

DATENBANKEN

Aufgabe 5



Gruppenmitglieder:	Tammo Post Johannes Theiner
Semester:	Wintersemester 2018/19
letzte Änderung:	21.12.2018

1

1.1 Nennen Sie 3 der wichtigsten Ziele, die man mit dem Einsatz von Datenbanken verfolgt.

- Alle Daten zentral organisieren
- Daten schützen vor unbefugtem Zugriff
- Daten vor Verlust schützen

1.2 Nennen Sie 3 der wichtigsten Eigenschaften, die den Erfolg von SQL begründen.

- Verhältnismäßig einfache Syntax
- Einfache Abbildung der Realität
- Darüber hinaus sehr viele Möglichkeiten für Experten

1.3 Nennen Sie zwei Vorteile und zwei Nachteile des Datenbankansatzes.

Vorteile:

- Von mehreren Benutzern gleichzeitig einsehbar
- Trennung von Daten und Anwendung

Nachteile:

- Bei Serverausfall kein Zugriff auf die Daten
- Aufwand beim Betrieb/Verwaltung einer Datenbank mit mehreren Benutzern

1.4 Warum kamen die Datenbanktheoretiker auf die Idee, Datenbanksysteme dreischichtig zu betrachten? Nennen Sie für jede Schicht einen Gesichtspunkt.

Interne Ebene für den Nutzer nicht unmittelbar relevant da nur wichtig für die Organisation der Daten auf der Festplatte/Server. Die Konzeptionelle Ebene beinhaltet alle Daten, die der Benutzer oder die Anwendung bekommt aber durch die externe Ebene immer nur eine eingeschränkte Sicht.

1.5 Welche Bedeutung spielt das Data Dictionary beim Datenbankentwurf?

Im Data Dictionary werden wichtige Informationen über eine Relation wie z.B. die beteiligten Entity-Typen oder die Komplexität gespeichert. Somit hilft das DD einen schnellen Überblick über die Struktur der Datenbank zu gewinnen.

1.6 Welche Ziele verfolgt man mit der Datenmodellierung?

Mithilfe der Datenmodellierung soll ein Problem analysiert und abstrahiert werden, damit je nach Einsatzzweck eine gute Datenbanklösung entwickelt werden kann.

1.7 Worin unterscheidet sich die logische von der physischen Datenunabhängigkeit?

Die logische Datenunabhängigkeit sagt aus das keine Daten von Anderen beeinflusst werden.

Physische Datenunabhängigkeit ist dann gewährleistet wenn das erstellen einer neuen Spalte/Tabelle keine Auswirkungen auf die Logische Struktur der Daten hat.

1.8 Worin unterscheidet sich eine partielle von einer disjunkten Generalisierung?

- Disjunkt: Ein Objekt taucht nur in einer Spezialisierung auf (Gegenteil= nicht disjunkt)

- Partiiell: Ein Objekt kann in einer Spezialisierung auftauchen, muss es aber nicht (Gegenteil=total)

1.9 Erläutern Sie zwei verschiedene Interpretationen des Begriffs „Relation“ in Verbindung mit relationalen Datenbanken.

Der Begriff Relation beschreibt zum einen die Beziehung zwischen zwei Objekten zum anderen bezeichnet man auch die Tabelle in denen die Daten stehen als Relationen.

1.9.1 Warum ist es sinnvoll, für diese zunächst verschiedenen Interpretationen denselben Begriff zu verwenden?

Von der Struktur her sehen sich beide jedoch sehr ähnlich.

1.10 Was versteht man unter einer Update-Anomalie?

Es kann zu Konsistenzproblemen kommen, wenn fehlerhafte/doppelte Daten in die Datenbank gelangen.

1.11 Warum stellt SQL eine nichtprozedurale Sprache dar?

SQL ist nichtprozedural, sondern eher objektorientiert, da durch die Befehle genau beschrieben wird welche Daten ausgewählt werden sollen.

1.12 Was versteht man in SQL unter einer korrelierten Unterabfrage? Nennen Sie ein Beispiel.

Die Unterabfrage ist mit der Hauptabfrage Verknüpft z.B welche Mitarbeiter haben einen Kollegen mit dem gleichen Gehalt?

1.13 Warum können korrelierte Unterabfragen "Laufzeitkiller" sein?

Das Ergebnis kann eine hohe Anzahl an Datensätzen ergeben und diese benötigt viel Zeit um ausgewertet zu werden.

1.14 Unter welcher Bedingung ist es (zumindest theoretisch) möglich, über eine View Daten einzugeben?

Es ist nur dann möglich über eine View Daten einzugeben, wenn sich dahinter keine aggregierten Daten befinden (z.B. kein AVG oder DISTINCT).

