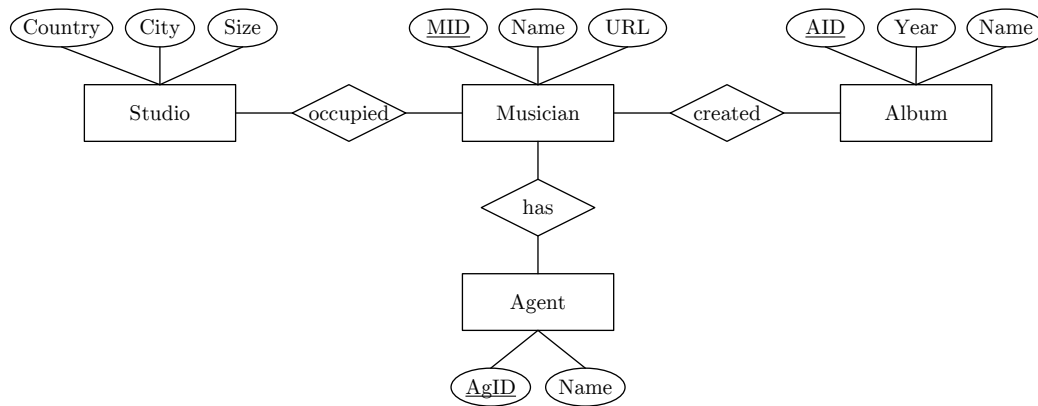


## Exam INF.01014UF Databases (Summer 2019, V1a)

**Important notes:** The working time is 90min, and lecture materials or any kind of mobile devices are not allowed. Please make sure to put your name and matriculation number on the top right of each piece of paper. You may give the answers in English or German, as well as directly write into the task description.

### Task 1 Data Modeling (25 points)



- (a) Given the above Entity-Relationship diagram, specify the cardinalities in Modified Chen notation based on the following information. (6 points)
- A musician might have created none or arbitrary many albums, and any album is created by at least one musician.
  - Every musician has exactly one agent, and an agent might be responsible for one to ten musicians.
  - Every musician occupies exactly one studio, and musicians never share a studio.
- (b) Map the given Entity-Relationship diagram into a relational schema, including data types, primary keys, and foreign keys. (9 points)

- (c) Assume the functional dependency  $City \rightarrow Country$ . Bring your schema in third normal form and explain why it is in third normal form. (10 points)

## Task 2 Structured Query Language (30 points)

Orders

<u>OID</u>	Customer	Date	Quantity	PID
1	A	'2019-06-22'	3	2
2	B	'2019-06-22'	1	3
3	A	'2019-06-22'	1	4
4	C	'2019-06-23'	2	2
5	D	'2019-06-23'	1	4
6	C	'2019-06-23'	1	1

Products

<u>PID</u>	Name	Price
1	X	100
2	Y	15
4	Z	75
3	W	120

- (a) Given the Orders and Products tables above, compute the results for the following three queries: (15 points)

Q1: `SELECT DISTINCT Customer, Date  
FROM Orders O, Products P  
WHERE O.PID = P.PID AND Name IN('Y', 'Z')`

Q2: `SELECT Customer, count(*) FROM Orders  
GROUP BY Customer  
ORDER BY count(*) DESC, Customer ASC`

Q3: `SELECT Customer, sum(O.Quantity * P.Price)  
FROM Orders O, Products P  
WHERE O.PID = P.PID  
GROUP BY Customer`

- (b) Given the Orders and Products tables above, write SQL queries to answer the following questions: (15 points)

- Q4: Which products were bought on 2019-06-22 (return the distinct product names)?

- Q5: Which customers placed more than one order?
  
- Q6: How much revenue (i.e.,  $\text{sum}(\text{O.Quantity} * \text{P.Price})$ ) did products with a price less than 90 generate (return set of product names, revenue)?

### Task 3 Query Processing (15 points)

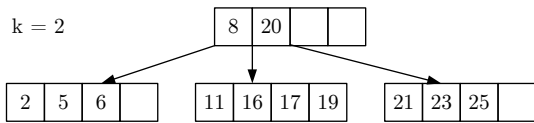
- (a) Assume tables R(a,b), S(c,d), and T(e,f,g), draw a logical query tree in relational algebra for the following query: (7 points)

```
Q7: SELECT R.a, sum(g)
      FROM R, S, T
      WHERE R.b = S.c AND S.d = T.e
            AND T.f > 100
      GROUP BY R.a
```

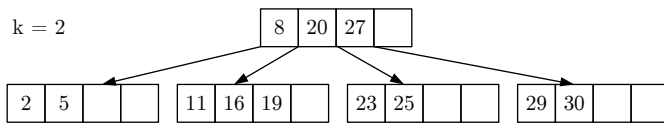
- (b) Describe the conceptual idea of a hash join operator, including the resulting time and space complexity. (8 points)

**Task 4 Physical Design (15 points)**

(a) Given the B-tree below, insert the key 9 and draw the resulting B-tree. (7 points)



(b) Given the B-tree below, delete the key 27 and draw the resulting B-tree. (8 points)



**Task 5 Large-Scale Analysis (15 points)**

(a) Explain Apache Spark’s abstraction of Resilient Distributed Datasets (RDDs), and discuss how they facilitate data-parallel computation and elasticity in terms of scale-out in cloud environments. (9 points)

(b) Describe means (i.e., techniques) of ensuring fault tolerance in distributed storage systems (like Hadoop Distributed File System) and distributed analysis frameworks (like Apache MapReduce or Apache Spark). (6 points)