

# Informationssysteme und Datenanalyse

## SS 2015

### Klausur 18.07.2015

---

#### Angaben

Nachname: \_\_\_\_\_

Vorname: \_\_\_\_\_

Matrikelnummer: \_\_\_\_\_

Fakultät / Studiengang (bitte ankreuzen):

- Fak IV - Bachelor Informatik
- Fak IV - Bachelor Wirtschaftsinformatik
- Fak IV - Bachelor Technische Informatik
- Fak IV - StuPO 90 Informatik
- BSc – Wirtschaftsingenieurwesen
- BSc – Wirtschaftsmathematik
- BSc – Mathematik
- MSc – Wirtschaftsingenieurwesen
- MSc – Wirtschaftsmathematik
- MSc – Mathematik
- Andere \_\_\_\_\_

Auf jedem Blatt bitte Namen und Matrikelnummer angeben!

#### Organisatorisches

Bearbeitungszeit: 75 Minuten

Erreichbare Punkte: 70

Zugelassene Hilfsmittel: Nur ein Wörterbuch (kein elektronisches!)

Weitere Hilfsmittel sind nicht zugelassen.

Aufgabe	Punkte	Erreicht	Korrektor
1	9,5		
2	5		
3	9		
4	6		
5	9		
6	3		
7	11		
8	17,5		
Punktsumme	70		

## Aufgabe 1: Relationales Datenmodell/Reverse Engineering (9,5 Punkte)

Gegeben sei ein Informationsmodell für eine Auto-Reparatur-Werkstatt, mit dem die Planung von Reparaturen innerhalb eines Tages incl. der Zuordnung von Mechanikern und Hebebühnen modelliert worden ist. Dieses Modell ist in die entsprechenden Tabellen einer relationalen Datenbank umgesetzt worden, von der man unten einen konkreten Zustand hier sieht:

Mechaniker	MechID	Name	Spezialisiert_auf	Auch_spezialisiert auf
	M1	Johansson	Volvo	Saab
	M2	Carlsson	Saab	Volkswagen
	M3	Lindström	Saab	Volvo
	M4	Persson	Peugeot	Renault
	M5	Lundquist	Volvo	Ford

Kunde	KundeID	KdName
	101	Meyer
	102	Andersson
	103	Winter
	104	Dahlgren
	105	Andersson
	106	Petersson

Auto	Kennzeichen	Marke	KundeID
	SKF 632	Volvo	101
	TVD 452	Volkswagen	101
	SXD 440	Volvo	103
	KJH 421	Saab	104
	BGN 333	Saab	105
	DGJ 124	Volvo	106

Reparatur	RepID	Kennzeichen	Problem	Arbeitsbühne	Anfang	Ende
	R1	SKF 632	Inspection	P1	10.07.2015-08:00	10.07.2015-11:00
	R2	TVD 452	Brake	P1	10.07.2015-12:00	10.07.2015-14:00
	R3	SXD 440	Engine	P2	10.07.2015-08:00	10.07.2015-14:00
	R4	KJH 421	Windscreen	P3	10.07.2015-08:00	10.07.2015-10:00
	R5	BGN 333	GearChange	P1	10.07.2015-15:00	10.07.2015-18:00
	R6	DGJ 124	Inspection	P3	10.07.2015-13:00	10.07.2015-15:00
	R7	DGJ 124	Brake	P3	10.07.2015-15:00	10.07.2015-18:00

Verantwortlich	RepID	MechID
	R1	M5
	R2	M2
	R3	M3
	R4	M1
	R5	M2
	R6	M1
	R7	M1

Assistiert	RepID	MechID
	R2	M5
	R3	M4
	R4	M2
	R5	M5
	R7	M4
	R5	M3

Name:

Matr.Nr.:

ISDA – Klausur – SS 2015 - 18.07.2015

---

### **Aufgabe 1.1(8 Punkte)**

Rekonstruieren Sie aus den vorliegenden Tabellen das zugrundeliegende ER-Typ-Diagramm (kein Glossar erstellen!). Versuchen Sie aus den Tabellen möglichst gut auf die Integritätsbedingungen im Original-Modell zurückzuschließen.

