

DATENBANKORGANISATION BEI DER HOECHST AKTIENGESELLSCHAFT

Otmar Saal, Diplom-Volkswirt, HOECHST AKTIENGESELLSCHAFT

Zusammenfassung

In einem generellen Rahmen wird zunächst aufgezeigt, von welchen Bedingungen und Überlegungen HOECHST bei der Planung von Datenbanksystemen ausgeht. Am Beispiel von Anforderungen seitens stark integrierter Abrechnungs- und Abwicklungssysteme werden dann ausgewählte Fragen aus der praktischen Anwendung von Datenbank- und Datenkommunikationssystemen erörtert.

DATENVERARBEITUNG IM SYSTEMVERBUND

Um den Rahmen der späteren Ausführungen verständlich zu machen, erscheint es zweckmäßig, zunächst einen kurzen Überblick über die Struktur unseres Unternehmens zu geben.

Die HOECHST AG legte für das Geschäftsjahr 1974 einen Weltabschluß vor, in dem über 400 in- und ausländische Gesellschaften konsolidiert sind, an denen das Unternehmen mit mindestens 50 % beteiligt ist. Weltweit wurde ein Umsatz von 20,2 Mrd. DM erzielt. Die Produktionspalette deckt mit etwa 50.000 verschiedenen Erzeugnissen fast vollständig das gesamte Gebiet der Chemie ab.

Das Gesamtunternehmen HOECHST wird in 3 Gruppen betrachtet, nämlich HOECHST Welt, HOECHST Konzern und HOECHST AG, wobei sich die nachfolgenden Ausführungen überwiegend auf die Muttergesellschaft mit insgesamt 13 inländischen Werken und einem Umsatzvolumen im Jahre 1974 von 9,7 Mrd. DM beziehen werden.

Betrachten wir das Zusammenwirken der einzelnen Unternehmenseinheiten der HOECHST AG (Werke, Konzerngesellschaften, Auslandsgesellschaften) mit den Funktionsbereichen (Ressorts, Bereiche, Unternehmensleitung) hinsichtlich der dabei anfallenden Aufgaben für die Datenverarbeitung, dann wird offensichtlich, daß die notwendigen

Daten für Abrechnungssysteme, Abwicklungs- und Dispositionssysteme sowie Informations- und Planungssysteme nur durch umfassende Systeme der Datenverarbeitung erfaßt, zugeführt, einheitlich aufbereitet, gespeichert und ausgewertet werden können.

Dementsprechend trägt ein den jeweils zentralen und/oder dezentralen Aufgaben entsprechendes System von Datenverarbeitungseinrichtungen in einem quasi hierarchisch organisierten Zusammenwirken dazu bei, die benötigten Daten zu erfassen und zu verarbeiten und/oder für eine weitere Stufe der Verarbeitung im Gesamtsystem bereitzustellen.

Unter der Bezeichnung Systemverbund HOECHST arbeiten wir an der Realisierung einer Konzeption, die die Probleme einer zentralen und dezentralen oder lokalen Datenverarbeitung in einer möglichst effizienten und betriebssicheren Weise lösen helfen soll. Unser Mehrrechnerverbundsystem von 3 Großrechnern in der Zentrale wird sinnvoll ergänzt durch einen Verbund angepaßter dezentraler Rechner- oder Terminalintelligenz, wobei weniger eine System-Distribution im Vordergrund steht, sondern vielmehr die Verwendung der jeweils geeignetsten Einrichtungen für die Aufgaben der lokalen Datenerfassung und Verarbeitung mit einer möglichst direkten Kommunikationseinrichtung zum zentralen System. Soweit möglich und sinnvoll werden vom lokalen Rechner aus auch die zentralen Ressourcen mittels "Remote-Job-Processing" genutzt, wofür ein entsprechendes Workstationprogramm zur Verfügung steht. Die Werke und Geschäftsstellen werden durch dieses Verbundsystem außerdem in die Lage versetzt, eigene werksbezogene Informationsbedürfnisse zu befriedigen, die normalerweise dort nur durch eine die wirtschaftlich sinnvolle Größenordnung übersteigende Datenverarbeitungsanlage erfüllt werden könnten.

Dieses Gesamtsystem der Datenverarbeitung in unserem Unternehmen kann sich aber nicht allein auf eine Weitergabe von Daten für die übergeordneten zentralen Datenverarbeitungsaufgaben durch die jeweils örtlich oder funktional getrennten Teileinheiten erstrecken, sondern verlangt gerade im Bereich der Datenspeicherung und Informationsauswertung eine einheitliche Architektur.

Damit ergab es sich fast zwangsläufig, daß die Datenverarbeitung bei HOECHST in konsequenter Verfolgung des Konzeptes einer integrierten Datenverarbeitung zu einer Datenbankorganisation kommen mußte, die eine Allgemeinverwendbarkeit der Datenbestände sowohl über die einzelnen Anwendungsbereiche als aber auch über die physischen Grenzen eines einzelnen Rechenzentrums hinaus sicherstellen kann.

Wie bei jedem Unternehmen, das schon sehr frühzeitig mit dem Einsatz der Datenverarbeitung begonnen hat, strebte auch HOECHST anfangs vorwiegend die Inte-

gration der Datenerfassung und einen effizienten Datenfluß zwischen den Anwendungsgebieten an.

Das soll aber keineswegs bedeuten, daß beim Aufbau der zentralen Dateien die Ganzheitlichkeit der Planung und die Beachtung der Gesamtzusammenhänge vernachlässigt worden ist. Es standen eben anfangs überwiegend Projekte der Massendatenverarbeitung in den Abrechnungs- und Administrationsbereichen an, die zuerst einmal primär für den eigentlichen Fachbereich aufgebaut wurden und die aufgrund der stapelorientierten Datenverarbeitung auch meist in der Dateiorganisation überwiegend auf die Belange der unmittelbaren Fachbereiche hin organisiert waren. Darüber hinausgehende, nicht fachbereichstypische Daten dienten vorwiegend der integrierten Datenerfassung und der zweckmäßigsten Weitergabe möglichst umfassend geprüfter Daten. Schon rein technisch gesehen hatten wir damals keinerlei Möglichkeiten hinsichtlich einer umfassenden, aber dennoch möglichst anpassungsfähigen und leicht zu handhabenden integrierten Datenbankorganisation.

