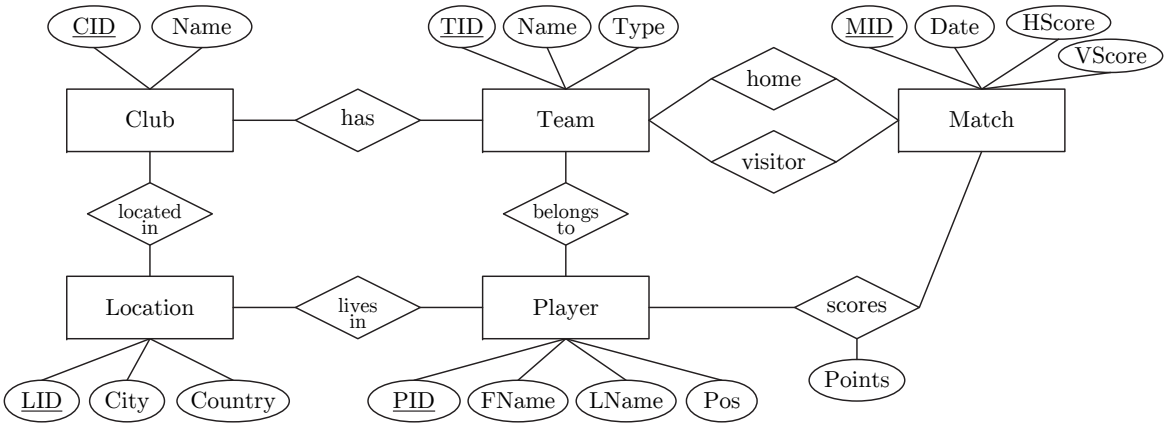
Univ.-Prof. Dr.-Ing. Matthias Boehm
Graz University of Technology
Computer Science and Biomedical Engineering
Institute of Interactive Systems and Data Science
BMK endowed chair for Data Management

July 05, 2021

Exam INF.01017UF Data Management (Summer 2021, 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 the first page of the task description, and each additional piece of paper. You may give the answers in English or German, written directly into the task description.

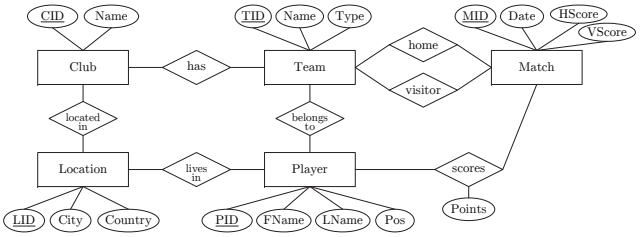
Task 1 Data Modeling (25 points)



(a) Given the above Entity-Relationship diagram, specify the cardinalities in Modified Chen notation based on the following information. (**12 points**)

- A sports club has at least 1 but potentially many teams, each of a certain type (e.g., men, women, U19, U17, etc). Each team belongs to exactly one club and teams consists of at least 11 but often more players. A player is part of at least one but potentially many teams (e.g., men, men II, and U19).
- Each club has a club area in exactly one location (city), each player lives in exactly one location, and any such location can, however, house an arbitrary number of clubs and players (including none).
- Teams compete in many matches but every match is conducted between exactly one home team and exactly one visiting team; the match results are reflected by the home and visit scores (HScore and VScore).
- Players might score points (contributing to the score of their teams) at arbitrary many games, but at most $2 \cdot (11 + 3) = 28$ players can score points at a single match.

(b) Map the given Entity-Relationship diagram into a relational schema in third normal form, including data types, primary keys, and foreign keys. Your schema should also ensure that each match references a home and a visiting team. Furthermore, note that there might be cities with the same name in different countries (no FD City → Country). (**13 points**)



Task 2 Structured Query Language (30 points)

Customers

CID	Name	Country
1	Red	AT
2	Orange	CH
3	Yellow	DE
4	Green	DE
5	Blue	AT
6	Violet	DE

Orders

CID	PID	Date	Qty
5	1	2021-06-29	4
6	3	2021-06-30	1
5	3	2021-06-30	2
2	3	2021-06-30	1
5	2	2021-06-30	10
3	1	2021-06-30	3
6	5	2021-06-30	1

Products

PID	Name	Price
1	Monitor	400
2	SSD	200
3	Laptop	2500
4	Tablet	600
5	Headphones	150

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

```
Q1: SELECT DISTINCT O.Date, P.Name
      FROM Orders O, Products P
      WHERE O.PID = P.PID
        AND P.Price > 200
```

```
Q2: SELECT C.Country, count(*)
      FROM Customers C, Orders O
      WHERE C.CID = O.CID
      GROUP BY C.Country
      ORDER BY count(*) DESC, C.Country
```

```
Q3: SELECT C.Name, C.Country, O.Date
      FROM Customers C LEFT JOIN Orders O
        ON C.CID = O.CID
      WHERE C.Country IN('AT','CH')
```

(b) Given the Customers, Orders, and Products tables above, write SQL queries to answer the following questions (in a way that is independent of the shown tuples): (**15 points**)

- Q4: Which unique products where ordered by customers from Austria, i.e., Country=AT? (return the product name and price)?