Rekonstruiertes ER-Typ-Diagramm:

Name:

Matr.Nr.:

ISDA – Klausur – SS 2015 - 18.07.2015

---

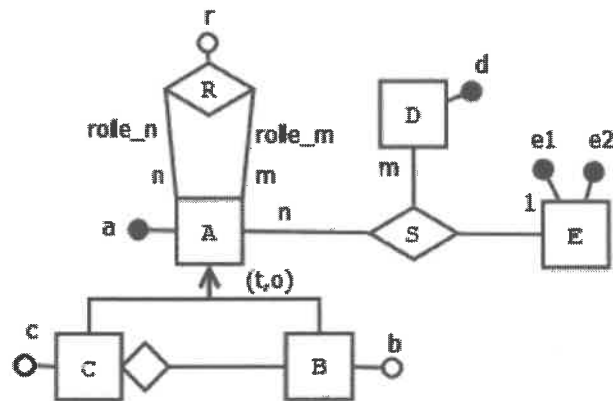
### Aufgabe 1.2 (1,5 Punkte)

Betrachten Sie die folgenden Integritätsbedingungen und prüfen Sie, ob sich die Bedingungen aus den gegebenen Tabellen ableiten lassen, bzw. ob die Bedingungen erfüllt sind.

<b>Integritätsbedingung</b>	<b>Erfüllt</b>	<b>Nicht Erfüllt</b>
Für jede Reparatur gibt es genau einen verantwortlichen Mechaniker		
An jeder Reparatur muss mindestens ein verantwortlicher Mechaniker mitwirken, der die jeweilige Automarke als 1. oder 2. Spezialisierung besitzt.		
Jede Arbeitsbühne darf zu einem Zeitpunkt nur von höchstens einem Auto belegt sein.		

## Aufgabe 2: Relationaler Datenbankentwurf (5 Punkte)

Gegeben sei das folgende abstrakte EER-Diagramm.

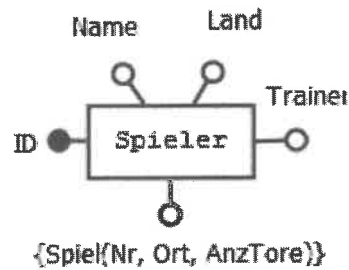


### Aufgabe:

Bilden Sie das gegebene abstrakte EER-Diagramm in ein relationales Schema ab. Fremdschlüssel müssen vorhanden sein, müssen aber nicht als solche gekennzeichnet werden. Fassen Sie die Relationen soweit wie möglich zusammen.

**Aufgabe 3: Normalisierung (9 Punkte)**

Gegeben sei folgendes ER-Diagramm mit einem Entity-Typ und eine Beispielinstantz der Relation:



Spieler	<u>ID</u>	Name	Land	Trainer	Spiel		
					Nr	Ort	AnzTore
1	Schürrle	DE	Löw	4	Salvador	0	
				1	Belo Horizonte	2	
				2	Porto Alegre	1	
2	Kroos	DE	Löw	1	Belo Horizonte	1	
				3	Rio de Janeiro	0	
				9	Rio de Janeiro	1	
3	Oscar	BRA	Scolari	7	Belo Horizonte	0	
4	Higuaín	ARG	Sabella	6	Brasília	1	
5	Oscar	COL	Pekerman	5	Belo Horizonte	1	

Die Relation 'Spieler' ist nicht in der 1. NF. Normalisieren Sie die Relation schrittweise bis zur BCNF. Sie müssen bei der Zerlegung die Instanzen nicht angeben, es reicht also das Relationen-Schema. Primärschlüssel müssen gekennzeichnet werden. Fremdschlüssel müssen ebenso vorhanden sein, müssen aber nicht als solche gekennzeichnet werden. Es müssen auch in jedem Schritt die funktionalen Abhängigkeiten angegeben werden, die die jeweilige Normalform verletzen.

