

Informationssysteme und Datenanalyse

Test

25.09.2017

Dies ist der Test der Lehrveranstaltung *Informationssysteme und Datenanalyse*. Bitte füllen Sie die Tabelle auf diesem Deckblatt aus und unterschreiben Sie den untenstehenden Hinweis.

Hinweise:

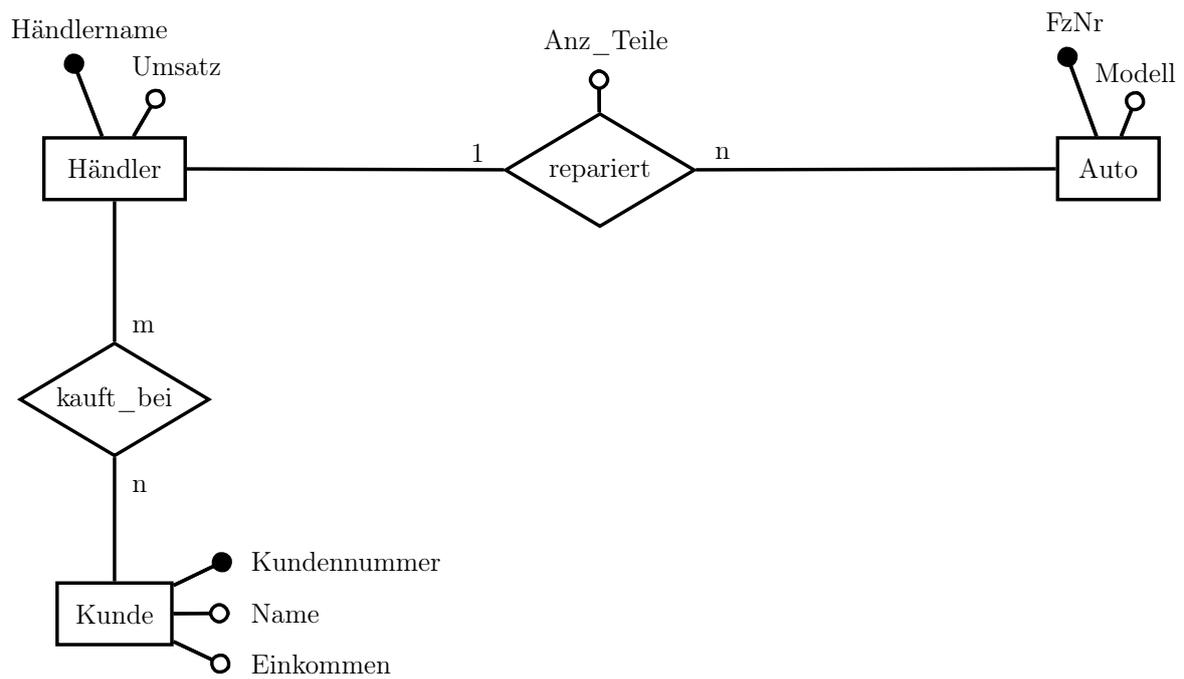
- Die Bearbeitungszeit für diesen Test beträgt 60 Minuten plus 10 Minuten Einlesezeit. Es können in 7 Fragen insgesamt 50 Punkte erreicht werden.
- Wenn Sie mehr als den zur Bearbeitung einer Aufgabe vorgesehenen Platz benötigen, können Sie ihre Antwort auf einer der freien Seiten fortsetzen. Machen Sie eine Weiterführung ihrer Antwort eindeutig kenntlich.
- Dieser Test besteht aus **16** Seiten. Bitte zählen Sie die Vollständigkeit der Seiten direkt nach Beginn der Einlesezeit.
- Bitte schreiben Sie außerdem direkt nach Beginn der Schreibzeit ihren Namen und ihre Matrikelnummer auf jede Seite.
- Die Verwendung von eigenem Papier ist **nicht** erlaubt. Zusätzliche leere Blätter werden auf Nachfrage ausgeteilt.
- Auf Ihrem Platz dürfen sich lediglich mehrere *dokumentenechte* Stifte sowie ihr Personal- und Studierendenausweis befinden. Einträge mit roten oder grünen Stiften sowie Füller und/oder Bleistift werden nicht gewertet. Weitere Hilfsmittel sind nicht zugelassen. Sämtliche elektronischen Geräte müssen sich ausgeschaltet in Ihrer Tasche befinden. Diese müssen Sie in der Reihe vor Ihnen oder anderweitig entfernt von Ihrem Platz abstellen.
- Klingelnde elektronische Geräte (Smartphones, Smartwatches o.Ä.) gelten als Täuschungsversuch.

