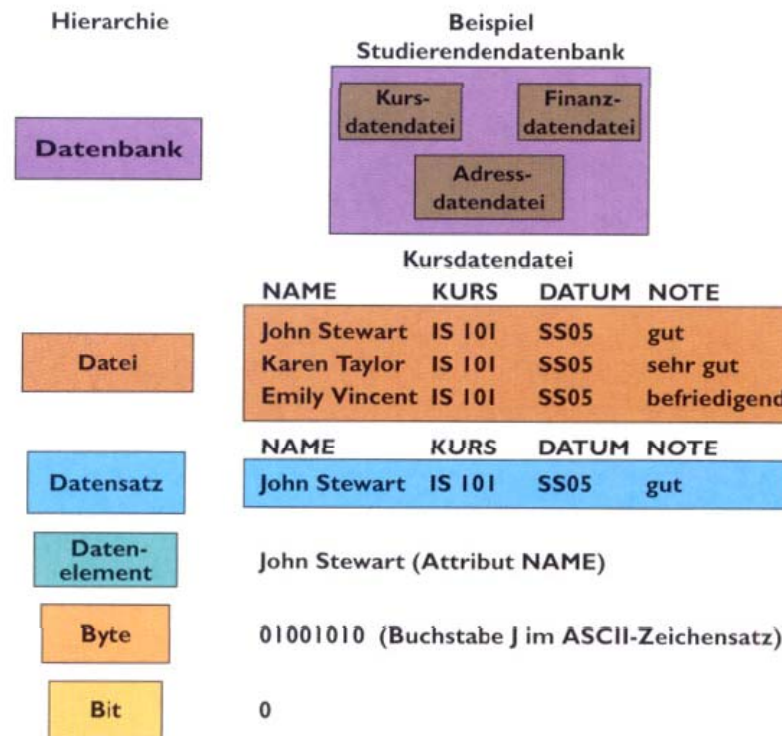


# IV. Datenbankmanagement

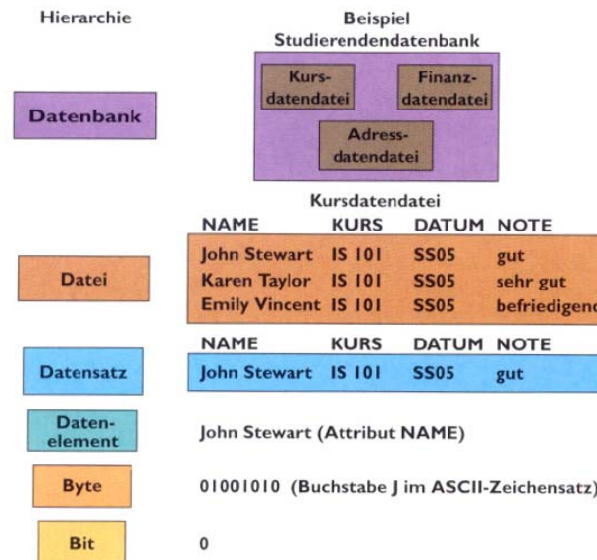
## Kapitel 1: Dateiorganisation und Datenbankansatz

- 1. Grundlagen der Dateiororganisation
- 2. Dateiansatz
- 3. Datenbankansatz



- Dateien werden in einer Hierarchie organisiert, deren unterste Ebene Bits bilden.
- Ein Bit repräsentiert entweder eine Null oder eine Eins.
- Buchstaben oder Ziffern werden durch 8 Bits (Byte) codiert.

- Ein Datenelement wird aus einer Folge von Bytes aufgebaut.
- Miteinander in Beziehung stehende Datenelemente können zu einem Datensatz gruppiert werden.
- Logisch zusammengehörige Datensätze können in einer Datei zusammengefasst werden.
- Zusammengehörige Dateien lassen sich in einer Datenbank zusammenstellen.



Quelle: Laudon, K.C., Laudon, J.P., Schoder, D. (2006)

- Weitere Grundbegriffe:

- Datenelement

- o Gruppierung von einem oder mehreren Zeichen zu einem Wort, einer Gruppe von Wörtern oder einer Zahl, z.B. Name.