*Hinweis:* 'AnzTore' steht für die Anzahl geschossener Tore des jeweiligen Spielers für das jeweilige Spiel!

Name:

Matr.Nr.:

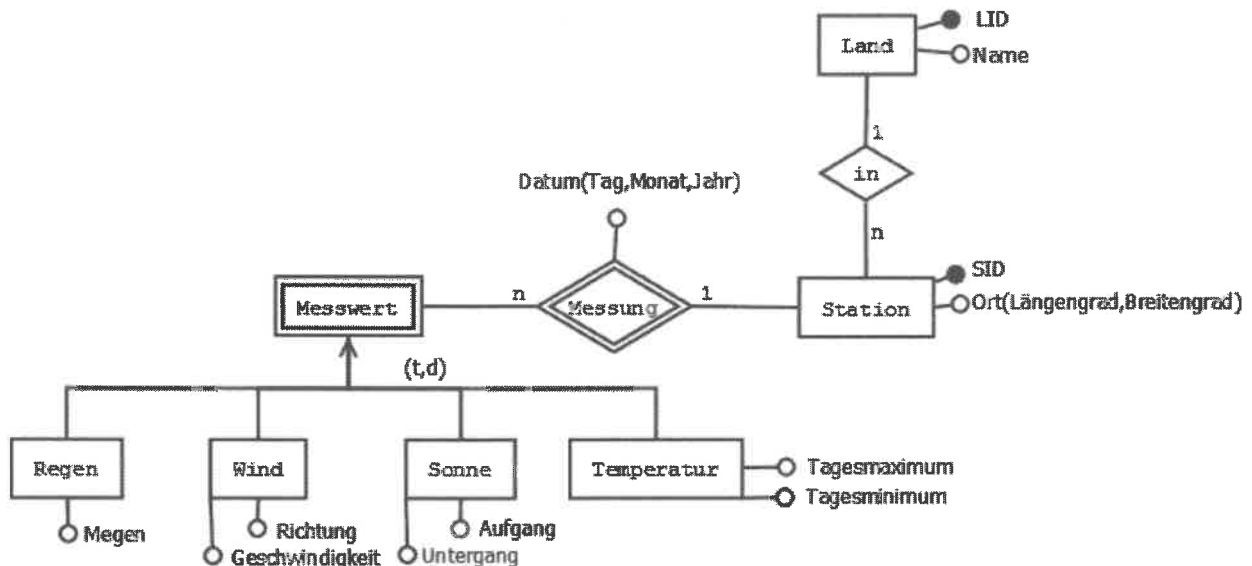
ISDA – Klausur – SS 2015 - 18.07.2015

---

## Aufgabe 4: Relationale Algebra (6 Punkte)

Gegeben sind das folgende EER-Typ-Diagramm und ein dazugehöriges relationales Schema:

### ER-Diagramm



### Relationales Schema:

Land(LID, Name)

Station(SID, Längengrad, Breitengrad, LID → Land)

Messwert(SID → Station, Tag, Monat, Jahr, Typ, Regenmenge, Windrichtung, Windgeschwindigkeit, Sonnenaufgang, Sonnenuntergang, Tagesmaximum, Tagesminimum)

Typ ∈ { Regen, Wind, Sonne, Temperatur }

Zur Vereinfachung können folgende Abkürzungen für die Messwertattribute verwendet werden:

Regenmenge **r**, Windrichtung **wr**, Windgeschwindigkeit **wg**, Sonnenaufgang **sa**, Sonnenuntergang **so**, Tagesmaximum **tm**, Tagesminimum **tmi**.

