

Aufgabe 1 – SQL (15 Punkte)

Das KIT ist damit beauftragt worden, die Ergebnisse und Spielpläne der neuen Universitäts-Fußballliga in einer SQL-Datenbank festzuhalten. Dazu wurden folgende Relationen modelliert:

`mannschaften (id, name, hochschule, ort),`

`spieler (id, vorname, name, mannschaft, tore, position),`

`spiel (heim, gast, tore_heim, tore_gast).`

Der Schlüssel ist jeweils unterstrichen. `mannschaft`, `heim` und `gast` sind jeweils Fremdschlüssel zu `mannschaften.id`.

Nehmen Sie die nachfolgend abgebildete Tabellenstruktur mit sinnvollen Datentypen an. Die Daten in den Tabellen sind exemplarisch; geben Sie für die folgenden Aufgaben immer generelle Lösungen an!

mannschaften			
id	name	hochschule	ort
1	'KIT FC'	'KIT'	'Karlsruhe'
2	'TUM FC'	'TU München'	'München'
3	'Berliner Einheit'	'TU Berlin'	'Berlin'
4	'Borussia HU'	'HU Berlin'	'Berlin'
5	'Concordia TU'	'TU Dortmund'	'Dortmund'
6	'Magdeburg 05'	'OVGU Magdeburg'	'Magdeburg'

spieler					
id	vorname	name	mannschaft	tore	position
1	'Erik'	'Salzberg'	1	0	'Torwart'
2	'Georg'	'Mueller'	1	0	'Verteidiger'
3	'Pavel'	'Meier'	1	4	'Angriff'
4	'Christine'	'Müller'	1	2	'Mittelfeld'
5	'Alfons'	'Saake'	5	2	'offensives Mittelfeld'
6	'Gunter'	'Kemper'	6	1	'Verteidiger'

spiel			
heim	gast	tore_heim	tore_gast
1	2	3	1
3	4	0	4
5	6	0	0
1	3	5	4
2	4	2	2

Hinweis: Alle Aufgaben lassen sich mit SQL-Konstrukten lösen, die in der Vorlesung vorgestellt wurden.

- a) Geben Sie den Namen der Mannschaft aus, in der 'Gunter Kemper' spielt.

(2 Punkte)

- b) Erstellen sie eine Sicht mit allen Heimsiegen aller Mannschaften. Ein Heimsieg bedeutet, dass die 'Heim'-Mannschaft das Spiel gewonnen hat, d.h. mehr Tore als die 'Gast'-Mannschaft geschossen hat. Die Sicht soll aus den Namen der beiden Mannschaften und der Tordifferenz bestehen. Sortieren Sie die Tupel entsprechend der Tordifferenz.

(3 Punkte)

- c) Geben Sie den Spielplan der Saison bestehend aus Hin- und Rückrunde aus. Dabei soll markiert werden, ob ein Spiel in der Hin- oder in der Rückrunde stattfindet, und welche Mannschaft bei dem jeweiligen Spiel die Heim- bzw. Gastmannschaft ist.

Hinweis: Der Spielplan besteht pro Saisonhälfte aus allen möglichen Paarungen der Mannschaften. Zusätzlich wechselt in der Rückrunde das Heimrecht. Spielt in der ersten Saisonhälfte der 'KIT FC' als Heimmannschaft gegen den 'TUM FC', so ist beim Rückspiel in der zweiten Saisonhälfte der 'TUM FC' die Heimmannschaft. Wie Sie die Paarungen der Hinrunde im SQL-Ausdruck wählen, ist Ihnen frei gestellt.

(3 Punkte)

- d) Es wird eine neue Regel eingeführt: 'Jede Mannschaft muss mindestens aus 13 und darf maximal aus 23 Spielern bestehen'. Erstellen Sie eine Anfrage, die alle Mannschaften ausgibt, die aktuell gegen diese Regel verstoßen.

(3 Punkte)

- e) Bei der Überprüfung des Schemas fiel dem Datenbankadministrator auf, dass die Relation `spieler` in Bezug auf das Attribut `position` Redundanzen enthält, da die Bezeichnungen der Positionen sehr umfangreich sind. Geben Sie eine sinnvolle Schema- und Datenbankänderung mittels SQL-Befehlen an, um diese Redundanzen aufzulösen und potentielle Inkonsistenzen durch Updates auf dieser Spalte zu vermeiden. Gehen Sie hierfür davon aus, dass im aktuellen Datenbestand alle möglichen Positionen enthalten sind und derzeit keine Inkonsistenzen durch Schreibfehler existieren.

