Datei vs. Datenbank

Datenbanken sind von zentraler Bedeutung für die Datenverarbeitung. Zum einen können viele verschiedene Programme mit dem gleichen Datenbestand arbeiten, und die Programme können gewartet und verändert werden, ohne dass die Daten ebenfalls neu organisiert werden müssen. Zum anderen werden die Anwendungen unabhängiger von dem Aufbau und der Struktur der Datenbestände.

In den 60er Jahren beherrschte vorwiegend eine verarbeitungsorientierte Sicht die Entwicklung von Anwendungssystemen, Die verarbeitenden Elemente, die Programme, benutzen Eingabedaten aus Dateien und liefern Ausgabedaten, die ihrerseits in Dateien abgespeichert werden und nachfolgenden Programmen als Eingabedaten dienen. Die Dateien sind in Form und Inhalt speziell auf die beteiligten Programme abgestimmt. Abb. l zeigt die allgemeine Struktur von Anwendungssystemen aus verarbeitungsorientierter Sicht. Abb. 2 enthält als Beispiel die Struktur eines Gehaltsabrechnungsprogrammsystems.

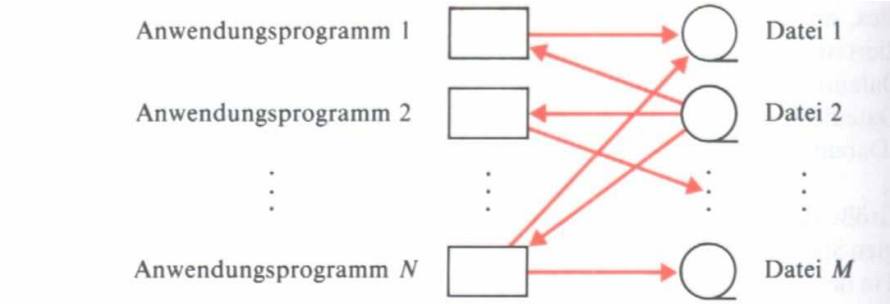


Abb. 1: Verarbeitungsorientierte Sicht eines Anwendungssystems

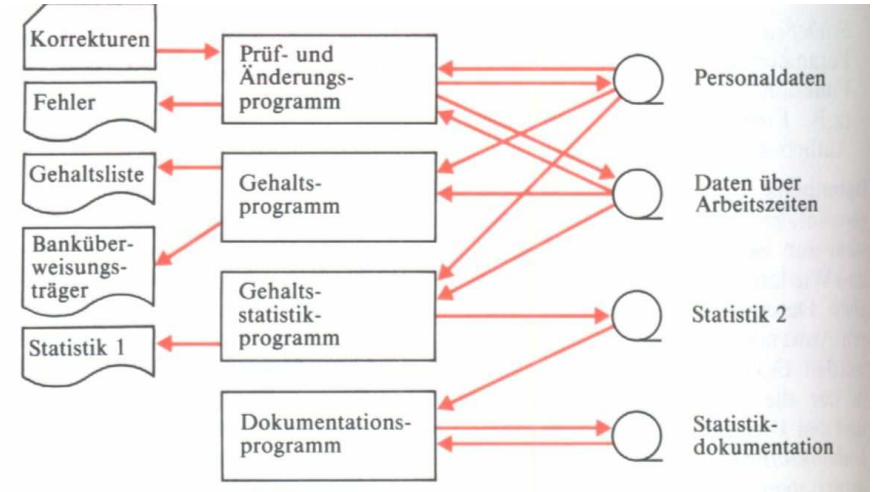


Abb. 2: Gehaltsabrechnungsprogrammsystem (verarbeitungsorientierte Sicht)

Obwohl dieses Konzept weit verbreitet ist, hat es einige gravierende Nachteile:
1) Viele Daten sind gleichzeitig in mehreren Dateien gespeichert (**Redundanz**), z.B. ist der Name eines Mitarbeiters in Abb. 2 sowohl in der Personaldatei als auch in der Datei über die Arbeitszeiten enthalten. Dadurch wird zusätzlich Speicherplatz benötigt und die Aktualisierung der Daten erschwert. So muss eine Namensänderung in allen Dateien gleichzeitig durchgeführt werden. Wird hierbei irrtümlich (z. B. durch einen Schreibfehler) in einer Datei etwas falsch geändert, so ist die gesamte Datenmenge inkorrekt (**inkonsistent**).
2) Bei jedem Programmablauf ist immer der gesamte Inhalt der bearbeiteten Dateien beteiligt (auch

die nicht benötigten Daten).
Beispiel: Eine Datei für ein Personalverwaltungssystem bestehe aus Datensätzen der Form

Pers.-Nr.	Name	Adresse	Chef	Eintrittsdatum	Abteilung

Ein Gehaltsabrechnungsprogramm kann sich derselben Datei bedienen, benötigt dann aber nur die rot markierten Felder. Trotzdem muss das Vorhandensein der anderen Felder programmtechnisch berücksichtigt werden. Dabei kann es auch zu Datenschutzproblemen kommen, etwa wenn nicht benötigte Felder als vertraulich eingestuft sind.
3) Der Aufbau einer Datei muss jedem Programm, das diese Datei benutzt, bekannt sein. Änderungen der Dateistruktur erfordern gleichzeitig Änderungen in allen beteiligten Programmen.
4) Maßnahmen zum Schutz von Daten (z. B. vor unbefugtem Zugriff) müssen in allen Programmen des Anwendungssystems einzeln realisiert werden.

