



**FernUniversität in Hagen**

**Fakultät für  
Wirtschaftswissenschaft**

Matrikelnr.

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Name

---

Vorname

---

Klausur: Modul 32711 – Business Intelligence

Termin: 13.09.2019, 11:30-13:30 Uhr

Prüfer: Univ.-Prof. Dr. Stefan Smolnik

### **Aufbau und Bewertung der Klausur**

| Aufgabe                       | 1  | 2  | 3  | 4  | Summe |
|-------------------------------|----|----|----|----|-------|
| Maximal erreichbare Punktzahl | 20 | 36 | 34 | 10 | 100   |
| Erreichte Punktzahl           |    |    |    |    |       |

Datum:

Note:



**Bitte lesen Sie die folgenden Hinweise sorgfältig, bevor Sie mit der Bearbeitung der Klausur beginnen!**

**Die Klausur umfasst vier Aufgaben mit der folgenden Punkteverteilung:**

| Aufgabe                       | 1  | 2  | 3  | 4  | Summe |
|-------------------------------|----|----|----|----|-------|
| Maximal erreichbare Punktzahl | 20 | 36 | 34 | 10 | 100   |

Für die Bearbeitung dieser Klausur stehen Ihnen 120 Minuten zur Verfügung. Die Klausur besteht aus einem Aufgabenheft und einem Lösungsbogen. Das Aufgabenheft umfasst sechs Seiten. Bitte prüfen Sie Ihre Klausur auf Vollständigkeit.

### **Hinweise zur Verwendung des Lösungsbogens**

1. Tragen Sie auf dem Deckblatt des Lösungsbogens Ihre Matrikelnummer sowie Ihren Namen und Vornamen ein. Versehen Sie zusätzlich jedes Lösungsblatt mit Ihrer Matrikelnummer.
2. Die Lösungen müssen in den vorgesehenen Raum auf dem Lösungsbogen eingetragen werden. Lösungen außerhalb des vorgesehenen Raumes werden nicht in die Bewertung einbezogen.
3. Markieren Sie bei Multiple-Choice-Aufgaben die von Ihnen gewählten Alternativen durch ein Kreuz (X) an der dafür vorgesehenen Stelle im Lösungsbogen.
4. Sorgen Sie für eindeutige Eintragungen im Lösungsbogen.
5. Notizen können auf den Rückseiten der Aufgabenblätter gemacht werden. Diese Anmerkungen werden nicht in die Bewertung einbezogen.
6. Sie dürfen das Aufgabenheft vom Lösungsbogen trennen. Das Aufgabenheft muss nicht abgegeben werden. Es müssen jedoch am Ende der Klausur alle Blätter des Lösungsbogens zusammengeheftet abgegeben werden. Trennen Sie keine einzelnen Blätter ab.

### **Hinweise zur Bewertung der Aufgaben**

1. Jede vollständig richtig gelöste Aufgabe oder Teilaufgabe wird mit der an Ort und Stelle angegebenen Punktzahl bewertet.
2. Für die Aufgabe 1 (Einfach-Wahlaufgaben 1 aus n) gilt: Es darf nur ein Kreuz pro Teilaufgabe gesetzt werden. Richtig gelöste Teilaufgaben werden mit der angegebenen Punktzahl bewertet. Nicht oder falsch beantwortete Teilaufgaben werden mit null Punkten bewertet.
3. Für die Aufgabe 4 (Richtig – Falsch) gilt: Richtig gelöste Teilaufgaben werden mit der anteiligen Punktzahl bewertet. Nicht oder falsch beantwortete Teilaufgaben werden mit null Punkten bewertet.
4. Für die Aufgaben 2 und 3 gilt: Teilweise richtig gelöste Aufgaben oder Teilaufgaben können mit einer entsprechend verminderten Punktzahl bewertet werden.

