Institut für Informationssysteme
Technische Universität Braunschweig

Technische Universität Braunschweig
Institut für Informationssysteme
http://www.ifis.cs.tu-bs.de

Silke Eckstein, Benjamin Köhncke, and Joachim Selke

SQL Lab: Assignment 3 (due to 11./12.01.2011)

General Information

This week, we will start writing and executing **SQL queries**. For doing so, a full dump of the IMDB database is provided to you. You may access and query the database as described in the next section. The schema of the database is really a pretty bad one – don't be tempted to create such a schema yourself (we do not know how IMDB could ever end up with such a horrible schema; probably they skipped their course on data modeling…). Furthermore, the schema does not contain any primary nor foreign keys. However, columns sharing the same name can safely be assumed to have the same semantics (e.g. actors can be associated with a movie via title_id).

Try to familiarize yourself as good as possible with the schema. IMDB uses Strings as ids for the different tables. These Strings contain all information about a movie, like e.g. title or year. The different parts of the id-String are already split up into single attributes in the tables. **Please note, the ids should only be used for table joins! For other tasks use the respective attributes!**

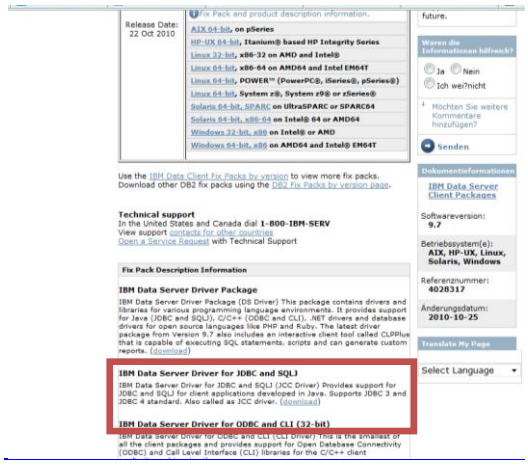
Example:

title_id:	*Terminator 3: Rise of the Machines (2003) (VG)*
title_title:	*Terminator 3: Rise of the Machines*
title_year:	*2003*
title_type:	*video game*

Setting up

First, you need to download the db2 driver from the IBM Website:

https://www-304.ibm.com/support/docview.wss?uid=swg24028317



The file you need is inside the downloaded archive (db2jcc4.jar).

For accessing the database, we still use the Netbeans IDE which needs to be configured as follows (if you prefer, feel free to use any other SQL client):

Register the downloaded driver with Netbeans as shown in Figure 1 and Figure 2.

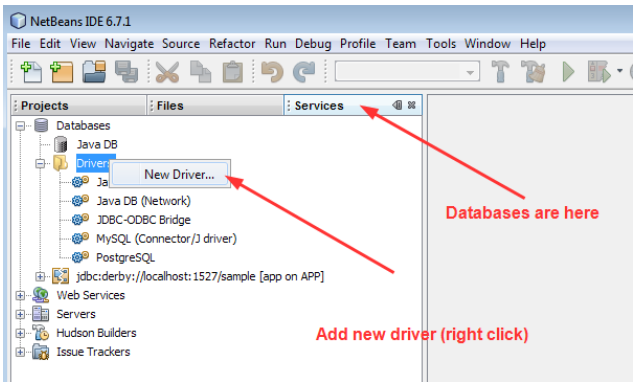


Figure 1 Add new driver in Netbeans

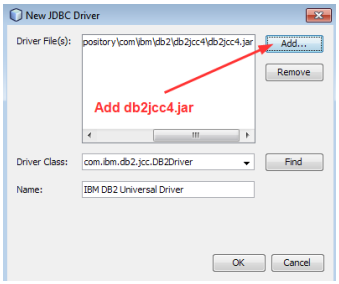


Figure 2 Add DB2 driver

Create a new database connection as shown in Figure 3 to Figure 5. Use the following JDBC connection URL (don't forget the **;**):

jdbc:db2://dblab.ifis.cs.tu-bs.de:50000/DBLAB:retrieveMessagesFromServerOnGetMessage=true;

Finally, select **IMDBRAW** as default schema.

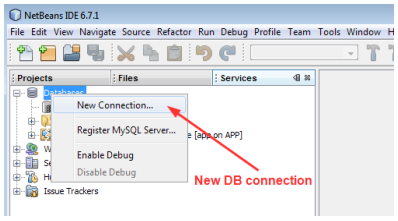


Figure 3 Create new connection

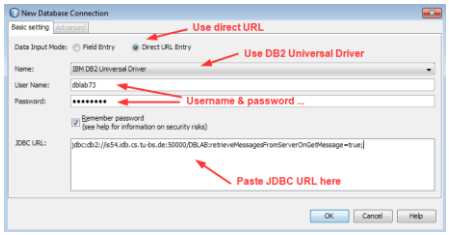


Figure 4 Enter the properties

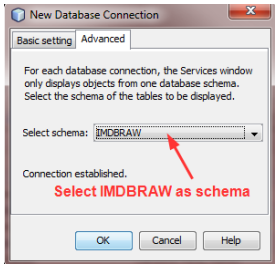


Figure 5 Select IMDBRAW

Everytime you start Netbeans, you have to connect to the database and execute the command **SET SCHEMA IMDBRAW** (Figure 6 - Figure 8).