### Aufgabe 4.1

Beschreiben Sie die folgenden Anfragen in relationaler Algebra:

1. Wie viele Messungen gibt es aus Deutschland? (Land.Name = „Deutschland“) (1 Punkt)

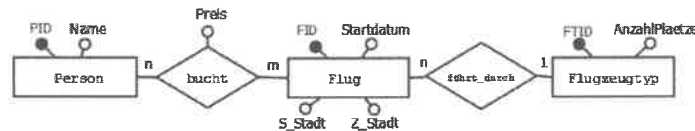




## Aufgabe 5: SQL Anweisungen (9 Punkte)

Gegeben sind das folgende EER-Typ-Diagramm und ein dazugehöriges relationales Schema mit einer Beispielinstantz.

### EER-Diagramm



### Relationales Schema mit Beispieldatensätze:

Person

PID	Name
1	Till
2	Franka
3	Karoline

bucht

PID	FID	Preis
2	4	300
3	4	280
2	6	100

Flug

FID	Startdatum	FTID	S_Stadt	Z_Stadt
4	03.04.2015	1	Berlin	Madrid
5	06.05.2015	1	Paris	Berlin
6	05.07.2015	2	Hamburg	Rom

Flugzeugtyp

FTID	AnzahlPlaetze
1	130
2	90

## Aufgabe

- Gegeben ist folgende Anfrage in SQL. Geben Sie das Ergebnis dieser Anfrage in tabellarischer Form (Vergessen Sie nicht die Attribute) an: (1 Punkt)

```
SELECT p.PID, COUNT(b.PID) as X
FROM Person p LEFT JOIN Bucht b on p.PID = b.PID
GROUP BY p.PID
HAVING COUNT(b.PID) < 2;
```

- Gegeben sind die folgenden zwei Anfragen in SQL. Geben Sie das Ergebnis dieser Anfragen in tabellarischer Form an (Vergessen Sie nicht die Attribute). Sind die Anfragen im allgemeinen äquivalent (liefern sie immer das gleiche Ergebnis)? (2 Punkte)

Anfrage 1:

```
SELECT *
FROM Person p
WHERE EXISTS (SELECT *
              FROM Bucht b, Flug f
              WHERE b.fid = f.fid AND
                    p.PID = b.PID AND
                    (f.Z_stadt = 'Madrid' OR f.Z_stadt = 'Rom'));
```

Anfrage 2

```
SELECT p.*
FROM Person p JOIN Bucht b on p.PID = b.PID JOIN Flug f on b.FID = f.FID
WHERE f.Z_stadt = 'Madrid'
UNION
SELECT p.*
FROM Person p JOIN Bucht b on p.PID = b.PID JOIN Flug f on b.FID = f.FID
WHERE f.Z_stadt = 'Rom';
```

Name:

Matr.Nr.:

ISDA – Klausur – SS 2015 - 18.07.2015

---

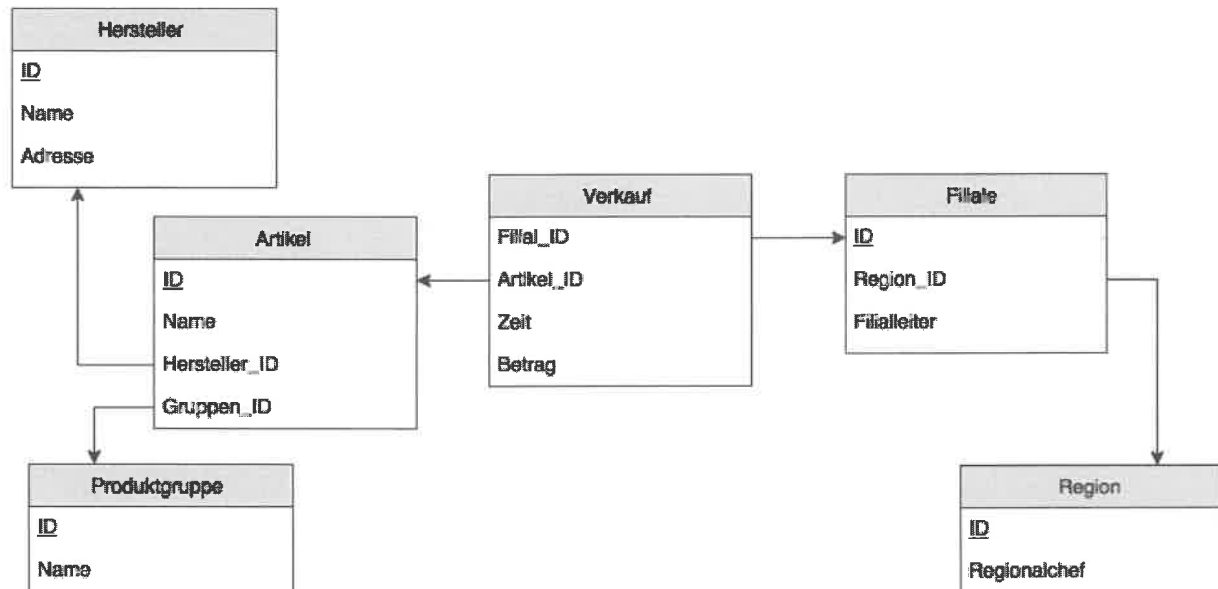
3. Gegeben ist folgende Anfrage in SQL. Geben Sie das Ergebnis in tabellarischer Form an (Vergessen Sie nicht die Attribute) und die umgangssprachliche Formulierung dieser Anfrage (Max. zwei Sätze, besch): (4 Punkte)

```
SELECT *
FROM (SELECT b.PID
      FROM BUCHT b join FLUG f on b.FID = f.FID
      WHERE f.Z_stadt = 'Madrid') as A,
      (SELECT b.PID
      FROM BUCHT b join FLUG f on b.FID = f.FID
      WHERE f.Z_stadt = 'Rom') as B
WHERE A.PID = B.PID;
```

1. Erstellen Sie die SQL Select-Anweisung, die die folgende Anfrage berechnet:  
Die Durchschnittliche Anzahl der Plätze aller Flüge nach Berlin. (2 Punkte)

## Aufgabe 6: DWH-Mehrdimensionale Modellierung(3 Punkte)

Betrachten Sie das folgende ER-Schema eines OLAP Würfels:



### Aufgabe:

1. Benennen sie die Fakten- und Dimensionstabellen: (0.5 Punkte)
2. In der Vorlesung wurden drei Ansätze vorgestellt, um einen mehrdimensionalen OLAP Würfel auf relationale Tabellen abzubilden. Welchem der vorgestellten Ansätze entspricht dieses Schema? (0.5 Punkte)
3. Wie sähe die Faktentabelle des Schemas aus, falls eine Fullfact-Darstellung verwendet worden wäre? (Geben Sie hierzu die neue Faktentabelle mit allen Attributen an.) (1 Punkt)
4. Nennen Sie je einen Vor- und Nachteil der Fullfact-Repräsentation gegenüber der vorhanden Darstellung: (1 Punkt)

Name:

Matr.Nr.:

ISDA – Klausur – SS 2015 - 18.07.2015

---

### Aufgabe 7: Datenanalyse (11 Punkte)

Gegeben ist der folgende Datensatz:

<b>X</b>	11	2	1	13	10	9	4	5	3	2	6
<b>Y</b>	-0,75	-0,5	0,5	0,25	-2,5	-3	-4	-4,5	-2,75	-1,75	-4,5
<b>Gruppe</b>	Oxi	Oxi	Oxi	Oxi	Oxi	Ne	Ne	Ne	Ne	Ne	Ne

#### Aufgabe 7.1 – Klassifikation: (1.5 Punkte)

Klassifizieren Sie die Attribute „X“, „Y“ & „Gruppe“ wie in der Vorlesung besprochen:

#### Aufgabe 7.2 – Basisstatistiken: (1.5 Punkte)

Berechnen Sie Median, Mittelwert und Modalwert (Modus) der X-Werte:

#### Aufgabe 7.3 – Scatterplot: (2 Punkte)

Skizzieren Sie einen (sinnvoll beschrifteten) Scatterplot der X und Y Werte. Markieren sie (z.B. durch Umkreisen) in ihrem Plot zudem, welche der Datenpunkte zur Gruppe „Oxi“ und „Ne“ gehören:

