



Hinweise zur Personalisierung:

- Ihre Prüfung wird bei der Anwesenheitskontrolle durch Aufkleben eines Codes personalisiert.
- Dieser enthält lediglich eine fortlaufende Nummer, welche auch auf der Anwesenheitsliste neben dem Unterschriftenfeld vermerkt ist.
- Diese wird als Pseudonym verwendet, um eine eindeutige Zuordnung Ihrer Prüfung zu ermöglichen.

Einsatz und Realisierung von Datenbanksystemen

Klausur: IN2031 / Probeklausur
Prüfer: Prof. Alfons Kemper, Ph. D.

Datum: Donnerstag, 20. Juli 2023
Uhrzeit: 11:00 – 12:30

	A 1	A 2	A 3	A 4	A 5	A 6	A 7	A 8	A 9	A 10	A 11	A 12
I												
II												

Bearbeitungshinweise

- There is an english translation of all exam questions in the back of the exam.
- Diese Klausur umfasst **16 Seiten** mit insgesamt **12 Aufgaben**.
Bitte kontrollieren Sie jetzt, dass Sie eine vollständige Angabe erhalten haben.
- Die Gesamtpunktzahl in dieser Klausur beträgt 90 Punkte.
- Das Heraustrennen von Seiten aus der Prüfung ist untersagt.
- Als Hilfsmittel sind zugelassen:
 - ein **analoges Wörterbuch** Deutsch ↔ Muttersprache **ohne Anmerkungen**
- Mit * gekennzeichnete Teilaufgaben sind ohne Kenntnis der Ergebnisse vorheriger Teilaufgaben lösbar.
- **Es werden nur solche Ergebnisse gewertet, bei denen der Lösungsweg erkennbar ist.** Auch Textaufgaben sind **grundsätzlich zu begründen**, sofern es in der jeweiligen Teilaufgabe nicht ausdrücklich anders vermerkt ist.
- Schreiben Sie weder mit roter / grüner Farbe noch mit Bleistift.
- Schalten Sie alle mitgeführten elektronischen Geräte vollständig aus, verstauen Sie diese in Ihrer Tasche und verschließen Sie diese.

Hörsaal verlassen von _____ bis _____ / Vorzeitige Abgabe um _____

Aufgabe 1 Allgemeine Fragen (5 Punkte)

Beantworten Sie die folgenden Fragen ausschließlich mit den folgenden Begriffen. Jede Frage benutzt exakt einen Begriff. Ordnen Sie genau jeder Frage genau eine Antwort zu.

Answer the following questions using only the following terms. Use exactly one term per question. There are more terms than questions given.

1. Nennen Sie einen Majority-Consensus Algorithmus. Name one majority consensus algorithm.
2. Nennen Sie ein binäres Datei-Format zur kompakten Speicherung von Daten.
Name one binary file format for compact data storage.
3. Nennen Sie die Anfragesprache die das Resource Description Frameworks (RDF) einsetzt.
Name the query language that the Resource Description Frameworks (RDF) uses.
4. Wofür steht die Abkürzung MVCC? What does MVCC stand for?
5. Nennen Sie einen Algorithmus der Hubs und Authorities berechnet.
State one algorithm that calculates hubs and authorities.

0	<input type="checkbox"/>	<input type="checkbox"/>
1	<input type="checkbox"/>	<input type="checkbox"/>
2	<input type="checkbox"/>	<input type="checkbox"/>
3	<input type="checkbox"/>	<input type="checkbox"/>
4	<input type="checkbox"/>	<input type="checkbox"/>
5	<input type="checkbox"/>	<input type="checkbox"/>

Tragen Sie die Nummer jeder Frage in exakt eine der unten stehenden Boxen ein.

Write the number of the question into the field of exactly one corresponding answer.

