

Nachrichten, Informationen, Daten und Redundanzen

- Daten

Daten sind die Darstellung von Tatsachen oder Konzepten in einer Form, die im Computer gespeichert und wieder gelesen werden können.

Beispiel: Eine Kundenadresse, eine Rechnung etc.

Heutzutage werden Unmengen von Daten jeglicher Art gesammelt. Daten können eine sehr kurze Lebensdauer haben (z.B. eine IP Adresse, welche beim Zugriff auf das Internet via Smartphone vergeben wird) oder sie können gesetzlich eine bestimmte Aufbewahrungszeit haben (z.B. Bankdaten) und müssen somit recht lange aufbewahrt werden.

Heute ist es sehr wichtig, dass die Daten strukturiert in flexiblen, gut strukturierten Datenmodellen abgespeichert werden, damit der Zugriff darauf möglichst einfach ist und keine „Datenleichen“ entstehen.

Daten können in verschiedenen Formen wie ASCII, Audio, Video, Bild oder anderen binären Formen wie z.B. PDF Dateien oder Word Dokumenten vorkommen.

- Informationen

Eine Information ist „die nutzbare Antwort auf eine korrekte Fragestellung“. Informationen werden meist aus kombinierten Rohdaten abgeleitet.

Beispiel: Die Kombination der Kundenadresse mit den Rechnungsdaten ergibt die Information wer zu welchem Zeitpunkt eine Rechnung erhalten hat, welche bestimmte Rechnungspositionen beinhaltet.

Von einer Information wird gesprochen, wenn auf eine bestimmte Frage eine Antwort geliefert wird, welche das Verständnis des Fragenden erhöht und ihn befähigt einem bestimmten Ziel näher zu kommen.

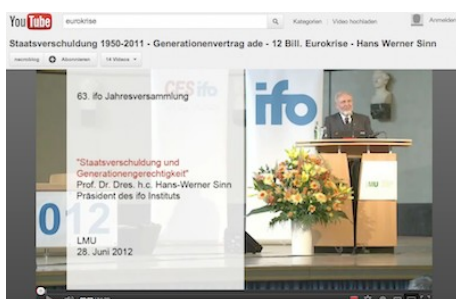
Beispiel: Wie viele Artikel eines bestimmten Typs wurden 2016 verkauft?

Die Begriffe Daten und Informationen werden oft vermischt und in falschem Zusammenhang verwendet.

Grundsätzlich kann davon ausgegangen werden, dass Informationen meist erst aus Daten aufgebaut bzw. zusammengesetzt werden müssen, Rohdaten hingegen müssen nicht zwingend Informationen beinhalten. Sie bilden oft die Basis für Informationen.

- Nachrichten

Bereits seit vielen Jahren wird der Begriff Nachricht, Meldung oder Mitteilung angewandt. Laut Wikipedia beschrieben die Gebrüder Grimm die Nachricht bereits im 18. Jahrhundert. *Als Nachricht bezeichnet man alle Meldungen, durch die wir auf eine Sache oder ein Ereignis aufmerksam gemacht werden.*



Eine Nachricht wird vom Sender an einen Empfänger gerichtet.

Wenn Sie im Internet Informationen beschaffen, Social Media nutzen, TV schauen oder ein Buch lesen, erhalten sie andauernd Nachrichten.

- Redundanzen

Redundanzen sind Daten, welche mehrfach, möglicherweise an unterschiedlichen Stellen, abgespeichert worden sind. Sie beinhalten keinen Neuigkeitswert und sind wertlos und nutzlos.

Redundanzen führen meistens früher oder später zu widersprüchlichen Datenbeständen (z.B. wenn Daten an einer Stelle aktualisiert wurden und an der anderen nicht). Wer kennt das nicht: Man notiert sich z.B. eine Adresse oder eine Telefonnummer für den schnellen Zugriff zusätzlich zu der Kontaktdatenbank z.B. in einem Excel Sheet. Die Kontaktdatenbank wird aktualisiert, das Excel Sheet jedoch nicht. Schon haben wir widersprüchliche, redundante Daten.