Hinweis: Bei dieser Aufgabe dürfen Sie Konstrukte wie `auto_increment` benutzen. `auto_increment` gibt für ein Attribut an, dass beim Einfügen eines neuen Tupels dieses Attribut automatisch einen Wert erhält, den noch kein Tupel bisher als Wert hat (nämlich: maximal bisher vergebener Wert für dieses Attribut + 1).

(4 Punkte)

Aufgabe 2 – ER-Modellierung (15 Punkte)

Gegeben ist das folgende ER-Diagramm in Abbildung 1. Die Schlüsselattribute sind durch Unterstreichen gekennzeichnet, die Kardinalitätsangaben sind Teilnehmerkardinalitäten.

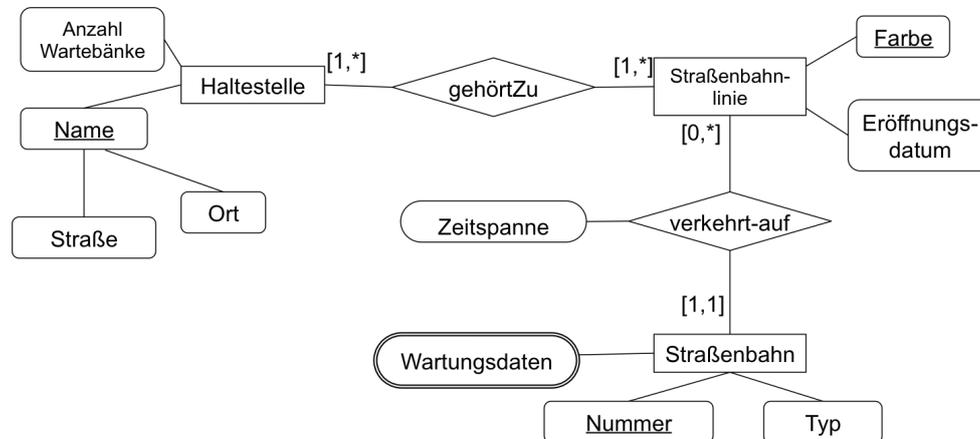


Abbildung 1: ER-Diagramm Straßenbahnen

- a) Geben Sie für die folgenden Aussagen auf dem Antwortblatt an, ob sie zutreffen oder nicht. Für jede richtige Antwort gibt es 1/2 Punkt für jede falsche wird 1/2 Punkt abgezogen. Diese Teilaufgabe wird mit mindestens 0 Punkten bewertet.

	RICHTIG	FALSCH
Im Schema in Abbildung 1 sind Straßenbahnen erlaubt, die zu keiner Straßenbahnlinie gehören.	■	■
Im Schema in Abbildung 1 können Straßenbahnen einer Straßenbahnlinie für mehrere Zeitspannen zugeordnet sein.	■	■
Im Schema in Abbildung 1 kann es für eine Straßenbahn mehrere Wartungsdaten geben.	■	■
Im Schema in Abbildung 1 wird eine Haltestelle eindeutig durch die Straße bestimmt.	■	■
Ein relationales Datenbankschema ist eine Menge von Attributen.	■	■
Jedes Attribut einer Relation ist Teil von mindestens einem Superschlüssel dieser Relation.	■	■
Jeder Kandidatenschlüssel kann als Primärschlüssel verwendet werden.	■	■
Eine Relation kann mehrere verschiedene Fremdschlüssel beinhalten.	■	■
Logische Datenunabhängigkeit ermöglicht es, das konzeptuelle Schema zu ändern, ohne auch das interne Schema ändern zu müssen.	■	■
Physische Datenunabhängigkeit erlaubt die Speicherstruktur zu ändern, ohne dabei auch das externe Schema ändern zu müssen.	■	■

(5 Punkte)

Name:

Matrikelnummer:

- b)** Überführen Sie das ER-Modell aus Abbildung 1 kapazitätserhaltend in ein relationales Modell. Optimieren Sie die Relationen dabei wie in der Vorlesung besprochen durch Verschmelzen.