<input type="checkbox"/>	CHORD	<input type="checkbox"/>	Multi-Version Concurrency Control
<input type="checkbox"/>	CSV	<input type="checkbox"/>	No-Random-Access Algorithm
<input type="checkbox"/>	Deadlock-Vermeidung Deadlock Avoidance	<input type="checkbox"/>	Parquet
<input type="checkbox"/>	HICS	<input type="checkbox"/>	PostgreSQL
<input type="checkbox"/>	HITS	<input type="checkbox"/>	Parquet
<input type="checkbox"/>	Hub Sort	<input type="checkbox"/>	RAFT
<input type="checkbox"/>	JSON	<input type="checkbox"/>	SPARQL
<input type="checkbox"/>	Minimum Viable Communication Control	<input type="checkbox"/>	Spark SQL
<input type="checkbox"/>	Model-View-Controller Communication	<input type="checkbox"/>	Threshold Algorithm

Aufgabe 2 Recovery (3 Punkte)

0	<input type="checkbox"/>	<input type="checkbox"/>
1	<input type="checkbox"/>	<input type="checkbox"/>

a)* Vor welchen Ereignissen muss das, was das WAL-Prinzip besagt, vollständig abgeschlossen sein? (Stichwort)

Before which events does what the WAL-principle stands for has to be completed? (Keyword)

0	<input type="checkbox"/>	<input type="checkbox"/>
1	<input type="checkbox"/>	<input type="checkbox"/>
2	<input type="checkbox"/>	<input type="checkbox"/>

b)* Nennen Sie die drei Phasen des Wiederanlaufs in chronologischer Reihenfolge. (3 Stichworte)

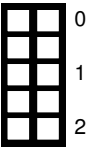
Name the three phases of the restart process in chronological order. (3 keywords)

Aufgabe 3 Historien (10 Punkte)

Gegeben sei die folgende Historie: Given the following History:

$$r_1(y) \rightarrow w_1(y) \rightarrow r_1(x) \rightarrow w_2(y) \rightarrow w_1(y) \rightarrow c_1 \rightarrow c_2 \rightarrow r_3(x) \rightarrow w_3(x) \rightarrow c_3$$

a)* Zeichnen Sie den Serialisierbarkeitsgraphen der Historie. Draw the serializability graph of the history.



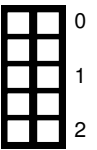
b)* Kreuzen Sie an, ob die Eigenschaften von der Historie erfüllt werden.

Tick the properties that are fulfilled by the history.

- Kann von einem strikten 2PL-Scheduler erzeugt worden sein. Could be generated by a strict 2PL-scheduler.
- Strikt Strict (ST)
- Kann von einem 2PL-Scheduler erzeugt worden sein. Could be generated by a 2PL-scheduler.

c)* Fügen Sie Commits in die folgende Historie H so ein, dass die Historie RC aber nicht ACA erfüllt:
Insert commits into the following history H such that the history fulfils RC but not ACA:

$$H = w_1(x) \rightarrow r_2(x) \rightarrow w_2(y)$$



Aufgabe 4 IT-Sicherheit (6 Punkte)

Gegeben sei die Tabelle PrAnmeldung mit einigen Beispielwerten. Die Tabelle speichert für alle Studenten, zu welchen Prüfungen sie angemeldet sind.

Given the relation PrAnmeldung with some example values. The relation stores for the exams each student has registered for.

<u>MatrNr</u>	Name	Pruefung (Exam)
03642000	Thuy	NetSec
03654321	Alex	Modern DBs
03636363	Anna	Robotics
⋮	⋮	⋮

Es gibt eine Website, in der man sich nach Eingabe der Matrikelnummer alle seine Prüfungsanmeldungen auflisten lassen kann. Leider hat das Entwicklungsteam vergessen, die Benutzereingabe auf SQL-Injections zu prüfen.

There is a website where you can see all exam registrations after entering the matriculation number MatrNr. Unfortunately, the developer team missed to test the user input for SQL injections.

Die im Formular genutzte Anfrage lautet: The user input looks as following:

```
SELECT Pruefung FROM PrAnmeldung WHERE MatrNr={INPUT};
```

Schreiben Sie SQL-Injections für das Feld {INPUT}:

Write SQL injections for the field INPUT

0				
1				
2				
3				
4				

a)* Sie haben vergessen, sich rechtzeitig für ERDB anzumelden. Schreiben Sie eine SQL-Injection, um sich mit Ihrer Matrikelnummer (MatrNr) und Ihrem Namen (Name) zur Prüfung (Pruefung) „ERDB“ anzumelden.