- Attribut

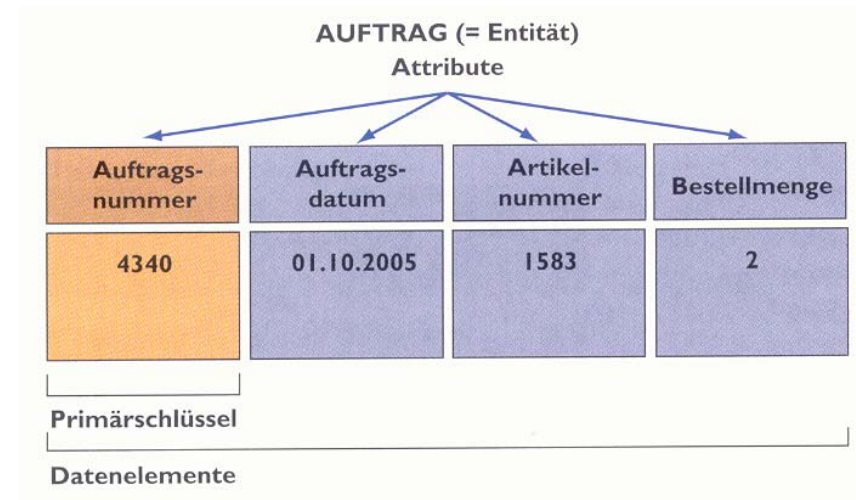
- o Eigenschaft, die eine bestimmte Entität beschreibt.

- Primärschlüssel

- o Eindeutig identifizierbares Attribut.

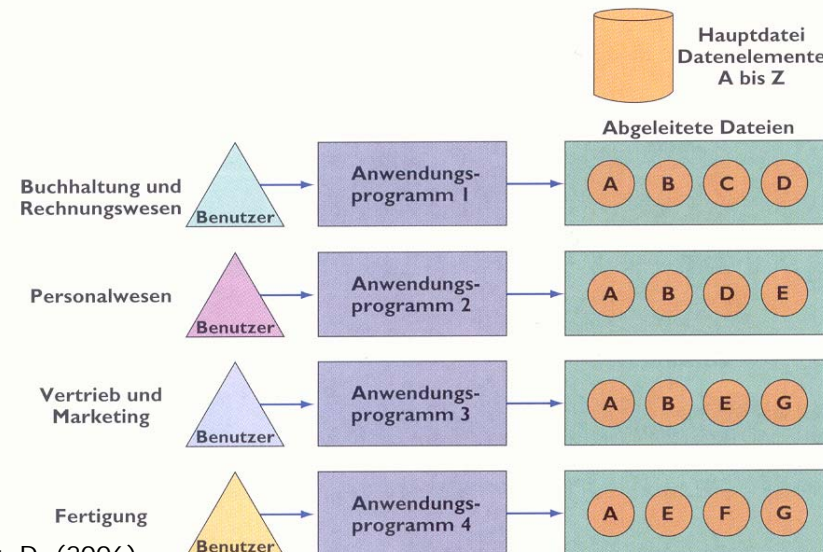
- Entität

- o Phänomen, z.B. eine Person, ein Ort, Ding oder Ereignis, über das Daten gespeichert werden sollen.



- 1. Grundlagen der Dateiorganisation
- 2. Dateiansatz
- 3. Datenbankansatz

- Einführung
  - Der Einsatz von Informationssystemen war ursprünglich nicht abteilungsübergreifend geplant.
  - Jeder Funktionsbereich entwickelte eigene Systeme, getrennt von anderen Funktionsbereichen
  - Die Systeme umfassten Anwendungen und die dazugehörige Dateiorganisation für die eigenen Dateien.



