



Datenbankmanagement

Übungen

Grundlagen

Übungsaufgaben zum 1. Tag

1. Benennen Sie mindestens 10 der sechs der 12 „goldenen“ Regeln von Codd für gutes Datenbankdesign (Lösung siehe hinten!)
2. Erläutern Sie die Begriffe Datenbasis, Data Dictionary, Datenbank, Datenbankmanagementsystem und Datenbanksystem (Lösung selber nachschlagen!)
3. Nennen Sie die Ihnen bekannten Datenbankmodelle und beschreiben Sie die Vor- und Nachteile
4. Erläutern Sie kurz die 3 Ebenen der Drei-Ebenen-Schemaarchitektur anhand eines selbst gewählten Beispiels
5. Nennen und Erläutern Sie die Ihnen bekannten Anforderungen an ein DBMS
6. Nennen und Beschreiben Sie die Phasen des Datenbankentwurfs
7. Erläutern Sie den Begriff OLAP.
8. Erläutern Sie den Begriff Data Mart.
9. Wie grenzen sich die Begriffe Data Warehouse und Data Mining voneinander ab?
10. Erläutern Sie das Grundprinzip eines Data Warehouse.
11. Wofür stehen OLTP und OLAP, grenzen Sie diese beiden Begriffe ab?
12. Erklären Sie den Begriff Drill-Down.
13. In welcher Beziehung stehen Data Mart und Data Warehouse zueinander?
14. Was versteht man unter Data Mining, und was ist der grundlegende Unterschied zu OLAP?
15. Grenzen Sie die Begriffe Anwendungsprogramm, Datenbank und Datenbankmanagementsystem (DBMS) und Datenbanksystem voneinander ab.

Lösungen

1.)

Regel 1: Darstellung von Informationen

Alle Informationen in einer relationalen Datenbank (einschließlich Namen von Tabellen und Spalten) sind explizit als Werte in Tabellen darzustellen.

Regel 2: Zugriff auf Daten

Jeder Wert einer relationalen Datenbank muß durch eine Kombination von Tabellennamen, Primärschlüssel und Spaltenname auffindbar sein.

Regel 3: Systematische Behandlung von Nullwerten

Das DBMS behandelt Nullwerte durchgängig gleich als unbekannte oder fehlende Daten und unterscheidet diese von Standardwerten.

Regel 4: Struktur einer Datenbank

Die Datenbank und ihre Inhalte werden in einem sogenannten Systemkatalog auf derselben logischen Ebene wie die Daten selbst - also in Tabellen - beschrieben. Demzufolge läßt sich der Katalog mit Hilfe der Datenbanksprache abfragen.

Regel 5: Abfragesprache

Zu einem relationalen System gehört mindestens eine Abfragesprache mit einem vollständigen Befehlssatz für Datendefinition, Manipulation, Integritätsregeln, Autorisierung und Transaktionen.

Regel 6: Aktualisieren von Sichten

Alle Sichten, die theoretisch aktualisiert werden können, lassen sich auch vom System aktualisieren.

Regel 7: Abfragen und Bearbeiten ganzer Tabellen

Das DBMS unterstützt nicht nur Abfragen, sondern auch die Operationen für Einfügen, Aktualisieren und Löschen in Form ganzer Tabellen.

Regel 8: Physikalische Datenunabhängigkeit

Der logische Zugriff auf die Daten durch Anwendungen und Ad-Hoc-Programme muß unabhängig von den physikalischen Zugriffsmethoden oder den Speicherstrukturen der Daten sein.

Regel 9: Logische Datenunabhängigkeit

Änderungen der Tabellenstrukturen dürfen keinen Einfluß auf die Logik der Anwendungen und Ad-Hoc-Programme haben.

Regel 10: Unabhängigkeit der Integrität

Integritätsregeln müssen sich in der Datenbanksprache definieren lassen. Die Regeln müssen im Systemkatalog gespeichert werden. Es darf nicht möglich sein, die Regeln zu umgehen.

