HW2 – 236620 – Documentation

Valentin Khais – 317348092 and Michael Leybovich – 310959317

July 8th, 2017

# Overview

The project makes use of previously computed item-item lifts (stored in HBase), in order to recommend items (movies) to users in a streaming fashion (using Storm topology). In parallel, a global recommendation quality score Q is tracked and plotted.

# HBase

Hw2HTablesCreator

* Responsible for all HBase-table-creation related operations.
* Provides the following public API:
  + void createTables() – creates all the necessary tables.
  + Table getTable(TableName name) – returns a Table object by name.
  + void cleanup() – closes all unnecessary resources (should be called when the tables are not needed anymore.

### HBase Tables

* Lift Table:
  + Calculated from *ratings.dat* file.
  + Created and populated during setup phase.
  + Holds Top-K poslift/neglift data for every rated movie.
  + Row keys – {movie\_id x, [pos|neg]}.
  + Column families – Lift.
  + Columns – Array<{movie\_id y, lift}>
* Users Table:
  + Created during setup phase.
  + Populated and maintained during Storm Topology execution.
  + Holds a sliding window of movies (as described in assignment description) for every user (and every aggregation method).
  + Row keys – user\_id.
  + Column families – Window.
  + Columns – sliding windows for every aggregation method.

# Storm

#### RecommendationsQualityStormTopology

shuffleGrouping

shuffleGrouping

shuffleGrouping

shuffleGrouping

fieldsGrouping by user\_id

fieldsGrouping by user\_id

input file

stream source

RatingsSpout

ratings.dat

#### RatingsSpout