Matrikelnummer	
Nachname(n)	
Vorname(n)	
Studiengang	
Hiermit bestätige ich, dass ich die oben genannten Hinweise verstanden haben und mich in der Lage fühle, diesen Test durchzuführen.	
Unterschrift:	

Aufgabe	Punkte	Erreicht	Korrektor
Datenbankentwurf	6		
Relationaler Entwurf	6		
Anfragesprachen	12		
Data Streams Management	5		
Data Warehousing	6		
Data Analysis	5		
Multiple Choice	10		
Summe	50		

Aufgabe 1: Datenbankentwurf (6 Punkte)

Gegeben sei das folgende Entity-Relationship-Diagramm für eine *Autohauskette*.

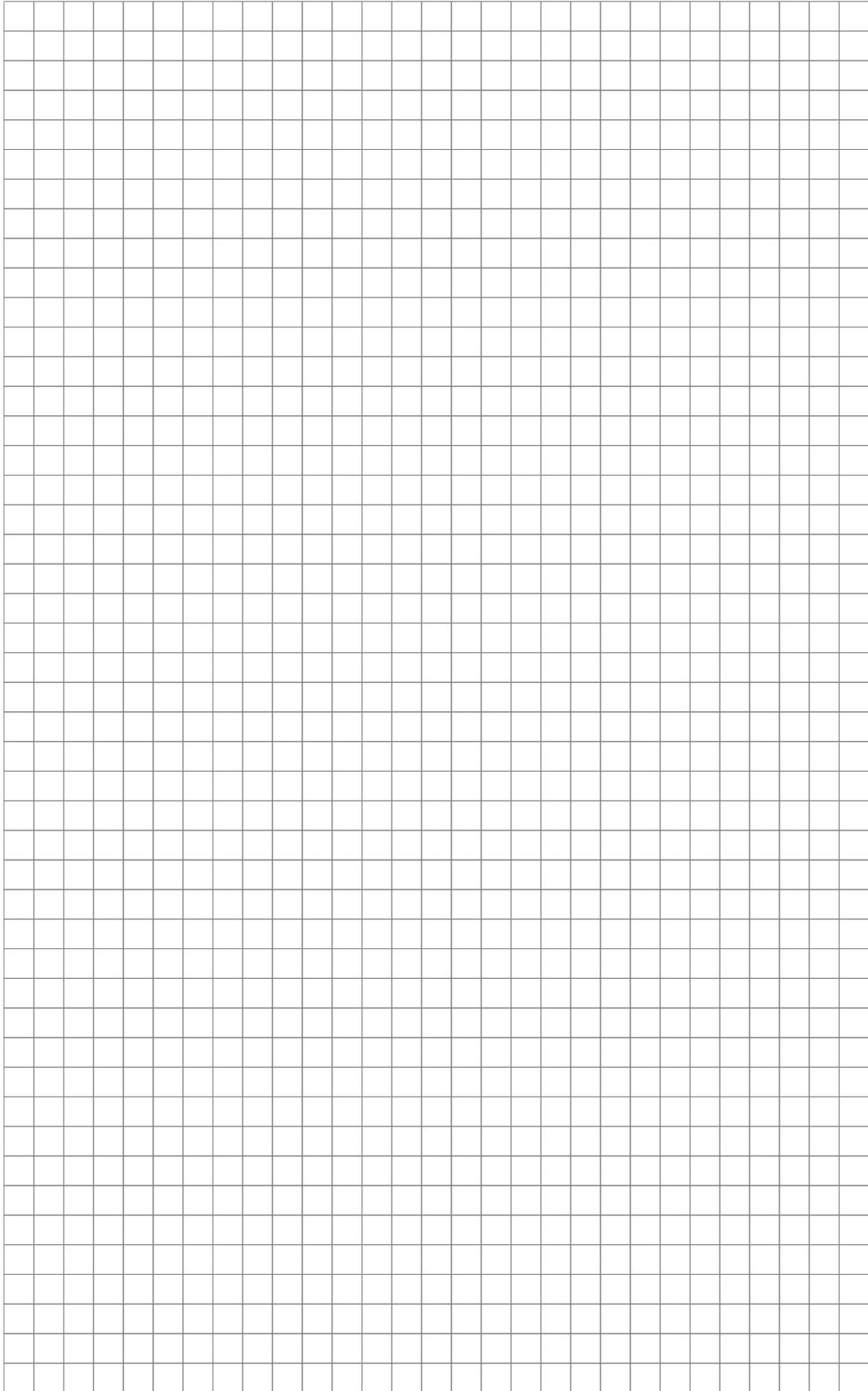


- 1.1. Ergänzen Sie das obenstehende Entity-Relationship-Diagramm um die folgenden Angaben. Achten Sie dabei auch auf mögliche Integritätsbedingungen . (1)
- a) Händler verschrotten Autos gegen eine Gebühr. Natürlich kann ein Auto nur einmal verschrottet werden. (1)
- b) *Jeder* Kunde kauft bei einem Händler. (0,5)
- 1.2. Gegeben seien außerdem die folgenden Relationen. Erweitern Sie das ER-Diagramm aus Aufgabe 1 durch Verwendung eines Abstraktionskonzeptes zu einem erweiterten ER-Diagramm (EER-Diagramm), indem Sie die Informationen aus den gegebenen Relationen verwerten. Weitere Datentupel als die angegebenen existieren nicht. Achten Sie dabei darauf, dass Ihre Modellierung nicht kapazitätserhöhend oder kapazitätsvermindernd ist. (3)

Auto	<u>FzNr</u>	Modell	Kombi	<u>FzNr → Auto</u>	Plaetze	ist_Elekt
	BMX3W	Lins		BMX3W	5	nein
	6J4V	Röns		5U743	7	ja
	5U743	Pus		VWT3N	8	ja
	53275	Sar				
	46P23	Nekk				
	VWT3N	Buli				
	21U73	Weer				

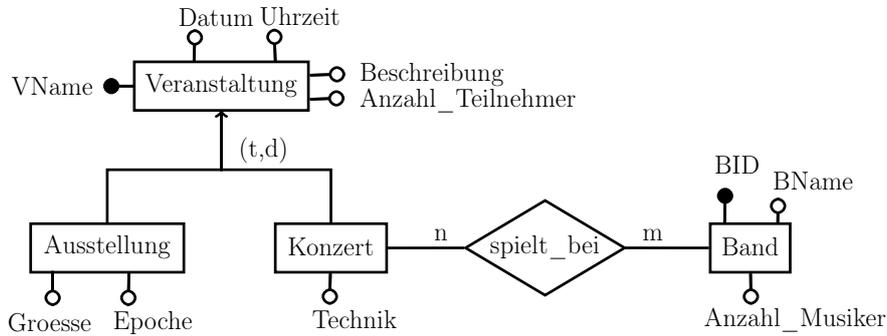
SUV	<u>FzNr → Auto</u>	CO2	Coupé	<u>FzNr → Auto</u>	{Ausstattung}
	46P23	4356		6J4V	{ABS, ESP, AutoPilot}
	21U73	9000			

- 1.3. Sind die folgenden Integritätsbedingungen im ER-Entwurf abgebildet ?
- a) Reparaturen vom selben Händler am selben Wagen können unterschieden werden. (0,5)
- Ja Nein
- b) Ein Auto kann von mehreren Händlern repariert werden. (0,5)
- Ja Nein
- c) Ein SUV muss mindestens 200g CO2 ausstoßen. (0,5)
- Ja Nein



Aufgabe 3: Anfragesprachen (12 Punkte)

Gegeben sei folgendes Datenbankschema eines *Kulturzentrums*. Beispielhafte Tupel sind bereits gegeben.

