

	DB - allgemeine Informationen		AEuP	v 1.0
	Name	Klasse	Datum	

# 1 Die Notwendigkeit von Datenbanksystemen

Den Begriff *Datenbank* haben viele schon einmal gehört – auch wenn man nicht aus der IT-Branche kommt. Wir verbinden mit diesem Begriff eine Ansammlung von Informationen, welche wir in irgendeiner Form abfragen, bzw. nutzen können. Die meisten gehen auch davon aus, dass die Informationen „irgendwie“ in Tabellen abgelegt sind. Den Begriff *Tabellen* kennen viele aus sog. *Tabellenkalkulationsprogrammen*, von denen Microsoft Excel der bekannteste Vertreter ist.

Die Frage ist, warum Excel und Co. nicht für alle datenrelevanten Anwendungen genutzt werden. Schließlich kann man Daten in derartige Programme doch sehr einfach eintragen und wieder auslesen! Das Problem ist, dass man mit diesen Programmen einige Nachteile in Kauf nehmen muss:

- Sie manipulieren oft die Daten bei Eingabe, da sie eingabezentriert sind.
- Die Verknüpfungsfunktionen sind umständlich.
- Die Verknüpfungen sind inperformant.
- Die Datenmenge ist begrenzt.
- Es gibt keine Möglichkeit der Indizierung.
- Es gibt kein durchgängiges Usermanagement.
- Der parallele Zugriff ist nur bedingt möglich.



„Wenn ein Tabellenkalkulationsprogramm ein Schweizer Taschenmesser ist, dann ist eine Datenbank eine Fabrik.“

Streng genommen handelt es sich bei einem Tabellenkalkulationsprogramm auch um eine Datenbank, wengleich der Ansatz von einem klassischen Datenbanksystem sich in wesentlichen Teilen unterscheidet.



Tabellenkalkulationsprogramme sind für größere Systeme nicht geeignet. Hier sind wir auf Datenbanksysteme DBMS angewiesen.

# 2 Definition der Grundbegriffe

.....

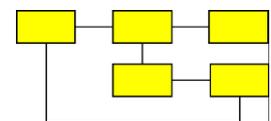
Meist die Bezeichnung des gesamten Systems. Streng genommen ist die Datenbank jedoch die Speicherstruktur, in der die Daten (oder auch Datenbasis) abgelegt werden. In MySQL finden wir diese meist im Ordner „Data“.

.....

Dies ist die Software, mit der die Datenbasis verwaltet wird. In der Regel handelt es sich hier um ein Serverprogramm, welches die Daten für mehrere Clients gleichzeitig zur Verfügung stellt.

.....

Bei der Relationalen Datenbank werden die Daten in Tabellen abgelegt, welche wiederum in Beziehung zueinander stehen. Die einzelnen Datensätze bilden die Zeilen in den Tabellen, die Attribute (Eigenschaftstypen) werden durch die Spalten abgebildet. Über die Attribute werden auch Beziehungen zwischen den Tabellen darstellbar. Der Zugriff auf die Relationalen Datenbanken erfolgt fast ausschließlich über SQL (siehe unten „Datenbanksprache“).



.....

Alle nicht relationalen Datenbankansätze werden hier zusammengefasst. NoSQL steht für „Not only SQL“ und soll die Abgrenzung zu den relationalen Ansätzen widerspiegeln. NoSQL Ansätze sind zumeist sehr speziell auf gewisse Problemstellungen ausgerichtet. Insofern wundert es nicht, dass verschiedene NoSQL Konzepte im dreistelligen Bereich existieren.

**Im Folgenden werden wir das relationale Datenbanksystem behandeln, also nicht NoSQL!**

.....

Der Server verwaltet die Daten und erlaubt den Clients den Zugriff hierauf. Bei relationalen Datenbanken ist es somit ein Programm, welches SQL-Statements erzeugt und zum DBMS sendet. Für den Enduser ist der Client fast immer eine Software, in der die SQL-Statements ohne Zutun des Users erzeugt werden. Der User bekommt somit nichts davon mit, dass seine Daten in einer Datenbank gespeichert sind. Programme müssen über geeignete Schnittstellen (bspw. ODBC) mit der Datenbank kommunizieren.

