

Name: \_\_\_\_\_ Vorname: \_\_\_\_\_ Matrikel-Nr.: \_\_\_\_\_

# Bergische Universität Wuppertal

## Schumpeter School of Business and Economics

**Bachelor of Science / Bachelor of Arts**

**WiSe 2023/2024**

---

Prüfungsgebiet:	BWiWi 2.8 Wissensbasierte Systeme und Informationstechnologien
Tag der Prüfung:	07.02.2024
Name des Prüfers:	Prof. Dr. Bock
Erlaubte Hilfsmittel:	Taschenrechner (nicht programmierbar) Aufgabenblock A: beigelegte Formelsammlung Aufgabenblock D: Anhang mit Anwendungsbeispiel.

---

Die Lösungen zu den Aufgaben sollen gegliedert und in vollständigen zusammenhängenden Sätzen dargestellt und Rechnungen mit ihren Zwischenschritten nachvollziehbar sein. Die Darstellungsform und die Systematik der Gedankenführung gehen in die Bewertung ebenfalls ein. In Klammern ist für jede Aufgabe die Anzahl der maximal möglichen Punkte angegeben, die bei einer richtigen und vollständigen Bearbeitung erreicht werden können. Sie entspricht in etwa dem erwarteten Zeitbedarf in Minuten.

### **Bearbeiten Sie Aufgabenblock A und einen weiteren Aufgabenblock!**

Insgesamt können **90 Punkte** erreicht werden. Für eine erfolgreiche Bearbeitung müssen wenigstens **45 Punkte** erworben werden. Bei Bearbeitung von mehr als zwei Aufgabenblöcken wird die Bearbeitung des Aufgabenblocks A und des ersten weiteren bearbeiteten Blocks gewertet. Sie haben **90 Minuten** Bearbeitungszeit zur Verfügung.

Die Klausur besteht mit diesem Deckblatt aus insgesamt 12 (zwölf) Seiten.

Unterschrift: \_\_\_\_\_

# Aufgabenblock A:

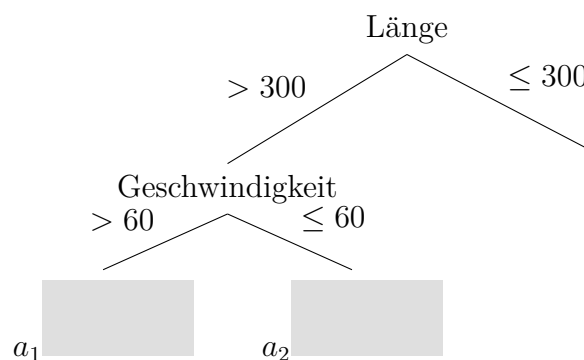
## Knowledge-based systems

### Aufgabe A.1: Entscheidungsbäume

[17 Punkte]

Zur Einschätzung der Beliebtheit von Achterbahnen werden diese in die Beliebtheitsgruppen *hoch* (H), *mäßig* (M) und *niedrig* (N) eingeteilt. In dem gegebenen Datensatz werden dazu die Attribute *Länge*, *Höchstgeschwindigkeit* und *Looping* betrachtet, wobei Looping ein binäres Attribut ist, das angibt, ob die Achterbahn über einen oder mehrere Loopings verfügt (1) oder nicht (0).

i	Länge	Geschw.	Looping	C
1	768	108	1	H
2	1320	117	0	H
3	200	7	0	N
4	1280	50	0	N
5	840	82	1	M
6	1170	97	1	H
7	250	12	0	N
8	1099	80	0	M
9	210	62	0	M