Regel 11: Verteilungsunabhängigkeit

Der logische Zugriff auf die Daten durch Anwendungen und Ad-Hoc-Programme darf sich beim Übergang von einer nicht-verteiltern zu einer verteilten Datenbank nicht ändern.

Regel 12: Kein Unterlaufen der Abfragesprache

Integritätsregeln, die über die Datenbanksprache definiert sind, dürfen sich nicht mit Hilfe von Low-Level-Sprachen umgehen lassen.

2.)

Die meisten modernen Datenbanksysteme bestehen aus drei Schichten bzw. Ebenen. Diese sind die konzeptuelle, logische und physische Ebene. Durch die Trennung ist es möglich, die Daten in den einzelnen Schichten unabhängig voneinander zu verändern. Der Aufbau eines Datenbanksystems besteht aus einer Datenbank (DB), welche die Datenbasis darstellt, und aus einem Datenbankmanagementsystem (DBMS). Ein Datenbankmanagementsystem abstrahiert die Details von Hard- und Software-Komponenten eines Systems. Dadurch können Datenbanken anwendungsunabhängig definiert, konstruiert und manipuliert (verändert) werden. Ein Datenbankmanagementsystem (DBMS) bietet schnelle und effiziente Zugriffsverfahren, um Daten verarbeiten zu können. Die meisten modernen Datenbanksysteme bestehen aus drei Schichten bzw. Ebenen. Diese sind die konzeptuelle, logische und physische Ebene. Durch die Trennung ist es möglich, die Daten in den einzelnen Schichten unabhängig voneinander zu verändern.

3.)

Hierarchisches Datenbankmodell

Pro: einfache Struktur, schneller Zugriff, Datenintegrität und -unabhängigkeit bleiben erhalten, auch für große Datenmengen geeignet

Contra: setzt Kenntnisse der Struktur voraus, jede Beziehung erfordert eigene Definition pro Satz nur ein Feld und eine Verknüpfung, nachträgliche Strukturänderungen kaum möglich

Netzartiges Datenbankmodell

Pro: flexibler als hierarchische Datenbankmodelle, leistungsfähiger als relationale Datenbankmodelle, gute Integrität

Contra: Wartung aufwändig und kompliziert, wird schnell unübersichtlich, Datenstruktur bestimmt über Aufbau

Relationales Datenbankmodell

Pro: einfach umzusetzen, Daten bleiben weitgehend unabhängig voneinander, SQL-fähig, sehr weit verbreitet (quasi Standard)

