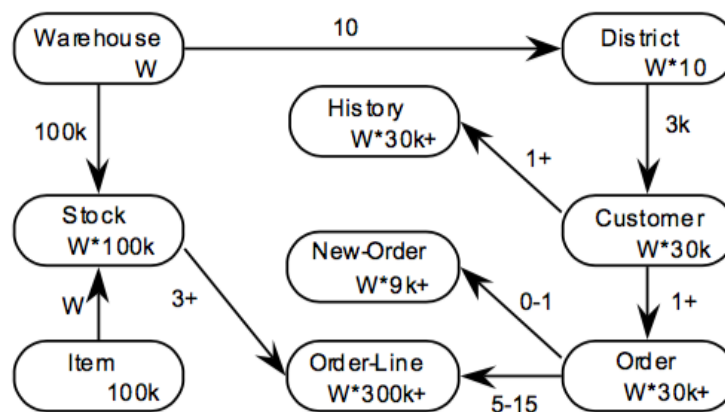




# Cloud Computing and Big Data Ecosystems Design

## HBase: TPC-C Benchmark

**Description:** This exercise will use some of the TPC-C benchmark tables. The TPC-C benchmark simulates the activity of any company that must manage, sell, and distribute products or services. TPC-C defines nine tables with the following cardinalities.



The schema of the tables and the primary are:

**Warehouse:** W\_ID, W\_NAME, W\_STREET\_1, W\_STREET\_2, W\_CITY, W\_STATE, W\_ZIP, W\_ID, W\_NAME, W\_STREET\_1, W\_STREET\_2, W\_CITY, W\_STATE, W\_ZIP, W\_TAX, W\_YTD

**Key:** W\_ID

**District:** D\_ID, D\_W\_ID, D\_NAME, D\_STREET\_1, D\_STREET\_2, D\_CITY, D\_STATE, D\_ZIP, D\_TAX, D\_YTD, D\_NEXT\_O\_ID

**key:** D\_W\_ID, D\_ID

**Item:** I\_ID, I\_IM\_ID, I\_NAME, I\_PRICE, I\_DATA

**Key:** I\_ID

**New\_order:** NO\_O\_ID, NO\_D\_ID, NO\_W\_ID

**Key:** NO\_W\_ID, NO\_D\_ID, NO\_O\_ID

**Orders:** O\_ID, O\_D\_ID, O\_W\_ID, O\_C\_ID, O\_ENTRY\_D, O\_CARRIER\_ID, O\_OL\_CNT, O\_ALL\_LOCAL

**Key:** O\_W\_ID, O\_D\_ID, O\_ID

**History:** H\_C\_ID, H\_C\_D\_ID, H\_C\_W\_ID, H\_D\_ID, H\_W\_ID, H\_DATE, H\_AMOUNT, H\_DATA

**Key:** table without a defined key

**Customer:** C\_ID, C\_W\_ID, C\_D\_ID, C\_FIRST, C\_MIDDLE, C\_LAST, C\_STREET\_1, C\_STREET\_2, C\_CITY, C\_STATE, C\_ZIP, C\_PHONE, C\_SINCE, C\_CREDIT, C\_CREDITLIM, C\_DISCOUNT, C\_BALANCE, C\_YTD\_PAYMENT, C\_PAYMENT\_CNT, C\_DELIVERY\_CNT, C\_DATA

**Key:** C\_W\_ID, C\_D\_ID, C\_ID

**Stock:** S\_I\_ID, S\_W\_ID, S\_QUANTITY, S\_DIST\_01, S\_DIST\_02, S\_DIST\_03, S\_DIST\_04, S\_DIST\_05, S\_DIST\_06, S\_DIST\_07, S\_DIST\_08, S\_DIST\_09, S\_DIST\_10, S\_YTD, S\_ORDER\_CNT, S\_REMOTE\_CNT, S\_DATA

**Key:** S\_W\_ID, S\_I\_ID

**Order line:** OL\_O\_ID, OL\_D\_ID, OL\_W\_ID, OL\_NUMBER, OL\_I\_ID, OL\_SUPPLY\_W\_ID, OL\_DELIVERY\_D, OL\_QUANTITY, OL\_AMOUNT, OL\_DIST\_INFO

**Key:** OL\_W\_ID, OL\_D\_ID, OL\_O\_ID, OL\_NUMBER

You can find a full description of the database tables following this link:

[http://www.tpc.org/tpc\\_documents\\_current\\_versions/pdf/tpc-c\\_v5.11.0.pdf](http://www.tpc.org/tpc_documents_current_versions/pdf/tpc-c_v5.11.0.pdf)

The goal of this project is to **develop a Java program using HBase to create and load the tables, and implementing the following queries:**

**Query1:** List the customers with an order from a given warehouse and district during time interval specified by a START\_DATE (included) and END\_DATE (excluded).

**Query2:** Insert/update (up to 6 times) the discount for a given customer, warehouse and district.

**Query3:** Show the latest 4 discounts for a given customer, warehouse and district.

**Query4:** List all the customers from a given list of districts in a specified warehouse.

Notes:

- A ready-to-use maven project, HBaseTPCC, is provided. It contains the skeleton of the application to be implemented. It is mandatory the use of Oracle Java 8 and the version of the libraries specified in the pom.xml of the project.
  - You can download it following this link:
    - <http://lsd11.ls.fi.upm.es/HBaseTPCC.7z>
- A set of CSVs file with data to be loaded (1 file per table)
  - You can download it following this link:
    - <http://lsd11.ls.fi.upm.es/CSVs.7z>

- To compile the project execute “mvn clean install” in a terminal. It will create the executable file called HBaseTPCC in the folder HBaseTPCC/target/HBase-1.0-SNAPSHOT-bin/HBase-1.0-SNAPSHOT/bin.
- Execution:
  - bin/HBaseTPCC zk\_host:zk\_port createTables
  - bin/HBaseTPCC zk\_host:zk\_port loadTables
  - bin/HBaseTPCC zk\_host:zk\_port query1 W\_ID D\_ID START\_DATE END \_DATE
    - bin/HBaseTPCC zk\_host:zk\_port query1 1 5 ‘2013-11-29 00:00:00.000’ ‘2013-12-05 00:00:00.000’
  - bin/HBaseTPCC zk\_host:zk\_port query2 W\_ID D\_ID C\_ID DISCOUNT\_LIST
    - bin/HBaseTPCC zk\_host:zk\_port query2 3 6 8 10,15,20,50,5,10
  - bin/HBaseTPCC zk\_host:zk\_port query3 W\_ID D\_ID C\_ID
    - bin/HBaseTPCC zk\_host:zk\_port query3 3 6 8
  - bin/HBaseTPCC zk\_host:zk\_port query4 W\_ID DISTRICT\_LIST
    - bin/HBaseTPCC zk\_host:zk\_port query4 2 3,4,5
- Outputs:
  - The methods in the skeleton are already prepared to print out the results of the queries.

You should implement all methods in the skeleton file HBaseTPCC/src/main/java/HBaseTPCC.java. In order to test your implementation compile the project and run the ./HBaseTPCC executable with the action desired.

### Submission:

- **Deadline:** 26nd January 2018 at 23:55
- **Where:** All the required files must be uploaded to Moodle by the deadline. The file must be named ID.rar (ID is the same id of the students provided by the instructor for Flink project). The structure of your delivery will be:
  - ID.tar.gz
    - HBaseTPCC
      - src/
      - pom.xml
- **Groups:** The project is implemented by the same groups that develop the Flink project.