Die Widersprüchlichkeit ist das grösste Problem von Redundanzen. Gibt es z.B. zwei Kunden mit dem gleichen Namen aber unterschiedlichen Adressen: Welche ist jetzt korrekt? Gibt es ev. Zwei Kunden mit dem gleichen Namen? Redundanzen können auch auf Datensatzebene auftreten, nicht nur in einem „falschen“ Datenmodell.

Ein weiterer Nachteil von Redundanzen in einem Datenmodell ist, dass Daten durch das mehrfache Speichern auch mehr Speicherplatz in Anspruch nehmen. Dieses Problem ist jedoch immer weniger tragisch, da der Speicherplatz immer günstiger wird.

In der Datenmodellierung muss generell darauf geachtet werden, dass Daten nicht redundant abgelegt werden. Probleme mit redundanten uU sogar inkonsistenten redundanten Daten können schnell sehr hohe Kosten verursachen, wenn die Redundanz behoben werden soll.

1. Eine Kollegin erzählt Ihnen eine Geschichte, die Sie bereits kennen.
 - a) Was hat die Geschichte Ihnen erzählt?
 - b) Wie lautet der Fachausdruck dazu?

[illegible]

-
- This image shows a full page of blank graph paper. The grid consists of thin, light gray horizontal and vertical lines that intersect to form small squares across the entire surface. There are no margins, text, or other markings on the paper.

Historisches

- Ab 18. Jahrhundert: Lochkarten eingesetzt bei Webstühlen. Insbesondere beim Jacquard Webstuhl.
- 1956: Erfindung der Festplatte mit erstem Filesystem.
- 1960er Jahre: Hierarchisches Datenmodell.
- Anfang 1970er Jahre: Erste Netzwerk Datenmodelle
- Anfang 1970er Jahre: Edgar Frank Codd: "A Relational Model of Data for Large Shared Data Banks". Das heutige relationale Datenbankmodell.
- 1974: IBM beginnt mit der Entwicklung von „System R“, einem relationalen Datenbankmanagement System (RDBMS)
- 1979: Erstes kommerzielles RDBMS System wurde durch die Firma Relational Software, heute ORACLE, zur Verfügung gestellt
- Ab 1980: Durchbruch des relationalen Datenmodells.

Die relationalen Datenbanksysteme verdrängen in den 1980er Jahren die hierarchischen und netzwerkartigen Systeme und der Großteil der Behörden, Konzerne, Institute und mittelständischen Unternehmen stellen ihre IT auf relationale Datenbanksysteme um.

- Objektbasierte Datenbanksysteme können sich bis heute nicht durchsetzen. RDBMS bewähren sich in allen Bereichen besser.
- Objektrelationale Datenbanksysteme werden vermehrt geprüft, sind aber für die Praxis noch zu unübersichtlich und komplex.

Einführung

Datenbanken entstanden in den 60er Jahren aus der Notwendigkeit die Zugriffe auf die immer grösser werdenden Datenmengen der Grossrechner zu vereinfachen und zu vereinheitlichen. In den 70er Jahren wurden die ersten relationalen Datenbankmodelle entwickelt (Codd). Mit diesen RDBMS (Relational Database Management System) Systemen wurde das Erstellen und der Zugriff auf die Daten massiv vereinfacht.

Nur ein wohlüberlegtes Datenbankmodell garantiert Flexibilität, problemlose Erweiterung, fehlerfreie Datenablage und schnelle Zugriffszeiten auf die Daten. Das Erstellen eines korrekten Datenmodells ist also die Basis für erfolgreiches Arbeiten mit Datenbanken.

Dieses Modul basiert auf dem am weitesten verbreiteten Datenbankmodell „Relationale Datenbank“.

Übungen und Beispiele werden an Hand des RDBMS Produkts Oracle Database unter Linux gemacht.