Zur Notation: Verwenden Sie für eine Relation R mit Attributen A, B, C und einem Primärschlüssel bestehend aus den Attributen A und B die Notation: $R(\underline{A}, B, C)$. Geben Sie nur die Attribute an, nicht aber deren Datentyp.

(7 Punkte)

- c)** Modellieren Sie mit einem EER-Modell folgenden Sachverhalt:

- Ein Verkehrsmittel wird beschrieben durch eine eindeutige Nummer, eine Farbe und das Datum der nächsten Verkehrstauglichkeitsprüfung.
- Verkehrsmittel können Bus, Straßenbahn oder Taxi sein.
- Ein Bus wird beschrieben durch die Anzahl Sitzplätze.
- Ein Taxi wird gekennzeichnet durch die Größe seines Kofferraums.
- Eine Straßenbahn hat einen Typ.

Verwenden Sie entweder Generalisierung, Spezialisierung oder Partitionierung, und begründen Sie kurz Ihre Wahl.

(3 Punkte)

Aufgabe 3 – Funktionale Abhängigkeiten und Normalformen (15 Punkte)

- a) Gegeben sei eine Relation $R(A, B, C, D, E, F)$ mit der folgenden Menge an funktionalen Abhängigkeiten:

$$F = \{ \begin{array}{l} A \rightarrow D E, \\ B \rightarrow F, \\ C E \rightarrow D, \\ E \rightarrow A, \\ F \rightarrow B D, \end{array} \}$$

Alle Attribute der Relation sind atomar.

- a1) Bestimmen Sie die Menge aller Schlüssel der Relation R . *Hinweis: Für jeden korrekten Schlüssel erhalten Sie $\frac{1}{2}$ Punkt, für jeden falschen Schlüssel wird Ihnen $\frac{1}{2}$ Punkt abgezogen. Die Aufgabe wird mit mindestens null Punkten bewertet.*

- a2) In welcher höchsten Normalform befindet sich R ? Begründen Sie, warum R nicht in einer höheren Normalform sein kann.

(3 Punkte)

- b) Gegeben sei eine Relation $R(A, B, C, D)$ mit der folgenden Menge an funktionalen Abhängigkeiten:

$$F = \{ \begin{array}{l} A D \rightarrow B, \\ B \rightarrow C, \end{array} \}$$

Alle Attribute der Relation sind atomar. R befindet sich in 2NF.

- b1) Erweitern Sie F zu einer neuen Menge an funktionalen Abhängigkeiten F_2 , so dass R sich höchstens in 1NF befindet. Fügen Sie genau eine neue funktionale Abhängigkeit hinzu. Es darf dabei keine funktionale Abhängigkeit aus F entfernt werden.

- b2) Erweitern Sie F zu einer neuen Menge an funktionalen Abhängigkeiten F_3 , so dass R sich in mindestens 3NF befindet. Fügen Sie genau eine neue funktionale Abhängigkeit hinzu. Es darf dabei keine funktionale Abhängigkeit aus F entfernt werden.

(3 Punkte)

- c) Histogramme zeigen die Häufigkeiten mit denen einzelne Werte auftreten.
- c1) Wofür können Histogramme in Datenbanksystemen verwendet werden? Erläutern Sie kurz den Einsatz.
- c2) Gegeben seien die folgenden Datenwerte eines Attributs:

1, 1, 1, 1, 1, 1, 2, 2, 3, 3, 4, 4, 5, 5, 5, 6, 6, 6

Geben Sie für ein Equi-Width- und ein Equi-Depth-Histogramm für dieses Attribut jeweils die Inhalte der Buckets an. Die Anzahl der Buckets soll jeweils drei sein. Füllen Sie hierfür die Tabelle auf dem Antwortblatt aus.