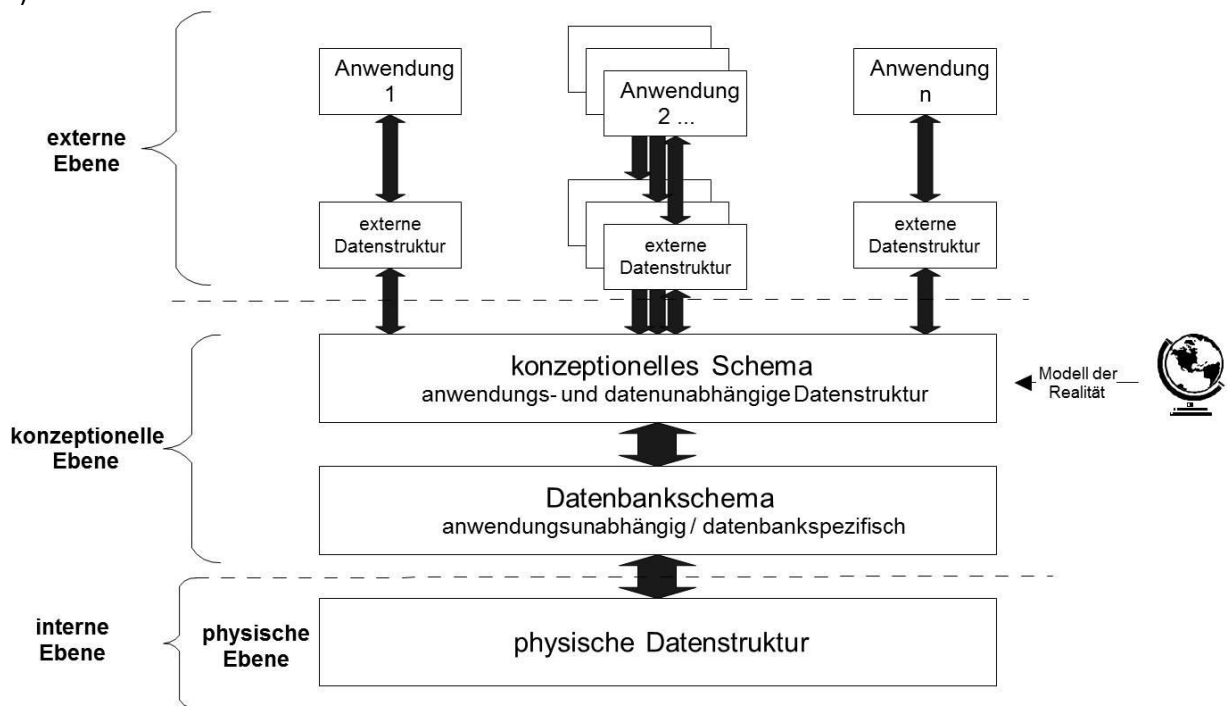
Contra: weniger leistungsfähig als andere Datenbankmodelle, keine Gewährleistung der Datenintegrität, fehler- und störungsanfällig

Objektorientiertes Datenbankmodell

Pro: Daten können flexibel repräsentiert werden, unterstützt mehrdimensionale Daten, mehrfache Verwendung von Objekten möglich

Contra: Implementierung recht kompliziert, geringe Geschwindigkeit

4.)



Beispiel: Datawarehouse

Extern: diverse DBs, Aggregationstaellen, ETL-Process, interne DWH-Tabellen

5.) DBMS:

Ein Datenbank-Management-System (DBMS) stellt unterschiedliche Werkzeuge bereit, mit welchen eine oder mehrere Datenbanken erstellt, mit Daten gefüllt und verwaltet werden können. Ein DBMS verfügt in der Regel über mindestens eine Benutzerschnittstelle und kann eine oder mehrere Datenbanken nur aufeinanderfolgend oder parallel verwalten. Ferner gibt es - explizit oder implizit - Regeln, wer zu welchem Zeitpunkt Zugriff auf und Änderungsrecht an den Daten hat. Schließlich können Sicherungsstrategien / Backup-Verfahren definiert und regelmäßig ausgeführt werden.

6.) Modellierungsphasen

Erste Phase: Semantische Modellierung

Zweite Phase: Logische Modellierung

Dritte Phase: Physikalische Modellierung

7. + 11.) OLAP / OLTP

OLAP (Online Analytical Processing) ist eine Technologie, mit deren Hilfe große Unternehmensdatenbanken organisiert und Business Intelligence unterstützt wird. OLAP-Datenbanken sind in einen oder mehrere Cubes unterteilt, wobei jeder Cube von einem Cubeadministrator organisiert und entworfen und so an Ihr Verfahren zum Abrufen und Analysieren von Daten angepasst wird.

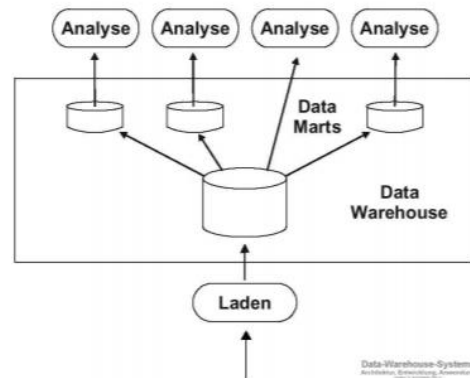
OLAP-Datenbanken (Online Analytical Processing) ermöglichen Business Intelligence-Abfragen. OLAP ist eine Datenbanktechnologie, die nicht für die Verarbeitung von Transaktionen, sondern für die Abfrage und Berichtserstellung optimiert wurde. Die Quelle für OLAP stellen OLTP-Datenbanken (Online Transactional Processing) dar, die meist in Data Warehouses gespeichert sind. OLAP-Daten

sind von diesen herkömmlichen Daten abgeleitet und werden in Strukturen aggregiert, die eine anspruchsvolle Analyse zulassen