- a) Notieren Sie für den Blattknoten  $a_2$  die resultierende Klassifikation in dem grauen Kästchen. (2 Punkte)
- b) Betrachten Sie den Knoten  $a_1$  des obigen Entscheidungsbaums. Bestimmen Sie mit Hilfe des *Gain* Kriteriums unter den im Folgenden aufgeführten Tests denjenigen, der für die Verzweigung  $a_1$  am geeignetsten ist. Testen Sie einerseits das Attribut Looping und verwenden Sie andererseits für die Attribute Länge und Geschwindigkeit jeweils folgende Tests:
- Länge  $\leq 1000$  / Länge  $> 1000$
  - Geschwindigkeit  $\leq 100$  / Geschwindigkeit  $> 100$
- Erweitern Sie anschließend den obigen Baum. Wird nun der Datensatz in diesem Teilbaum vollständig und fehlerfrei klassifiziert? (8 Punkte)
- c) In welchem Fall empfiehlt sich die Verwendung des *Gain Ratio* Kriteriums anstelle des *Gain* Kriteriums? (2 Punkte)
- d) Angenommen, Sie betrachten einen allgemeinen Datensatz  $M$ . Für die Generierung eines  $M$  klassifizierenden Entscheidungsbaums verwenden Sie das Gain-Kriterium und erhalten für ein Attribut  $A_i$  den Wert  $gain(M, A_i) = info(M)$ . Wie lässt sich dieses Ergebnis interpretieren? Nehmen Sie begründet Stellung. (5 Punkte)

**Aufgabe A.2: Boosting mit AdaBoost****[14 Punkte]**

- a) In der Veranstaltung haben Sie die Varianten des AdaBoost-Verfahrens *AdaBoost.M1* und *AdaBoost.M2* kennengelernt. Beschreiben Sie kurz, welche Nachteile von *AdaBoost.M1* durch *AdaBoost.M2* ausgeglichen werden. (4 Punkte)
- b) Im Folgenden wurde der Algorithmus *AdaBoost.M2* auf einen nicht näher spezifizierten Datensatz angewandt, der die Daten in die zwei Klassen **0** und **1** unterteilt. Unten dargestellt sind die Ergebnisse nach einer Iteration des *AdaBoost.M2*-Verfahrens. Angegeben sind die Rückgabewerte des Schätzers  $h_1(i, j)$  sowie die Gewichte  $D_1(i, j)$  für  $j \in \{0, 1\}$  und alle  $i = 1, \dots, 5$ .

$i$	$h_1(i, 0)$	$h_1(i, 1)$	$D_1(i, 0)$	$D_1(i, 1)$
1	0.8	0.3	-	0.2
2	0.5	0.6	0.2	-
3	0.2	0.7	0.2	-
4	0.9	0.3	-	0.2
5	0.7	0.4	0.2	-

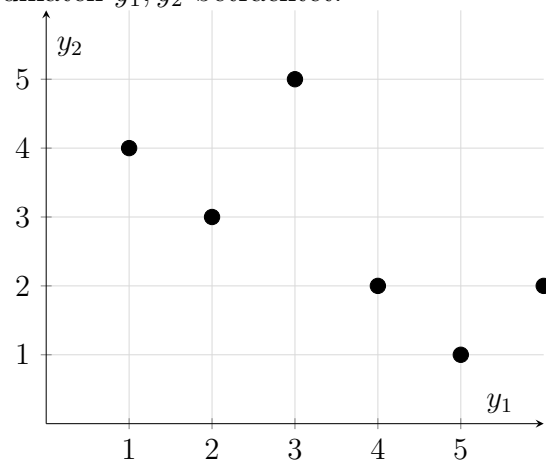
1. Bestimmen Sie  $\beta_1$  basierend auf den gegebenen Daten. (5 Punkte)
2. Bestimmen Sie die Klassifikation des Gesamtschätzers (the final prediction of all generated predictors) für  $i = 1, \dots, 5$  nach der ersten Iteration des Algorithmus. Geben Sie ebenfalls an, ob die Fälle jeweils korrekt klassifiziert werden. (5 Punkte)

### Aufgabe A.3: Clustering und nächste Nachbarn

[14 Punkte]

Im folgenden werden die Datenpunkte  $x_i$  mit Koordinaten  $y_1, y_2$  betrachtet.