Definition von Datenbanken

Eine Datenbank ist ein System zur Beschreibung, Speicherung und Wiedergewinnung von umfangreichen Datenmengen, die von mehreren Anwendungen (– gleichzeitig –) benutzt werden können. Sie besteht aus zwei Hauptteilen:

1. Den eigentlichen Daten und
2. dem Datenbank-Management-System (DBMS), das gemäss einer vorgegebenen Beschreibung Daten speichern, suchen, löschen oder ändern kann.

Eigenschaften von Datenbanken

- Die Daten werden klar strukturiert. Es gibt keine Redundanzen.
- Die Datenstruktur wird zentral gespeichert und ist unabhängig von der Applikation.
- Die Anwendungsprogramme sind "datenunabhängig", d.h. interne Reorganisationen innerhalb des Datenbanksystems tangieren die Anwendungsprogramme nicht.
- Die Datenbank ist leicht erweiterbar.
- Die Integrität der Daten wird von der Datenbank garantiert.
- Die Daten sind zeitlich persistent und nicht an die Lebensdauer eines Programms gebunden.
- Den verschiedenen Benutzern stehen spezifische Datensichten zur Verfügung.

Anforderungen an eine Datenbank

An Datenbanken werden hohe Anforderungen gestellt.

Unterschiedliche Personengruppen haben unterschiedliche Anforderungen an Datenbanken. Hier ein paar (nicht abschliessende) Beispiele:

Anwender

- Die Antwortzeit der Datenbank muss so schnell wie möglich sein
- Die Datenbank muss konsistent sein
- Die Daten müssen zu einem genau definierten Zeitpunkt gespeichert- oder rückgängig gemacht werden können (save- oder cancel Button)
- Die Datenbank muss multiuserfähig sein

Management

- Die Datenbank muss zu den definierten Zeiten verfügbar sein.
 - z.B. 7/24h oder zu Bürozeiten
- Die Datenbank muss nach einem Systemausfall „ohne“ Datenverlust wieder hergestellt werden können
- Der Zugriff auf die Daten muss eingeschränkt werden können, damit nicht alle alles sehen
- Die Datenbank soll wirtschaftlich mit dem Speicherplatz umgehen um nicht unnötig hohe Kosten zu verursachen

Entwickler

- Die Datenbank soll einfach erweiterbar sein
- Der Zugriff auf die Daten soll so durchsichtig wie möglich sein
- Die Datenbank soll auch bei einer grossen Datenmenge noch performant Antworten zurück liefern

Datenbank Administratoren

- Die Datenbank soll einfach zu verwalten sein
- Die DB soll sprechende Fehlermeldungen ausgeben
- Die Datenbank muss vollumfänglich überwacht werden können um mögliche Fehler zu erkennen
- Die DB soll einfach und zuverlässig zu sichern sein

Datenbankmodelle

Heute wird im Wesentlichen zwischen vier verschiedenen Datenbankmodellen unterschieden:

- Hierarchische Datenbanken
- Netzwerk Datenbanken
- Relationale Datenbanken
- Objektorientierte Datenbanken

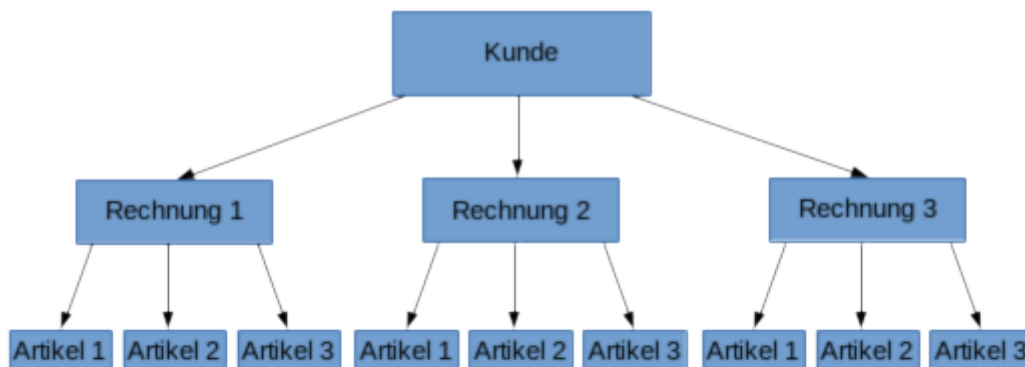
Der Unterschied der verschiedenen Datenbankmodellen liegt in der physischen Speicherstruktur der Daten.