**Aufgabe 7.4 – Klassifikation: (2 Punkte)**

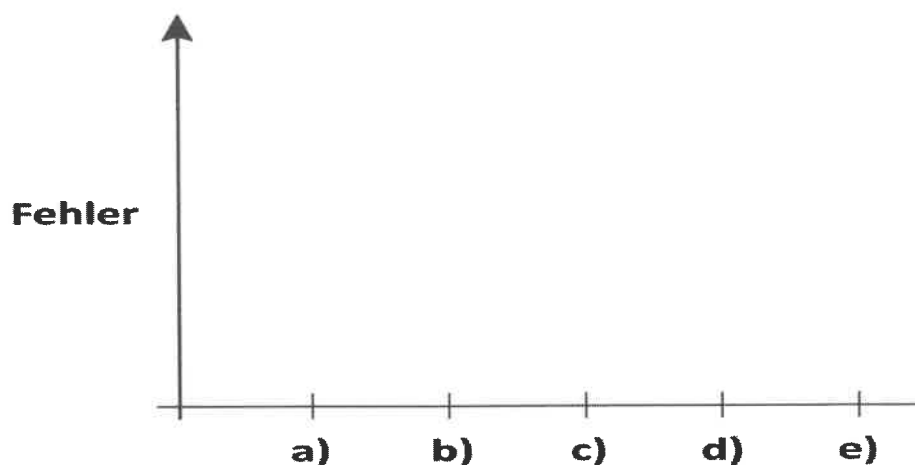
Tragen Sie die folgenden drei Hyperebenen in ihren Plot aus Aufgabe 7.3 ein. Welche eignet sich am besten zur Klassifikation der Daten nach Oxi / Ne?

- a)  $y + 2 = 0$
- b)  $x - 9 = 0$
- c)  $0,5 \cdot x + y = 0$

**Aufgabe 7.5 – Regression: (2 Punkte)**

Sie möchten nun mittels einer Regressionsanalyse den Zusammenhang der Variablen X und Y modellieren. Hierzu soll die Variable X zunächst in einen höheren Raum transformiert werden. Skizzieren Sie für die fünf angegebenen Transformationen den für ein lineares Regressionsmodell zu erwartenden Trainings- und Testfehler. Welche Transformationen würden Sie wählen? Erläutern Sie kurz (1 Satz) ihre Entscheidung.

- a)  $X \rightarrow [X]$
- b)  $X \rightarrow [X, X^2]$
- c)  $X \rightarrow [1, X, X^2]$
- d)  $X \rightarrow [1, X, X^2, X^3, X^4]$
- e)  $X \rightarrow [1, X, X^2, X^3, X^4, \dots, X^{15}]$



Name:

Matr.Nr.:

ISDA – Klausur – SS 2015 - 18.07.2015

---

**Aufgabe 7.6 – Modellauswertung: (2 Punkte)**

Sie evaluieren ein binäres Klassifikationsmodell. Auf den Testdaten erhalten Sie die folgenden Werte:

	<b>Klasse = 1</b>	<b>Klasse = 0</b>
<b>Vorhersage = 1</b>	8	2
<b>Vorhersage = 0</b>	92	998

Berechnen Sie Precision & Recall für dieses Modell. Ist das Modell eher gut oder eher schlecht? Erläutern Sie kurz (1 Satz).

**Aufgabe 8: MC (17,5 Punkte)**

Diese Aufgabe umfasst **21 MC-Fragen**. Sie bestehen aus jeweils drei Antwortalternativen, wobei mindestens eine Alternative richtig und mindestens eine Alternative falsch ist. Im Allgemeinen gibt es für jede Aufgabe, in welcher alle korrekten Aussagen markiert worden sind, zwischen 0,5 und 3 Punkte. Wurden nicht alle richtigen Aussagen identifiziert oder falsche Aussagen markiert, wird die jeweilige Aufgabe mit 0 Punkten bewertet.

