI-Gesamtkatalog. Inzwischen ist auch er durch eine CD-ROM bzw. Datenbank abgelöst worden.

Der 16-mm-Mikrofilm kann auch dazu benutzt werden, in Filmtaschen eingeschoben zu werden, die auf einem Din-A-6-Format in fünf Reihen aufgebracht sind. Dadurch entsteht ein sog. **Mikrofilm-Jacket**. Dieses Jacket kann wiederum benutzt werden, um eine Kopie zu ziehen, so dass im Endeffekt ein Mikrofiche herauskommt.

Neben diesen selbstständigen Informationsträgern lassen sich diese Mikroformen auch mit anderen Informationsträgern kombinieren, so etwa mit dem Speichermedium **Lochkarte**, in das ein Mikrofilm einmontiert wird.

Der Mikrofilm hat inzwischen als Informationsträger im Bibliotheksbetrieb zunehmende Konkurrenz durch die elektronischen Datenbanken bekommen. Dennoch verfügen zahlreiche Bibliotheken über große Mikroformen-Sammlungen (im Bibliothekssystem der Freien Universität wird die Zahl der Mikroformen auf 2 Millionen geschätzt), die weiterhin benutzbar gehalten werden müssen. Unter dem Gesichtspunkt der Datenspeicherung hat er jedoch gegenüber allen elektronischen Verfahren so große Vorteile hinsichtlich seiner Langlebigkeit und einfachen Reproduzierfähigkeit (denkt man an die Parameter Zugriffszeit, Speicherkapazität, Permanenz und Preis), dass er seine Bedeutung in der dauerhaften Archivierung⁵⁹ nicht verlieren wird.

Diesen Einsatz hat vor allem auch die Möglichkeit gefördert, direkt aus dem Computer heraus die Mikrofilme zu beschriften (**COM** = Computer-output on microform). Damit wurden z. B. die Zeitschriftendatenbank und die Schlagwortnormdatei produziert.

Die Wiedergabe der Mikroformen erfolgt über einfache Lesegeräte oder auch Kombinationen von Lesern und Druckern, den **Readerprintern**.

Inzwischen gibt es Geräte, die, anstelle ein Mikrofiche-Duplikat anzulegen, vorhandene Mikroformen auf optische Speicher übertragen. Bei Verwendung von DVD, auch im Blue-Ray-Format, lassen sich so größere und größte Mikroformen-Sammlungen auf ein relativ handliches, in Computernetzen einzuspielendes Format umsetzen und damit die nutzungsaufwendigen Klippen der Mikroformen-Leser und Mikroformen-Printer umschiffen.

Copyright 2013 © Prof. Dr. Ulrich Naumann: ulrich-naumann@gmx.de

⁵⁹ S. auch Grube, Michael: Kulturgutschutz und Bergungsorte in Deutschland Ost & West, online verfügbar unter der URL <http://www.lostplaces.de/barbarastollen-kulturgutschutz.html> [URL geprüft 5.11.2013]. - S. auch Etzenberger, Klaus: Technische Normen und Anweisungen für die Verfilmung von Archivgut. In: Verwahren, Sichern, Erhalten: Handreichungen zur Bestandserhaltung in Archiven / Hrsg. von Mario Glauert und Sabine Ruhnau. – Potsdam: Landesfachstelle für Archive und öffentliche Bibliotheken im Brandenburgischen Landeshauptarchiv, 2005, S. 147-180.