(4 Punkte)

- d) Gegeben sei die Relation $R(A, B, C, D)$ mit der folgenden Menge an funktionalen Abhängigkeiten:

$$F = \left\{ \begin{array}{l} A \rightarrow C D, \\ B D \rightarrow C, \\ D \rightarrow A, \end{array} \right\}$$

Geben Sie eine minimale, verbundtreue, abhängigkeitserhaltende Zerlegung von R an, welche in dritter Normalform (3NF) ist. Benutzen Sie für das Ergebnis Ihrer Zerlegung die Notation:

$$S = \{(R_i, K_i), \dots, (R_n, K_n)\}$$

wobei R_i für ein Relationenschema und K_i für die entsprechende Schlüsselmenge stehen. Verwenden Sie zur Berechnung das aus der Vorlesung bekannte Syntheseverfahren. Geben Sie für jede Umformung das Zwischenergebnis und abschließend die Zerlegung von R an, die sich aus der Anwendung des Syntheseverfahrens ergibt.

(5 Punkte)

Aufgabe 4 – Nebenläufigkeit und Transaktionen (15 Punkte)

a) Gegeben sei der folgende Ausschnitt einer History:

Schritt	T_1	T_2
1	Read(A, a1)	
2	a1 := a1 + 1	
...	Auf dem Lösungsblatt auszufüllen	
3 + i		Read(A, a2)
4 + i		a2 := a2 - 2
5 + i		Write(A, a2)
6 + i		commit
...	Auf dem Lösungsblatt auszufüllen	

Erstellen Sie drei Histories H_1, H_2 und H_3 , indem Sie die Transaktion T_1 jeweils in den vorgegebenen Kästchen zwischen den Schritten 2 und $3 + i$ sowie nach dem Schritt $6 + i$ ergänzen. $i (\geq 0)$ steht hierbei für die Anzahl an Operationen, die Sie jeweils in das erste Kästchen zwischen den Schritten 2 und $3 + i$ eintragen. Ergänzen Sie die Histories so, dass

- H_1 ein Beispiel für „Lost Update“ ist,
- H_2 ein Beispiel für „Dirty Read“ ist und
- H_3 ein Beispiel für „Non-Repeatable Reads“ ist.

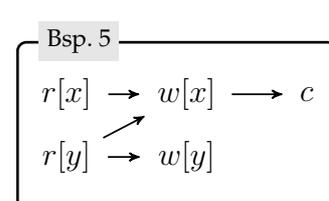
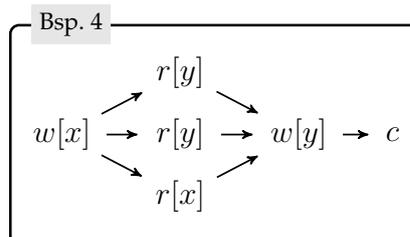
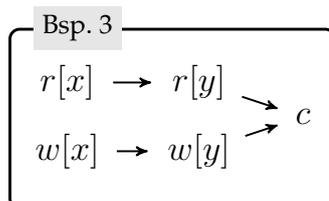
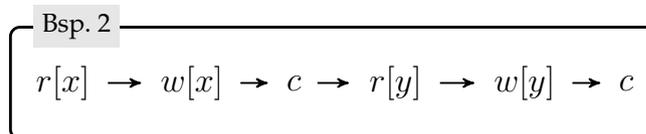
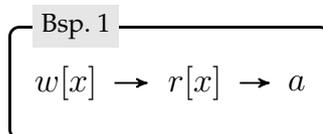
Geben Sie außerdem an, ob ihre Histories serialisierbar sind. Achten Sie hierfür darauf, dass Sie explizit angeben, ob T_1 committed oder aborted.

Hinweis: Die Notation entspricht der aus den Vorlesungsfolien. So bedeutet beispielsweise Read(A, a1), dass das Datenobjekt A gelesen und dessen Wert in der lokalen Variable a1 gespeichert wird.

(4,5 Punkte)