1.15 Was versteht man unter einer Transaktion?

Falls alle Befehle einer Transaktion direkt nacheinander ausgeführt werden die Datenbank wieder in einen konsistenten Zustand überführt wird.

1.16 Erläutern Sie den Begriff „Serialisierbarkeit“.

Ein Schedule ist serialisierbar wenn es einen seriellen Schedule gibt, der denselben Ausgangszustand und die selben Ergebnisse für die einzelnen Transaktionen liefert.

1.17 Grenzen Sie die Begriffe Integrität und Konsistenz gegeneinander ab.

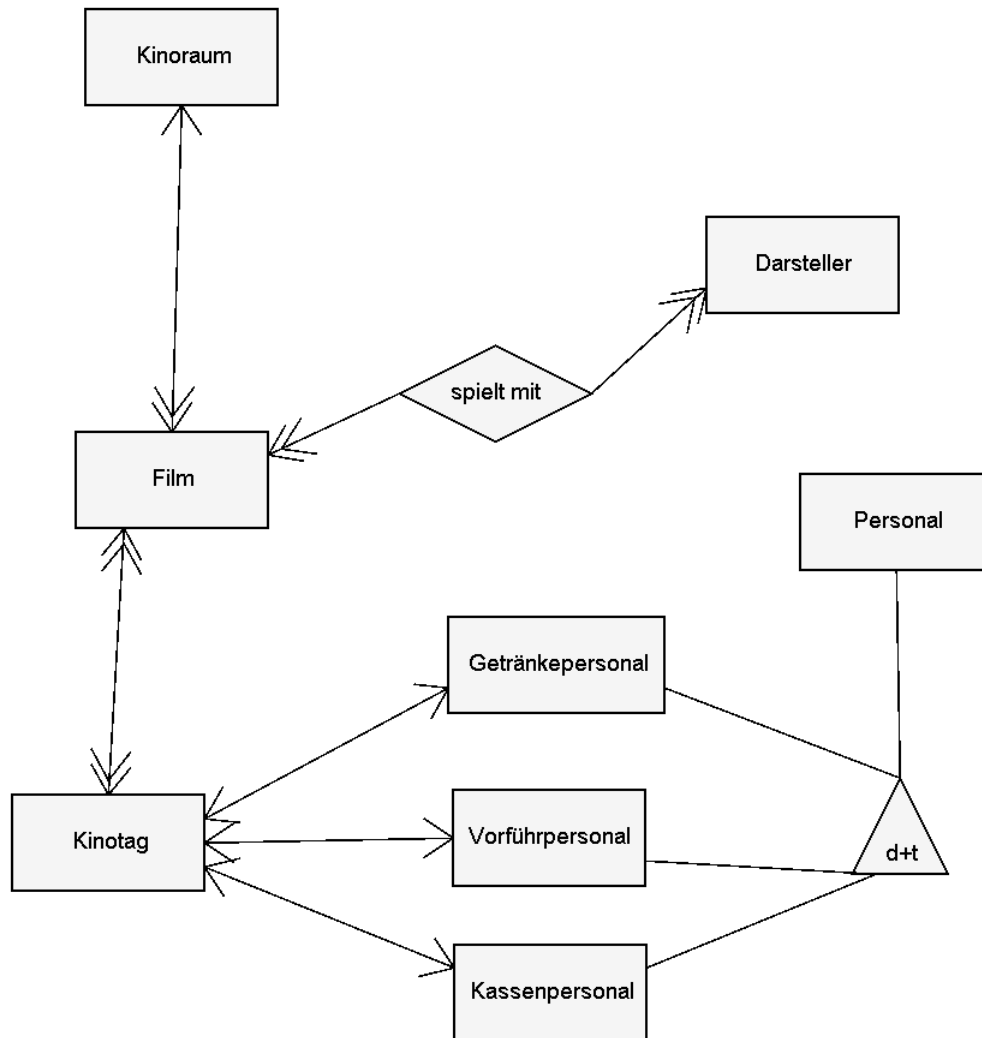
Beim Löschen eines Datensatzes muss dafür gesorgt werden das alle verknüpften Datensätze ebenfalls geändert/gelöscht werden. Konsistenz bezieht sich auf das Format der Daten bzw. ob der Wertebereich eingehalten wurde, und es darf keine doppelten Datensätze geben.

1.18 Grenzen Sie die Begriffe Datenschutz und Datensicherheit gegeneinander ab.

Datenschutz sichert die Daten vor unbefugten Zugriff, Datensicherheit sorgt dafür das keine Daten verloren gehen.

2

2.1 Geben Sie für diese Modellbeschreibung ein ERM an, das so eng wie möglich an der oben angegebenen Spezifikation bleibt.