8. + 13.) Ein **Data Mart** ist ein Auszug bzw. eine Kopie eines Teildatenbestandes innerhalb eines Data Warehouse und dient für eine bestimmte Anwendung oder einen organisatorischen Bereich zur Analyse. Data Marts bieten den Vorteil der schnelleren Abfrage von Daten. Da sie in der Regel nur eine Teilmenge des Data Warehouse enthalten, ist die Datenmenge eines Data Marts folglich erheblich geringer.

9. + 10) DWH und Data Mining

Um aus Daten Informationen zu gewinnen muss man sie mit verschiedenen Werkzeugen analysieren können. Die Operativen Systeme der Unternehmen sind oftmals dafür nicht geeignet, da dort meist nur die zur Verarbeitung nötigen Daten gespeichert werden. Um Analysen mit stärkerer Aussagekraft zu erhalten muss man jedoch auch Daten aus anderen Systemen mitbetrachten. Aus mehr Daten lassen sich auch genauere Informationen gewinnen.



Data Mining ist ein Analyseansatz und zielt darauf ab, Beziehungsmuster, wie z.B. Regelmäßigkeiten und Auffälligkeiten, in den zugrunde liegenden Daten zu ermitteln und durch logische oder funktionale Beziehungszusammenhänge abzubilden.

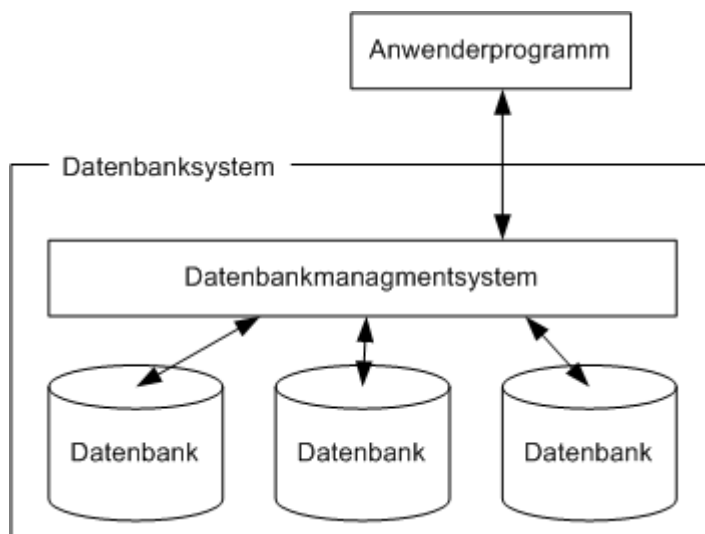
12.) Drill-Down

Als Drill-down wird im Allgemeinen die Navigation in hierarchischen Daten bezeichnet.

14.) **OLAP und Data-Mining** sind Werkzeuge zur Informationsanalyse, die beide auf grossen Datenbanken laufen müssen, um einen Nutzen zu erzielen. Beide Verfahren haben einen unterschiedlichen Ansatz:

OLAP ist Top-down gesteuert, -> abfragegesteuert (query-driven), Data Mining ist Bottom-up (Entdeckungsgesteuert); Mit anderen Worten: OLAP beantwortet die Frage: Ist etwas richtig; Data Mining beantwortet die Frage: Wie ist das passiert?

15.) Abgrenzung DBS, DBMS, DB



Datenbanksystem: Bezeichnung für ein System eines Herstellers, z., B. MySQL, Oracle, MS SQL