Zu erwähnen sind auch die neueren „NoSQL“ Datenbanken. (Not Only SQL)

Die relationalen Datenbanken sind heute der Quasi-Standard für Datenbanken. Die neueren Objektorientierten Datenbanken konnten sich nie richtig durchsetzen.

Hierarchische Datenbanken

Hierarchisches Datenbankmodell



Das hierarchische Datenbankmodell ist das älteste Datenbankmodell.

Der logische Aufbau entspricht einer Baumstruktur. Der Zugriff erfolgt immer über die Wurzel in Richtung des gesuchten Datensatzes.

Der Vorteil von hierarchischen Datenbanken ist, dass der Zugriff sehr schnell ist, Schreiboperationen dagegen sind eher langsam.

Ein weiterer Nachteil hierarchischer Datenbanken:

- Da der Zugriff immer via Wurzel erfolgt, kann jeweils nur ein Baum abgefragt werden. Somit können nur 1:1 und 1:n Beziehungen abgebildet werden. n:m Beziehungen können nur via Redundanzen abgebildet werden.

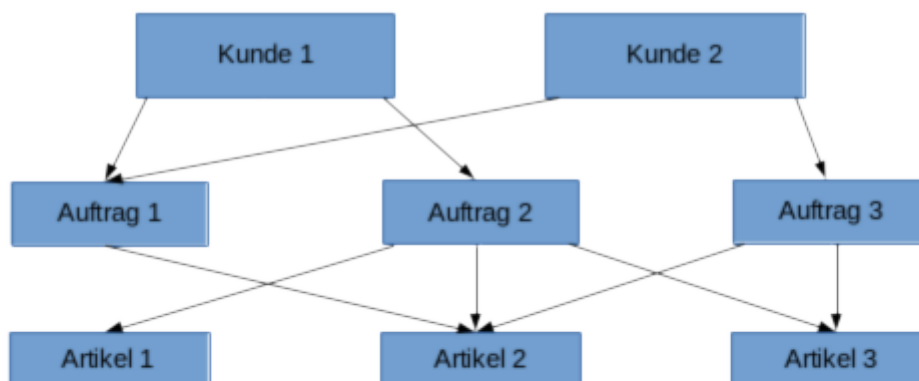
Netzwerk Datenbanken

Netzwerkartige Datenbanken verwenden keine Baumstruktur sondern eine Art Netzwerk von Tabellen, die miteinander in Verbindung stehen. Dies macht die Datenbank viel flexibler für Änderungen als die hierarchischen Datenbanken, erhöht jedoch die Komplexität des Aufbaus. Dadurch, dass mehr als eine Basis der Daten zur Verfügung steht, sind die Abfragen jedoch weniger schnell als beim hierarchischen Datenmodell.

Mit dem netzwerkartigen Datenmodell können jetzt n:m Beziehungen abgebildet werden, da ein Datensatz mehrere Root Datensätze haben kann.

Beide Systeme sind für heutige Anwendungen, welche sehr schnell weiter entwickelt werden und daher flexibel sein müssen, nicht mehr geeignet.

Netzwerk Datenbankmodell



Relationale Datenbanken

Eine relationale Datenbank besteht ausschliesslich aus Tabellen, die meist zu anderen Tabellen in einer Beziehung (Relation) stehen. Das Ändern oder Hinzufügen einer Tabelle ist sehr einfach, darum werden heutzutage meistens relationale Datenbankmodelle eingesetzt.

Für den Zugriff auf die Daten müssen oft mehrere Tabellen abgefragt werden. Im Datenmodell muss berücksichtigt werden, dass ein Zugriff auf mehrere Tabellen performant gemacht werden kann.