Erst die technologischen Entwicklungen der Datenverarbeitung ließen uns etwa ab 1967 durch geeignete externe Speicher mit wahlfreiem Zugriff und durch die neuartigen Kommunikationsformen im Rahmen einer Echtzeitverarbeitung an die Planung und Realisierung von umfassenderen und dateiorganisatorisch stärker integrierten Anwendungssystemen herangehen.

AUFGABEN DER DATENBANKEN IM GESAMTINFORMATIONSSYSTEM

Seit etwa 6 - 7 Jahren befinden sich Art und Struktur unserer Anwendungen in einem starken Wandel. Die reinen Abrechnungs- und Administrationssysteme konnten jetzt durch Datenbankkonzeptionen und direkten Zugriff über Datenfernverarbeitungseinrichtungen zu Dispositions- und Informationssystemen ausgebaut werden. Die Datenspeicherung kann von dem bisher überwiegend inaktiven Zustand auf den Magnetbändern in eine aktive, jederzeit von den Benutzern ansprechbare Speicherungsform auf Magnetplattenspeicher überführt werden. Die Datenerfassungs- und Administrationssysteme für die Fachbereiche selbst wurden dadurch sowohl im Hinblick auf die Art der maschinellen Durchführung durch die Echtzeitverarbeitung viel rationaler gestaltet als aber auch durch die direkte Verknüpfbarkeit zu anderen Datenbanken aussagefähiger. Denn nun konnten die für den Fachbereich notwendigen Informationen durch Zugriff auf die Datenbankorganisation anderer Fachbereiche oder auf zentral geführte Datenbanken leichter zu wirksamen Teilinformationssystemen ausgebaut werden. Zusätzlich ermöglicht die integrierte Datenbankorganisation einen direkten Abruf von Daten aus fachbereichsbezogenen Teilsystemen zur Bearbeitung in Schwerpunktsystemen mehrerer Funktionsbereiche oder gar in zentralen Informations-

systemen.

Wenn wir den Begriff "zentrales Informationssystem" oder auch "zentrales Berichtssystem" anstatt der vielfach üblichen Bezeichnung "Management Information System" benutzen, dann hat das seinen Grund.

Uns scheint MIS zu stark auf eine Informationsgewinnung nur für höhere Führungsebenen festgelegt, wodurch der Eindruck erweckt wird, daß das entsprechende System der Datenverarbeitung und Datenspeicherung primär unter diesem Gesichtspunkt konzipiert wurde. Wir sind vielmehr der Ansicht, daß die Informationsbedürfnisse von der operativen Ebene bis hin zur höchsten Führungsebene unbedingt aus jeweils gemeinsamen, zentral geführten Datenbanken abgedeckt werden müssen. Diese Datenbanken selbst können dann durch verschiedene Teilinformationssysteme erstellt werden und dienen zunächst einmal primär zur Bewältigung der Aufgaben in Systemen, die für die operative Ebene erstellt wurden. Daß diese Datenbanken darüber hinaus auch in der Lage sein müssen, die Anforderungen von übergeordneten Informationssystemen abdecken zu können, das ist im wesentlichen eine Frage einer planvollen und flexiblen Datenbankstruktur.

Eine planvolle und auf die Erfüllung aller zentralen Informationsbedürfnisse ausgerichtete Datenspeicherung erfordert aber zunächst einmal ein einheitliches System der Datendefinition und Datenverschlüsselung. Dementsprechend wird auch bei HOECHST vom Beginn der Datenverarbeitung an ein sehr großer Wert auf die zentrale Definition, Entwicklung und Pflege aller Schlüsselbegriffe und Ordnungskriterien gelegt, die in einem zentralen Schlüsselbuch der Datenverarbeitung für das gesamte Unternehmen verbindlich festgelegt und ergänzt werden.

Unter der Voraussetzung einer klaren Datendefinition und Verschlüsselung ist es dann prinzipiell kein allzu schwieriges Problem mehr, die Daten den jeweiligen Anforderungen der Informationssysteme entsprechend bereitzustellen, zu verdichten, zu verknüpfen und auszuwählen. Wenn man ein geeignetes Datenbank-Management-System zur Verfügung hat, kann durch einen universellen Aufbau der Datenbanken viel elastischer und unmittelbarer auf wechselnde Informationsbedürfnisse des Managements reagiert werden als dies bei dem starren Rahmen eines einmal vorgedachten und im Dateieninhalt festgelegten MIS möglich wäre.

Damit gehen wir bei HOECHST eindeutig den Weg, zunächst einmal sehr umfassende Teilinformationssysteme aufzubauen und das eigentliche MIS als ein quasi übergeordnetes "zentrales Berichtssystem" durch die gemeinsame, zentrale Organisation der Datenbanken jederzeit aussagefähig zu halten.

DATENBANKEN IN EINEM INTEGRIERTEN TEILINFORMATIONSSYSTEM

Durch praktische Beispiele möchte ich die bisherigen generellen Aussagen etwas konkreter werden lassen. Um aber auch hierfür zunächst den Gesamtzusammenhang verständlich zu machen, werde ich einen Überblick über ein umfangreiches und stark verzahntes System im Bereich des Verkaufs und der Produktion geben, das seinerseits aus einer Anzahl für sich allein wirksamer Teilsysteme besteht.

Die derzeit engste Verzahnung im Datenverbund haben wir zwischen den Teilsystemen im Bereich der Auftragsabwicklung, der Lagerbestandsführung und Disposition, der Produktionsdatenerfassung, der Versanddisposition und -Abwicklung, der Einkaufsdisposition sowie der Kontokorrentführung. In den wesentlichen Bestandteilen arbeiten alle diese Teilsysteme im Echtzeitbetrieb und greifen dabei weitestgehend auch auf zentrale Datenbanken zurück.