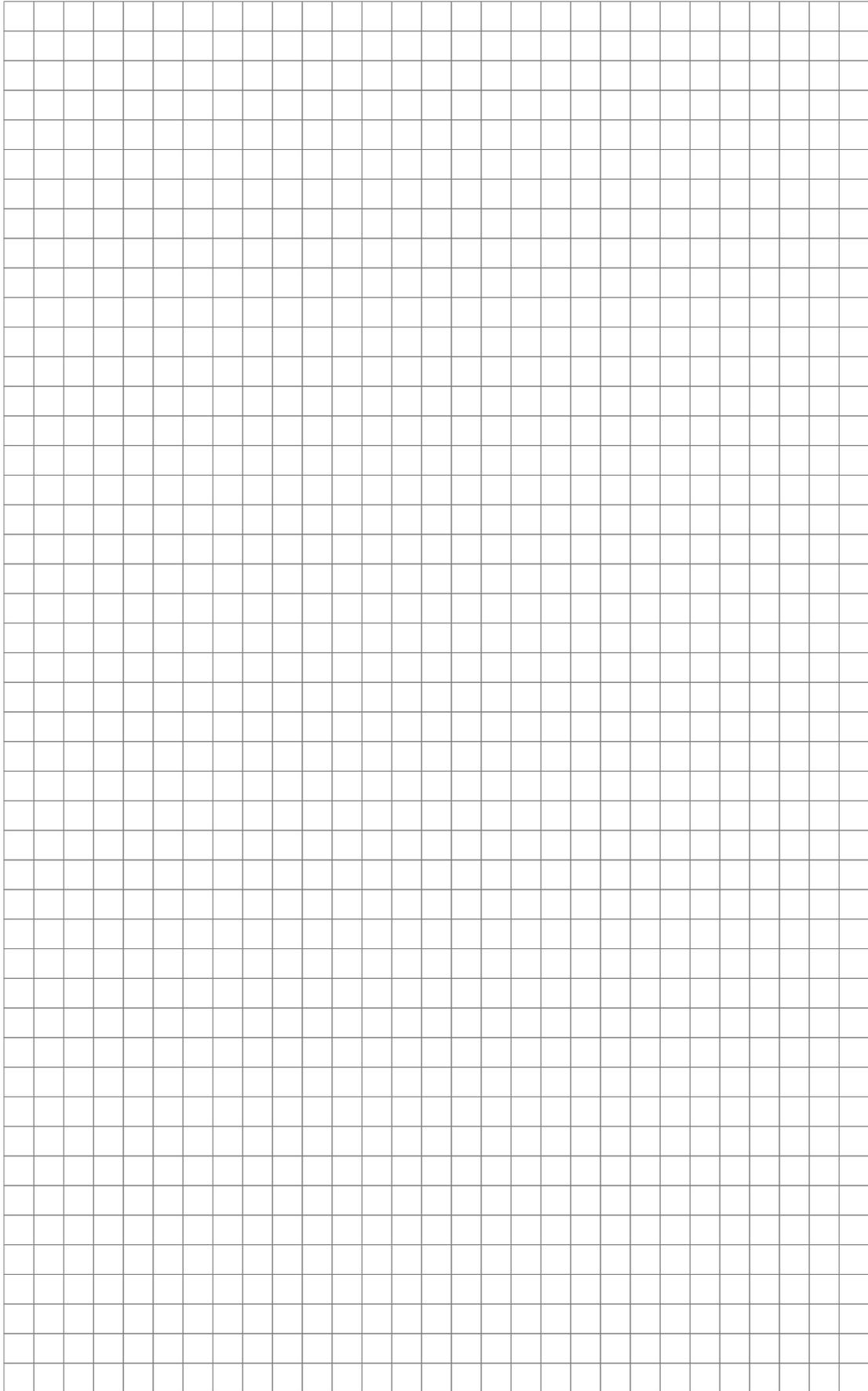


spielt_bei	<u>BID</u>	<u>VName</u>	Band	<u>BID</u>	<u>BName</u>	<u>Anzahl_Musiker</u>
	1	Lollapalooza		1	Rise Against	4
	2	Eurovision SC		2	Alligatoah	1
	2	Lollapalooza		3	Von Wegen Lisbeth	5
	3	Eurovision SC		4	Helene Fischer	3
	4	Musikantenstadl		5	Rammstein	6
	4	Rammstein Live		6	Michael Jackson	3
	5	Rammstein Live				
	2	Musikantenstadl				
	5	Lollapalooza				