2.2 Geben Sie für dieses Modell ein Relationenschema in dritter Normalform an.

Filme (FNr, Titel, Regisseur, Länge, Verleiher_Nr)

Darsteller = (DNr, Name, Vorname)

Kinoraum = (RNr, Raumbez)

Kinotag = (KTNr, Datum)

Personal = (PNr, Name, Vorname)

R_spieltMit = (FNr, DNr)

R_Kinotag_Raum_Film = (RNR, FNr, KTNr)

Verleiher = (Verleiher_Nr, Name)

3

3.1 Erstellen Sie für folgende Anfragen SQL-Ausdrücke

3.1.1 Welche (Film-)Längen gibt es (Ausgabe Filmlaenge, ohne Doppelnennungen, absteigend sortiert)?

3.1.2 Welche Filme mit einem Darsteller Meyer (Name) gibt es (Ausgabe: FNr, Film_Titel)?

3.1.3 Wie viele Darsteller haben die einzelnen Filme (Ausgabe: FNr, Titel, Anzahl)?

3.1.4 Welche Filme haben mehr als 10 Darsteller (Ausgabe: FNr, Filmtitel, Darsteller_Anzahl)?

3.1.5 Welcher Film bzw. welche Filme (Ausgabe: FNr, Titel) haben die meisten Darsteller?

3.1.6 Welche Durchschnittslänge (Ausgabe: DLänge) haben die Filme mit mehr als 10 Darstellern?

3.1.7 Gibt es Darsteller (Name), die in allen Filmen mitspielen (Ausgabe: ja bzw. nein)?

3.2 Was liefern folgende SQL-Ausdrücke?

3.2.1 `SELECT COUNT(*) AS Frage FROM F WHERE FNr = ANY (SELECT FNr FROM F_D);`

Gibt die Anzahl der Filme die in der Relation F_D auftauchen an. Somit auch die Anzahl der Schauspieler die mindestens in einem Film mitspielen.

3.2.2 SELECT COUNT(*) AS Frage FROM F F1 WHERE F1.FNr IN (SELECT FNr FROM F_D WHERE FNr = F1.FNr);

Dieser Befehl ist gleichbedeutend mit 3.2.1.

3.3 Vereinfachen Sie den SQL-Ausdruck aus 3.2.2.

4

4.1 Untersuchen Sie den Schedule auf 2 verschiedene Arten auf Serialisierbarkeit

T1	T2
	Read A
	a:=A+1
Read A	
	Write A
a:=A+1	
Write A	

Serielle Schedules: $T1 \rightarrow T2$, $T2 \rightarrow T1$, $A = 8$

Neues Schedule: $A = 7$

Nicht serialisierbar!

Serialisierbarkeit über Konfliktoperationen:

Log(A)

T2 Read	T2 Read vor	T1 Read
T1 Read	T1 Read vor	T2 Write
T2 Write	T2 Write vor	T1 Write
T1 Write	T2 Write vor	T1 Write

Nicht serialisierbar, weil die Richtung wechselt

4.2 Untersuchen Sie den Schedule auf 2 verschiedene Arten auf Serialisierbarkeit

T1	T2
Read A	
A:=A+1	
	Read A
	Write A
Write A	

Serielle Schedules: $T1 \rightarrow T2$, $T2 \rightarrow T1$, $A = 2$

Neues Schedule: $A = 1$

Nicht serialisierbar!

Serialisierbarkeit über Konfliktoperationen:

$\text{Log}(A)$

T2 Read	T1 Read vor	T1 Read
T2 Read	T2 Read vor	T1 Write
T2 Write	T2 Write vor	T1 Write
T1 Write		

Nicht serialisierbar, weil die Richtung wechselt

4.2.1 Zu welchen Ergebnissen kommen Sie?

Bei beiden Transaktionen tritt ein lost update auf.

4.3 Überführen Sie die Relation schrittweise in die dritte Normalform

unnormalisiert

$R = (\underline{A}, B, C, D^*, E, F, G)$

$C = (C1, C2)$

$D = (D1, D2^*, D3)$

$B \rightarrow E$

$F \rightarrow G$

1. Normalform

$R = (\underline{A}, B, C.C1, C.C2, E, F, G)$

$R_D = (\underline{D1}, \underline{LFNR_1}, D2, D3)$

$R_BE = (\underline{B}, \underline{E})$

$F \rightarrow G$

2. Normalform

$R = (\underline{A}, \underline{B}, C.C1, C.C2, F)$

$R_D = (\underline{D1}, \underline{LFNR_1}, D2, D3)$

$R_BE = (\underline{B}, \underline{E})$

$R_FG = (\underline{F}, \underline{G})$

3. Normalform

$R = (\underline{A}, \underline{B}, C.C1, C.C2, F)$

$R_D = (\underline{D1}, \underline{LFNR_1}, D2, D3)$

$R_BE = (\underline{B}, \underline{E})$

$R_FG = (\underline{F}, \underline{G})$