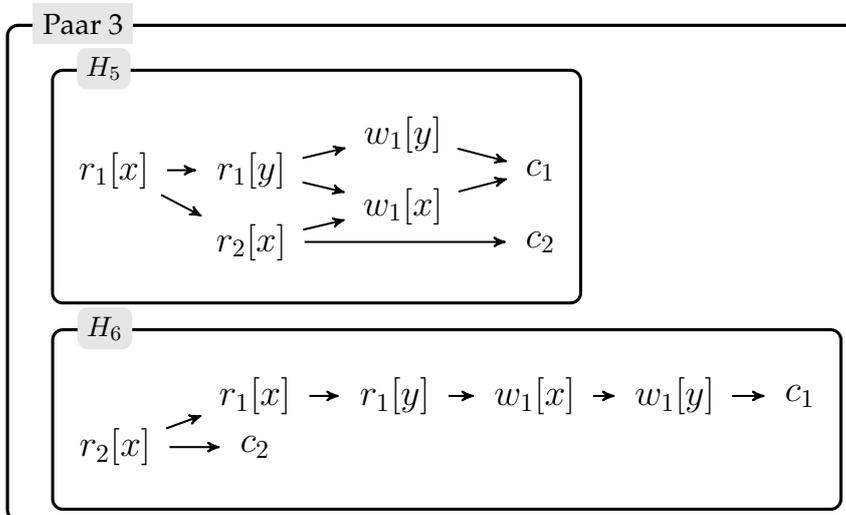
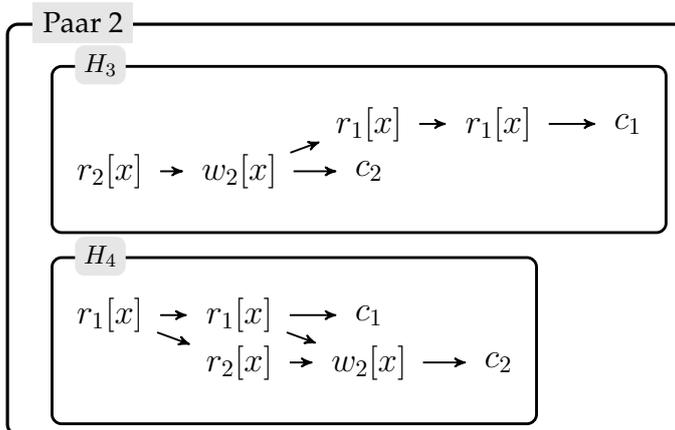
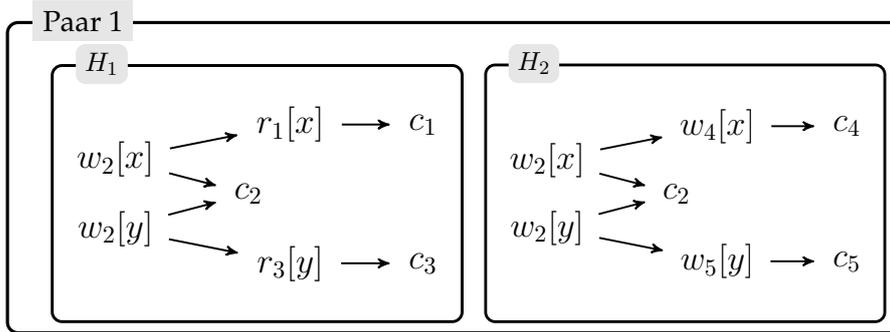
b) Geben Sie für jedes der folgende Beispiele an, ob es sich hierbei um eine Transaktion gemäß der Definition aus der Vorlesung handelt.

Hinweis: Für jede korrekte Antwort erhalten Sie $\frac{1}{2}$ Punkt, für jede falsche Antwort wird Ihnen $\frac{1}{2}$ Punkt abgezogen. Die Aufgabe wird mit mindestens null Punkten bewertet.



(2,5 Punkte)

- c) Geben Sie für jedes der folgenden History-Paare an, ob diese konflikt-äquivalent sind. Falls zwei Histories *nicht* konflikt-äquivalent sind, begründen Sie warum.



(3 Punkte)

- d) Gegeben seien die folgenden Histories, die auf vier Transaktionen $T_1 - T_4$ beruhen:

$$H_1 = r_1[z] r_3[x] r_1[x] w_2[y] w_4[y] r_4[z] w_1[z] w_3[x] c_3 r_2[x] r_4[y] c_4 w_1[z] c_1 c_2$$

$$H_2 = r_1[z] w_4[y] r_3[x] w_3[x] c_3 r_1[x] w_1[z] w_2[y] r_2[x] r_4[z] c_2 r_4[y] c_4 w_1[z] c_1$$

Konstruieren Sie jeweils für H_1 und H_2 die Serialisierbarkeitsgraphen $SG(H_1)$ und $SG(H_2)$. Begründen Sie für jede History, ob sie serialisierbar ist.

(5 Punkte)