## **Hinweise zur Verwendung eines Taschenrechners**

Die Verwendung eines Taschenrechners ist dann und nur dann erlaubt, wenn dieser einer der folgenden Modellreihen angehört:

- Casio fx86 oder Casio fx87
- Texas Instruments TI 30 X II
- Sharp EL 531

Die Verwendung anderer Taschenrechnermodelle wird als Täuschungsversuch gewertet und mit der Note „nicht ausreichend“ (5,0) sanktioniert.

Ob ein Taschenrechner einer der Modellreihen angehört, können Studierende selbst überprüfen, indem sie die vom Hersteller auf dem Rechner angebrachte Modellbezeichnung mit den oben angegebenen Bezeichnungen vergleichen: Bei vollständiger Übereinstimmung ist das Modell erlaubt. Ist die auf dem Rechner angebrachte Modellbezeichnung umfangreicher, enthält aber eine der oben angegebenen Bezeichnungen vollständig, ist das Modell ebenfalls erlaubt. In allen anderen Fällen ist das Modell nicht erlaubt. Eventuelle Vorgänger- oder Nachfolgemodelle, die nicht in der oben aufgeführten Liste enthalten sind, sind ebenfalls nicht erlaubt.

**Wir wünschen Ihnen viel Erfolg bei der Lösung der Aufgaben!**

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

**Aufgabe 1****(20 P)**

Überprüfen Sie die folgenden Aussagen auf ihre Richtigkeit. Markieren Sie im Lösungsbogen die zutreffende Aussage. Es ist nur **EINE** Aussage korrekt, d. h. Sie müssen sich für **EINE** Aussage entscheiden und dürfen nur **EIN** Kreuz setzen.

- a) Welcher der nachfolgenden Bereiche ist Bestandteil der Business Model Ontology nach Osterwalder (2004)? (5 P)
- A Service
  - B Angebot
  - C Intelligence
  - D Strategie
- b) Im erweiterten Phasenmodell nach Simon sind die Schritte ... (5 P)
- A ... Steuerung, Feedback und Lernen typisch für die Intelligence-Phase.
  - B ... Analyse, Erkenntnisgewinn und Informationsselektion typisch für die Design-Phase.
  - C ... Synthese, Alternativenermittlung und Bewertung typisch für die Review-Phase.
  - D ... Aktion, Transformation und Überwachung typisch für die Choice-Phase.
- c) Im Zuge der Datenaufbereitung und Datenbereinigung haben Sie verschiedene Fehlerklassen kennengelernt. Welche Aussage bzgl. Fehlerarten ist nicht korrekt? (5 P)
- A Verschmutzung und Redundanz zählen zu den semantischen Fehlern.
  - B Fehlende und unzulässige Werte zählen zu den Coverage-Fehlern.
  - C Unregelmäßigkeiten zählen zu den syntaktischen Fehlern.
  - D Syntaktische Fehler können im Gegensatz zu semantischen Fehlern durch Konsistenzprüfungen automatisch erkannt werden.
- d) Der zeitliche Abstand zwischen dem Eintreten eines Ereignisses und der Reaktion darauf kann als Latenz bezeichnet werden. Es wird dabei zwischen verschiedenen Latenzen unterschieden. Welche der folgenden Aussagen ist diesbezüglich korrekt? (5 P)
- A Die Analyselatenz ist die Zeit, die nach Eintreten eines Ereignisses benötigt wird, um die das Ereignis betreffenden Daten für die weitere Analyse bereitzustellen.
  - B Die Datenlatenz ist die Zeit, die benötigt wird, um die Daten zu analysieren und die Analyseergebnisse den richtigen Personen bereitzustellen.
  - C Die Entscheidungslatenz ist die Zeit, die benötigt wird, um die generierten Informationen zu verarbeiten und auf ihrer Basis eine Handlung zu initiieren.
  - D Alle zuvor getroffenen Aussagen sind korrekt.

