

Hussain Hussain, Adrian Spataru

Based on previous exercises from Prof. Matthias Böhm

Graz University of Technology

Computer Science and Biomedical Engineering

Institute of Interactive Systems and Data Science

1 Data Management WS22/23: Exercise 01 – Data Modeling

Published: October 10th, 2022

Deadline: October 30th, 2022, 11.59pm

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 zip archive named `DBExercise01-<student_ID>.zip`, containing the partial results of the individual sub-tasks, submitted in TeachCenter.

1.1 ER Modeling (15/25 points)

Create an ER diagram in Modified Chen (MC) notation—including entity types, relationship types, attributes, cardinalities, and keys (create surrogate keys if natural unique identifiers are missing)—for managing a music social network¹. Please use a digital tool for designing the ER diagram. We strongly recommend draw.io² since it has a toolbar for ER elements. There are multiple correct ways of modeling (alternatives, and level of details), however the diagram should capture the essential information of the following discourse:

- The platform contains *users* who register their age, gender and country. A user also has a timestamp (date and time) of their registration. Each user can form friendships with other users.
- The dataset contains several *artists*, described by their names. Each user can listen to different artists on the platform. We register the user-artist playcount, that is the number of times a user listened to an artist on the platform.
- The artist plays one or more *genres* of music.
- The platform also keeps record of upcoming music *festivals*. Each festival takes place in a certain country on a certain date and lasts for one or more days. A festival can either be indoors or outdoors, and may have a known standard ticket price. A festival features one or more genres.
- Each user can buy a number of tickets to one or more festivals. The set of tickets bought by a user can be either standard or VIP tickets.
- *Countries* are further described by their population, and GDP (gross domestic product) per capita.

Partial Result: ERDiagram.pdf

¹Dataset link: <https://github.com/tug-db/datasets/tree/main/LFM-1b/raw>. A cleaned version of the dataset will be made available for Exercise 2 after additional data preparation and cleaning.

²<https://app.diagrams.net/>

1.2 Mapping ER Diagrams into the Relational Model (10/25 points)

Create a relational schema for the ER diagram designed in Task 1.1 and bring it into third normal form, assuming functional dependencies from country codes to their names. This schema should include the relations and typed attributes, as well as all primary and foreign keys. It is up to you if you provide either a SQL script (`CreateSchema.sql`) with `CREATE TABLE` statements, or provide a text schema (`Schema.txt`) in the following text notation (one line per table):

`<Table>(<Attribute 1>:<type>(PK), <Attribute 2>:<type>, ..., <Attribute n>(FK))`

Here, PK and FK indicate primary and foreign keys, where multiple attributes with (PK) represent a composite primary key. If an attribute is both foreign key and (part of) a primary key, use `<Attribute m>(PK,FK)`. Please, adhere to this notation with unchanged parentheses and delimiters, and limit yourself to common data types (i.e., `int`, `numeric(p,s)`, `char(n)`, `varchar(n)`, or `date`) because this sub-task is subject to automated grading.

Partial Result: `Schema.txt`, or `CreateSchema.sql`