1. Welche/er der folgenden Operatoren aus der relationalen Algebra ist KEIN Basisoperator? (0,5 Punkte)
  - Differenz
  - Schnittmenge
  - Natural Join
  
2. Was für ein Prinzip muss bei der Ausführung von Transaktionen eingehalten werden? (0,5 Punkte)
  - ACID-Prinzip
  - Merge-Prinzip
  - Serialisation-Prinzip
  
3. Welche Eigenschaften erfüllen alle Normalformen (1. NF bis BCNF)? (0,5 Punkte)
  - Abhängigkeitstreue
  - Die Anzahl der Relationen bleibt bei jedem Umformungsschritt gleich
  - Verlustfreiheit
  
4. Welche Aussagen zu einem Primärschlüssel sind richtig? (0,5 Punkte)
  - Ein Primärschlüssel muss minimal sein.
  - Ein Primärschlüssel darf kein Fremdschlüssel sein.
  - Ein Primärschlüssel-Wert darf nicht NULL sein.
  
5. Welche der folgenden Mengen von Operatoren besteht nur aus Basisoperatoren? (0,5 Punkte)
  - $\{\pi, \bowtie, \cap, \sigma, \times, \rho\}$
  - $\{\sigma, \pi, \cup, \times, -, \rho\}$
  - $\{/, \times, \pi, \sigma, \bowtie, \cup\}$
  
6. Welche Reihenfolgen der folgenden SQL-Anweisungen sind möglich? (0,5 Punkte)
  - SELECT ... FROM (SELECT ... FROM ... WHERE ...)
  - SELECT ... GROUP BY ... FROM
  - SELECT ... FROM ... WHERE ... GROUP BY ...
  
7. Bei welchem Konstrukt ist die Reihenfolge irrelevant und Duplikate sind erlaubt? (0,5 Punkte)
  - Mengen
  - Multimengen
  - Supermengen



8. Durch welchen Ausdruck in relationaler Algebra wird die nachstehende SQL-Anfrage richtig ausgedrückt? (1 Punkt)

```
SELECT *
FROM Film, Studio
WHERE Film.StudioNr = Studio.ID AND Film.Länge > 120
```

- $\pi_{\text{StudioNr}}(\sigma_{\text{Länge} > 120}(\text{Film} \times \text{Studio}))$
- $\sigma_{\text{Film.Länge} > 120 \text{ AND } \text{Film.StudioNr} = \text{Studio.ID}}(\text{Film} \times \text{Studio})$
- $\sigma_{\text{Film.Länge} > 120}(\text{Film} \bowtie_{\text{Film.StudioNr} = \text{Studio.ID}} \text{Studio})$

9. Was sind Vorteile der Normalisierung? (0,5 Punkte)

- Anfragen können immer schneller bearbeitet werden.
- Anomalien werden weitestgehend beseitigt.
- Es werden weniger Superschlüssel benötigt.

10. Welche der folgenden Aussagen gelten?(0,5 Punkte)

- Die Funktionale Abhängigkeit definiert einen Constraint für alle möglichen Instanzen einer Relation.
- Die Funktionale Abhängigkeit definiert einen Constraint für eine konkrete Instanz einer Relation.
- Ein Fremdschlüssel bestimmt meistens alle anderen Attribute funktional.

11. Wie lautet die natürlichsprachliche Beschreibung der nachstehenden SQL-Anfrage? (3 Punkte)

```
SELECT a.Vorname, a.Nachname
FROM Angestellter a
WHERE NOT EXISTS (
  SELECT *
  FROM Projekt p
  WHERE p.ProjektOrt <> 'Köln'
  AND NOT EXISTS (
    SELECT *
    FROM ArbeitetIn ai
    WHERE ai.PID = p.PID
    AND ai.AID = a.AID));
```

- Finde alle Angestellten, die an allen Projekten in Köln arbeiten.
- Finde alle Angestellten, die an allen Projekten arbeiten, die nicht in Köln sind.
- Finde alle Angestellten, die an irgendeinem Projekt arbeiten, das nicht in Köln ist.

