

# Hochschule Karlsruhe – Data Engineering

## WS 2021/2022

### DSCB330 – Übungsklausur

Name:

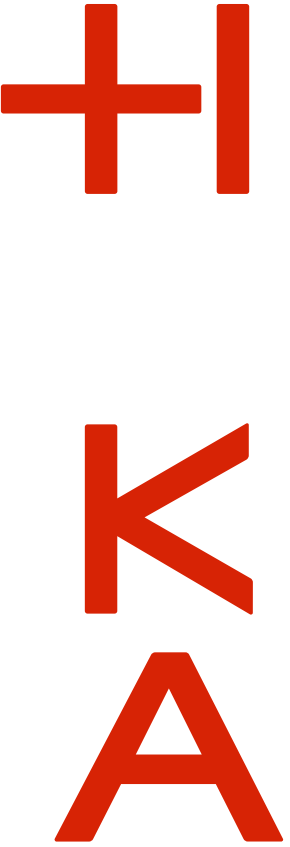
Vorname:

Matrikel-Nr:

Maximale Bonuspunkte: 10

Punktzahl der Klausur: 100

Zum Bestehen sind 50 Punkte notwendig.



# Name:



- + Möglichkeiten zur Bearbeitung der Übungsklausur
- Sie beantworten die Fragen direkt im PDF-Dokument
- Sie beantworten die Fragen auf Papier



Name:



**1) Welche Eigenschaften hat Big Data?**

+ 2 Punkte

**2) Wie wird Smart Data definiert?**

+ 2 Punkte



# Name:



### 3) Genauigkeit der Zeitstempel

- + 3 Punkte
- + python datetime:
- + pandas:
- + numpy:





# Name:



## 5) Auslesen und Umwandeln von Binärdaten aus einem SQL-Server

- + 4 Punkte
- + In einer Tabelle sind Messwerte (RawStreamData) eines IIoT-Sensors im Binärformat gespeichert. Der Sensor erzeugt bei jeder Messung 512 Messwerte im Datenformat Int16. Diese 512 Messwerte werden binär als Stream in der Spalte [RawStreamData] gespeichert.
- + Aus wie vielen Bytes besteht ein Stream?
- + Schreiben Sie ein SQL-Statement zum Auslesen einer beliebigen Zeile aus der Tabelle [timeseries\_c464e3fe9e76lcsstream] mit anschließender Umwandlung der Binärdaten in ein Array von Messwerten.

dbo.timeseries_c464e3fe9e76lcsstream	
Spalten	
Id	(PS, bigint, Nicht NULL)
timestamp	(datetime2(7), NULL)
UUID	(varchar(max), NULL)
streamlength	(int, NULL)
RawStreamData	(varbinary(max), NULL)

# Name:



## 6) SPARQL

- + 2 Punkte
- + Welche Informationen werden durch folgende Abfrage ausgegeben?
- Q11344: chemical element
- P1086: atomic number

```
SELECT ?element ?elementLabel ?symbol ?number
WHERE
{
  ?element wdt:P31 wd:Q11344;
           wdt:P246 ?symbol;
           wdt:P1086 ?number.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]". }
}
ORDER BY ?number
```

Name:



**7) Beschreiben Sie die MQTT Architektur**

+ 4 Punkte





## 8) AMQP

- + 4 Punkte
- + Folgender (unvollständiger) Code implementiert das Empfangen einer Nachricht mit RabbitMQ
- + Was bewirkt die Zeile  
„ch.basic\_ack(delivery\_tag=method.delivery\_tag)”?

```
[...]
cnxn= pyodbc.connect('DRIVER={ODBC Driver 17 for SQL Server};SERVER='+server+';DATABASE='+database+';UID='+username+';PWD='+ password)
cursor=cnxn.cursor()
connection = pika.BlockingConnection(
    pika.ConnectionParameters(host='localhost'))
channel = connection.channel()

channel.queue_declare(queue='task_queue', durable=True)
print(' [*] Waiting for messages. To exit press CTRL+C')

def callback(ch, method, properties, body):
    print(" [x] {} Received {}".format(datetime.datetime.now(),body.decode()))
    # save work to SQL database
    try:
        cursor.execute("INSERT INTO MessageFromAMQP(timestamp,message) VALUES (?,?)",timestamp,body.decode("UTF-8"))
        cnxn.commit()
        ch.basic_ack(delivery_tag=method.delivery_tag)
        print(" [x] {} Saved message {} to SQL server.".format(datetime.datetime.now(),body.decode("UTF-8")))
    except:
        print("Insert to SQL Server failed.")

# give one message to a worker at a time, i.e. don't dispatch a new message to a worker until it has processed and acknowledged the previous one
channel.basic_qos(prefetch_count=1)
channel.basic_consume(queue='task_queue', on_message_callback=callback)
channel.start_consuming()
```