Eine Auftragsabwicklung von der Auftragsannahme über die verschiedenen Dispositions- und Abwicklungsstufen bis hin zur Rechnungsverbuchung im Kontokorrent läßt sich nur voll automatisiert durchführen, wenn auch die jeweils relevanten und wirklich aktuellen Daten aus den tangierten anderen Teilsystemen im direkten Zugriff zur Verfügung gestellt werden können.

Dementsprechend benötigt bereits das Teilsystem der Auftragsabwicklung und Rechnungsschreibung umfassenden Zugriff auf Informationen aus den Bereichen Kunden, Produkt, Lagerbestandsführung, Produktionsplanung, Transportmittel und andere. Aber diese gegenseitige Bereitstellung von Daten von und für andere Arbeitsgebiete darf keineswegs nur im engen Rahmen eines lokal orientierten Systems erfolgen, sondern muß dem Gesamtverbund der Abwicklung und Abrechnung über einzelne Unternehmenseinheiten hinweg Rechnung tragen.

So kann die Definition der Kundenaufträge sowohl in dem Stammhaus als auch in den Geschäftsstellen erfolgen, und die Auslieferung ist von Außenlagern, von Zentral- lagern oder von den verschiedenen Betriebsstätten in Deutschland aus möglich.

Anhand einiger stark vereinfachter Frage- und Aufgabenstellungen aus diesem Bereich der Auftragsabwicklung bei HOECHST werde ich nun versuchen, Überlegungen, die zur entsprechenden Datenbankorganisation geführt haben, zu verdeutlichen. Dazu stelle ich drei stark integrierte Datenbanken heraus, nämlich für:

- A = Aufträge (Lager, Werke, interne Lieferungen)
- B = Bestände (Istbestand, Dispositionsbestand, Prod.-Plan, Bestellung)
- C = Kunden und Lieferanten (Offene Posten, Bestellungen)

Zunächst soll durch Auftragsdefinition in dem Bestand der Datenbank A ein Neuzugang gebildet werden. Dazu bedarf es aber bereits bei der Auftragsannahme folgender Feststellungen:

- Kann die Ware zur Zeit überhaupt geliefert werden und zu welchen Konditionen, oder wenn nicht, wann und von welchem Lager oder Produktionsbetrieb kann wieder geliefert werden?
- Ist der Kunde bezüglich seines Kreditlimits noch belieferbar, oder aber, falls der Kunde gleichzeitig auch als Lieferant vorkommt, wie sieht die Differenz zwischen Kundenobligo und unseren Verbindlichkeiten oder Bestellwerten aus?

Um diese Fragestellungen beantworten zu können, müssen für die erste Frage Informationen aus der Datei B und für die andere Frage Daten aus der Datenbank C zur Verfügung stehen.

Die Veränderungen der Bestandsdatei (B) werden, soweit es sich um Betriebsbestände handelt, durch Aktivitäten bewirkt, die nicht allein durch das Verkaufsgeschehen verursacht werden, sondern ebenso durch Zu- oder Abgänge im Produktionsprozeß, denn bei der Auftragsabwicklung in den Produktionslagern ist der Kundenauftrag nur einer von vielen statusverändernden Vorgängen. Eine ständige Dispositionsbereitschaft erfordert nämlich noch eine Reihe anderer Daten. Je nachdem, ob die Produktion auftragsorientiert, lagerorientiert, kontinuierlich oder diskontinuierlich abläuft, müssen die entsprechenden Dispositionssysteme auch auf aktuelle Bestandsdaten, Auftragsdaten, Produktionsplandaten, Anforderungen aus Produktion und Bestelldaten zugreifen können.

Ausgehend von einer einfachen Fragestellung nach der Lieferbereitschaft für ein Produkt können wir jetzt bereits eine beachtliche Verknüpfung verschiedener Teilsysteme erkennen, die alle einen gemeinsamen Integrationspunkt in der Bestandsdatenbank haben.

Auch bei der Beantwortung der anderen Frage nach der Bonität des Kunden erkennen wir Abhängigkeiten vom Zahlungseingang, von der Auftragsannahme durch Disponenten anderer Verkaufsbereiche und schließlich von den eigenen Bestellanforderungen sowie der Begleichung unserer Lieferantenrechnungen, falls dieser Kunde auch gleichzeitig uns gegenüber Lieferant ist.

In meinen bisherigen Ausführungen wurde absichtlich diese organisatorische Umgebung und andeutungsweise auch die funktionalen Zusammenhänge der Teilsysteme mit

aufgezeigt, um klar erkennen zu lassen, daß die Dateiorganisation vor allem unter dem Aspekt der Gesamtzusammenhänge gesehen werden muß. Datenbankorganisation muß sich nämlich von der früher vorherrschenden Dateiorganisation dadurch unterscheiden, daß eine universelle Verwendbarkeit der Daten für eine Vielzahl von Anwendungen auch über den primären Anwendungsbereich hinweg erreicht werden kann.

Wenn man zusätzlich die bekannten Postulate für den Einsatz von Datenbanken erfüllen will, nämlich Aktualität der Daten für alle Benutzer, Redundanzfreiheit und Zugriff zu den Daten nach verschiedenen Kriterien, dann war dies genau die Ausgangssituation unserer Überlegungen, als wir etwa Anfang 1968 an die Systemplanung für unser erstes Auftragserfassungs- und -Abwicklungssystem im Echtzeitbetrieb herangingen und uns nach einer geeigneten Datenbank-Software umsahen. Unsere Anforderungen an die einzusetzende Datenbank-Software betrafen aber nicht nur ein Instrument für die eigentliche Datenbankverwaltung, sondern wir suchten ein insgesamt flexibles und ausbaufähiges, aber auch in seiner weiteren Entwicklung abgesichertes System.