12. Wenn dieselbe Lese-Anfrage in einer Transaktion nacheinander unterschiedliche Ergebnisse liefert, dann ist das ein Beispiel für ein ... (0,5 Punkte)

- Non-Repeatable Read
- Phantom-Problem
- Dirty Read

## 13. Folgender Ausschnitt einer Table Definition ist gegeben: (1 Punkt)

```
CREATE TABLE X (
...
, ref INTEGER FOREIGN KEY REFERENCES Y(id) ON DELETE CASCADE
, ...)
```

Die Tabelle Y wird nun komplett gelöscht. Dann gilt:

- Die Tabelle X wird ebenfalls komplett gelöscht
- Elemente aus X, bei denen 'ref' gesetzt ist werden ebenfalls gelöscht
- Das Attribut 'ref' wird aus dem Schema entfernt

14. Gegeben sei die Relation R (A, B, C, D, E) mit den funktionalen Abhängigkeiten:  $A \rightarrow B$ ,  $B \rightarrow C$  und  $(B, C) \rightarrow D$ . Welche der folgenden Möglichkeiten ist ein möglicher Primärschlüssel für R? (2 Punkte)

- A
- (A, E)
- (A, C)

## 15. Welche/r der folgenden Begriffe ist keine Eigenschaft des ACID-Prinzips?(0,5 Punkte)

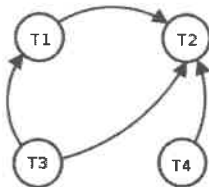
- Consistency.
- Durability.
- Invocation.

## 16. Welche Anomalie könnte in dem gegebenem Schedule auftreten?(0,5 Punkte)

T1	T2
R(A)	
	R(A)
W(A)	
	W(A)

- Phantom-Problem.
- Lost-Update.
- Non-Repeatable Write.

## 17. Gegeben sei zu einem Schedule S1 folgender Graph (1 Punkt)



Welche Aussagen sind richtig?

- S1 ist konfliktserialisierbar, ein konfliktäquivalenter serieller Schedule zu S1 ist:  $T4 \rightarrow T3 \rightarrow T1 \rightarrow T2$ .
- S1 ist konfliktserialisierbar, ein konfliktäquivalenter serieller Schedule zu S1 ist:  $T3 \rightarrow T1 \rightarrow T4 \rightarrow T2$ .
- S1 ist nicht konfliktserialisierbar, da der Graph ein Zyklus enthält.

18. Bei der Anfrageoptimierung ... (0,5 Punkte)

- ... sollten Joins immer vor Selektion ausgeführt werden.
- ... sollten Selektionen vor Joins ausgeführt werden.
- ... ist die Reihenfolge von Joins und Selektionen egal.

19. Bei folgenden Anfragen in relationaler Algebra gilt: (1 Punkt)

(1)  $\sigma_C(X \bowtie Y)$  und

(2)  $\sigma_C(X) \bowtie \sigma_C(Y)$  gilt:

- (1) ist immer gleich gut oder performanter als (2).
- (2) ist immer gleich gut oder performanter als (1).
- (2) ist immer performanter als (1).

20. Welche der folgenden Gleichungen gelten in der relationalen Algebra? (1 Punkt)

- $\pi_A(R \cup S) = \pi_A(R) \cup \pi_A(S)$
- $\sigma_{R.A1=S.A2}(R \times S) = R \bowtie_{R.A1=S.A2} S$
- $\pi_A(R \cap S) = \pi_A(R) \cap \pi_A(S)$

21. Seien folgende Relationen  $R(A, B, C)$  und  $S(A, B, D)$ . Welche Gleichungen gelten? (1 Punkt)

- $\pi_{A,B}(R \bowtie S) = \pi_{A,B}(R) \cap \pi_{A,B}(S)$
- $\pi_{A,B}(R \times S) = \pi_{A,B}(R) \cap \pi_{A,B}(S)$
- $R \bowtie S = S \bowtie R$