Quelle: Laudon, K.C., Laudon, J.P., Schoder, D. (2006)

- Beispiele:
  - o Der Funktionsbereich Personalwesen verfügt u.a. über eine Datei mit Personalinformationen, eine Lohnbuchungsdatei, eine Krankenversicherungsdatei und eine Rentenversicherungsdatei. Für die Verwendung dieser Dateien in anderen Anwendungen, mussten diese abgeleitet und neu aggregiert werden.
  - o Vorlesungsunterlagen zu Informationssystemen – Architekturansätze für Informationssysteme. Stichwort: Isolierte Informationssysteme
- Dieser Ansatz führt zu erheblichen Problemen:
  - o Redundanzen und Inkonsistenzen
  - o Abhängigkeiten zwischen Anwendung und Daten
  - o Fehlender Datenaustausch und mangelnde Flexibilität
  - o Mangelnde Datensicherheit

- Unter Datenredundanz wird das mehrfache Vorkommen identischer Datenelemente in verschiedenen Dateien verstanden.
  - Mehrfaches Speichern derselben Datei an unterschiedlichen Speicherorten (z.B. unterschiedliche Unternehmensbereiche erfassen oder ändern Daten unabhängig voneinander).
  - Folgen: Verschwendung von Speicherressourcen sowie möglicherweise Dateninkonsistenz
  
- Unter Dateninkonsistenz wird die Zuordnung unterschiedlicher Werte für dasselbe Attribut verstanden.

- In der Datei eines gespeicherten Auftrags kann die Bestellmenge eines Artikels bereits aktualisiert worden sein, während die Änderung an anderen Speicherorten nicht entsprechend nachvollzogen wurde und zugehörige Anwendungssysteme mit falschen Werten operieren.
- Unterschiedliche Funktionseinheiten können für dieselben Entitäten bzw. Attribute unterschiedliche Bezeichnungen und Codierungen verwenden.
  - o Produktnummer vs. Artikelnummer
  - o Extra groß vs. XL

- Bei einer Abhängigkeit zwischen einer Anwendung und den dazugehörigen Daten wirken sich Änderungen der Anwendung direkt auf die Daten aus.
  - Beispiel:
    - Eine Anwendung wird dahingehend geändert, dass sie anstatt 5-stelligen Postleitzahlen nun 9-stellige Postleitzahlen verarbeitet. In der ursprünglichen Datendatei sind daraufhin alle 5-stelligen Postleitzahlen durch 9-stellige Postleitzahlen ersetzt worden. Dies hat zur Folge, dass andere Anwendungen, die lediglich fünfstellige Postleitzahlen verarbeiten können, nicht mehr auf die Daten zugreifen können.

- In Dateien gespeicherte Daten sind unflexibel gegenüber
  - Auswertungen,
  - Verknüpfungenda der Zugriff auf mehrere Attribute in unterschiedlichen Dateien zeitaufwendige Such- und Sortierverfahren erfordert.
  
- Der Austausch von Daten zwischen verschiedenen Funktionsbereichen eines Unternehmens gestaltet sich schwierig und findet in der Regel nicht statt.

- Keine Kontrolle über die aufwändige Verwaltung von Dateien an mehreren unterschiedlichen Speicherorten einzelner Funktionsbereiche.
- Überblick über letzten Dateizugriff oder letzte Dateiänderung nicht möglich.

→ Vollständige Datensicherung nicht möglich.

- 1. Grundlagen der Dateiorganisation
- 2. Dateiansatz
- 3. Datenbankansatz

- Datenbanksysteme bestehen grundsätzlich aus einem Datenbankmanagementsystem (DBMS) und einer Gruppe von Daten (der Datenbank), zwischen denen logische Abhängigkeiten bestehen.
- Eine Datenbank ist eine Sammlung von Daten, die darauf ausgelegt ist, einen effizienten und gleichzeitigen Zugriff mehrerer Benutzer oder Anwendungen auf möglichst redundanzfreie Daten zu gewährleisten.

- Statt Daten für einzelne Anwendungen in verschiedene Dateien zu speichern, werden sie so gespeichert, dass es für den Benutzer scheint, als befänden sich die Daten an einem gemeinsamen Ort.
  
  - Beispiel:
    - Eine gemeinsame Datenbank mit Daten über den Mitarbeiter wie
      - o Personenummer, Arbeitsstunden, Stundenlohn, Steuerklasse, ...
      - o Arbeitslosen-, Renten-, oder Pflegeversicherung, ...
- kann von der Personalabteilung oder von der Lohnbuchhaltung abgefragt werden.