Da auch unsere ersten Datenbank-Anwendungen nur als Teilssysteme konzipiert werden konnten und selbst innerhalb der Einzelsysteme lediglich in Entwicklungsphasen zu realisieren sind, mußte das einzusetzende Datenbanksystem ebenfalls in seiner Struktur recht anpassungs- und ausbaufähig sein und im Datenbankverwaltungsteil leicht und sicher die Integration weiterer Anwendungsprogramme ermöglichen.

IMS als Datenbank-Software

Noch bevor uns IMS bekannt wurde, haben wir unter dem Stand der Erkenntnisse und Möglichkeiten Anfang 1968 versucht, einen eigenen Datenbankprozessor zu entwickeln. Ausgehend von der Dateiorganisation des Stücklistenprozessors sollte die notwendige Strukturierung der Dateien möglichst und über entsprechende Makros der universelle Zugriff zu den Datenelementen realisiert werden.

Doch noch während der Entwicklung dieses eigenen Datenbankprozessors erhielten wir Vorabinformationen über das Information Management System (IMS) und entschlossen uns nach einem umfangreichen Systemtest zum Einsatz dessen Datenbankteiles (DL1).

Für die Datenbankorganisation mit IMS sprach vor allem der aufgrund der Baumstruktur gegebene flexible Aufbau mit maximal 256 Segmenttypen und einer variablen

Segmentanzahl auf 15 verschiedenen Stufen. Darüber hinaus bot das IMS, durch die programmunabhängigen Datenbankbeschreibungen und durch die Einrichtung, für die einzelnen Benutzerprogramme nur jeweils erforderliche Segmente als sensitiv zugänglich machen zu können, eine uns ideal erscheinende Möglichkeit, programmunabhängige und dem wachsenden Integrationsgrad gut anpaßbare Datenbanken aufzubauen.

Konnten wir uns in den beiden ersten Anwendungsjahren noch mit der stark linearen Struktur unserer IMS-Datenbanken zufrieden geben, so brachte der wachsende Integrationsgrad durch unterschiedliche Anwendungssysteme und das ansteigende Informationsbedürfnis die Notwendigkeit, die Datenbanken unabhängig von ihrer physischen Speicherung auch logisch strukturieren zu können, wie dies dann ab IMS Version 2 auch möglich wurde.

Ebenfalls mit Version 2 wurde auch der Datenkommunikationsteil des IMS zu unserem zentralen System für die Nachrichtenannahme, -Steuerung- und Verwaltung von inzwischen 140 Datenstationen übernommen, so daß heute schon ein System /370-168 fast nur für den IMS-Betrieb eingesetzt werden muß. Darüber hinaus wird IMS aufgrund seiner zentralen Datenbankverwaltung und Nachrichtensteuerung jetzt auch in Verbindung mit GIS eingesetzt und außerdem mit STAIRS verknüpft. Somit wird IMS heute von HOECHST als ein umfassendes Datenbankverwaltungs- und Nachrichtensteuerungssystem angesehen.

Aufgrund der zentralen Bedeutung der mit IMS organisierten Datenbanken und eines recht hohen Nachrichtenaufkommens mit einem starken Anteil änderungswirksamer Vorgänge in den Datenbanken mußten wir besonders beim Datenbankdesign sowie bei der Programmstruktur und bei der Wahl der Zugriffsbefehle einen großen Wert auf Schnelligkeit und Sicherheit legen.

IMS läßt dem Benutzer einen großen Spielraum bei der Organisation und Strukturierung der Datenbanken. Deren Design aber beeinflußt ganz entscheidend die Verarbeitungsgeschwindigkeit der zugehörigen Anwendungsprogramme und kann sich auch im Hinblick auf den Gesamtdurchsatz im IMS-System spürbar bemerkbar machen.

Da zu Beginn der Anwendung von IMS weder Erfahrungen vorlagen noch in irgendeiner Weise ein Verfahren zur Simulation des Zeitverhaltens der Datenbanken bei unterschiedlicher Strukturierung zur Verfügung stand, mußten viele grundlegende Erkenntnisse von uns zuerst einmal im Rahmen spezieller Testuntersuchungen gesammelt und dann im praktischen Betrieb ergänzt und angepaßt werden. Allerdings mußten im Laufe der Zeit manche unserer dabei gewonnenen Regeln infolge wesentlicher Änderungen von Hard- und/oder Software wieder neu überdacht und verändert werden.

So hat die Verfügbarkeit über preiswertere Plattenspeicher mit erheblich verbesserter Speicherkapazität einerseits und die immer aufwendiger werdende Kommunikation zwischen Benutzer- und Verwaltungssystem bei den neuen Betriebssystemen andererseits dazu geführt, vom Konzept der tieferen Strukturierung mit feiner Segmentierung wieder abzugehen. Es wird dabei zwangsläufig mit größeren Informationseinheiten (Segmenten) gearbeitet, die jedoch oft nicht voll genutzt werden und entsprechend mehr externen Speicherplatz benötigen. Der Mehraufwand bei der Datenbereitstellung im Anwenderprogramm zwischen einem größeren und einem kleineren Segment ist verschwindend gering gegenüber dem zweimaligen Kommunizieren zwischen Anwender- und Kontrollprogramm.

Ähnliche Einsparungen erlauben die im Laufe der Weiterentwicklung von IMS eingeführten Syntaxverbesserungen. Während früher im Regelfall mehrere Segmente angefordert wurden und die Auswahl im Anwenderprogramm erfolgen mußten, erlauben es jetzt die booleschen Verknüpfungen verschiedener Kriterien in den Suchanweisungen, die gewünschten Informationen mit weniger Aufrufen vom System auswählen zu lassen.

In der Organisationsform der IMS-Dateien streben wir heute überwiegend Verfahren an, bei denen zum Auffinden des Satzes nicht mehr das aufwendige Durchsuchen der Indextafeln erforderlich ist, sondern durch ein Umrechnungsverfahren aus dem Sortierschlüssel eine direkte Adresse ermittelt werden kann. Allerdings ist es oft schwierig ein Verfahren zu finden, das gleichmäßig über den gegebenen Bereich verteilt. Dies gilt besonders für sogenannte sprechende Schlüssel, die keinerlei Rücksicht auf eine Speicherorganisation nehmen. Obwohl bei diesen Umrechnungsverfahren in der Regel die Sortierfolge verloren geht, interessiert uns der schnellere Zugriff für die Echtzeitverarbeitung erheblich mehr als der vermehrte Aufwand für ein gelegentliches sequentielles Verarbeiten dieser Datenbestände.