$i$	$y_1$	$y_2$
1	1	4
2	2	3
3	4	2
4	6	2
5	3	5
6	5	1



Hinweis: Verwenden Sie in allen Teilaufgaben das *Manhattan-Distanzmaß* (*rectangular distance measure*).

- a) Bestimmen Sie die Cluster  $C_1$  und  $C_2$  für die Zentren  $c_1 = (1, 5)$  und  $c_2 = (6, 0)$  wie im  $k$ -means Verfahren.

*Hinweis: Für die Ermittlung der Distanzen können Sie untenstehende Tabelle verwenden. Eine Rechnung ist nicht notwendig.*

Manhattan-Distanzen						
$c_i$	$d(c_i, x_1)$	$d(c_i, x_2)$	$d(c_i, x_3)$	$d(c_i, x_4)$	$d(c_i, x_5)$	$d(c_i, x_6)$
$c_1$						
$c_2$						

(3 Punkte)

- b) Bestimmen Sie die im Rahmen des  $k$ -means Verfahren aus den Clustern  $C_1$  und  $C_2$  resultierenden neuen Zentren  $c'_1$  und  $c'_2$ . (2 Punkte)
- c) Betrachten Sie die folgenden Klassen für die Datenpunkte  $x_i$  des gegebenen Datensatzes:

$i$	1	2	3	4	5	6
$C(i)$	4	2	8	13	6	14

Klassifizieren Sie den Punkt  $x_7 = (4, 4)$  mit Hilfe einer geeigneten 3-nearest-neighbor Variante. (4 Punkte)

- d) Skizzieren Sie ein Beispiel, für das DBSCAN und  $k$ -means für  $k = 2$  unterschiedliche Cluster bilden (d.h. es sollen bei Ihrem Beispiel verschiedene Cluster und/oder eine unterschiedliche Anzahl an Clustern entstehen). Zeigen Sie anhand Ihrer Skizze, welche Cluster das jeweilige Verfahren bilden würde und erklären Sie die Unterschiede. (5 Punkte)

## Formelsammlung zum Aufgabenblock A

$$\text{info}(M) = - \sum_{i=1}^k \frac{\text{freq}(C_i, M)}{\text{freq}(M)} \cdot \log_2 \left( \frac{\text{freq}(C_i, M)}{\text{freq}(M)} \right)$$

$$\text{info}(M, A_j) = \sum_{i=1}^l \frac{\text{freq}(T_i)}{\text{freq}(M)} \cdot \text{info}(T_i)$$

$$\text{prob}(C_i, M) = \frac{\text{freq}(C_i, M)}{\text{freq}(M)} = \frac{\text{freq}(C_i, M)}{|M|}$$

$$G(b_l) = \frac{\#(b_l, l)}{m} \cdot \left( 1 - \sum_{i=1}^k N(i, b_l, l)^2 \right) + \frac{\#(b_l, r)}{m} \cdot \left( 1 - \sum_{i=1}^k N(i, b_l, r)^2 \right)$$

$$P(X|Y) = \frac{P(Y|X) \cdot P(X)}{P(Y)}$$

$$d^\alpha(\mathbb{X}_1, \mathbb{X}_2) = \left( \sum_{j=1}^d |x_{1,j} - x_{2,j}|^\alpha \right)^{1/\alpha}$$

$$h_t = \text{argmin} \{ \epsilon_j \mid \epsilon_j = \sum_{i=1}^m D_t(i) [y_i \neq h_j(x_i)] \wedge h_j \in \mathcal{X} \}$$

$$\epsilon_t = \frac{1}{2} \sum_{(i,y) \in B} D_t(i, y) \cdot (1 - h_t(x_i, y_i) + h_t(x_i, y))$$

$$w_{i,j} = \frac{1}{\sum_C^{k=1} \left( \frac{\|x_j - \mu_i\|}{\|x_j - \mu_k\|} \right)^{\frac{2}{m-1}}} = \frac{1}{\|x_j - \mu_i\|^{\frac{2}{m-1}} \cdot \sum_C^{k=1} \|x_j - \mu_k\|^{\frac{-2}{m-1}}}$$

