Big Data Course Term Project: Running Proof  
Recommendation System - Amazon Books

Ashwini Kulkarni, Sindhu Balakrishnan

Computer Science Department, Indiana Universtiy-Purdue University

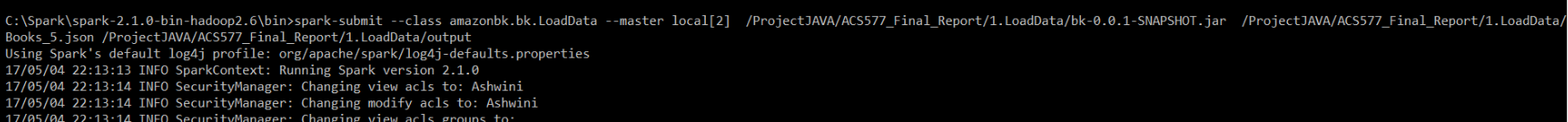
Fort Wayne, IN, 46805,

{ kulka02, balas01}@students.ipfw.edu

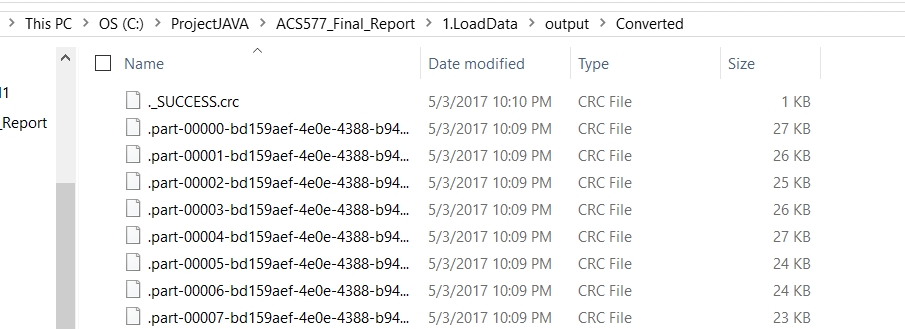
**1.Load Data (Data Prepprocessing)**

Command:

spark-submit --class amazonbk.bk.LoadData --master local[2] /ProjectJAVA/ACS577\_Final\_Report/1.LoadData/bk-0.0.1-SNAPSHOT.jar /ProjectJAVA/ACS577\_Final\_Report/1.LoadData/input/Books\_5.json /ProjectJAVA/ACS577\_Final\_Report/1.LoadData/output



Output is available at /ProjectJAVA/ACS577\_Final\_Report/1.LoadData/output/Converted



Note: There are many files generated as it has 142 milion records:

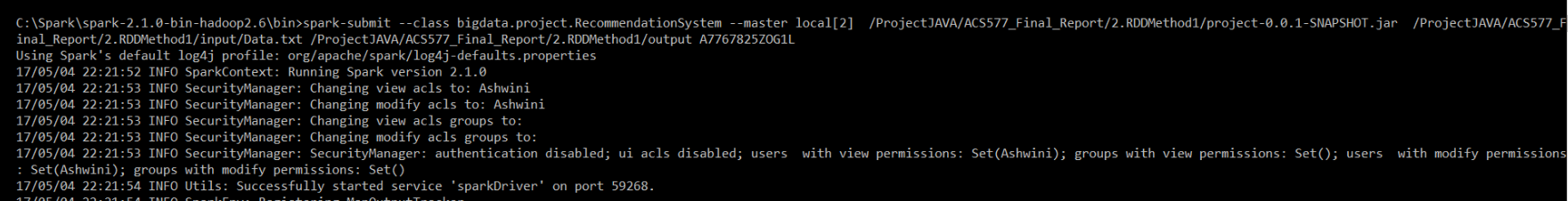
Attaching following one of the file from Load Data prprocessing



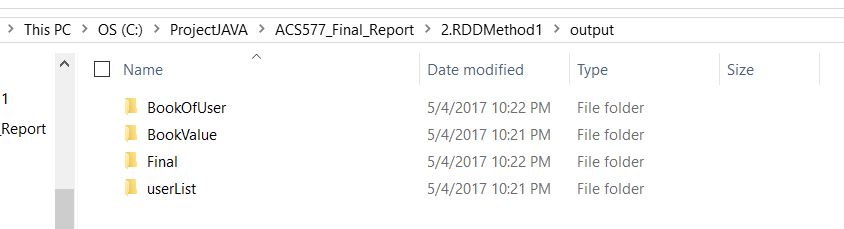
**2.RDD Approach**

Command:

spark-submit --class bigdata.project.RecommendationSystem --master local[2] /ProjectJAVA/ACS577\_Final\_Report/2.RDDMethod1/project-0.0.1-SNAPSHOT.jar /ProjectJAVA/ACS577\_Final\_Report/2.RDDMethod1/input/Data.txt /ProjectJAVA/ACS577\_Final\_Report/2.RDDMethod1/output A7767825ZOG1L