Heute sind bei HOECHST ca. 60 % aller online-Dateien nach direkten Zugriffsverfahren organisiert. Die restlichen Dateien konnten wegen ihrer Schlüsselstruktur und häufigen sequentiellen Verarbeitung noch nicht umgestellt werden. Problematisch bei den indexorientierten Verfahren sind Neuzugänge, da IMS für sie Überlaufketten bildet, was die Performance ganz erheblich senkt. Gerade bei einem online-System werden die neuen Sätze in mehreren Phasen geprüft, verarbeitet und weitergeleitet, wobei jeweils ein aufwendiges Lesen erforderlich ist. In jeder Nacht reorganisieren wir die meisten dieser Dateien, wobei benötigte Auswertungen und Statistiken erstellt werden und als Datensicherung eine Kopie anfällt.

VSAM als verbesserte indexorientierte Zugriffsform des Betriebssystems wird zur Zeit bei uns getestet. An einen produktiven Einsatz ist aber erst zu denken, wenn wir vom sicheren fehlerfreien Funktionieren im Zusammenspiel mit IMS überzeugt sind.

Durch praktische Erfahrungen wurde auch ein gewisser Wandel bei der Gestaltung umfangreicher zentraler Datenbanken ausgelöst; speziell dann, wenn von verschiedenen Benutzern oft recht unterschiedliche Anforderungen hinsichtlich Inhalt und Umfang aufzunehmender Daten gestellt werden. Die in solchen zentralen Stammdatenbanken, wie beispielsweise der Kunden- und Lieferantendatenbank, für einzelne spezielle Abwicklungs- oder Abrechnungsprogramme zu speichernden Informationen, können häufig über das hinausgehen, was für die restlichen Benutzer jemals von Bedeutung sein kann.

Wir hatten dieses Problem vor allem bei typischen branchenbezogenen Daten in unserer Kundendatei, die beispielsweise bei Arzneimittelkunden zu einer völlig anderen Ausfüllung des einheitlichen Strukturrahmens führte als bei Kunden des Industriebereichs. Hinzu kommen häufig abweichende Anforderungen hinsichtlich Aktualisierung der Informationen, Zuständigkeit im Änderungsdienst und Aufnahme neuer Daten, wodurch trotz allen Komforts der Datenbankverwaltungssysteme doch immer wieder Unruhe auch in diejenigen Benutzergruppen solcher Datenbanken getragen wird, die primär von der Erweiterung oder einer kleineren Änderung in der Struktur nicht betroffen sind.

Aus überwiegend pragmatischen Gründen haben wir uns daher in einigen Fällen weniger an die reine Theorie eines universell verwendbaren und redundanzfreien Datenbankkonzeptes gehalten und mehr die flexible und benutzerfreundliche Handhabbarkeit sowie ein sicheres Verhalten der Datenbank in der Systemumwelt in den Vordergrund unserer Überlegungen gestellt.

Darum wurden einige bisher schon recht komplexe Datenbanken nicht mehr in dem an und für sich durch neu hinzukommende Anwendungsgebiete erforderlich werdenden Umfang erweitert. Wir gingen nun verstärkt auf das Prinzip der Auslagerung spezieller Daten in dedizierte Dateien über. Diese Subdateien bleiben logisch in gewissem Umfang noch von der Mutterdatenbank abhängig, weil der Stammteil der Informationen von dorthin eingespeist wird. Auf der anderen Seite müssen veränderte Daten aus der Subdatei völlig unabhängig von der Mutterdatenbank übernommen werden können. Physisch werden diese Subdateien völlig unabhängig von der Mutterdatenbank geführt und erhalten dort im Rootsegment lediglich Vermerke über Eröffnung, Änderungshinweise oder Löschungen. Logisch bleiben sie dadurch voneinander abhängig, denn ohne einen bereits eröffneten Stammteil in der Mutterdatei kann auch die Subdatei nicht eröffnet werden. Ebenso dürfen Basisdaten in der Mutterdatei nicht gelöscht werden, solange entsprechende Daten in den Subdateien noch benötigt werden.

Derartige Aufteilungen einer Datenbank in Mutterdatei und Subdateien können nicht nur aus organisatorischen Gründen erfolgen, sondern müssen auch durch die

Verhaltensweise der Hard- und Softwaresysteme in Erwägung gezogen werden; denn bei sehr viele Anwendungsgebiete umfassenden Datenbanken kann es schon allein durch die Zugriffshäufigkeiten auf den gleichen Plattenstapel zu erheblichen Engpässen kommen. Andererseits bereitet uns die temporäre Sperrung der Datenbanken während der Update-Vorgänge einige Zeitprobleme, vor allem dann, wenn die Datenbankänderungen sinnvollerweise im Stapelbetrieb auf einem anderen Rechner durchgeführt werden.

Durch die Aufteilung in dedizierte Subdateien ermöglichen wir es jedoch, daß im Bereich der voneinander unabhängigen Daten einer logischen Gesamtdatenbank verschiedene Anwendungsprogramme weitgehend ungestört und gleichzeitig arbeiten können. Synchronisationspunkte und ein aufeinander abgestimmtes System von Hinweisvermerken in den Einzeldateien sorgen dann dafür, daß der Gesamtzusammenhang der Datenbank erhalten bleibt.