Quelle: Laudon, K.C., Laudon, J.P., Schoder, D. (2006)

- Eine spezielle organisatorische Funktion für die Verwaltung der Datenressourcen eines Unternehmens, die für die Richtlinien bezüglich der Datenplanung, der Datenerfassung, der Datenqualitätsstandards, der Pflege sowie der Nutzung und Weitergabe von Informationen zuständig ist.
- DBMS können nutzbringend eingesetzt werden, wenn
  - Unternehmen die strategische Bedeutung von Informationen erkennen,
  - Informationen als Ressource aktiv verwalten,
  - Informationen in ihren Planungen berücksichtigen.

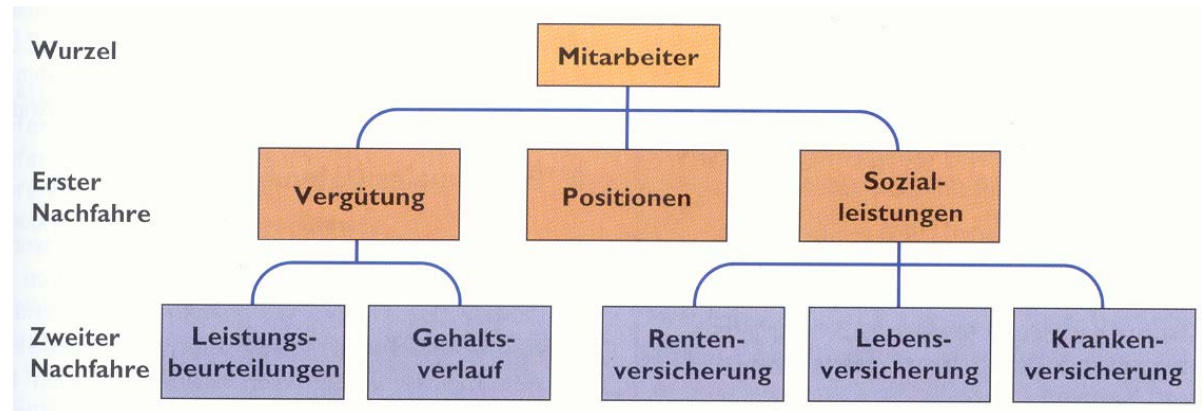
- Dazu werden folgende Festlegungen benötigt:
  - Planung der Datenerfassung
  - Koordination des logischen Datenbankentwurfs
  - Entwicklung von Richtlinien für die Informationsverwendung
  - Überwachung der Datennutzung

- Ein DBMS ist wie folgt definiert:
  - Sammlung von Programmen zum Erstellen, Verwalten und Nutzen einer Datenbank, die es mehreren Anwendungen gleichzeitig ermöglicht, die von ihnen benötigten Daten zu speichern, zu extrahieren und zu manipulieren, ohne jeweils eigene Dateien erstellen zu müssen
  
- Der Einsatz von Datenbanken setzt ein Datenmanagementsystem voraus.
  
- Ein DBMS trennt die physische und logische Datenstruktur.

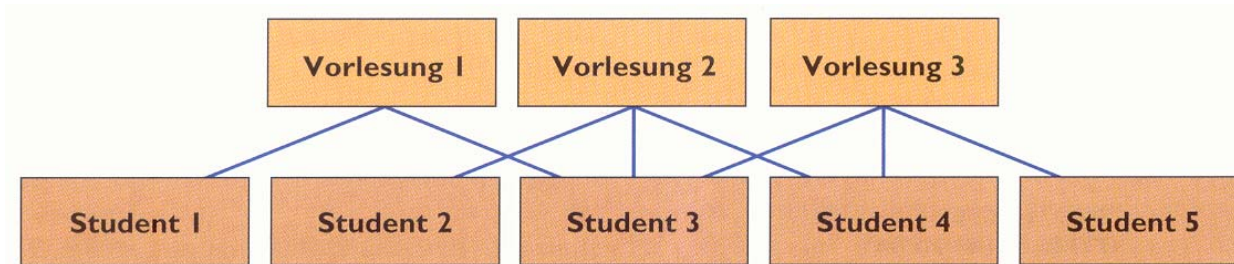
- **Physisch:**
  - Die physische Sicht zeigt, wie die Daten tatsächlich auf den physischen Speichermedium angeordnet sind.
- **Logisch**
  - Die logische Sicht beschreibt die Wahrnehmung des Anwenders durch die Verwendung logische Konzepte.
- Die logische Beschreibung einer Datenbank (sämtliche Datenelemente und deren Beziehungen untereinander) entspricht dem konzeptionellen Datenbankschema, während die Spezifikation, wie die im konzeptionellen Schema beschriebenen Daten auf den physischen Datenträger gespeichert werden, als physisches Schema bezeichnet werden.

