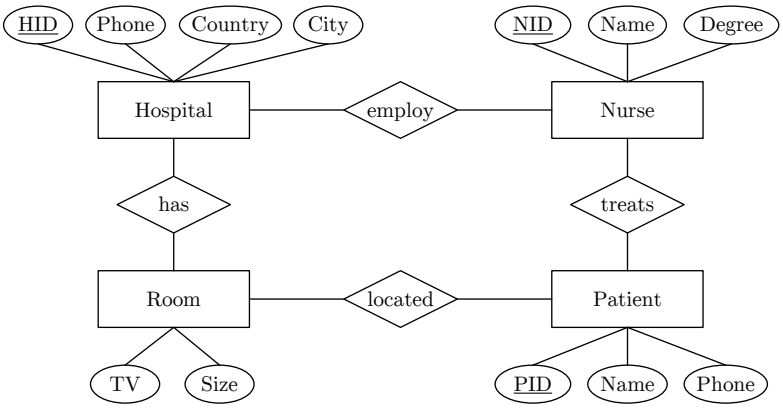
Univ.-Prof. Dr.-Ing. Matthias Boehm
Graz University of Technology
Computer Science and Biomedical Engineering
Institute of Interactive Systems and Data Science
BMVIT endowed chair for Data Management

June 27, 2019

Exam INF.01014UF Databases (Summer 2019, V2a)

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 (30 points)



(a) Given the above Entity-Relationship diagram, specify the cardinalities in Modified Chen notation based on the following information (if unspecified allow general case). (8 points)

- A hospital employs at least 4 nurses and has at least 8 patient rooms.
- A nurse works in exactly one hospital and treats up to 16 patients.
- A patient is treated by at least one but potentially many nurses.
- Every patient has a room, a rooms belongs to exactly one hospital, and rooms are never shared by multiple patients.

(b) Map the given Entity-Relationship diagram into a relational schema, including data types, primary keys, and foreign keys. (10 points)

(c) Let Hospital.Phone and Patient.Phone be multi-valued attributes (lists of phone numbers), and assume the functional dependency City → Country. Bring your schema in third normal form, describe the modifications or the full schema, and explain why it is in third normal form. (12 points)

Task 2 Structured Query Language (30 points)

Orders

OID	Customer	Date	Quantity	PID
1	A	'2019-06-25'	3	2
2	B	'2019-06-25'	1	3
3	A	'2019-06-25'	1	4
4	C	'2019-06-26'	2	2
5	D	'2019-06-26'	1	4
6	C	'2019-06-26'	1	1

Products

PID	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 where bought on 2019-06-25 (return the distinct product names)?

- Q5: Which customers placed only one order?

- Q6: How much revenue (defined as sum(O.Quantity * P.Price)) did products with a price less then 90 generate (return a set of (product name, revenue) tuples)?

Task 3 Query Processing (15 points)

(a) Assume tables R(a,b), and S(c,d,e), draw a logical query tree in relational algebra for the following query: (5 points)

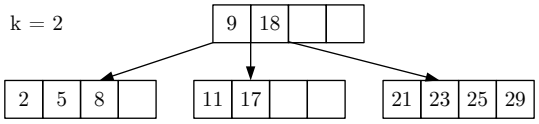
```
Q7: SELECT R.a, S.d FROM R, S
      WHERE R.b = S.c AND S.e < 3
    UNION ALL
    SELECT R.a, S.d FROM R, S
      WHERE R.b = S.c AND S.e = 7
```

(b) Given the schema and query above, which attribute or attributes are good candidates for secondary indexes and how could they be exploited during query processing? (4 points)

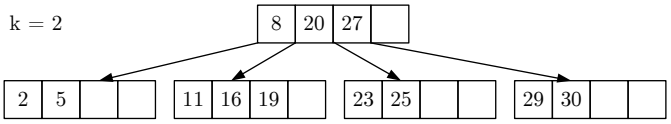
(c) Describe the volcano (open-next-close) iterator model by example of a selection operator and discuss the space complexity of this selection operator. (6 points)

Task 4 Physical Design (15 points)

(a) Given the B-tree below, insert the key 19 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 NoSQL Systems (10 points)

Describe the concept and system architecture of a key-value store, including techniques for achieving high write throughput, and scale-out in distributed environments. Please focus specifically on aspects of physical design such as index structures, and distributed data storage.