Veranstaltung	<u>VName</u>	Datum	Uhrzeit	Anzahl_Teilnehmer	Beschreibung
	Lollapalooza	2017-08-09	20:00:00	30000	Festival
	Musikantenstadl	2017-03-12	17:00:00	2500	Tolle Hits!
	Eurovision SC	2016-05-10	20:15:00	9999	Wettbewerb
	Rammstein Live	2016-12-07	16:00:00	100	Tourneestart
	Vernissage Wedding	2017-07-20	08:00:00	42	Hipsterstuff
	Sommerausstellung	2017-07-01	08:30:00	1337	Action Painting

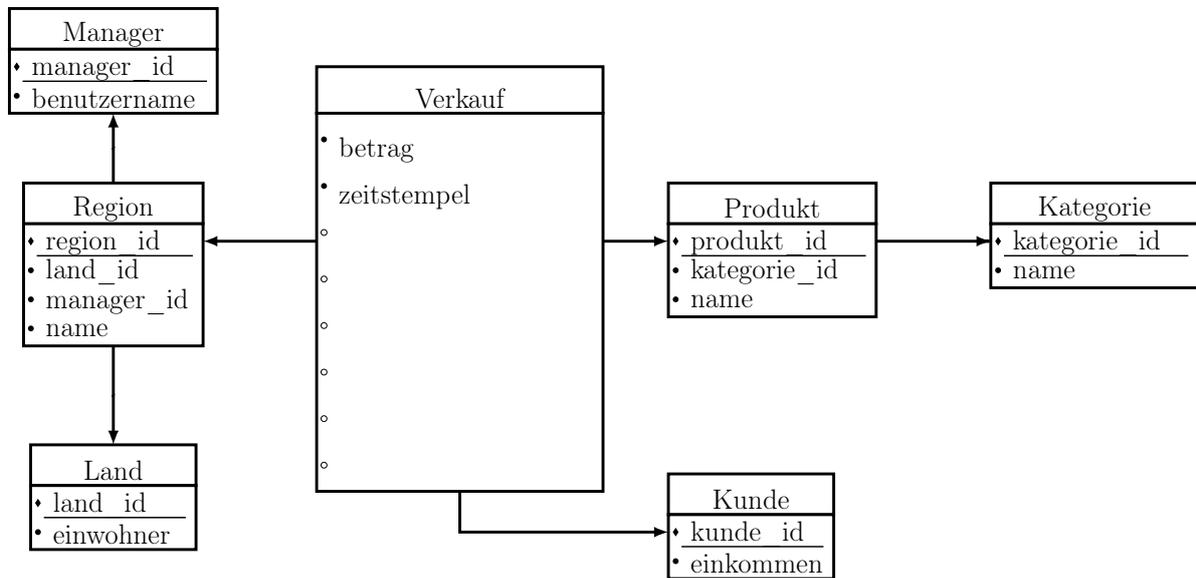
Ausstellung	<u>VName</u>	Epoche	Groesse
	Sommerausstellung	Gegenwart	klein
	Vernissage Wedding	NULL	NULL

Konzert	<u>VName</u>	Technik
	Lollapalooza	Soundmaster XL
	Musikantenstadl	Soundmaster XL
	Eurovision SC	Stereoanlage ZZZ
	Rammstein Live	Dosentelefon Nofeletnesod



Aufgabe 5: Data Warehousing (6 Punkte)

5.1. Gegeben sei das folgende relationale Diagramm eines OLAP-Würfels :



a) Ergänzen Sie das oben gegebene Schema, sodass daraus ein Fullfact-Schema entsteht. Vervollständigen Sie dazu die Attributliste der Faktentabelle und ergänzen Sie etwaige Verbindungen zwischen den einzelnen Relationen. Unterstreichen Sie dabei die Schlüsselattribute aus Verkauf. (1,5)

b) Geben Sie die Anzahl der Relationen an, die bei einer Umwandlung des obigen relationalen Schemas in ein Sternschema entstehen. (0,5)

.....

c) Nennen und begründen Sie einen Nachteil der Verwendung des Fullfact-Schemas kurz in höchstens vier Sätzen. (1)

.....