Ein anderer in der praktischen Arbeit nicht zu unterschätzender Vorteil von dedizierten Subdatenbanken liegt in der erheblich besseren Reaktionsfähigkeit bei Fehlersituationen oder sonstigen Störungen im System. So nehmen wir ggf. innerhalb dieses Subdatenbanksystems lieber eine gewisse Datenredundanz in Kauf, als daß wichtige Programme im Falle eines Ein-/Ausgabefehlers oder anderer technischer Behinderungen im Zugriff zu lange auf die Durchführung umfangreicher Wiederherstellungsmaßnahmen für die Gesamtdatenbank warten müssen. Dies gilt im Prinzip auch dann, wenn die nachts im Stapelbetrieb durchzuführenden Änderungs- und Reorganisationsläufe zentraler Datenbanken eine Störung im Ablauf erfahren und nicht mehr rechtzeitig bis zum Anlaufen des Echtzeitbetriebs bereitgestellt werden können. Zugehörige Subdatenbanken hingegen bleiben von diesen Störungen oft völlig unberührt oder können auch ohne die anstehenden Datenbankänderungen weiterhin benutzt werden.

Andererseits können aber auch Auswirkungen von Programmzusammenbrüchen auf ein Gesamtdatenbanksystem durch dedizierte Subdateien geringer gehalten werden. Man muß dann nicht unbedingt die Gesamtdatenbank und damit alle tangierten Anwendungsprogramme stoppen, sondern kann wegen geringer gegenseitiger Abhängigkeiten viel gezieltere Maßnahmen zur schnellen Behebung der Fehlersituation einleiten.

Gerade dem Problem der Vermeidung einer gewissen Anfälligkeit gegenüber Störeinflüssen, sowohl durch die Benutzersysteme als aber auch nach wie vor durch Soft- und Hardware, wird heute vom Hersteller noch viel zu wenig Aufmerksamkeit geschenkt; denn welche Vorteile soll eine theoretisch sehr sinnvoll strukturierte und auf alle Informationsbelange eingerichtete Datenbankorganisation bringen,

wenn diese nicht absolut benutzungsfreundlich und zuverlässig angelegt ist? Das Ziel integrierter Datenbank- und Informationssysteme muß es vielmehr sein, daß, bei aller wünschenswerten Bewahrung der Gesamtzusammenhänge, die Teilsysteme in ihrer speziellen Funktion von Störungen verbundener Systeme unbeeinträchtigt und soweit wie nur möglich operationsfähig erhalten bleiben.

Da Auswirkungen durch Aufbau und Anwendungen von Datenbanken sowohl hinsichtlich des Systemverhaltens einzelner Anwendungsprogramme als auch im Hinblick auf die Belastung für das Gesamtsystem und die grundsätzlichen Verfügbarkeitsaspekte nur noch aus übergeordneter Sicht beurteilt werden können, wurde bei HOECHST eine spezielle Koordinationsstelle eingerichtet. Diese hat zur Aufgabe, bereits während der Planungsphase sowohl die datenbanktechnischen Gesamtaspekte zu beachten und auf das einzelne Datenbankdesign entsprechend einzuwirken, als auch durch Empfehlung geeigneter Strukturierungen der Anwendungsprogramme zu einem zeitlich und sicherheitsmäßig günstigen Ablauf im Gesamtsystem rechtzeitig beizutragen.

Darüber hinaus werden von diesen Spezialisten allgemein gültige Normen und Anwendungsbeispiele für Datenbankdesign erarbeitet und in jeweils geeigneter Form durch Merkblätter und Informationsseminare den Anwendern von Datenbanksystemen zugänglich gemacht.

HILFSMITTEL FÜR DATENBANKDESIGN UND -VERWALTUNG

Während wir uns in der Vergangenheit sehr tastend und mit teilweise recht aufwendigen Testversuchen an eine endgültige Struktur einer Datenbank heranbewegt haben, bemühen wir uns heute beim Design sowohl mehr um die Anwendung generell gesicherter Erkenntnisse aus der praktischen Erfahrung als auch um den Einsatz geeigneter Hilfsmittel für eine wirksame Datenbankmodellierung.

Einerseits helfen uns zur Erreichung dieses Zieles eine Reihe von Hilfsprogrammen, die rein organisatorisch die Struktur der Datenbank und deren Inhalt transparenter gestalten und als Modellierungshilfe sehr einfach notwendige Veränderungen im Design ermöglichen; andererseits können zusätzlich noch Programme zur Simulation der Datenbanken eingesetzt werden, die eine geeignete Struktur unter dem Gesichtspunkt der Mengengerüste, der Zugriffshäufigkeit und der Kombinationsfähigkeit der Datenelemente herausfinden helfen.

Allerdings möchte ich einschränkend sagen, daß wir umfassende Simulationen wegen der sehr aufwendigen Vorarbeit für die Beschaffung der quantitativen Angaben und

wegen des Aufwandes für die Beschreibungen der vielfältigen Zugriffsfunktionen seitens der Benutzerprogramme bisher noch nicht durchgeführt haben. In der Zukunft jedoch werden zuverlässigere Planungen unter Anwendung verbesserter Simulationsverfahren schon deshalb unerlässlich werden, weil die Datenverarbeitung nicht nur wegen des allgemeinen Kostendrucks, sondern auch wegen der überhöhten Systembeanspruchung durch spezielle Anwendungsprogramme nicht ständig das Gesamtsystem erweitern kann.

Andere, ganz dringend notwendige Hilfsmittel, sowohl für die Design-Phase als auch für die laufende Verwaltung der Datenbanken, sind geeignete Dokumentationsprogramme. Ohne derartige, im englischen Sprachraum mit "Data Dictionary and Directories" bezeichnete Systeme kann man eine effektive Datenbankplanung und einen laufenden Überblick über Struktur, Querverbindungen im Daten- und Benutzerbereich sowie über den jeweiligen Status der Datenbank nicht mehr zuverlässig erreichen.