**Übertragen Sie Ihre endgültige Lösung auf den Lösungsbogen!**

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

**Aufgabe 2****(36 P)**

Clusteranalyse wird z. B. im Marketing zur Bildung von Kundensegmenten verwendet. Das Ziel einer Clusteranalyse ist die Gruppierung von Instanzen gemäß objektiv nachvollziehbarer Kriterien. Häufig wird für die Bildung der Cluster auf eine Distanzfunktion  $dist$  zurückgegriffen. Die Distanzen können in einer Distanzmatrix  $D$  dargestellt werden, d. h. für zwei Instanzen  $x_i, x_j$  wird  $D_{i,j} = dist(x_i, x_j)$  gesetzt, sofern  $j < i$ , sonst gilt  $D_{i,j} = 0$ . Tabelle 1 zeigt ein Beispiel für eine Distanzmatrix.

| i \ j | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|-------|-------|-------|-------|-------|-------|
| $x_1$ | 0     | 0     | 0     | 0     | 0     |
| $x_2$ | 5     | 0     | 0     | 0     | 0     |
| $x_3$ | 7     | 3     | 0     | 0     | 0     |
| $x_4$ | 11    | 15    | 5     | 0     | 0     |
| $x_5$ | 9     | 12    | 8     | 2     | 0     |

Tabelle 1: Beispielhafte Distanzmatrix

**2.1)** Hierarchisch-agglomerative Clusteringalgorithmen basieren darauf, aus einer Distanzmatrix von Instanzen sukzessive Distanzmatrizen von Clustern zu berechnen. Beschreiben Sie kurz die fünf Schritte, die bei einem hierarchisch-agglomerativen Clustering durchgeführt werden müssen mit je zwei Sätzen. (10P)

**2.2)** Erstellen Sie für die Daten aus der nachfolgenden Tabelle 2 die Distanzmatrix  $D$ . Verwenden Sie hierfür die Manhattan-Distanz, die gegeben ist in der Form:

$$dist(x_i, x_j) = \sum_{q \in Q} |x_{i,q} - x_{j,q}|$$

Führen Sie anschließend ein hierarchisch-agglomeratives Clustering nach dem Single-Linkage-Verfahren durch. (20 P)

| Kunde | Zufriedenheit | Marken-treue | Zahlungs-moral |
|-------|---------------|--------------|----------------|
| A     | 2             | 6            | 1              |
| B     | 1             | 9            | 6              |
| C     | 3             | 2            | 8              |
| D     | 4             | 3            | 7              |
| E     | 5             | 4            | 9              |

Tabelle 2: Kundenmerkmale als Ausgangsdaten für hierarchisches Clustering

**2.3)** Diskutieren Sie mögliche Probleme von Cluster-Verfahren. Beleuchten Sie dabei insbesondere, welchen Einfluss die Struktur der Daten oder das gewählte Distanzmaß auf die Ergebnisse haben kann. (6 P)

**Übertragen Sie Ihre endgültige Lösung auf den Lösungsbogen!**

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

**Aufgabe 3****(34 P)**

Eine Einzelhandelskette betreibt Filialen in den Regionen Nord, Ost, Süd und West. Ihr Produktsortiment teilt sich in die Kategorien Lebensmittel, Drogerieartikel, Getränke und Elektronik. Bei sämtlichen Geschäftsaktivitäten fallen Daten an, die in einem Datenbankmanagementsystem (DBMS) gespeichert werden. Nachfolgend finden Sie einen beispielhaften Auszug aus einer der Tabellen des DBMS sowie einen exemplarischen OLAP-Würfel, der aus dieser Tabelle erstellt wurde:

