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

1 Database Systems SS19: Exercise 01 – Data Modeling

Published: March 11, 2019 (last update: March 11)
Deadline: April 02, 2019, 11.59pm CET

This exercise on data modeling aims to provide practical experience in Entity-Relationship (ER) modeling, ER-relational mapping, and relational normalization. The expected result is a PDF file named DB_Exercise01_<student_ID>.pdf, submitted in TeachCenter.

1.1 ER Modeling: Soccer World Cup (12/25 points)

Create an ER diagram in Modified Chen (MC) notation—including entity types, relationship types, attributes, cardinalities, and keys[1]—for managing the data of the soccer world cups from 1954 through 2014. The schema should capture the following discourse information:

- A world cup *tournament* is a competition contested every 4 years by national soccer *teams*—which represent their respective *countries*—hosted by one or more *countries* (e.g., the world cup 2002 was jointly organized by South Korea and Japan), and organized in up to 8 groups with up to 4 *teams* per group.

- Each *team* consists of up to 23 *players*, each having a position (with a minimum of three goalkeepers per *team*), a unique jersey number, and a current *club* affiliation. Each *club* has a unique name and is associated with at most one *country*.

- The *tournament* is conducted in phases: one or more group phases, and the finals. In a group phase, *teams* play all against all within each group ($n(n-1)/2$ matches for n teams per group), and the top-1 or top-2 teams per group advance. The finals are conducted as play offs with an additional *match* for the third place.

- Individual *matches*[2] between two *teams* (home and visitor) are played at a certain date and location (*stadium*, which is located in one of the host *countries*), have a type (e.g., final, semi-final, group), last 90min plus overtime, and are decided based on a score of goals. For each goal, the *player* name and time of the match are optionally recorded.

- In the group phase, draws are awarded one point, while wins are awarded three points. However, in *tournaments* before 1994, wins were only awarded two points.

---

[1] Please, use surrogate keys in case of missing unique identifiers or non-integer primary keys.

[2] Note that for simplicity and due to the lack of goal statistics for all world cups, you can model *match* as an entity type as well as the home and visit scores as attributes.

Partial Results: (1) ER diagram (it's up to you if you draw this by hand or use existing tools for data modeling) in Modified Chen (MC) notation, and (2) an additional list of all relationship types in (min,max)-notation using the following notation:

```
<entity1> (min,max) - <relationship> - (min,max) <entity2>
```

1.2 Mapping ER Diagrams into the Relational Model (8/25 points)

Create a relational schema for the ER diagram designed in Task 1.1. This schema should include the relations and typed attributes, all primary and foreign keys, and NOT NULL-constraints. Furthermore, please name additional semantic or domain constraints for ensuring integrity (e.g., see relationship types in (min,max)-notation, restricted naming of groups, etc).

Partial Results: (3) Relational schema (it's up to you if you draw this by hand, use existing tools for data modeling, or provide a SQL DDL script), and (4) a list of at least 4 additional semantic/domain constraints.

1.3 Relational Normalization (5/25 points)

Bring the relational schema from Task 1.2 into third normal form. Furthermore, please explain in detail—with specifics of your particular schema—why this schema is in 3rd normal form.

Partial Results: (5) List of schema changes, and (6) explanation as bullet points.