Umso mehr müssen wir es als Anwender komplexer Datenbanksysteme bedauern, daß bisher vom Anbieter der Datenbanksysteme dieses schwierige Problem der Datenbank-Dictionary-Systeme so sehr schleppend bearbeitet wurde und die Benutzer meist eigene und wegen des hohen Aufwandes oft unzureichende Teillösungen für ihre Datenbankdokumentation und Administration erarbeiten mußten. In dieser, für eine weitere und gesicherte Fortentwicklung von Anwendungen mit Datenbanken so entscheidenden Frage müssen wir an IBM die dringende Aufforderung richten, die Kunden in ihrer Datenbankverwaltungsarbeit durch ein umfassenderes und benutzerfreundliches "Data Dictionary System" zu entlasten und zu einer weitgehend maschinellen Dokumentation der eingesetzten Datenbanken beizutragen.

Neben diesen Administrationshilfen für eine leichtere Gestaltbarkeit und Verwaltung von Datenbanksystemen ist auch der permanente Einsatz von Hilfsprogrammen zur Beobachtung des arbeitenden Systems und zur Auswertung statistischer Kenngrößen unerlässlich, um dadurch sowohl die Arbeitsweise einzelner Programme als auch das gesamte Systemverhalten beurteilen und anpassen zu können. Hierfür können wir aber auf ausreichende Daten aus IMS und SMF zurückgreifen und geeignete Monitoren zu deren Auswertung einsetzen.

Geringe Belastung durch einzelne Anwender führt bei IMS wegen der Verzahnung der Abläufe innerhalb der online-Kontrollregion zu einem insgesamt besseren Performance-Verhalten. Es ist daher erforderlich, sowohl das Verhalten einzelner Programme als auch ihr Zusammenspiel miteinander zu überprüfen. Dazu benutzen wir Programme, die auf der Auswertung von Logbandsätzen basieren. Die Tagesstatistik

zeigt die Aktivitäten eines Programmes während eines ganzen Tages. Daraus kann man erkennen, ob einzelne Programme im Verhältnis zu den verarbeiteten Nachrichten eine überdurchschnittliche Rate von Datenbankzugriffen haben.

Die Tagesstatistik ermittelt keinen Eindruck vom Verhalten eines Programms innerhalb einer gewissen Umgebung (z. B. zu Spitzenzeiten), von der Reihenfolge der Aktivitäten innerhalb eines Programmdurchlaufs sowie von der Dauer einzelner Datenbankzugriffe. Für diese Zwecke gibt es den DC-Monitor, der auf besondere Anforderung entsprechende Informationen mitschreibt. Systemspezialisten werten die damit gewonnenen Listen aus und können den verantwortlichen Programmierer zu geschickteren Datenbankaufrufen veranlassen bzw. allgemeine Richtlinien herausgeben. Wir haben auf diesem Wege schon wesentliche Verbesserungen im Ablauf der Programme erreichen können. Es ist bei einem online-System, in dem ca. 40 Anwendungsprogramme miteinander konkurrieren, natürlich kaum möglich, gleiche Konstellationen oder Ablaufreihenfolgen zu wiederholen. Der Einfluß von kleinen Änderungen kann daher meist nicht in exakten Zahlen ausgedrückt werden. Bewertungsmaßstäbe sind daher allenfalls Gesamtzahl der Zugriffe, insgesamt verbrauchte Rechnerzeit, Durchsatzrate von Nachrichten zu bestimmten Tageszeiten usw.

GEGEBENE UND NOTWENDIGE KOMMUNIKATIONSFORMEN

Aus den bisherigen Ausführungen war zu erkennen, daß die entscheidende Notwendigkeit zum Aufbau einer umfassenden Datenbankorganisation durch die Echtzeitprojekte ausgelöst wurde. Echtzeitverarbeitung mit Öffnung der Datenverarbeitung unmittelbar hin zum Arbeitsplatz des eigentlichen Systembenutzers, der in einer interaktiven Betriebsweise mit dem System und seinen Datenbanken kommunizieren soll, erfordert aber die Anwendung eines umfassenden Nachrichtensteuerungs- und Verwaltungssystems.

Die wesentlichen hier vorgestellten Anwendungen sind von ihrem Typ her sogenannte Teilhabersysteme, bei denen in einem vorgegebenen Anwendungssystem aufgrund der einzelnen vom Benutzer ausgelösten Transaktionen fest zugeordnete Prozeduren aktiviert werden. Für diesen Typ der Nachrichtensteuerung und Programmkontrolle bietet IMS mit seinem Datenkommunikationsteil die für ein Informationssystem notwendige Ergänzung des Datenbankteils. Auch die für das Gesamtsystem des IMS vorhandenen Sicherheitseinrichtungen mit einem umfangreichen Logging sowohl der Nachrichten als auch der datenbankwirksamen Aktivitäten und einem wirksamen Prüfpunkt und Wiederanlaufverfahren, bestärken uns zusätzlich in der Ansicht, daß wir mit IMS im Prinzip das richtige Softwareprodukt für unsere Informationssysteme zur Verfügung haben.

Richtigerweise ist die Systemsteuerung von IMS recht umfassend ausgelegt, so daß wir inzwischen unter dessen Kontrollprogramm nicht nur die transaktionsbedingten "Message Control Programme" laufen lassen, sondern auch stapelorientierte Datenfernverarbeitungsprogramme, das GIS und das "Information Retrieval System" STAIRS mit umfangreichen Dokumentationsdatenbanken zur Anwendung bringen.

Das erwähnte GIS hat allerdings zur Zeit noch eine völlig untergeordnete Bedeutung und soll erst nach Ablauf einer erfolgreichen Erprobungszeit in zukünftige Planungen einbezogen werden. Trotzdem können wir schon aufgrund der ersten Probenanwendungen erkennen, daß es eine interessante Ergänzung zu den Datenbanksystemen darstellen kann. Ob es allerdings voll geeignet ist, um unmittelbar vom Endbenutzer sporadisch auftretende Anfragen an bestehende Datenbanken schnell formulieren zu lassen, scheint noch ungewiß. Es wäre unseres Erachtens besser, für die Endbenutzer eine einfachere und in deutscher Sprache formulierbare Abfragesprache zu haben und dafür GIS für erfahrene Benutzer noch weiterhin auszubauen, um beispielsweise auch durch Feld- und variable Indizierung noch bequeme Anfragen an IMS-Datenbanken richten zu können. Dabei wäre es ebenfalls von Vorteil, wenn aus den vorhandenen IMS-Datenbankbeschreibungen auch automatisch die IMS-Dateibeschreibung erzeugt würde, oder aber durch ein übergeordnetes Datenbankmanagement IMS und GIS gemeinsam bedient würden.

