ifis
Institut für Informationssysteme
Technische Universität Braunschweig

15.12.2011

SQL Lab: Assignment 3
(due to 10. / 11.01.2011)

General Information

This week, we will start writing and executing **SQL queries**. For doing so, a full dump of a chemical digital library is provided to you. You may access and query the database as described in the next section.

Setting up

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

https://www-304.ibm.com/support/docview.wss?uid=swg24028317
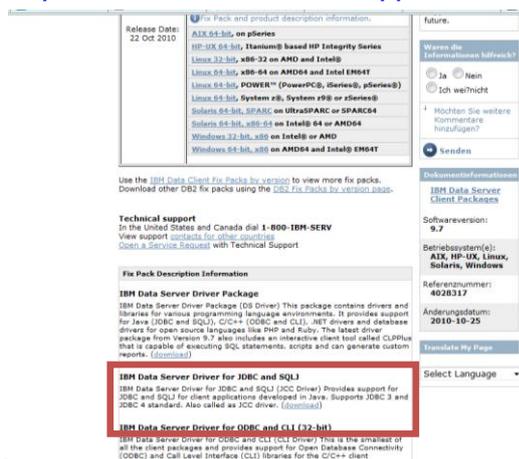
The file you need is inside the downloaded archive (db2jcc4.jar).

For accessing the database, we will use the Netbeans IDE (http://netbeans.org/downloads/index.html) 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.
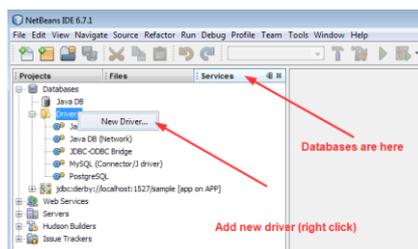

Figure 1 Add new driver in Netbeans

Figure 2 Add DB2 driver

Create a new database connection as shown in Figure 3 to Figure 6. Use the following JDBC connection URL (don't forget the **;**):

jdbc:db2://dblab.ifis.cs.tu-bs.de:50000/DBLAB:retrieveMessagesFromServerOnGetMessage=true;
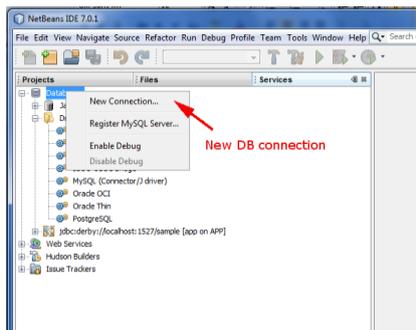Finally, select **CHEMDB** as default schema.
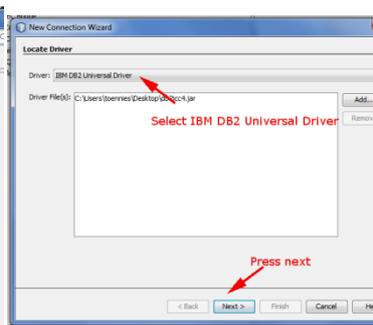

Figure 3 Create new connection

Figure 4 Choose driver

Figure 5 Enter Connection String, Username and Password
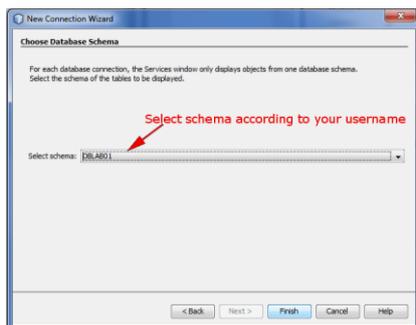

Figure 6 Select the default schema

Everytime you start Netbeans, you have to connect to the database and open the execute command window (Figure 7 – Figure 8). After that, you may start executing queries (Figure 9).
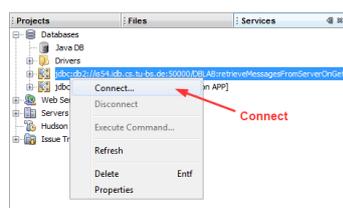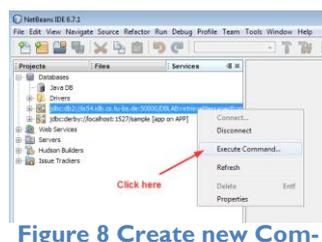

Figure 7 Connect
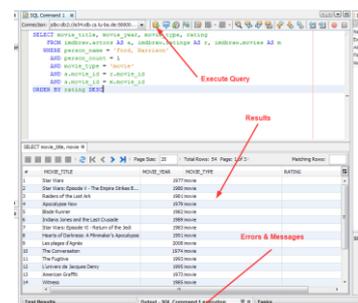

Figure 8 Create new Command

Figure 9 Execute Queries

Tasks

The underlying database schema is the physical model from the chemical digital library schema you already know from the last assignment (Model 2 from the website). Write a SQL query to perform the following tasks:

a. Show the number of different dcRights of documents contained in the CHEMDB.
b. List the top 10 authors published most documents.
c. List the 10 authors published least documents in this collection. Important: There can be authors published none documents!
d. List all NormalizedReaction names which have more than 2 synonyms.
e. The NormalizedReaction table has a special structure. Due to this structure, it is possible to store trees within the table. Here, we have the tree structure from the RXNO ontology (http://www.rsc.org/ontologies/RXNO/index.asp). Please execute the following select statement and try to understand the structure:

   `SELECT * FROM NormalizedReaction N where lft is not null order by lft;`

   Please describe your findings and create a statement where it is possible to retrieve the whole path from the given reaction name (here 'Suzuki-Miyaura coupling') up to the root.

f. Count the number of chemical entities found in the documents which have a thumbnail. Remark: A chemical entity was found in a document if there is a relation between 'Entities' and 'Names'.
g. List all Document IDs which have assigned Tags AND Keywords
h. Count the number of Documents not having any keywords assigned.
i. List all documents with all attributes from the document table and the respective Journal Name and Publication Year and the publisher. A screenshot of the first 10 elements and the total number of resulting documents is enough.
j. Select all Entites which have exactly 70 H atoms in there formula, this means the formula contains 'H70'). The list should contain the ID, the chemical formula and the entity name.
k. List all author first names which are not empty ('') and can also be a lastname. Order the list alphabetically.

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

![ifis logo] **ifis**
Institut für Informationssysteme
Technische Universität Braunschweig

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

Wolf-Tilo Balke, Sascha Tönnies, and Benjamin Köhncke

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). In addition to the SQL queries, **your solutions should also include a screenshot** of the resulting tuples, if you are asked to show them.

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.