You forgot to register for the ERDB exam on time. Write an SQL-Injection to register yourself with your matriculation number (MatrNr) and your name (Name) for the exam (Pruefung) „ERDB“.

0				
1				
2				

b)* Die Manipulation ist aufgefallen und Sie verwischen Ihre Spuren. Schreiben Sie eine SQL-Injection, um die gesamte Tabelle zu löschen.

The manipulation was noticed and you want to cover your tracks. Write an SQL-Injection to delete the entire table.

Aufgabe 5 Datalog (7 Punkte)

Gegeben seien die Fakten voraussetzen und vorlesungen aus der Universitätsdatenbank (hier nur in Ausschnitten dargestellt):

Given the facts voraussetzen (required) and vorlesungen (lectures) of the university database (here only an excerpt is visualized)

```

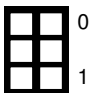
%%voraussetzen(VorgNr, NachfNr)
voraussetzen(5001, 5041).
voraussetzen(5001, 5043).
voraussetzen(5043, 5022).
...
%%Vorlesungen(VorlNr, Titel, SWS, gelesenVon)
vorlesungen(5001, grundzuege, 4, 2137).
vorlesungen(5022, wissenschaftstheorie, 3, 2126).
vorlesungen(5041, ethik, 4, 2125).
vorlesungen(5043, erkenntnistheorie, 3, 2126).
...

```

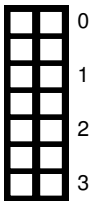
Gegeben sei der folgende Ausdruck in Domänenkalkül: Given the following expression in domain calculus:

$$\{[t] \mid \exists v, s, g([v, t, s, g] \in \text{Vorlesungen} \wedge \exists n([v, n] \in \text{voraussetzen} \wedge \exists s2, g2([n, \text{Ethik}, s2, g2] \in \text{Vorlesungen})))\}$$

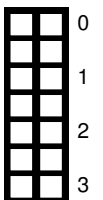
a)* Nennen Sie den Zweck des Ausdrucks. State the purpose of the expression.



b)* Übersetzen Sie den Ausdruck in Datalog. Translate the expression into Datalog.



c)* Gegeben seien die folgenden Datalog-Prädikate: Given the following Datalog predicate:



```

.decl vorgaenger(v: number, n: number)
vorgaenger(V,N) :- voraussetzen(V,N).
vorgaenger(V,N) :- vorgaenger(V,N), voraussetzen(V,N).

.decl nichtVoraussetzen(v: number, nv: number)
nichtVoraussetzen(V,NV) :- vorlesungen(V,_,_,_), vorlesungen(NV,_,_,_), !vorgaenger(V,NV).
.output nichtVoraussetzen

```

Begründen Sie stichpunktartig, ob das Prädikat nichtVoraussetzen stratifiziert ist.

Reason with bullet points whether the predicate nichtVoraussetzen is stratified.

Aufgabe 6 Fragmentierung (8 Punkte)

<u>MatrNr</u>	<u>Name</u>	<u>Note (Grade)</u>	<u>Standort (Location)</u>
10101	Philipp	1,0	München
10102	Magdalena	1,0	Garching
10103	Alex	1,0	Garching
10104	Maxmilian	1,0	München

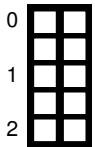
Für eine verteilte Datenbank soll die Tabelle *Klausur* mit Schlüssel *MatrNr* geeignet fragmentiert werden. Ziel ist, Namen mit Standort der Studenten lokal und die Noten getrennt abzuspeichern.

For a distributed database the table *Klausur* with the key *MatrNr* should get fragmented in a suitable way. The objective is to store the name and the location of the students separately from the grades.