- Zur Erstellung, Manipulation und Auswertung von Datenbanken stellen DBMS mehrere Komponenten bereit.
  - Datendefinitionssprache (DDL)
    - o Definiert die Struktur des Datenbankinhalts
  - Datenmanipulationssprache (DML)
    - o Dient zur Bearbeitung der, in der Datenbank enthaltenen, Daten (z.B. SQL)
  - Datenwörterbuch (DD)
    - o Hilfsmittel für die Verwaltung von Datenbeständen
    - o Generiert Berichte als Überblick über die, in einer Datenbank enthaltenen, Daten

- Es existieren unterschiedliche Ansätze zur Darstellung des konzeptionellen Datenbankschemas
  
- Zu unterscheiden sind:
  - Hierarchisches Datenmodell
  - Netzwerkdatenmodell
  - Relationales Datenmodell
  - Objektorientiertes Datenmodell



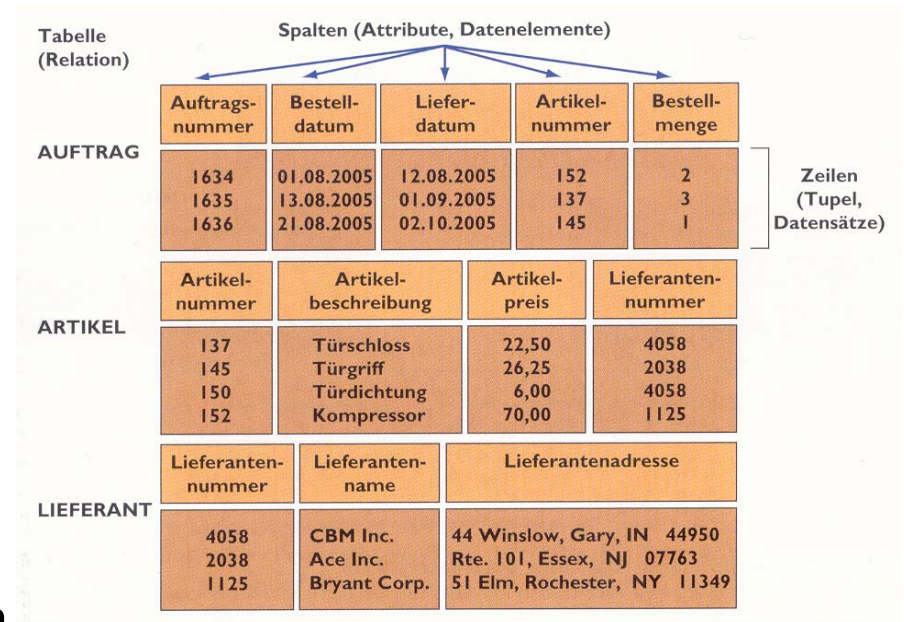
- Ähneln einem Organigramm oder Stammbaum
- Verfügt über ein Wurzelement, das mit untergeordneten Segmenten verbunden ist (Baumstruktur)
- 1:n Beziehung zwischen Entitäten
- Jeder Datensatz wird als Knoten repräsentiert.
- Die Kanten repräsentieren die Beziehungen.



- Repräsentation logischer n:m-Beziehungen
- Übergeordnete Entitäten können Beziehungen zu vielen untergeordneten Entitäten haben.
- Eine untergeordnete Entität kann zu vielen übergeordneten Entitäten Beziehungen haben.

- Hierarchische und Netzwerkdatenbank fanden früher Anwendung bei Banken und Versicherungen (rechenintensive und umfangreiche Transaktionen)
  - Sie sind unflexibel,
  - Erlauben keine Ad-hoc Abfragen,
  - Beginnen immer mit der Wurzel.
  