$$\mu_i = \sum_{j=1}^n \frac{(w_{i,j})^m}{\sum_{j=1}^n (w_{i,j})^m} \cdot x_j = \frac{1}{\sum_{j=1}^n (w_{i,j})^m} \cdot \sum_{j=1}^n (w_{i,j})^m \cdot x_j$$

$$c_i = \frac{1}{|C_i|} \cdot \sum_{x \in C_i} x$$

$$p(x') = \frac{D(x')^2}{\sum_{x \in X} D(x)^2}$$

$$h_{fin}(x) = \text{argmax} \left\{ \sum_{t=1}^T \log \left( \frac{1}{\beta_t} \right) \cdot h_t(x, y) \mid y \in Y \right\}$$

# **Aufgabenblock B:**

## **Computer Hardware und Systembetrieb**

### **(Grundlagen der Rechnerarchitektur und Informationsverarbeitung)**

#### **Aufgabe B.1: Superskalare CPUs** **[5 Punkte]**

- a) Was ist eine superskalare CPU? (3 Punkte)
- b) Mit welchen Architekturmerkmalen erreicht man Superskalarität? (2 Punkte)

#### **Aufgabe B.2: Instruction Reordering** **[10 Punkte]**

- a) In welchen Fällen ist Instruction Reordering sinnvoll? (5 Punkte)
- b) Welche Voraussetzungen an die Prozessorhardware müssen dabei erfüllt sein? (3 Punkte)
- c) Was ist zu beachten, um keine falschen Resultate zu erhalten? (2 Punkte)

#### **Aufgabe B.3: Speculative Execution** **[10 Punkte]**

- a) Was versteht man unter Speculative Execution? (3 Punkte)
- b) Welche Aktionen sind erforderlich, wenn Befehle aufgrund falscher Sprungvorhersagen spekulativ fälschlich ausgeführt wurden? (3 Punkte)
- c) Welche Arten von Operationen sind dabei besonders problematisch? (4 Punkte)

#### **Aufgabe B.4: SMT vs. Multi-Core** **[10 Punkte]**

Erläutern Sie die Hauptunterschiede zwischen einem Prozessor mit 2 Kernen und einem Prozessor mit einem Kern und 2-Weg SMT.

#### **Aufgabe B.5: Copy-on-Write** **[10 Punkte]**

- a) Erläutern Sie kurz die Idee hinter Copy-on-Write. (6 Punkte)
- b) Wo wird dieses Konzept typischerweise eingesetzt? (4 Punkte)

# Aufgabenblock C:

## Kommunikationssysteme (Internet-Technologien)

### Aufgabe C.1: Internet

[15 Punkte]

- a) Was versteht man unter dem Internet aus physischer Sicht bzw. aus logischer Sicht?  
(4 Punkte)
- b) Skizzieren das Internet-Referenzmodell und erläutern Sie die Aufgaben und Eigenschaften der einzelnen Schichten.  
(4 Punkte)
- c) Wie ist eine Internet-Adresse (IPv4) aufgebaut?  
Was ist eine Netzwerkmaske und wofür wird sie eingesetzt?  
Was ist unter der Netzwerk-Basis-Adresse zu verstehen?  
Was ist eine Broadcast-Adresse und wofür wird sie eingesetzt?  
(4 Punkte)
- d) Nennen Sie die drei Hauptziele der Sicherheit im Internet und erläutern Sie diese in jeweils einem Satz.  
(3 Punkte)

### Aufgabe C.2: E-Mail

[15 Punkte]