| ID     | Kategorie    | Produkt       | Datum      | Filiale | Menge | Preis  | ... |
|--------|--------------|---------------|------------|---------|-------|--------|-----|
| ...    |              |               |            |         |       |        |     |
| 501337 | Lebensmittel | Brot          | 25.07.2019 | Ost-3   | 1     | 4,59   |     |
| 501338 | Elektronik   | LED-Fernseher | 25.07.2019 | Süd-1   | 1     | 599,99 |     |
| 501339 | Getränke     | Limonade      | 25.07.2019 | Süd-2   | 3     | 2,22   |     |
| 501340 | Drogerie     | Waschmittel   | 25.07.2019 | West-1  | 2     | 4,87   |     |
| 501341 | Getränke     | Rotwein       | 25.07.2019 | Süd-3   | 4     | 8,99   |     |
| 501342 | Drogerie     | Shampoo       | 25.07.2019 | Nord-7  | 1     | 2,95   |     |
| 501343 | Elektronik   | Fön           | 25.07.2019 | Nord-3  | 1     | 52,99  |     |
| ...    |              |               |            |         |       |        |     |

Tabelle 3 Beispielhafter Auszug aus einer der Tabellen des DBMS

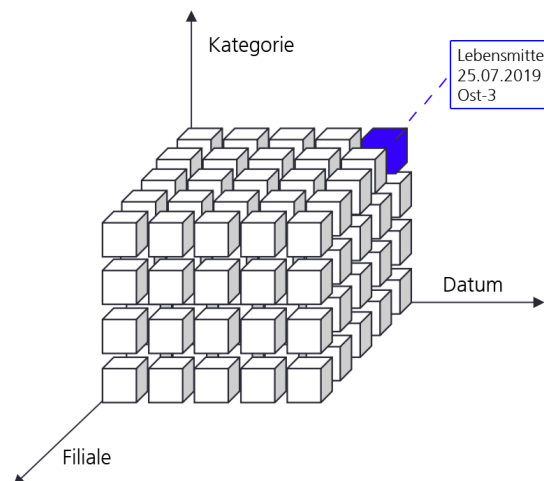


Abbildung 1: Exemplarischer OLAP-Würfel

**3.1)** Wofür stehen die Abkürzungen OLAP, MOLAP, ROLAP und HOLAP? (2,5 P)

**3.2)** Beschreiben Sie kurz die OLAP-Operationen Pivotierung (auch bekannt als Rotation), Roll-up und Drill-down sowie Slice und Dice. (12,5 P)

Ihre Matrikelnummer:

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

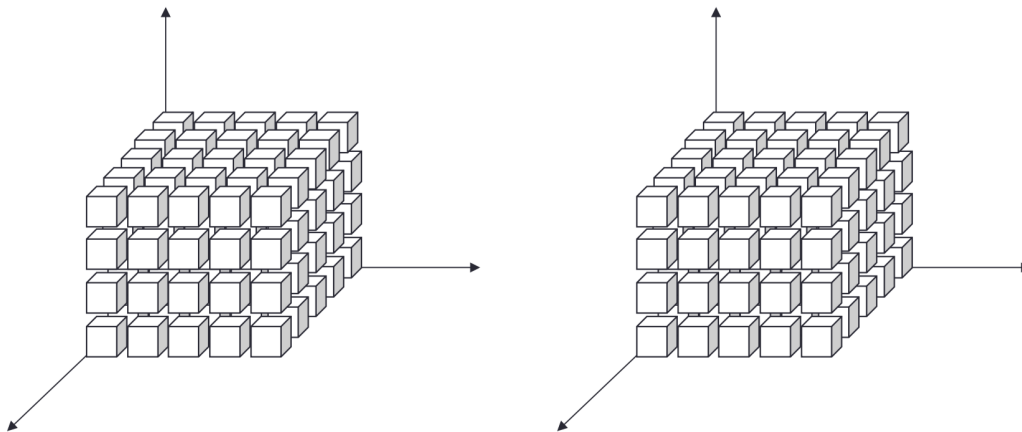
Klausur Business Intelligence

SS 2019

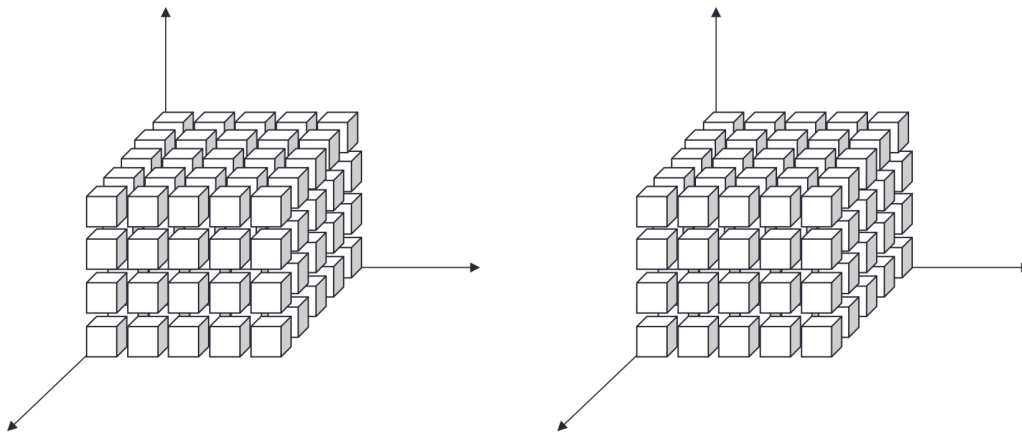
4

**3.3)** Nachfolgend finden Sie mehrere Abbildungen von OLAP-Würfeln. Der linke Teil stellt dabei jeweils eine Ausgangssituation dar, die angegebene Operation ist auf dem linken Würfel anzuwenden und auf dem rechten Würfel einzuzichnen. Wenden Sie die geforderten Operationen an und tragen Sie Ihr Ergebnis in den **Lösungsbogen** ein. Es ist dabei wichtig, dass aus Ihrer Lösung die Veränderung, zu der die Operation führt, nachvollziehbar ist (z. B. durch Änderung von Achsenbeschriftung, Markierungen von Würfelbestandteilen, Pfeile zum Anzeigen einer Drehrichtung). Orientieren Sie sich für Beispieldaten an obiger Tabelle. Hinweis: Übertragen Sie Ihre endgültige Lösung auf den Lösungsbogen! (15 P)

a) Pivotierung:



b) Roll-up:



Ihre Matrikelnummer:

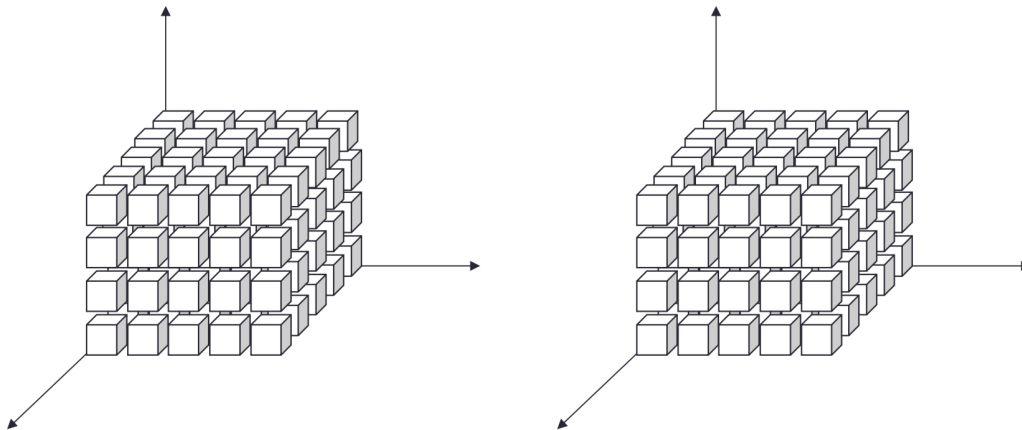
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