Fragmentieren Sie die Relation *vertikal*. Fragment the relation *vertically*.



a)* Geben Sie das Schema für die zwei resultierenden Relationen *KlausurV₁* und *KlausurV₂* an.
Specify the schema of the two resulting relations *KlausurV1* and *KlausurV2*.

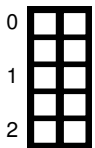


b)* Geben Sie in SQL-92 die zwei resultierenden Relationen *KlausurV1* und *KlausurV2* mittels *with* an.
Specify in SQL-92 the two resulting relations *KlausurV1* and *KlausurV2* using *with* clauses.

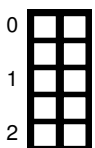
Die geeignetere der beiden resultierenden Relationen soll *horizontal* fragmentiert werden.
The more appropriate of both relations should be fragmented *horizontally*.



c)* Geben Sie das Prädikat der Selektion an, mit dem fragmentiert wird.
State the predicate of the selection used for fragmentation.



d)* Geben Sie in SQL-92 die resultierenden Relationen *KlausurH1* und *KlausurH2* als mittels *with* an.
Specify in SQL-92 the two resulting relations *KlausurH1* and *KlausurH2* using *with* clauses.



e) Schreiben Sie eine SQL-Abfrage, die die Ursprungsrelation aus den Teilrelationen zusammensetzt.
Write a SQL query that builds the original relation from the partial relations.

Aufgabe 7 Window-Functions (10 Punkte)

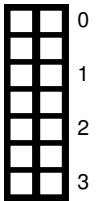
Betrachten Sie die folgende Tabelle Waren mit verkauften Produkten in einem Supermarkt. Die Spalte verkauft besagt, wieviele Einheiten des jeweiligen Produktes verkauft worden sind.

Given is the table Waren (goods) with sold products in a supermarket. The column sold tells you how many units of the respective products have been sold.

Name	Preis (Price)	Kategorie (Category)	Verkauft (Sold)
Brot	1.00	Backwaren	8128
Butter	0.80	Kühlwaren	496
Grill	60.00	Haushalt	6
Steak	8.00	Kühlwaren	28
...

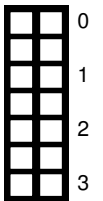
a)* Ermitteln Sie in SQL mittels Fensterfunktionen (Windowfunctions) den prozentualen Umsatzanteil jedes Produktes innerhalb seiner Kategorie.

Determine in SQL with window functions the percentage share of all products in their respective category.



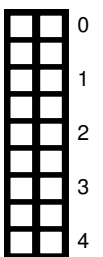
b)* Ermitteln Sie in SQL mittels Fensterfunktionen (Windowfunctions) für jedes Produkt das Mittel der Verkaufszahlen aus den 5 besser verkauften (höhere Verkaufszahlen) Produkten geordnet nach Verkaufszahlen.

Determine in SQL with window functions for each product the mean of the sale figures of the five products that have better sales figures than the product itself, ordered by sales figures.



c)* Ermitteln Sie in SQL mittels Fensterfunktionen (Windowfunctions) die drei Produkte mit dem meisten Umsatz pro Kategorie.

Determine in SQL with window functions for each category the three products with the most sales.



Aufgabe 8 Skyline (6 Punkte)

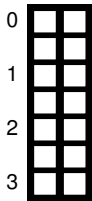
Gegeben sei die Relation Zuege:

Given the relation Zuege (Trains):

Baureihe (Model)	Gewicht (Weight)	Vmax
401	795	280
402	455	280
403	459	330
406	488	330
407	495	320
411	402	230
412	670	250
415	311	230

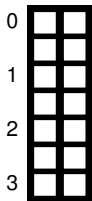
Wir betrachten die Skyline über das Minimum des Attributs *Gewicht* und das Maximum über das Attribut *Vmax*.

We look at the skyline over the minimum of the attribute *weight* (*Gewicht*) and the maximum of the attribute *Vmax*.



a)* Geben Sie die Anfrage, die die oben genannte Skyline in SQL-92 berechnet, an (d.h. ohne Skyline-Operator).