### 3 Anforderungen

Ganz allgemein werden an Datenbankprogramme folgende **Anforderungen** gestellt:

#### Normierter Zugriff:

.....

Hierbei ist folgendes wichtig:

- **Redundanzvermeidung**

- **Datenvollständigkeit**

- **Datenidentifikation**

- **Referenzielle Integrität**

- **Datenschutz und -sicherheit**

#### Mehrnutzerbetrieb:

#### Datenunabhängigkeit:

#### Zentrale Verwaltung:

Neben diesen allgemeinen Anforderungen gibt es noch weitere Anforderungen, welche wir zu einem späteren Zeitpunkt näher beleuchten werden, bspw. Skalierbarkeit, Anpassungsfähigkeit, Leistungsfähigkeit.

Eine Datenbank ist ein System zur Datenorganisation mit dem Zweck, die Daten \_\_\_\_\_, \_\_\_\_\_ und \_\_\_\_\_ zu verwalten.

### 4 Datenbanksprache

Relationale Datenbanken werden in aller Regel mit SQL (Structured Query Language) angesprochen. Die Kommunikation der Clients mit dem DBMS erfolgt ausschließlich über SQL. Hierbei können vier wesentliche Aufgabenbereiche definiert werden:

**Datendefinition:**

.....  
.....  
.....

**Datenabfrage:**

.....  
.....

**Datenmanipulation:**

.....  
.....

**Datenschutz:**

.....

Ein weiterer Wichtiger Begriff für die Verarbeitung von Daten ist das Akronym CRUD:

C.....

R.....

U.....

D.....

Diese Aktionen müssen bei einer Datenbank normiert möglich sein.

## 5 NoSQL Datenbanken

Folgende Anforderungen haben zur Entwicklung von NoSQL Datenbanken geführt:

- Extrem hoher Datendurchsatz bei „günstiger“ Hardware
- Horizontale Skalierbarkeit
- Reduktion von Komplexität (bspw. für die API-Anbindung)

NoSQL Datenbanken lassen sich grob in folgende Kategorien einteilen:

.....

Ablegen der Daten in meist frei definierbaren Strukturen, sog. „Dokumenten“ – bspw. als JSON-Struktur. Das Schema kann, muss aber nicht fest vorgegeben sein. Eine recht weit verbreitete Dokumentendatenbank ist „MongoDB“ (von engl. „humongous“, also „enorm“ oder „riesig“).

.....

Werte werden jeweils über einen eindeutigen Schlüssel identifiziert, vergleichbar mit assoziativen Arrays. Sie zeichnen sich oft durch eine extrem hohe Performance aus. Die derzeit (2025) am meisten verbreitete Key-Value Datenbank ist Redis.

.....

Datenpunkte werden als Knoten und Kanten dargestellt. Diese werden, anders als bei relationalen Datenbanken, als Daten eingetragen, nicht als (Tabellen-)Struktur. Somit ist der Ansatz für sich flexibel verändernde Datenstrukturen geeignet und werden damit gerne bei Social-Media Applikationen genutzt, um die vernetzten Kontakte zu modellieren. Ein Vertreter unter der GPL Lizenz wäre hier Neo4j, welche in Java implementiert wurde.

.....

Hier werden im Vergleich von relationalen Datentypen Spalten und Zeilen einfach vertauscht. Ein „Datensatz“, wie er in einer relationalen Datenbank als Zeile zu finden ist, wird in einer spaltenorientierten DB zu einer Spalte. Spalten können wiederum als einzelne Einheit gespeichert werden. Dadurch kann man bei bestimmten Verarbeitungsszenarien (bspw. Big Data Berechnungen) schneller auf manche Daten zugreifen, bzw. ist eine Verteilung auf mehrere Server besser möglich. Ein Beispiel wäre die „Cassandra“ Datenbank, welche ursprünglich im Rahmen von Facebook entwickelt wurde und 2009 von der Apache Foundation weiterentwickelt wird.