5

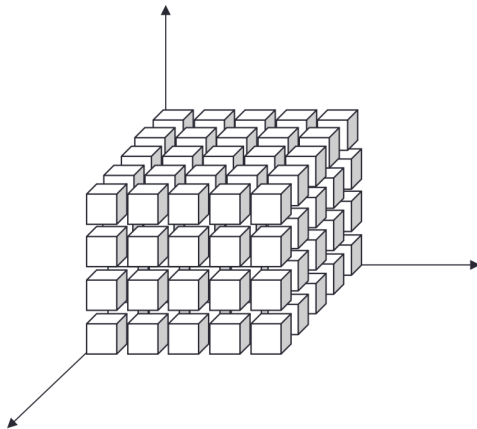
Klausur Business Intelligence

SS 2019

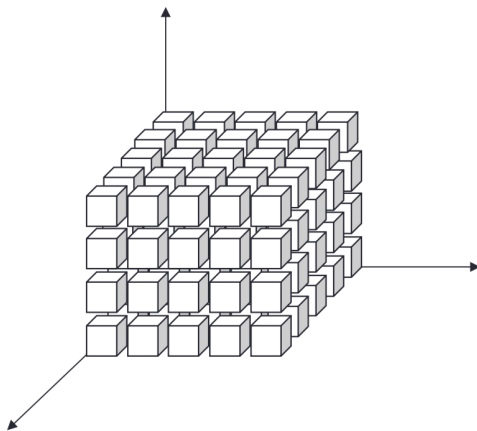
c) Drill-down:



d) Slice: Bitte beschriften Sie zunächst die Achsen und markieren dann die relevanten Daten für die Anfragen: „Alle Daten aus der Region Nord“.



e) Dice: Bitte beschriften Sie zunächst die Achsen und markieren dann die relevanten Daten für die Anfragen: „Alle Daten der Kategorie Drogerie aus den Filialen Ost-1 und West-2 im 3. und 4. Quartal des Jahres 2019“.



**Übertragen Sie Ihre endgültige Lösung auf den Lösungsbogen!**

**3.4)** Erklären Sie den Unterschied zwischen OLAP und OLTP.

(4 P)



Ihre Matrikelnummer:

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Klausur Business Intelligence

SS 2019

6

#### **Aufgabe 4**

**(10 P)**

Überprüfen Sie die folgenden Aussagen auf ihre Richtigkeit. Kennzeichnen Sie im Lösungsbogen uneingeschränkt zutreffende Aussagen mit einem Kreuz bei „Richtig“ und alle anderen Aussagen mit einem Kreuz bei „Falsch“. Bitte beachten Sie auch die Hinweise zur Bewertung.

- a) Werden Fehler im Zuge der Datenbereinigung erkannt, bereinigt und Maßnahmen zur künftigen Vermeidung sowie Überwachung der betreffenden Fehlerquelle eingeführt, wird von einem reaktiven Vorgehen gesprochen.
- b) Bevor eine Normalisierung durchgeführt werden kann, muss sichergestellt werden, dass die Daten normalverteilt sind.
- c) Principle Component Analysis (PCA) wird hauptsächlich zur Dimensionsreduktion eingesetzt.
- d) Bei Clustering-Verfahren kann eine Instanz immer genau einem Cluster zugeordnet werden.
- e) Aus einer guten Trainingsperformance eines Modells mit formal geringer Komplexität folgt nicht automatisch eine gute Testperformance.
- f) Die Abkürzung OLAP steht für Offline Analytical Processing.
- g) Data Mining, Text Mining und Predictive Analytics zählen zu Verfahren aus dem Gebiet Advanced Analytics.
- h) Ein neuronales Netz verbindet mehrere Neuronen miteinander, indem die Ausgänge von Neuronen mit den Eingängen anderer Neuronen verbunden werden.
- i) Der Apriori-Algorithmus kann z. B. im Rahmen einer Warenkorbanalyse dazu eingesetzt werden, zu erklären, warum bestimmte Waren gemeinsam gekauft werden.
- j) Das 0,5-Perzentil entspricht dem arithmetischen Mittel.