OutPut Available at /ProjectJAVA/ACS577\_Final\_Report/2.RDDMethod1/output



OutPut Details (Attached Files as object in this document):

1st RDD :- BookValue



2nd RDD :- UserList



3rd RDD :- BooksOfUser



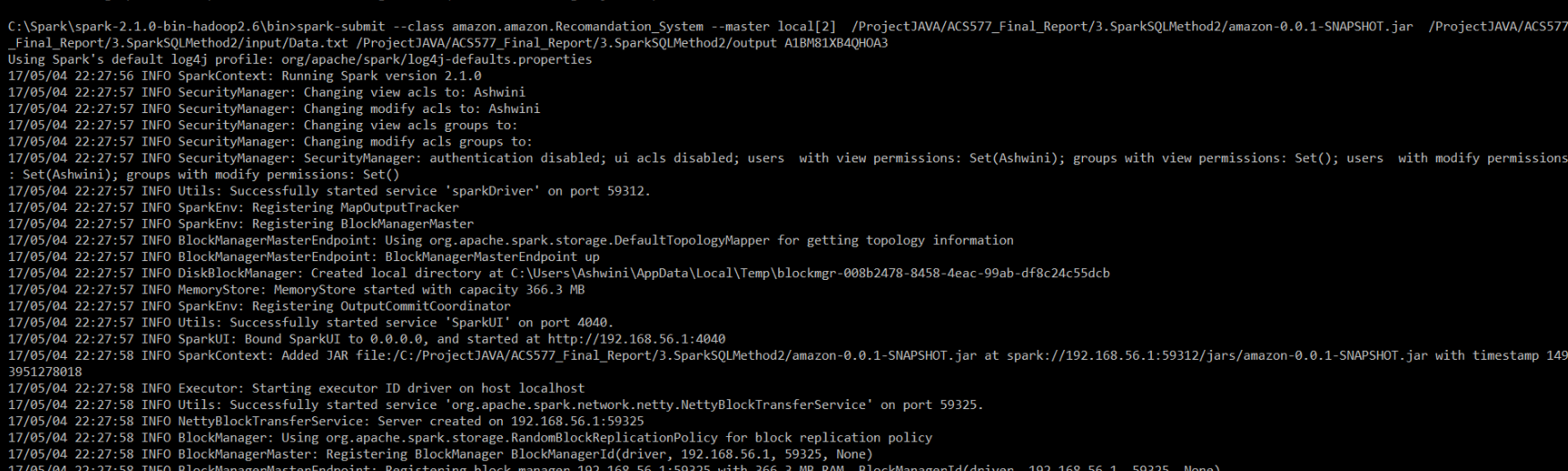
4th RDD :- Final



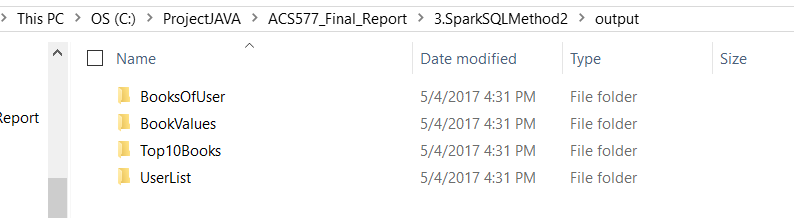
**3.SparkSQL Approach (This is alternate approach for RDD validation)**

Command:

spark-submit --class amazon.amazon.Recomandation\_System --master local[2] /ProjectJAVA/ACS577\_Final\_Report/3.SparkSQLMethod2/amazon-0.0.1-SNAPSHOT.jar /ProjectJAVA/ACS577\_Final\_Report/3.SparkSQLMethod2/input/Data.txt /ProjectJAVA/ACS577\_Final\_Report/3.SparkSQLMethod2/output A1BM81XB4QHOA3



Output:



OutPut Details:

1st Query:- BookValue



2nd Query:- UserList



3rd Query:- BooksOfUser



4th Query:- Top10Books



\*\*\*