**Item-based Collaborative Filtering Assignment 3: ReadMe File**

Input Parameters for MapReduce Program:

* Input Directory. Eg: “/home/gaurang/ItemCf/input”
* Output Directory. Eg: “/home/gaurang/ItemCf/output”

Input Parameters for Sorting Program:

* Input File Path.

Eg: “/home/gaurang/ItemCf/output/output2.txt/part-r-00000”

Input Files for the Program:

* Data file

Sample record:

<User ID> <tab> <Movie ID> <tab> <Rating> <tab> <Timestamp>

The MapReduce Program consists of the following classes:

1. **Driver Class**

The driver class sets the configuration parameters (Input / Output file, Mapper class, Reducer class, Partitioner class etc.) for MapReduce jobs through the Job object. Hadoop uses these configurations to run the MapReduce program.

This Assignment has 2 jobs:

* Job 1 uses 1 Mapper class (Mapper1.class) and 1 Reducer class (Reducer1.class)
* Job 2 uses 1 Mapper (Mapper2.class), 1 Reducer (Reducer2.class).

1. **Mappers and Reducers Classes**

Mapper1.class

This Mapper takes Data file as the input and generates key value pairs with key = User ID and value = Movie details (Movie ID, Rating)

Reducer1.class

The Reducer1 class takes input from Mapper1.class and further processes, i.e., generates a single output line for multiple values of the same key (User ID).

Mapper2.class

This Mapper takes the output from Reducer1.class as the input and processes it to generate key value pairs. In this Mapper, the key is the co-rated movie pairs and the value is the corresponding ratings for the movies.

Reducer2.class

The Reducer2 class takes input from Mapper2.class and further processes, i.e., generates a single output line (similarity score) for multiple values of the same key (Co-rated movie pairs).

1. **Sorting Class**

The Sorting class takes as input the output file generated from Reducer2.class and further sorts the results based on descending values of similarity score.

Eg Input: “/home/gaurang/ItemCf/output2.txt/part-r-00000”