- Q5: Compute the revenue (i.e., `sum(Orders.Qty*Product.Price)`) per day (return the date, revenue; sorted descending by revenue)?

- Q6: Compute for all customers—including customers with zero orders—how many orders they placed (return the customer name and count)?

Task 3 Query Processing (20 points)

(a) Given a relation $R(x, y, z)$ with four tuples (a, b, c), (d, e, f), (a, b, d), and (d, e, g) indicate in the table below for each relational algebra expression (row) the number of output tuples in set and bag (multiset) semantics, respectively. (**4 points**)

RA Expression	**Set Semantics**	**Bag Semantics**
$\pi_{x,y}(R)$		
$R \cup (\sigma_{x=a}R)$		

(b) Draw a logical query tree for the following query. (**6 points**)

```
(SELECT DISTINCT Name
   FROM Products
   WHERE Price > 750)
UNION ALL
(SELECT P.Name
   FROM Orders O, Products P
   WHERE O.PID = P.PID
   GROUP BY P.Name
   HAVING sum(O.Qty) >= 5)
```

(c) Given a join $R \bowtie S$ with cardinalities $N = |R|$ and $M = |S|$, enter the space and time complexity of the following physical join operators (in the open-next-close iterator model) in the table below. (**4 points**)

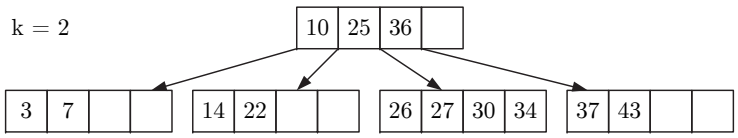
Operator	Time Complexity	Space Complexity
Nested Loop Join		
Sort-Merge Join (unsorted inputs)		
Sort-Merge Join (sorted inputs)		
Hash Join		

(d) Given relations $R(a, b, c)$ and $S(d, e)$, indicate in the table below whether or not the two relational algebra expressions per row are equivalent in bag semantics (✓for equivalent, × for non-equivalent). For non-equivalent expressions briefly explain why. (**6 points**)

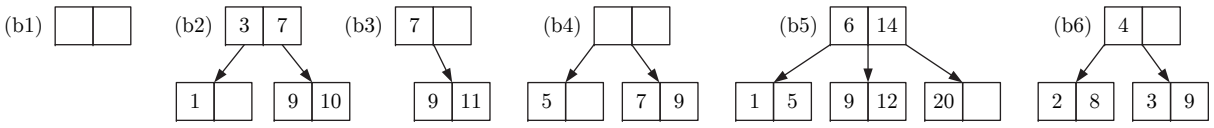
Expression 1	Expression 2	Equivalent? Why Not?
$R \bowtie_{a=e} S$	$\sigma_{a=e}(R \times S)$	
$\pi_{b,d}(R \bowtie_{a=e} S)$	$(\pi_{a,b}(R)) \bowtie_{a=e} (\pi_{d,e}(S))$	
$(\sigma_{b<3}(R)) \cup (\sigma_{b>3}(R))$	$\sigma_{b=3}(R)$	
$\sigma_{c=3}(\sigma_{b=7}(\pi_{b,c}(R)))$	$\pi_{b,c}(\sigma_{b=7}(\sigma_{c=3}(R)))$	
$\delta(\pi_a(R))$	$\pi_a(\gamma_{a,\text{count}}(R))$	
$R \cap (\sigma_{b<4 \wedge b=7}(R))$	R	

Task 4 Physical Design (13 points)

(a) Given the B-tree with k=2 below, delete the key 7, then insert key 35, and draw the resulting final B-tree (or both trees individually). (**7 points**)



(b) Which of the following trees are valid—i.e., satisfy the constraints of—B-trees with k=1. Mark each tree as valid (✓), or invalid (×) and name the violations. (**6 points**)



Task 5 Transaction Processing

Explain the concept of a database transaction log, and how it ensures Atomicity and Durability of changes made by uncommitted and committed transactions in failure scenarios. (**7 points**)

Task 6 Stream Processing

Describe the basic concept and execution model of continuous queries. In this context, also explain stream joins with tumbling window semantics. (**5 points**)