Read Me - Restaurant Recommendation System using ALS

Please find detailed instructions about how to execute the project and where to execute listed below.

How to Execute project?

The source code for executing the project is present in recommendation_Source_Code.scala file.
Please execute the same on spark shell with 4 nodes at least in order to get optimal performance.
Please make sure that the program has access to urls mentioned below. This would be required in order to run the program. For running code on complete dataset of 2gb, please download dataset from Yelp from following link:

Source Code File name:
- Yelp_Review_ALS.scala

Sample Input Data files:
- /user/kpw150030/review_200mbfile.json
- /user/kpw150030/input.json (1mb sample data)

Input.json is also present along with the source code file in the zip submitted through the portal.

Actual Input data file Link:
- https://www.yelp.com/dataset_challenge

Statistics on data

MSE: Double = 0.12136101448410759

time taken for 200 mb file

start time=> 10:53:14

end time =>10:55:40

total time taken : 2 min 26 seconds

Datasets used

1. Sample Review data of 1mb
    a. Output MSE - 0.032029814106296135
2. Actual review data of 204 mb
    a. Output MSE - 0.12136101448410759