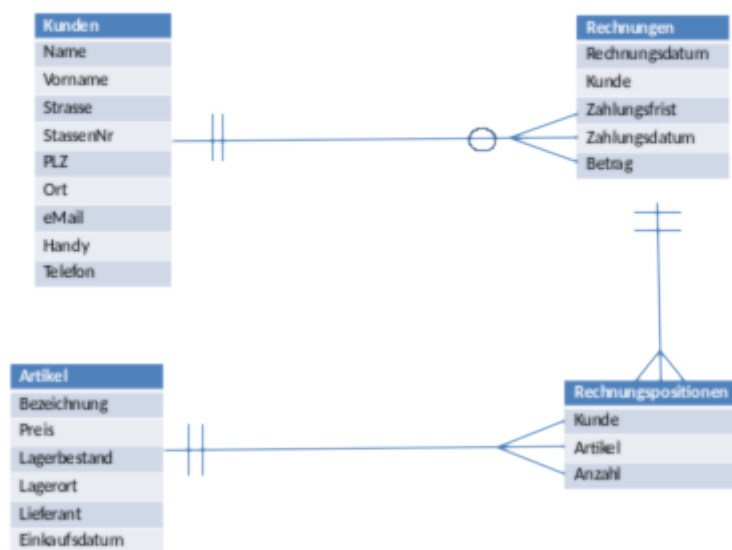
Beispiele kommerzieller relationaler Datenbanken:

- Oracle Database
- Microsoft MSSQL Server
- IBM DB2

Beispiele von relationalen Opensource Datenbanken:

- MySQL
- PostgreSQL

Relationales Datenbankmodell



Objektorientierte Datenbanken

Eine objektorientierte Datenbank besteht aus Objekten. Ein Objekt kann z.B. eine Person, eine Abteilung einer Firma oder ein realer Gegenstand (CD, Buch etc) sein.

Der objektorientierte Ansatz wie z.B. Klassen, Vererbung etc. Konnte sich auf Datenbankebene nie richtig durchsetzen weil das Datenmodell so sehr aufwändig und komplex- und dadurch unflexibel wird.

Was heute durchaus verwendet wird ist eine Mischform von objektorientierten- und relationalen Datenbanken. Man nennt sie „objektrelationale“ Datenbanken. In Objektrelationalen Datenbanken werden komplexe Objekte in relationalen Datenbanktabellen abgelegt.

Beispiele von objektrelationalen Datenbanken:

- Oracle Database
- IBM DB2
- PostgreSQL

Einleitung in das relationale Datenbankmanagementsystem (RDBMS)

Grober Aufbau eines RDBMS

Ein RDBMS besteht aus 3

Hauptkomponenten:

1. Datenbank

Die Datenbank (RD, relational Database) ist die physische Repräsentation der Daten.

In der Datenbank sind die Daten in Tabellen gespeichert. Es gibt Systemtabellen, welche durch das RDBMS verwaltet werden und Benutzerdefinierte Tabellen, die z.B. durch Applikationen bearbeitet werden.

Je nach RDBMS Produkt gibt es eigene Systemdatenbanken, in welchen die Systeminformationen abgelegt werden oder Systemtabellen innerhalb einer Datenbank.

2. Managementsystem:

Mit dem Management System (MS)

Hier steht eine mengenrelationale Sprachschnittstelle zur Datenbank bereit.

Der Datenbankbenutzer kann die Datenbank mit der relationalen

Datendefinitions- und

Datenmanipulationssprache (SQL)

bearbeiten. Es stehen aber auch

Dienstfunktionalitäten wie Datensicherung und Wiederherstellung und Tools zum importieren und exportieren von Tabellen zur Verfügung.

Je nach RDBMS Produkt stehen noch weitere Funktionen im Bereich Security oder Hochverfügbarkeit zur Verfügung.

3. Instanz:

Wird ein RDBMS gestartet, werden verschiedene Prozesse gestartet, die uA Memorybereiche zum Cachen der Daten zur Verfügung stellen.

Der Memorybereich einer Datenbank wird Instanz genannt.

