Cloud Computing and Big Data Ecosystems Design: HBase

HBase: Scrabble Games

Description:
The goal of this project is to design the schema of an HBase table and develop a Java program that implements a set of queries on that HBase table. The table, called ScrabbleGames, stores information of a scrabble tournament (http://www.cross-tables.com/). The ScrabbleGames table stores this information according to https://github.com/fivethirtyeight/data/tree/master/scrabble-games:

Column name	Definition
Gameid	A numerical game ID
Tourneyid	A numerical tournament ID
Tie	A binary variable indicating if the game ended in a tie
Winnerid	A numerical ID for the winning player
Winnername	The name of the winning player
Winnerscore	The score of the winning player
Winneroldrating	The winner's rating before the game
Winnernewrating	The winner's rating after the game
Winnerpos	The winner's position in the tournament
Loserid	A numerical ID for the losing player
Loserscore	The score of the losing player
Loseroldrating	The loser's rating before the game
Losernewrating	The loser's rating after the game
Loserpos	The loser's position in the tournament
Round	The round of the tournament in which the game took place
Division	The division of the tournament in which the game took place
Date	The date of the game
lexicon	A binary variable indicating if the game's lexicon was the main North American lexicon (False) or the international lexicon (True)

The project is **a Java program that uses *HBase* to create and load the table, and implements the following queries:**

Query1: Returns all the opponents (*Loserid*) of a given *Winnername* in a tournament (*Tourneyid)*.

Query2: Returns the *ids* of the players (winner and loser) that have participated more than once in all tournaments between two given *Tourneyids.*

.

Query3: Given a *Tourneyid*, the query returns the *Gameid*, the ids of the two participants that have finished in tie.

Notes:

- A READY-TO-USE maven project, HBaseScrabble, is in Moddle. It has the skeleton of the application to be implemented. You must use Oracle Java 8 and the version of the libraries specified in the pom.xml of the project.
- Compile the project executing "mvn clean install" in a terminal. It will create the executable file called HBaseScrabble in the folder HBaseScrabble/target/HBase-1.0-SNAPSHOT-bin/HBase-1.0-SNAPSHOT/bin.
- Execution:
    - bin/HBaseScrabble zk_host:zk_port createTable
    - bin/ HBaseScrabble zk_host:zk_port loadTable
    - bin/ HBaseScrabble zk_host:zk_port query1 tourneyId winnerName
        - bin/ HBaseScrabble zk_host:zk_port query1 3 "Paul Avrin"
    - bin/ HBaseScrabble zk_host:zk_port query2 firstTouneryId lastTourneyId
        - bin/ HBaseScrabble zk_host:zk_port query2 100 150
    - bin/ HBaseScrabble zk_host:zk_port query3 tourneyId
        - bin/ HBaseScrabble zk_host:zk_port query3 235
- Outputs:
    - The methods in the skeleton are already prepared to print out the results of the queries.

All methods in the skeleton file HBaseScrabble/src/main/java/HBaseScrabble.java should be implemented. In order to test your implementation compile the project and run the ./HBaseScrabble executable with the action desired.

Submission:

- **Deadline: 17th January 2019 (23:55)**
- **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 the delivery must be:
    - ID.tar.gz
        - HBaseScrabble
            - src/
            - pom.xml
- **Groups:** The project is implemented by the same groups that developed the Flink project.