- Heute gelten sie als veraltet.

- Heute am weitesten verbreitetes Konzept zur Organisation von Datenbanksystemen
- Geeignet für Ad-hoc Abfragen
- Flexibel beim Hinzufügen neuer Daten bzw. dem Zusammenführen von Daten aus unterschiedlichen Quellen
- In relationalen Datenmodellen werden alle in einer DB enthaltenden Daten als Tabellen (Relationen) mit einer festen Anzahl an Spalten, aber variablen Anzahl an Zeilen dargestellt.



Quelle: Laudon, K.C., Laudon, J.P., Schoder, D. (2006)

- Die Spalten repräsentieren Datenelemente bzw. Attribute, welche die Entitäten beschreiben.
- Die Zeilen repräsentieren konkrete Datensätze. Die Bezeichnung für eine Zeile in einer Relation wird Tupel genannt.
  
- Relationen wären: Auftrag, Artikel, Lieferant
  
- Tupel wären:
  - 137; Türschloss; 22,50; 4058
  - 145; Türgriff; 26,25; 2038
  - ...

- Motiviert durch Schwächen des relationalen Datenmodells in
  - Ingenieurwissenschaftlichen Entwurfsanwendungen
  - Multimedia-Anwendungen
  - Architektur etc.
- Einzelne Beschränkungen relationaler Datenmodelle aufgehoben:
  - Z.B. Verwendung komplexer Objekte
    - o Definition komplexer Objekte, deren Attribute wieder aus Objekten bestehen können
    - o Attribute können einfache Datentypen sein oder sich aus beliebigen Datentypen zusammensetzen.
- Lösung: objektorientierte DBMS (OODBMS), wie beispielsweise Matisse, POET oder Object Store

- Die Aufgabe der Datenverwaltung ist die Koordination des Datenbankentwurfs. Datenbankadministratoren sind für die Definition und Strukturierung der Datenbankinhalte sowie für die Pflege und Verwaltung der Datenbank verantwortlich.
- Das Vorgehen wird in zwei Prozesse unterteilt:
  - Logischer Datenbankentwurf
  - Physischer Datenbankentwurf

- Der Entwurf einer Datenbank wirft einige Fragen auf:
  - Welche Beziehungen bestehen zwischen den Daten?
  - Welche Datentypen sind notwendig?
  - Wie sollen die gespeicherten Daten verwendet werden?
  
- Der logische Datenbankentwurf sieht ein abstraktes Modell der zu speichernden Daten vor. Dieses beinhaltet:
  - Relevante Entitäten
  - Beziehungen zwischen Datenelementen
  
- Werkzeug zur Modellierung: ER-Diagramme

Quelle: Laudon, K.C., Laudon, J.P., Schoder, D. (2006)

- Der physische Datenbankentwurf adressiert folgende Fragen:
  - Wie sollen die Daten physisch gespeichert werden?
    - o Zentrale Datenbank
    - o Verteilte Datenbanken
      - logisch zusammenhängend, physisch an unterschiedlichen Orten
    - o Fragmentierte bzw. partitionierte Datenbanken
      - Tupel einer Relation werden an unterschiedlichen physischen Speicherorten gespeichert.
    - o Replikation als Alternative zu Fragmentierten Datenbanken

- Entwicklung einer Datenbank erfordert neben dem konzeptionellen und konkreten Datenbankentwurf eine grafische Benutzeroberfläche (GUI):
  - Besteht aus Formularen zum Anzeigen und Eingeben von Daten
  - Ermöglicht unterschiedliche logische Sichtweisen auf eine physische Datenbank, z.B.
    - o Unterschiedliche Abteilungen (Personalwesen, Lohnabrechnung) greifen auf unterschiedlichen Daten zu.

- Brüggemann, T. (2003) „Datenorganisation“ Institut für Wirtschaftsinformatik, Wirtschaftswissenschaftliche Fakultät, Universität Hannover
- Laudon, K.C., Laudon, J.P., Schoder, D. (2006) „Wirtschaftsinformatik – Eine Einführung“