Die Lösung dieser Probleme bildet das Konzept der Datenbank, in der alle Daten des Anwendungssystems zusammengefasst und nach einheitlichen Regeln abgespeichert werden (**Datenintegration**). Der Zugriff der Anwendungsprogramme zu den Daten erfolgt stets über ein Kontrollprogramm. Kontrollprogramm und Datenbestand bilden zusammen die Datenbank. Das Kontrollprogramm nennt man **Datenbankmanagementsystem**, den Datenbestand **Datenbasis**.
Abb. 3 zeigt die allgemeine Struktur eines auf einer Datenbank aufbauenden Anwendungssystems, Abb. 4 die Realisierung des Gehaltsabrechnungsprogrammsystems unter Verwendung einer Datenbank.

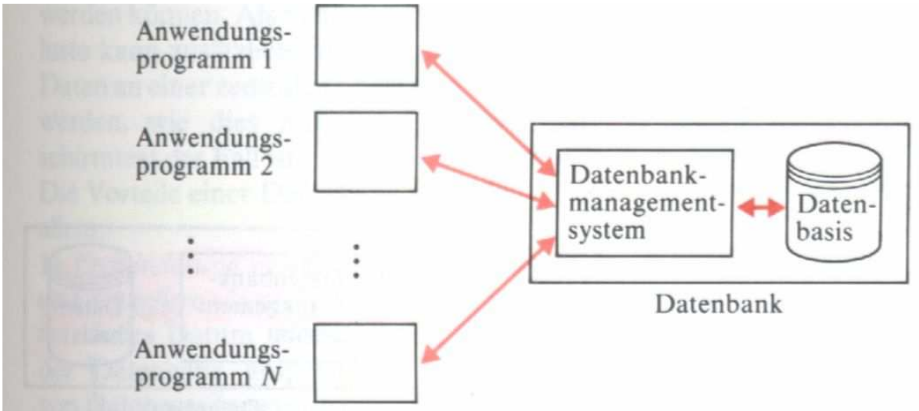


Abb. 3: Struktur eines auf einer Datenbank aufbauenden Anwendungssystems

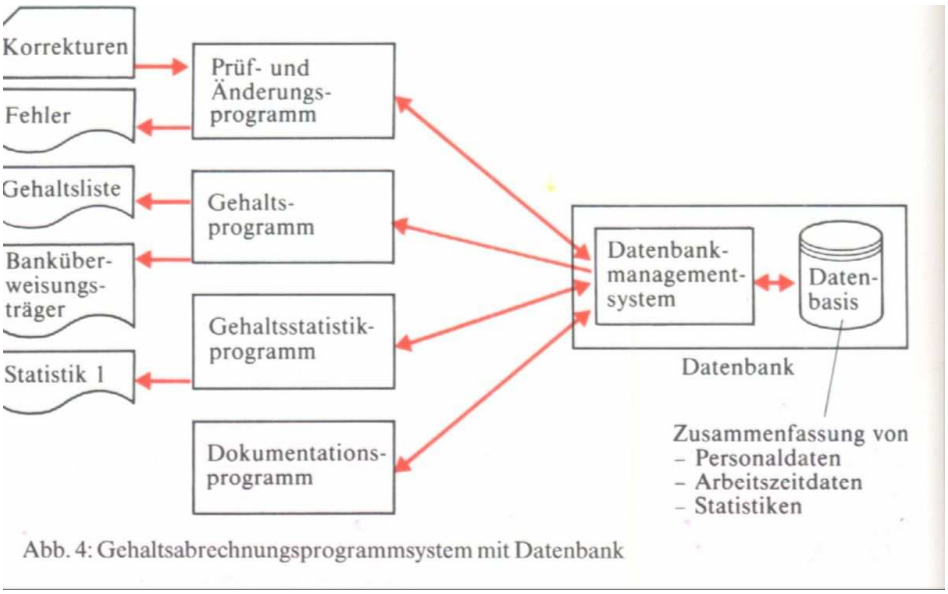


Abb. 4: Gehaltsabrechnungsprogrammsystem mit Datenbank

Die Vorteile einer Datenbank sind vor allem:
1) *Verminderung der Datenredundanz und schnellere Aktualisierung der Daten*: Jedes Datum wird nur einmal in der Datenbank gespeichert. Kopien von Daten werden einzig und allein aus Sicherheitsgründen gehalten und nicht etwa. weil zwei verschiedene Anwendungsprogramme das Datum in zwei verschiedenen Darstellungen benötigen
2) *Einhaltung der Datenintegrität*: Die einheitliche Kontrolle erlaubt eine einfachere Überprüfung der Daten auf Korrektheit und Vollständigkeit (**Datenkonsistenz**). Bei Veränderungen werden notwendige Folgeänderungen automatisch vorgenommen, so dass der Datenbestand möglichst keine sich widersprechenden Informationen enthält
3) *Verbesserter Schutz der Daten vor unberechtigtem Zugriff*: Grundsätzlich fördert der einheitliche kontrollierte Zugang aller Anwender zur Datenbasis die Einhaltung von Datenschutzvorschriften. Durch die Integration und Zentralisierung der Daten verbunden mit der Möglichkeit, auf jedes beliebige Datum in Sekundenschnelle zugreifen und Querbeziehungen zwischen Daten herstellen zu können, entstehen neue Einsatzbereiche, aber auch neue Datenschutzprobleme (Beispiel: Rasterfahndung der Polizei).
4) *Datenunabhängigkeit*: Die Änderung der physikalischen Organisation der Daten erfordert keine Änderung der Anwendungsprogramme.

(Quelle: Duden Informatik, Dudenverlag 1993)