5.2. Zur Durchführung einer Wahlanalyse sollen die Datensätze aus den verschiedenen Datenbanken in ein zentrales Data Warehouse geladen werden. Dazu wird als Austauschformat XML verwendet. Im Folgenden sehen Sie einen Auszug aus einer dieser XML-Dateien.

```
<?xml version="1.0" encoding="UTF-8"?>
<wahl name="bundestag" jahr="2017">
  <partei name="RotGelbWeiss" id="p1">
    <mitglied alter="25">Meier</mitglied>
    <mitglied alter="64">Soenne</mitglied>
  </partei>
  <partei id="p2" name="OrangeGruen">
    <mitglied alter="74">Hoenne</mitglied>
    <mitglied alter="21">Aref</mitglied>
    <mitglied alter="34">Kondler</mitglied>
  </partei>
  <partei id="p3" name="BlauSchwarzLila">
    <mitglied alter="34">Pahler</mitglied>
    <mitglied alter="52">Lehner</mitglied>
  </partei>
  <wahlbezirk id="wb1-1" stadt="Berlin">
    <kandidat name="Meier" alter="25"/>
    <kandidat name="Schneider" alter="54" hatSitz="Ja"/>
    <partei id="p1">34,2</partei>
    <partei id="p2">48,5</partei>
    <partei id="p3">17,3</partei>
  </wahlbezirk>
  <wahlbezirk id="wb1-2" stadt="Berlin">
    <kandidat name="Aref" alter="21" hatSitz="Ja"/>
    <kandidat name="Pahler" alter="34"/>
    <partei id="p2">75,5</partei>
    <partei id="p3">24,5</partei>
  </wahlbezirk>
  <wahlbezirk id="wb2-1" stadt="Hamm">
    <kandidat name="Kondler" alter="34" hatSitz="Ja"/>
    <kandidat name="Soenne" alter="64"/>
    <partei id="p1">34,2</partei>
    <partei id="p2">65,8</partei>
  </wahlbezirk>
</wahl>
```

Sie sollen nun zur manuellen Validierung der Daten einzelne Informationen aus der oben stehenden exportierten XML-Datei extrahieren. Entwickeln Sie zu diesem Zweck zu den folgenden Aufgabenstellungen XPath-Anfragen:

a) Die Mitglieder aller Parteien. (0,5)

.....

b) ID der Wahlbezirke, in denen mindestens drei Parteien angetreten sind. (1)

.....

c) Die Namen aller Kandidaten, die im XML-Dokument nach dem Kandidaten mit Name "Schneider" stehen und einen Sitz gewonnen haben. (1,5)

.....

Aufgabe 7: Multiple Choice (10 Punkte)

Diese Aufgabe umfasst 10 Multiple-Choice-Fragen. Sie bestehen aus jeweils mehreren Antwortmöglichkeiten, wobei mindestens eine Antwortmöglichkeit richtig und mindestens eine Antwortmöglichkeit falsch ist. Jede Frage, in der alle richtigen Antwortmöglichkeiten angekreuzt und alle falschen Antwortmöglichkeiten nicht angekreuzt sind, wird mit genau einem Punkt bewertet. Sollten nicht alle richtigen Antwortmöglichkeiten angekreuzt worden sein oder wurde mindestens eine falsche Antwortmöglichkeit angekreuzt, wird die Frage mit 0 Punkten bewertet.

7.1. Welche Aussagen gelten in Bezug auf die Abbildung ins relationale Schema? (1)

- Im ER-Stil entsteht eine Relation für jede Generalisierung und Spezialisierung.
- Bei Anwendung des objektorientierten Stils (total, überlappend) entstehen 2^n Relationen.
- Beim NULL-Wert-Stil muss immer ein Typ-Attribut hinzugefügt werden.

7.2. Gegeben sei die folgende Relation $T(A, B, C)$. (1)

Welche der folgenden funktionalen Anhängigkeit(en) (FDs) ist/sind möglicherweise erfüllt?