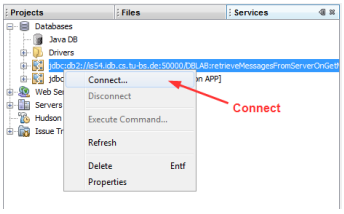


Figure 6 Connect

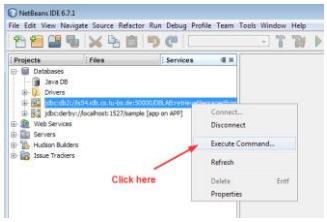


Figure 7 Create new Command

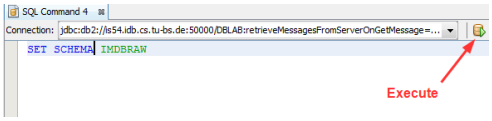


Figure 8 Execute set Schema

After that, you may start querying the DB and review results (Figure 9).

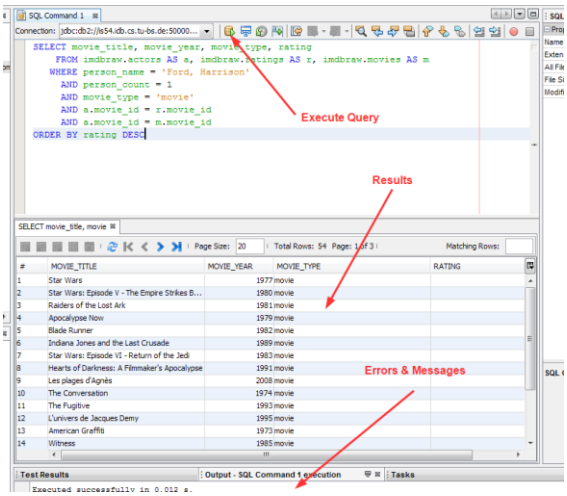


Figure 9 Execute Queries

Tasks

Write an SQL query to perform the following tasks:

a. List the different types of movies contained in IMDB.
b. Which people played a role in the cinema movie (title_type = ‘film’) named Alien (title_title = ’Alien’), but never played a role in any TV movie?
c. Show all countries in descending order where more than 5000 movies have been produced.
d. Show title and year of all action cinema movies that have been produced in Panama?
e. Return the names of all male actors who did the most cinema movies together with Cate Blanchett ('Blanchett, Cate' in IMDB). Order the list by the number of joint titles and return only the top-ten results.
f. Show all titles where Cate Blanchett and Brad Pitt played together that are no cinema movies.

g. Generate an ordered list of all directors who directed more than 10 movies (not series, TV films, etc) overall which are, in average, rated worse than 4.0. The list should contain the name or the ID of the director, the average rating of his/her movies and the number of movies he made.
h. Return the names of all male actors whose movies (not tv series etc.) got an overall average rating of more than 6.0, but who also acted in a movie (!) directed by Greydon Clark ('Clark, Greydon' in IMDB).
i. Generate a list of all movies and their directors (not series, TV films, etc) which were produced, written, and directed by the same person.
j. Which actresses portrayed themselves most often in a cinema movie (!)? (Character is 'Herself' in IMDB). Return only the top-ten.
k. Print the credits of the cinema movie Star Wars (title_title = 'Star Wars') in the IMDB order.

Please note that we limited the database engine to cancel any query which consumes more than two minutes of raw computation time; if your query is ever canceled by a time out error, your approach was either wrong or very inefficient. Try another query!

For additional reference on DB2 SQL, please refer to either the DB2 online documentation (http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/index.jsp?topic=/com.ibm.db2.luw.sql.ref.doc/doc/c0004100.html) or take a look at http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/topic/com.ibm.db2.luw.messages.sql.doc/doc/rsqlmsg.html for SQL error codes.

Sending in Results

Please send your results **via email to your HiWi**. Additionally, **print your queries** on paper and hand them over to your HiWi. For creating the result documents, please use a word processor and format your queries in a readable fashion (e.g. you may want to try the following online formatter: http://www.dpriver.com/pp/sqlformat.htm).
If you fail to either send your solution via email or do not format them nicely on your result document, you **will be graded with -1** !

Please note, that typically there is a high load on the database server on the day before the due date. Therefore, we strongly recommend starting early working on this assignment.