**Übertragen Sie Ihre endgültige Lösung auf den Lösungsbogen!**

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

**Aufgabe 1** (ankreuzen)

|           | <b>A</b> | <b>B</b> | <b>C</b> | <b>D</b> |
|-----------|----------|----------|----------|----------|
| <b>a)</b> |          |          |          |          |
| <b>b)</b> |          |          |          |          |
| <b>c)</b> |          |          |          |          |
| <b>d)</b> |          |          |          |          |

**Lösungsbogen für Aufgabe 2:**

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

**Lösungsbogen für Aufgabe 3:**

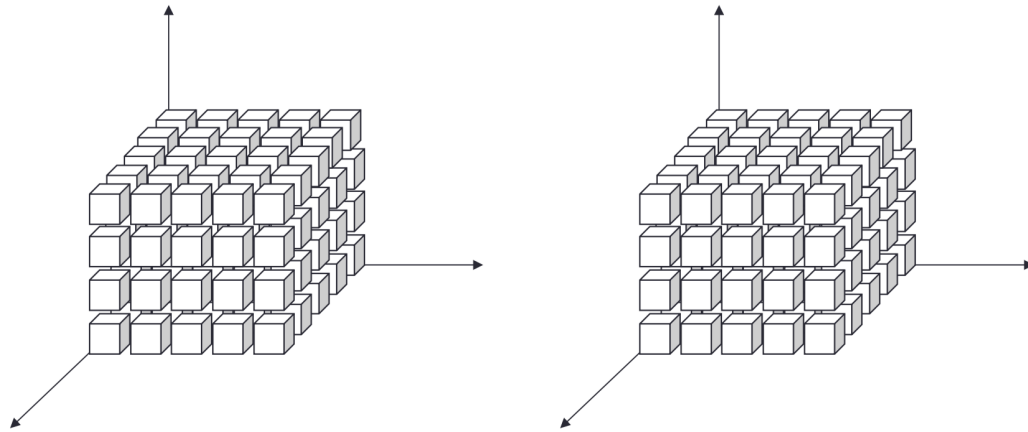
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|



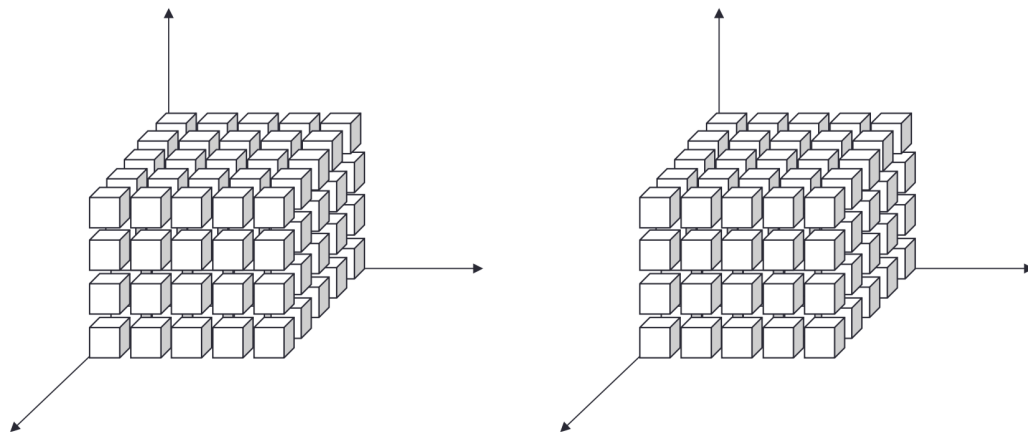
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

