HOMEWORK-4

1

IMPLEMENTATION OF MARKET BASKET ANALYSIS ON GOOGLE CLOUD PLATFORM

HOMEWORK REQUIREMENTS

- Design Implement MapReduce Application that processes MBA algorithm on Google Cloud Platform Cluster.

- Your application should process 100 records and generates an output similar to what is shown on slide #15 in this lecture.

- Each record of your input records should represent a ternary mix of maximum of 5 different products (A, B, C, D, E).

- Each product of your 5 products should be represented at least 10 times in your input data set. Your input format looks like the following:
  - TransactionId, product1 product2 product3

- Example of Input:
  1, A B C
  2, C D E
  3, A B E

DELIVERABLES

- Screenshot of Your Cluster on GCP (25% of HW grade).

- Your MapReduce Java Files (25% of HW grade).

- Your Input Dataset (25% of HW grade).

- Screenshot of Your Output (25% of HW grade).

SUPPORTING MATERIALS

- GCP Guide
  - http://people.cs.pitt.edu/~cloudcmp/Assignments/Project/gcp.pdf