AALBORG UNIVERSITY

Cassiopeia

Database Systems – Spring 2014
DPT Group, Aalborg University
Handed out: 07.03.2014

Teacher: Katja Hose
Self study 2: mini-project part 2
Deadline: 12.03.2014

The overall goal in the whole mini-project is to create and work with a movie database in five incremental steps. The first step is to identify the information that is going to be represented as well as special requirements. Then, the information is modeled in an ER diagram and mapped to a relational schema. The database schema is then refined and normalized, so that it can be instantiated in PostgreSQL. After filling database with data, it will be queried and optimized.

Database Modeling

The database should store information about actors, directors, writers, movies, genres, awards, ratings, and users providing ratings. In general, you may use the Internet Movie Database (IMDB `http://www.imdb.com`) for inspiration. But be aware that IMBD stores a large variety of information. You should not try to incorporate all this information into your database design! For most aspects, it will be sufficient to have 5 attributes instead of 20. Please choose the most interesting ones.

As a guideline, your final design should not have more than 10–15 relations.

The created relational schema will be the basis for the following self study slots.

In self study 1, you have already designed this database based on what you already knew back then. Now try it again based on what you have learned so far in the course.

Steps

(a) Identify relevant data (consider the preview of a future self study slot at the end of this document)

(b) Create an ER Diagram with additional information about cardinality (Chen and [min,max]), roles if necessary, primary, candidate, and foreign keys.
You should also add a (short!) description explaining non-trivial design choices that you have made.

(c) Create an appropriate relational model – you do not yet have to consider functional dependencies and normalization.
Only a few null values are expected, hence do not create extra relations if an alternative solution with fewer relations is possible!
You may ignore attribute domains but do not forget to mark the primary keys and foreign keys.

(d) Compare your solution to the results of self study 1 and reflect on the differences.

(e) Write and hand in a report.

Report

- Extend your previous report with a new chapter – each report you hand in should hence contain the complete "history" of what you have done in earlier self studies.

- ER diagram with short additional explanation if appropriate
- A relational database schema clearly defining its relations
  - Names of relations
  - Attribute names and data types
  - Cardinality constraints (Chen and [min,max])
  - Primary keys and candidate keys
  - Foreign key constraints
- Reflections on the initial design from self study 1. Questions you should consider are:
  - What are the differences to your initial design?
  - Why did these differences occur?
  - What advantages does the new design have over the old one?
  - ...

Course goals covered by this self study

- Conceptually design a database (ER model, conceptual design)

Practical hints and preview

Tools for ER modeling

There is a variety of tools available. Many use different notations. Please use one with exactly the notation we discussed in the lecture! Example tools are:

- Windows: Dia, Visio
- Linux: Dia
- Mac: OmniGraffle

You do not need to use a tool that specializes in ER modeling; a program that allows you to draw regular shapes is all you need.

Preview of future self study

In one of the future self studies you will be asked to run the following queries on your database. Hence, please make sure that you cover all necessary information in your design.

- How many Danish language movies are in the database?
- For each year, what is the total number of reviews to movies from that year?
- Which movies have John Travolta and Uma Thurman starred in together?
- How many actors and directors have a first name starting with "Q"?
- How many users rated at least 3 movies?
- What is the name and birth year of all actors in "Pulp Fiction"? Your query should list the actors in increasing order of birth year.
- What are the titles and years of all movies from the 1980s that John Travolta starred in?
- What are the top-2 highest rated movies (average) from the 1990s according to the users?
- What are the top-2 highest rated movies (average) from the 1990s according to at least 2 users?
- In 1994, what was the average rating of a movie for each language?
- Which actors in Pulp Fiction have never, before or after, starred in the same movie as one of the other actors in "Pulp Fiction"?
- Which movie starring John Travolta has the highest user ratings?
- How many actresses have not been alive at the same time as Charles Chaplin?
- What is the average rating of movies from each genre?

- What is the average rating of movies from each genre? List only genres with at least 2 ratings.
- Which movie has the largest number of 2-link references? (If A refers to B, and B refers to C, then we say that A has a 2-link reference, through B, to C. If there are several paths leasing from A to C, we count all of them.)
- How many actors have also been active as director of at least one movie?
- Which two genres are most often linked to the same movie? (Note that each movie has a set of genres.)

Self study: 07.03.2014
The report must be handed in via Moodle no later than
12.03.2014, 23:55 CET