Specify the request that computes the described skyline from above in SQL-92 (without using the skyline operator).



b)* Geben Sie alle Tupel an, die in der Skyline enthalten sind. Es reicht, wenn Sie die zutreffenden Baureihen angeben.

Check all tuples that are part of the skyline. It is sufficient to name the model series (*Baureihe*) of the tuples.

Aufgabe 9 Hauptspeicher-Datenbanken (10 Punkte)

Sie sollen für die Alexander-Maximilians-Universität (AMU) ein Hauptspeicherdatenbanksystem optimieren. In dem System sind die Daten aller Studenten gespeichert.

You have to optimize the main-memory database system of the Alexander-Maximilians-Universität (AMU). The system stores all student information.

Relationen: Relations:

Studenten (Students): MatrNr (8 Byte), Name (48 Byte), Studiengang (4 Byte), Semester (4 Byte)

MatrNr ist der Primärschlüssel der indiziert ist. Eine Cacheline hat 64 Byte.

MatrNr is a primary key and is indexed. One cache line is 64 Byte.

Schätzen Sie für jede der untenstehenden Anfragen einzeln, ob ein Row- oder Column-Store besser geeignet ist. Wenn beides gleich gut geeignet ist, kreuzen Sie beides an.

Tick the boxes for each of the queries below, whether a row- or column-store is better suited. If both are suited equally well, tick both.

a)* select * from Studenten;

- Row-Store
- Column-Store

b)* select Semester, count(*) from Studenten group by Semester;

- Row-Store
- Column-Store

c)* select Name, Studiengang, Semester from Studenten where MatrNr = 42;

- Row-Store
- Column-Store

d)* select Studiengang from Studenten where MatrNr = 42;

- Row-Store
- Column-Store

e)* select * from Studenten where Semester < 5;

Semester < 5 ist nicht selektiv, d.h. es qualifizieren sich sehr viele Tupel.

Semester < 5 ist not selective (very many tuples qualify).

- Row-Store
- Column-Store

f)* select * from Studenten where Semester = 25;

Semester = 25 ist sehr selektiv, d.h. es qualifizieren sich sehr wenige Tupel.

Semester = 25 ist very selective (only very few tuples qualify).

- Column-Store
- Row-Store

g)* insert into studenten values(4242, Max Meyer, Info, 7);

- Row-Store
- Column-Store

Aufgabe 10 XQuery (11 Punkte)

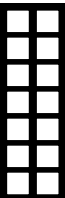
Als Grundlage für die Aufgabe wird eine leichte Abwandlung des Uni Schemas genutzt, von dem hier ein Ausschnitt geben ist. Fakultaeten ist der Wurzelknoten des Dokuments unifak:


This task uses a slightly altered university schema (here only an excerpt is visualized). Fakultaeten (Faculties) is the root node of the document unifak:


```

<Fakultaeten>
  <Fakultaet>
    <FakName>Theologie</FakName>
    <ProfessorIn PersNr="P2134">
      <Name>Augustinus</Name>
      <Rang>C3</Rang>
      <Raum>309</Raum>
      <Vorlesungen>
        <Vorlesung VorlNr="V5022">
          <Titel>Glaube und Wissen</Titel>
          <SWS>2</SWS>
        </Vorlesung>
        ...
      </Vorlesungen>
    </ProfessorIn>
  </Fakultaet>
  <Fakultaet>
    <FakName>Philosophie</FakName>
    <ProfessorIn ID="P2126" PersNr="P2126">
      <Name>Russel</Name>
      <Rang>C4</Rang>
      <Raum>232</Raum>
      <Vorlesungen>
        <Vorlesung ID="V5043" VorlNr="V5043"
          Voraussetzungen="V5001">
          <Titel>Erkenntnistheorie</Titel>
          <SWS>3</SWS>
        </Vorlesung>
        ...
      </Vorlesungen>
    </ProfessorIn>
  </Fakultaet>
</Fakultaeten>

```

- 0  a)* Geben Sie in XPath die Namen der Professoren aus, die eine Vorlesung mit dem Titel Maeeutik halten.
 1 Write an XPath expression to return the name of the Professoren (professors) that give a Vorlesung (lecture)
 2 with the Titel (title) Maeeutik.
 3

