Name: Shehroz Khan

Roll #: 19I-1751

Section: BS(DS)-N

**DATA WAREHOUSE AND BUSINESS INTELLIGENCE**

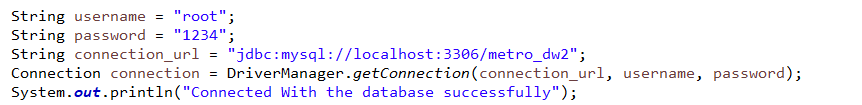
**Semester Project**

**Report**

* **Reading Database (DB) data**

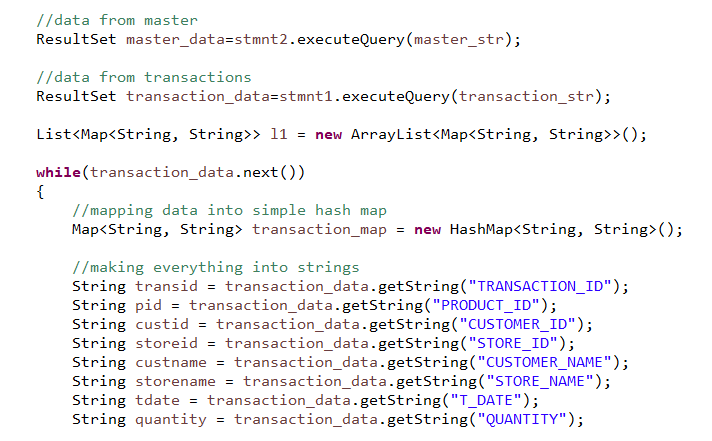
I first ran the required SQL files for Transaction/Master data generator as well as the createDW sql file.

Then I initialized my sql connectors with required parameters.



I then ran the SQL query and stored the ‘next 50’ transaction table results in transaction\_str and the ‘next 10’ master table results in master\_str variables.

And then looped over the transaction\_str where I stored the attribute values in variables.

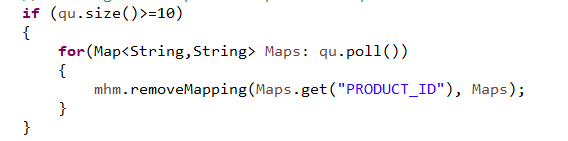


* **MeshJoin implementation**

This project simulates a Data Warehouse so while the data is not actually in stream, we can assume it is so we iterate over the ‘next 10’ master data tuples and ‘next 50’ transaction data tuples. MultiHashMaps, HashMaps, Queues and Lists were used to

store the required tuples accordingly.

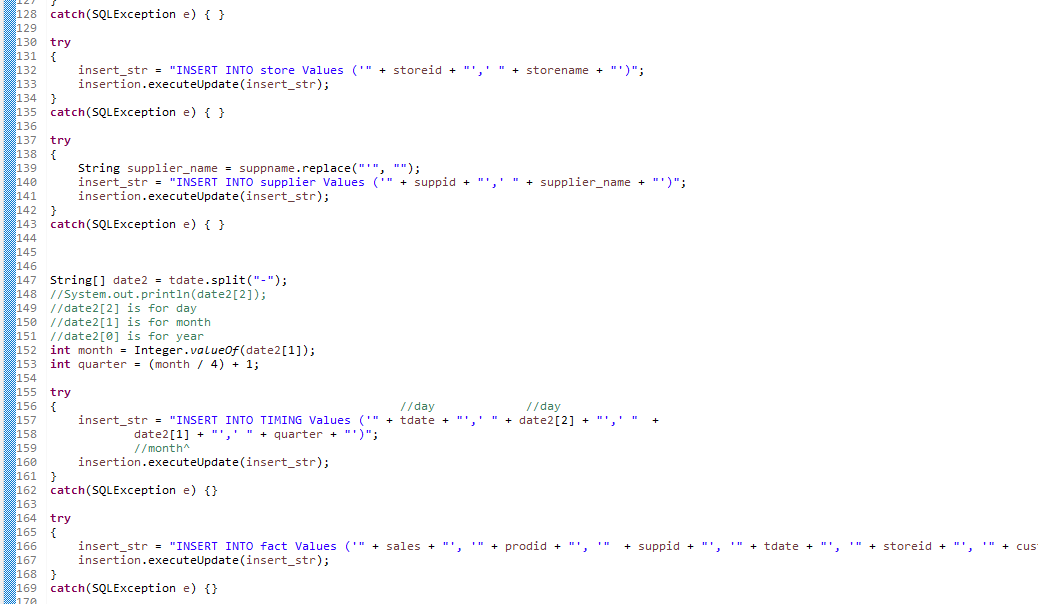
This part re adjusts the size of the Queue that stores the partitions. And when the queue is full it removes the mapping



* **DDL**

For the creation of the Star Schema, I used the createDW sql to create the FACT and Dimension tables. And to insert the data inside them I used the following code inside Java.





* **Problems faced**

One problem was that there was a supplier in Suppliers that had an apostrophe “ **‘** “ in his name.

* **Draw backs of MeshJoin**

One issue with MeshJoin is that while joining, some transactions are skipped due to the nature of the streams and the buffer sizes.   
 Another is that it doesn’t use memory to store the most frequent items and so it will have to repeatedly make system calls which can be time consuming and costly on the system.

**STAR SCHEMA**