Name:



**9) Vergleichen Sie MQTT, AMQP und OPC UA**

+ 6 Punkte



Name:



**10) Beschreiben Sie die OPC UA-Struktur**

+ 3 Punkte



# Name:



## 11) ETL

- + 8 Punkte
- + In welche Teilprozesse kann der ETL-Prozess unterteilt werden?
- + Wozu dienen die einzelnen Teilprozesse?



# Name:



## 12) Airflow - DAG

- + 3 Punkte
- + In Airflow können z.B. ETL-Prozesse mit Directed Acyclic Graphs (DAGs) beschrieben werden.
- + In einem DAG werden drei Tasks t1, t2 und t3 definiert.
- + Was bedeutet der Befehl `t1 >> [t2,t3]`



Name:



**13) Airflow - Parallele Ausführung der Tasks / Remote Execution**

- + 3 Punkte
- + Welche Möglichkeiten bestehen mit Airflow, Tasks parallel bzw. remote ausführen zu lassen?



## 14) Verteilte Web-Anwendung – Fehlerbehandlung

- + 8 Punkte
- + In einer verteilten Web-Anwendung speichert ein Microservice das Ergebnis einer Modell-Berechnung in einer SQL-Datenbank:

```
def gpmodel_todb(config,function_id:int,fnc:str,rmse:float):  
    drv = pyodbc.drivers()  
    DRIVER_NAME=drv[3]  
    SQL_STR="Driver={"+DRIVER_NAME+"};SERVER="+config['SERVER']+'';DATABASE='+config['DBGP']+'';UID='+config['UID']+'';PWD='+config['PWD']  
    cnxn=pyodbc.connect(SQL_STR,autocommit=True)  
    cnxn.cursor().execute("UPDATE gpresult SET [scheduled]=?,[ready]=?,[function]=?,[root_mean_squared_error]=? WHERE [FunctionId]=?",\  
        False,True,fnc,rmse,function_id)  
    cnxn.close()
```

- + Zählen Sie mindestens zwei mögliche Fehlerfälle auf und beschreiben Sie, wie diese behandelt werden können.

# Name:



## 15) Verteilte Web-Anwendung – Cache

- + 4 Punkte
- + Wie kann ein In-Memory-Cache bzw. ein Distributed Cache die Performance einer Web-Anwendung steigern?





Name:



**16) gRPC**

- + 4 Punkte
- + Welche API-Varianten bietet gRPC?



# Name:



## 17) Web-Architektur mit Microservices

+ 12 Punkte

- A) Beschreiben Sie die Web-Architektur mit Microservices und gehen Sie dabei auf Unterschiede zum monolithischen Ansatz ein.
- B) Wie können die Grenzen von Microservices bestimmt werden?
- C) Welche Möglichkeiten gibt es, den Grad der Kopplung von Microservices zu bewerten?



Name:



**17) Web-Architektur mit Microservices (Fortsetzung)**



# Name:



## 18) Data Warehouse

- + 24 Punkte
- + Ein Data Warehouse für eine Spezialchemie-Anlage hat
  - fünf Dimensionen: Lieferant, Kunde, Herstellungsdatum, Produkttyp, Qualitätsstufe (Ganze Zahl zwischen 1 und 3) und
  - drei assoziierte numerische Werte (Measures): Anzahl produzierte Chargen, Produktionsdauer und produzierte Menge (Masse in Tonnen)
- + Erstellen Sie Queries für folgende Fragestellungen:
  - Gesamte produzierte Menge nach Produkttyp im Jahr 2021
  - Year-to-date produzierte Menge nach Produkttyp in den Jahren 2019 bis 2021 aufgeschlüsselt nach Monaten
  - Produzierte Menge aufgeschlüsselt nach a) Quartalen, b) Kunde, c) Produkttyp und d) Qualitätsstufe
- + Welche Hierarchie(n) ist/sind in den Fragestellungen enthalten?
- + Zeichnen Sie ein Star-Schema für das Data Warehouse.



Name:



**18) Data Warehouse (Fortsetzung 1)**



Name:



**18) Data Warehouse (Fortsetzung 2)**