**Lösungsbogen für Aufgabe 3.3:**

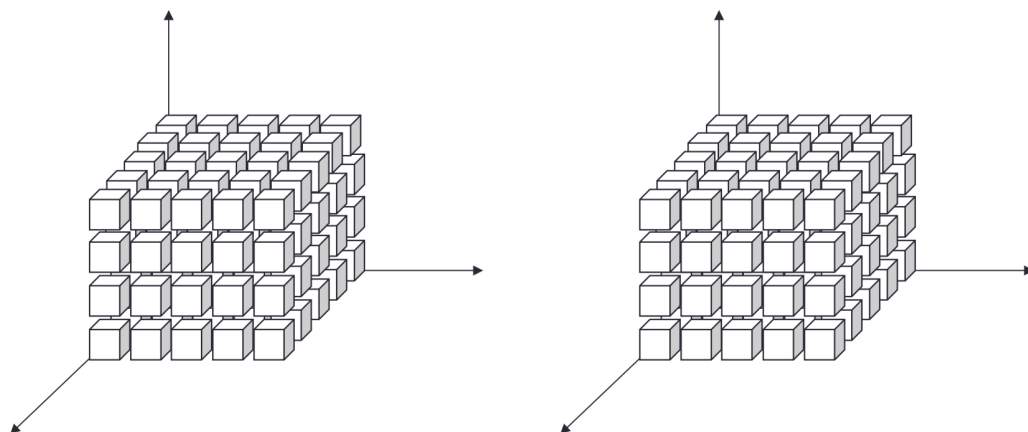
a) Pivotierung:



b) Roll-up:

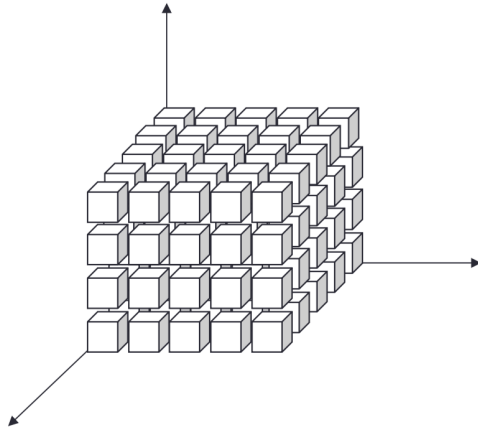


c) Drill-down:

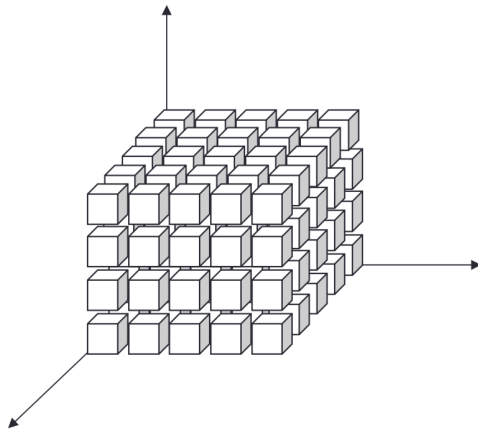


|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

d) Slice: Bitte beschriften Sie zunächst die Achsen und markieren dann die relevanten Daten für die Anfragen: „Alle Daten aus der Region Nord“.



e) Dice: Bitte beschriften Sie zunächst die Achsen und markieren dann die relevanten Daten für die Anfragen: „Alle Daten der Kategorie Drogerie aus den Filialen Ost-1 und West-2 im 3. und 4. Quartal des Jahres 2019“.



|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

**Aufgabe 4** (Ein Kreuz bei Richtig oder Falsch eintragen)

|           | <b>Richtig</b> | <b>Falsch</b> |
|-----------|----------------|---------------|
| <b>a)</b> |                |               |
| <b>b)</b> |                |               |
| <b>c)</b> |                |               |
| <b>d)</b> |                |               |
| <b>e)</b> |                |               |
| <b>f)</b> |                |               |
| <b>g)</b> |                |               |
| <b>h)</b> |                |               |
| <b>i)</b> |                |               |
| <b>j)</b> |                |               |