- a) Skizzieren und beschreiben Sie das Speichervermittlungsverfahren für die Übermittlung einer E-Mail vom E-Mail-Client (Mail User Agent) des Absenders zum E-Mail-Client des Empfängers.  
(7 Punkte)
- b) Warum garantiert die Verwendung von SSL/TLS für die Verbindung der Mail-Clients mit Posteingangs- und Postausgangsservern keine ausreichende Sicherheit bei der Übertragung einer E-Mail vom Absender zum Empfänger?  
Wie lässt es sich dennoch erreichen, dass eine E-Mail sicher und vertraulich ist?  
(4 Punkte)
- c) Skizzieren Sie auf Anwendungsebene (Layer 4) die Nutzung eines Web-Mail-Dienstes per Webbrowser, wobei der Web-Mailer sowohl für den Abruf empfangener E-Mail von einem Posteingangsserver als auch für das Versenden von E-Mail über einen Postausgangsserver genutzt wird. Geben Sie in der Skizze die jeweils genutzten Anwendungsprotokolle zwischen den beteiligten Anwendungen und Diensten an.  
(4 Punkte)

### Aufgabe C.3: World Wide Web

[15 Punkte]

- a) Erläutern Sie kurz die Funktion der drei Kernstandards HTTP, HTML und URL sowie der Standarderweiterungen HTTPS, CSS und JavaScript der WWW-Architektur.  
(6 Punkte)
- b) Erläutern Sie kurz drei typische Anwendungen im WWW, die mit DHTML und nicht mit HTML realisiert werden können.  
(6 Punkte)

c) Welche Aufgaben erfüllen die drei Standards SOAP, WSDL und UDDI beim Einsatz von Web-Services? (3 Punkte)



## Aufgabenblock D:

### Datenorganisation (Datenbankmanagementsysteme)

Bearbeiten Sie 3 der folgenden 4 Aufgaben! (insgesamt 45 Punkte)

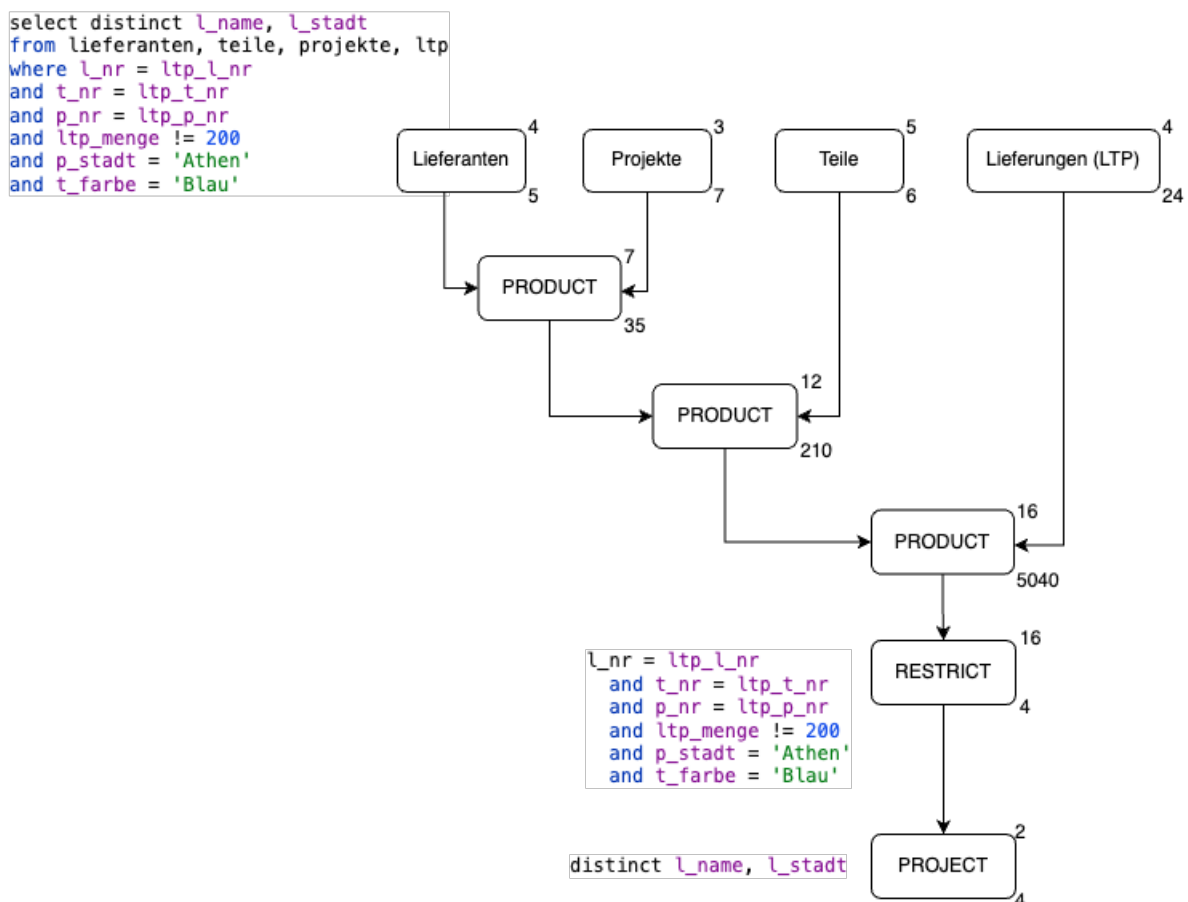
Sollten Sie alle vier Aufgaben bearbeiten, werden nur Ihre Lösungen zu den ersten drei Aufgaben des Aufgabenblattes bewertet.