T	A	B	C
	a_1	b_1	c_1
	a_2	b_1	c_2
	a_3	b_2	c_4
	a_4	b_2	c_1

- $B \rightarrow C$
- $A \rightarrow B$
- $A \rightarrow C$

7.3. Gegeben seien die Relationen R und S sowie der Ausdruck $R \cup S$. Die Relation R und S sind kompatibel, somit ist sichergestellt, dass der Operator in jedem Fall angewandt werden kann. Bitte geben Sie alle korrekten Aussagen über die mögliche Kardinalität des Ausdrucksergebnisses an. (1)

- Die Kardinalität ist nie höher als $|R| \cdot |S|$.
- Die Kardinalität ist höchstens $|R| + |S|$.
- Die Kardinalität ist immer höher als die Kardinalität von $R \cap S$.

7.4. Gegeben seien die folgenden Relationen $R(A, B)$ und $S(B)$. (1)

Welches Ergebnis erhält man bei der Division R/S ?

R	A	B
	x	a
	y	b
	x	b

S	B
	a
	b

R/S	A	B
	x	a
	x	b

R/S	A
	y

R/S	B
	b

R/S	A
	x

7.5. Wofür steht das Akronym SQL im Kontext von Datenbanksystemen? (1)

- Search Quintessential Links
- Structured Query Language
- Super Questions Listed

7.6. Welche(s) Mehrbenutzerproblem(e) tritt/treten in folgendem Schedule auf? (1)

t	T_1	T_2
1	BOT	
2	read(A, a_1)	
3	$a_1 := a_1 - 300$	
4		BOT
5		read(A, a_2)
6		$a_2 := a_2 \cdot 1.19$
7		write(A, a_2)
8		Commit
9	write(A, a_1)	
10	read(B, B_1)	
11	$b_1 := b_1 + 300$	
12	write(B, b_1)	
13	Commit	

- Dirty Read
- Non-repeatable Read
- Lost Update
- Phantom-Problem

7.7. Welche Aussage(n) in Bezug auf das *MapReduce*-Programmiermodell trifft/treffen zu? (1)

- Die Funktion 2. Ordnung ruft eine Funktion 1. Ordnung mit unabhängigen Teildatenmengen auf.
- Die Daten werden von Tripeln aus *primary key*, *key* und *value* gehalten.
- Typen von *Values* können im allgemeinen auch Typen für *Keys* sein.
- MapReduce*-Programme haben für gewöhnlich folgende statische Struktur:
Eingabe → *Map* → *Shuffle* → *Reduce* → *Ausgabe*

7.8. Gegeben sei ein Sternschema mit insgesamt d Tabellen. Wie viele Join-Operationen werden maximal benötigt, um einen OLAP-Würfel aus den Tabellen zu konstruieren? (1)

- $d \cdot d \cdot d = d^3$
- $d - 1$
- $2^d - 1$
- d

7.9. Gegeben sei eine Anzahl Observationen $x_1, \dots, x_n \in \mathbb{R}$ und dazugehörigen Labels $y_1, \dots, y_n \in \mathbb{R}$ wie lautet das Optimierungsproblem für Least-Squares Linear Regression? (1)

- $\operatorname{argmin}_{a,b} \sum_{i=0}^n (a \cdot x_i + b - y_i)^2$
- $\operatorname{argmin}_{a,b} \sum_{i=0}^n |a \cdot x_i + b - y_i|$
- $\operatorname{argmax}_{a,b} \sum_{i=0}^n |a \cdot x_i + b - y_i|$
- $\operatorname{argmin}_{a,b} \sqrt{\sum_{i=0}^n (a \cdot x_i + b - y_i)^2}$

7.10. Welche der folgenden Aussagen über den k -Nearest-Neighbor-Algorithmus sind korrekt? (1)

- Wird der Wert k zu hoch gewählt, kann Underfitting auftreten.
- Wird der Wert k zu hoch gewählt, kann Overfitting auftreten.
- Wird der Wert k zu niedrig gewählt, kann Overfitting auftreten.
- Wird der Wert k zu niedrig gewählt, kann Underfitting auftreten.