Das hier ebenfalls erwähnte STAIRS wird bei HOECHST als umfassendes Dokumentations- und Information Retrieval System eingesetzt. Unter der Nachrichten- und Programmsteuerung von IMS wird STAIRS bisher für umformatierte Datenbanken im Bereich der medizinischen Literaturdokumentation, der Forschungsdokumentation und zur Patentedokumentation eingesetzt. Sämtliche Fragestellungen und Suchvorgänge in zur Zeit auf 16 Magnetplattenspeicher IBM 3330-11 gespeicherten Dokumenten erfolgen in Echtzeitverarbeitung über Bildschirmterminals unmittelbar im System-Benutzerdialog.

Die bisher erläuterte umfassende Nutzung des IMS als Gesamtsteuerungssystem bringt allerdings ein überproportionales Ansteigen der CPU-Belastung durch vermehrten systeminternen Verwaltungsaufwand mit sich, so daß wir uns jetzt Gedanken machen, bis zu welchem Zeitpunkt das IMS bei unserer geplanten Vermehrung der Datenstationen - selbst bei einer /370-168 - noch in der Lage sein kann, alle Anforderungen in einem System zu bedienen.

Sollte demzufolge die Leistungsfähigkeit und die jetzige Arbeitsweise des IMS seitens IBM nicht entscheidend geändert werden, dann bliebe nur eine verhältnismäßig unwirtschaftliche Aufteilung der IMS-Anwendungen auf zwei Systeme. Dem

sind allerdings sowohl durch die Datenbankverwaltung des IMS als auch durch die Anwendungssysteme eindeutige Grenzen gesetzt. Ein wesentlich sinnvoller Weg scheint uns hingegen in der Realisierung einer Konzeption zu liegen, die eine Auslagerung geeigneter Funktionen in intelligente und mit dedizierten Dateien ausgestattete Datenstationen ermöglicht. Dadurch kann das Zentralsystem sowohl von vermeidbaren Transaktionen entlastet werden als aber auch die Funktionsfähigkeit des Systems durch eine zumindest temporär mögliche unabhängige Arbeit an den peripheren Datenstationen verbessert werden.

In diese Richtung weisende Konzeptionen wurden ja auch beispielsweise mit den SNA-Systemen IBM 3790 oder auch 3770 angekündigt. Allerdings erwarten wir dann auch im IMS sowohl hinsichtlich seiner Datenbankkonzeption als auch bei der Nachrichtensteuerung eine volle Integration der Möglichkeiten dieser Terminalcomputer im Sinne eines echten hierarchisch gegliederten Systemverbundes. Wünschenswert wäre dann nämlich ein voller Einbezug der vom Vorrechner geführten Datenbestände in das Datenbankverwaltungssystem des IMS, so daß alle entsprechenden Dateiveränderungen in der Mutter-Datenbank auch sofort für die dedizierte Datei mit ausgelöst würden und umgekehrt.

Wenn wir die übliche Definition eines Teilhabersystems, bei dem verschiedene Teilhaber am System voneinander abhängig sind und über ein gemeinsames Informationssystem miteinander verbunden sind, hinsichtlich einer umfassenden Datenbankorganisation in unserem Unternehmen betrachten, dann sind viele Arbeiten, die im Remote Job Processing von den Werken aus im Rahmen des Systemverbundes betrieben werden, eher Teilhaber als Teilnehmersysteme. Obwohl die Datenübertragung stapelorientiert erfolgt und in der Regel auch völlig unabhängige Programme aufgerufen werden, gibt es doch auch im RJE-Betrieb viele Anwendungen, die auf gemeinsame zentrale Prozedur- und Datenbanken zugreifen.

Diese Arbeiten können derzeit aber nur im Rahmen der RJE-Prozeduren auf unserem ASP-System abgewickelt werden, das einerseits auf einer ganz anderen Anlage gefahren wird als das IMS, andererseits aber auch eine völlig andere Übertragungstechnik benutzt, so daß nicht einmal eine gemeinsame Leitungsbenutzung von IMS und RJE möglich ist, obwohl beide Systeme Bestandteile des gleichen Gesamtinformationssystems und der gemeinsamen Datenbankorganisation sind. Hier ist es sowohl aus wirtschaftlichen als auch aus organisatorischen Gründen dringend erforderlich, umgehend eine gemeinsame Leitungssteuerung und möglichst auch Datenbankverwaltung für IMS und RJE herbeizuführen.

Ein solches Paket von Verfahren für eine einheitliche Datenübertragungssteuerung

ist ja inzwischen von IBM als "System Network Architecture" angekündigt. Allerdings scheint uns diese Bezeichnung wenigstens bisher noch ein vielversprechendes Schlagwort zu sein, das vor allem hinsichtlich des Wortes "Network" noch mit sehr viel Inhalt ausgefüllt werden muß; denn wir benötigen in unserem Unternehmen im Rahmen des Systemverbundes nicht nur das einheitliche Konzept für die Datenübertragungssteuerung und die hierarchisch geordnete Kommunikation zwischen Rechner und untergeordneten Datenstationen, sondern wir erwarten vor allem aus Gründen einer erhöhten Sicherheit und Verfügbarkeit ein Netzwerksystem zwischen gleichberechtigten Systemen mit gemeinsam benutzbaren Programmbibliotheken und Datenbanken.

Ich hoffe, daß ich trotz des überwiegend allgemein gehaltenen und nur auf praktische Erfahrungen oder konkrete Planungsansätze bei HOECHST ausgerichteten Referates auch den anwesenden Wissenschaftlern und Software-Architekten einige Bestätigungen ihrer Auffassungen oder auch einige Anregungen zum Thema Datenbanken und Datenkommunikation geben konnte.