Jede der Aufgaben kann entweder in deutsch oder englisch beantwortet werden. Im Anhang finden Sie das in der Vorlesung eingeführte Anwendungsbeispiel.

#### Aufgabe D.1: Anfrageverarbeitung

[15 Punkte]

- a) In der folgenden Abbildung sehen Sie eine *SQL SELECT*-Anweisung sowie den vom Datenbanksystem generierten initialen logischen Anfrageplan (iLAP). Erstellen Sie einen verbesserten (präferierten) logischen Anfrageplan (pLAP). (12 Punkte)
- b) Benennen Sie die von Ihnen angewendeten Optimierungsmethoden und/oder Steuerungsheuristiken, die Sie für die Erstellung des verbesserten LAP verwendet haben. (3 Punkte)



**Aufgabe D.2: Aktuelle Entwicklung****[15 Punkte]**

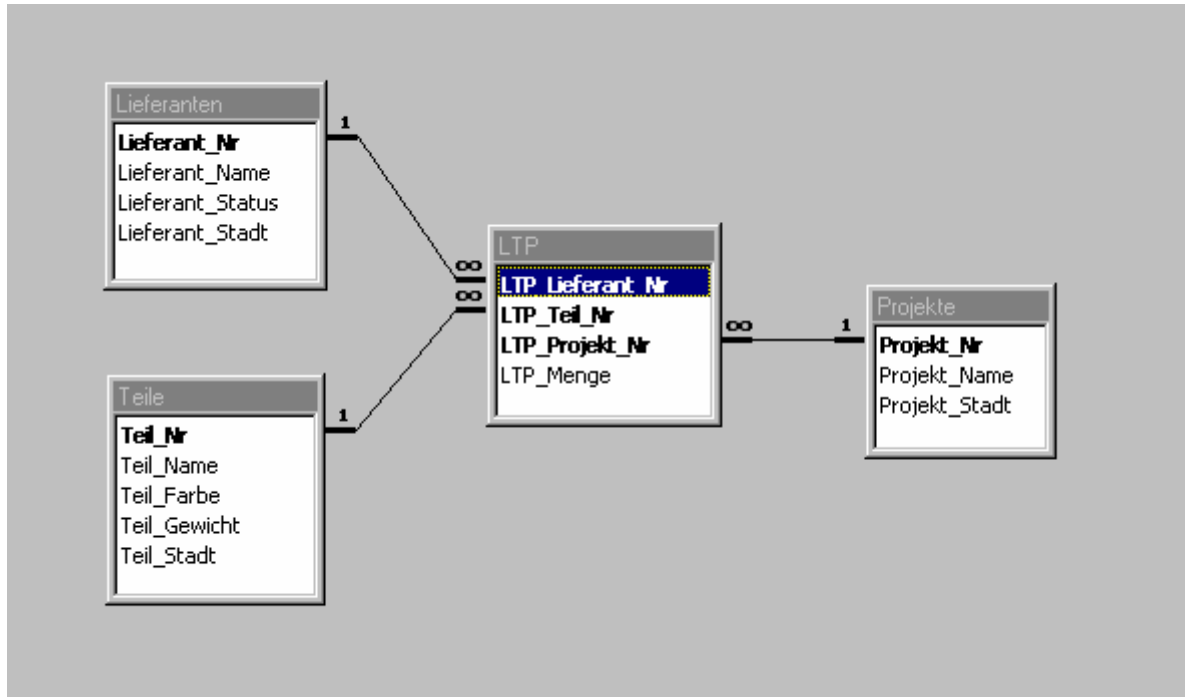
- a) Grenzen Sie NoSQL Datenbanksysteme von Relationalen Datenbanksystemen (RDBS) ab. (5 Punkte)
- b) Erläutern Sie den Ansatz von *Document Databases* und benennen Sie typische Anwendungsgebiete. (5 Punkte)
- c) Erläutern Sie den Ansatz von *Graph Databases* und benennen Sie typische Anwendungsgebiete. (5 Punkte)

