

GRUNDLAGEN DES DATENBANKENTWURFS, ER-MODELL, NORMALISIERUNG

Phasen der Datenbankentwicklung.

Der Aufbau einer Datenbank gelingt nicht in einem Schritt. Dazu gehören mehrere Phasen:

1) EXTERNE PHASE . Ermittlung der Informationsstruktur : Die DB soll einen Ausschnitt aus der realen Welt im Rechner abbilden. Diese Abbildung erfolgt durch die Beschreibung der Daten. Dazu ist der Informationsbedarf der Benutzer zu ermitteln und zu strukturieren. Das Ergebnis dieses ersten Schrittes - auch als Spezifikations- & Anforderungsanalyse bezeichnet - ist eine informelle Beschreibung des Fachproblems.

2) KONSEPTIONELLE PHASE . Aufstellung des semantischen Modells : Ziel des konzeptionellen Entwurfs ist die formalisierte Beschreibung des betrachteten Sachverhalts. Es existieren verschiedene Ansätze zur Erzeugung einer solchen Gesamtsicht . Das bekannteste Modell ist das s.g. ENTITY RELATIONSHIP MODELL (ER-MODELL). Das Ergebnis dieses Schrittes ist das Fachkonzept der Datenbank.

3) LOGISCHE PHASE · Erstellung des logischen Datenmodells. Ziel ist die Übertragung des semantischen Datenmodells in ein logisches Datenmodell, zB in ein „relationales Datenmodell“ (Tabellenform). Diese Phase umfasst zwei Schritte:

3.1) Im ersten Schritt muss eine Transformation des konzeptuellen Schemas (ER-Modell) in das Datenbankschema erfolgen. Die ist automatisierbar.

3.2) Im zweiten Schritt erfolgt eine Optimierung des relationalen Schemas, zB. die Durchführung einer Normalisierung^(*) der Tabellenstruktur.

$$(*) \quad z_i = \frac{x_i - \bar{x}}{sx}$$

4) PHYSISCHE PHASE · Implementierung der Datenbank. Am Ende dieser Phase sollte die leere Datenbank existieren. Dazu sollte das logische Modell unter Verwendung einer Datenbanksprache (zB SQL) in ein konkretes Datenbankschema übersetzt werden.

Es müssen also:

- Datentypen
- Wertebereiche
- Relationen

& Hierarchie

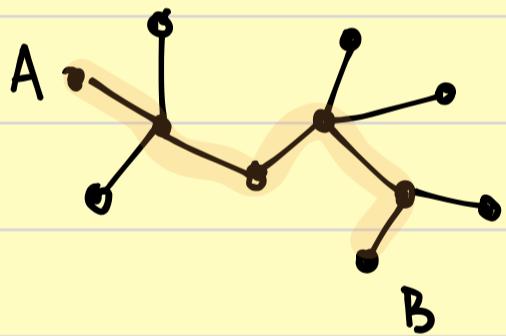
Festgelegt werden

RELATIONALES DATENMODELL

Es existieren verschiedene Datenmodelle. Ausgangspunkt sind es hierarchische Beziehungen zw. Merkmalen (zB Produktmerkmale: AUTO → MODELL → FARBE → MOTOR → ...)

⋮ ⋮ ⋮ ⋮ ⋮

Dies ist heute wegen beschränkten Anwendbarkeit kaum noch in Datenbankmanagementsysteme (DBMS) zu finden. Vielmehr wird ein NETZWERK MODELL bevorzugt um die Leistung der Großrechner auszuschöpfen.



Der Average Path length von einem Netzwerk verhält sich besser als die reine Hierarchy.

Das am meisten verwendete Modell ist das relationale Datenmodell, welche auf folgende Elemente basiert:

- A) • TABELLEN (Relationen)/Wechselwirkungen
- B) • OPERATOREN, d.h. Rechenoperationen für Tabellen
- C) • REGELN, d.h. Bedingungen für die Werte bestimmter Spalten.

TABELLEN A)

Folgende Eigenschaften gelten für eine Relation:

- 1) Es tritt kein Tupel doppelt auf, d.h. zu keinem Zeitpunkt gibt es zwei Tupel, die den gleichen Wert haben.
- 2) Die Tupel-Reihenfolge ist nicht definiert, d.h. man darf sich nicht auf eine bestimmte Reihenfolge der Zeilen der Tabelle verlassen.
- 3) Die Attributfolge (Reihenfolge der Spalten) ist nicht definiert.

OPERATOREN B)

Die wichtigsten Operationen mit Relationen, die ein DBMS zur Verfügung stellen muss:

- 1) Auswahl von Zeilen & Spalten
- 2) Aneinanderfügen von Tabellen
- 3) Verbind von Tabellen

Beispiel:

Attribute (Spalte)		
Name	Wohnort	Straße
Hans	Berlin	In der Kirsche
Harald	Hamburg	Kreuzstraße
Dietmar	München	Lange Gasse
Martin	Stuttgart	Marktplatz

REGELN C)

Regeln, bzw. Integritätsbedingungen sorgen dafür,

Auss die DB keine widersprüchliche Daten enthält.
 Lokale Integritätsbedingungen gelten jeweils für eine Tabelle, zB gilt für den Primärschlüssel: Jeder Wert in der Spalte darf nur einmal vorkommen. Dagegen, gelten globale Integritätsbedingungen (referentielle Integrität) für mehrere Tabellen. Ein Beispiel für eine globale Integrität: Zu jedem Wert, der ein Fremdschlüssel in einer Tabelle ist, existiert genau ein Wert in einer weiteren Tabelle und ist dort Primärschlüssel: zB

kurs Nr	Name	Vorname	Matrikel Nr.
1	Müller	Marion	1234
2	Maier	Klaus	2341
3	Schulz	Sven	3412
4	Schneider	Lisa	4123
1	Beck	Lothar	1423

kursNr	Bezeichnung
1	Mathematik
2	Deutsch
3	EDV
4	Recht

Beispiel für globale Integrität

Das Entity Relationship Model (ER-Modell)

Das Schema, in dem man eine DBStruktur konzipiert, nennt sich ER Modell. In einem ER-Modell wird die grundlegende Tabellen- & Beziehungsstruktur einer DB entworfen & abgebildet. Die Darstellung im ER-Modell (grafisch) und die Abbildung der DBStruktur mithilfe von Tabellen sind äquivalent.

DEFINITIONEN ER-Modell:

- Attribute (Eigenschaften). Dabei handelt es sich um die Felder eines Datensatzes.
- Entität. Handelt es sich um ein Objekt der jeweiligen Tabelle, also ein Datensatz. Die Entitäten unterscheiden sich untereinander durch unterschiedliche Werte der Eigenschaften.
- Entitätsmengen. Mehrere Entitäten mit gleichen Attributen (nicht Attributwerte) werden zu einer Entitätsmenge
- Beziehungen. Die einzelnen Tabellen sind nicht voneinander losgelöst. Zu den Entitäten werden Verknüpfungen hergestellt. Eine Beziehung kann einem sogenannten Beziehungstyp zugeordnet werden:

1:1 Beziehung	1:n Beziehung	n:m Beziehung
---------------	---------------	---------------