- 0  b)* Geben Sie in XPath alle Professoren aus, die mindestens zwei Vorlesungen halten.
 1 Write an XPath expression to return the Professoren (professors) that give at least 2 Vorlesung (lectures).
 2
 3

- 0  c)* Erstellen Sie in XQuery ein Vorlesungsverzeichnis geordnet nach Vorlesungstitel (aufsteigend) wie
 1 nachfolgend gezeigt.
 2 Write an XQuery expression for a Vorlesungsverzeichnis (lecture overview) ordered by Vorlesungstitel (title
 3 of the lecture) ascending as follows:
 4
 5

```

<Vorlesungsverzeichnis>
  <Vorlesung Titel="Erkenntnistheorie"/>
  <Vorlesung Titel="Glaube und Wissen"/>
  ...
</Vorlesungsverzeichnis>

```

Aufgabe 11 TF-IDF (7 Punkte)

Berechnen Sie für die folgenden drei Dokumente die TF-IDF Werte. Dabei sind alle Worte relevant.

Calculate the TF-IDF values for the three following documents. Each of the words is relevant. (*The actual content of the three german sentences is irrelevant.*)

1. ERDB macht echt viel Spaß
2. Die Klausur ist sicher machbar
3. Wir wünschen euch allen viel Erfolg bei der ERDB Klausur

Welches Ranking ergibt sich für die Anfrage: "ERDB Klausur"?

Which Ranking does the query "ERDB Klausur" have?

Hilfswerte (gerundet): Pre-Calculated Values (rounded):

$$\log(3) = 1/2$$

$$\log(2, 5) = 2/5$$

$$\log(2) = 1/3$$

$$\log(1, 5) = 1/6$$

$$\log(1) = 0$$

a)* Berechnen Sie zunächst die TF-IDF Werte der Worte "ERDB" und "Klausur" getrennt voneinander. Geben Sie alle Werte als Bruch an und verwenden Sie die Hilfswerte.

Calculate the TF-IDF values for the words "ERDB" und "Klausur" independently. Write all values as fractions and use the pre-calculated values.

0
1
2
3
4
5

b) Welches Ranking ergibt sich für die gemeinsame Anfrage: "ERDB Klausur"? Berechne Sie die Werte auf 3 Nachkommastellen genau.


Which Ranking does the query "ERDB Klausur" have? Calculate the result with three significant digits and use the (rounded) precalculated values from the task if needed.

0
1
2





Aufgabe 12 Spark DataFrame API (7 Punkte)


Formulieren Sie die folgende SQL-Query in der Spark DataFrame API mithilfe der Transformationen und Aktionen, die in der Vorlesung vorgestellt wurden. Sie können davon ausgehen, dass alle Relationen des Unischemas bereits als DataFrames mit gleichen Namen in die Spark-Shell geladen wurden.



Translate the following SQL-Query into the Spark DataFrame API using the transformations and actions presented in the lecture. All relations of the Unischema are already imported into the spark shell with the same name and fields as their SQL counterpart.

0  a)* Formulieren Sie die folgende SQL-Query in der Spark DataFrame API und speichern Sie das Ergebnis in einer Variable.

1  Translate the following SQL-Query into the Spark DataFrame API and save the result in a variable.

2  WITH anzahlVorlesungen AS (
3  SELECT persnr, COUNT(vorlnr) AS vlAnzahl
4  FROM professoren, vorlesungen
5  WHERE persnr = gelesenvon
GROUP BY persnr
)

0  b)* Formulieren Sie die folgende SQL-Query in der Spark DataFrame API und geben Sie das Ergebnis aus.
Translate the following SQL-Query into the Spark DataFrame API and print out the result.

1  SELECT persnr
2  FROM anzahlVorlesungen
WHERE vlAnzahl > 2

Zusätzlicher Platz für Lösungen. Markieren Sie deutlich die Zuordnung zur jeweiligen Teilaufgabe. Vergessen Sie nicht, ungültige Lösungen zu streichen.