**Aufgabe D.3: Interne Ebene****[15 Punkte]**

- a) Erläutern Sie den Begriff eines Index. Wofür wird ein Index in einem Datenbanksystem verwendet? (5 Punkte)
- b) Erläutern Sie den Begriff des Multi-Level Index. Woraus setzt sich ein Multi-level Index zusammen? Geben Sie ein Beispiel für einen Multi-level Index. (10 Punkte)

**Aufgabe D.4: Relationenmodell****[15 Punkte]**

- a) Erläutern Sie die folgenden Operationen der Relationalen Algebra: **UNION**, **MINUS** und **INTERSECT**. Verwenden Sie hierzu jeweils ein selbstgewähltes Beispiel. (12 Punkte)
- b) Für welche der vorgenannten Operationen gilt das Kommutativgesetz? (1 Punkte)
- c) Welche Voraussetzung muss für alle 3 Operationen jeweils gegeben sein? (2 Punkte)



### Lieferanten

Lieferant_Nr	Lieferant_Name	Lieferant_Status	Lieferant_Stadt
L1	Smith	20	London
L2	Jones	10	Paris
L3	Blake	30	Paris
L4	Clark	20	London
L5	Adams	30	Athen

### Teile

Teil_Nr	Teil_Name	Teil_Farbe	Teil_Gewicht	Teil_Stadt
T1	Mutter	Rot	12	London
T2	Bolzen	Grün	17	Paris
T3	Schraube	Blau	17	Rom
T4	Schraube	Rot	14	London
T5	Nocken	Blau	12	Paris
T6	Zahnrad	Rot	19	London

### Projekte

Projekt_Nr	Projekt_Name	Projekt_Stadt
P1	Sortierer	Paris
P2	Bildschirm	Rom
P3	OCR	Athen
P4	Konsole	Athen
P5	RAID	London
P6	EDS	Oslo
P7	Bandlaufwerk	London

### LTP - Lieferungen

LTP_Lieferant_Nr	LTP_Teil_Nr	LTP_Projekt_Nr	LTP_Menge
L1	T1	P1	200
L1	T1	P4	700
L2	T3	P1	400
L2	T3	P2	200
L2	T3	P3	200
L2	T3	P4	500
L2	T3	P5	600
L2	T3	P6	400
L2	T3	P7	800
L2	T5	P2	100
L3	T3	P1	200
L3	T4	P2	500
L4	T6	P3	300
L4	T6	P7	300
L5	T1	P4	100
L5	T2	P2	200
L5	T2	P4	100
L5	T3	P4	200
L5	T4	P4	800
L5	T5	P4	400
L5	T5	P5	500
L5	T5	P7	100
L5	T6	P2	200
L5	T6	P4	500

---