**Mahout and Spark Assignment 4: ReadMe File**

Item Based Collaborative Filtering using Mahout:

* Class files:

FileGenerator.java

Generates comma separated input file from the raw data file which can then be passed to the ItemRecommend.java for processing using Mahout libraries.

Inputs:

* + - * + Data file (input1.txt)

ItemRecommend.java

An Item-based recommender class that generates top 100 related movies list based on similarity scores. (Similarity measure used = LogLikelihoodSimilarity)

Inputs:

* + - * + csv file generated from FileGenrator.java (movies.csv)
        + movie name-id mapping file (u.item)

LDA using Spark:

* Class files:

DocGenerator.java

Generates 1500 documents based on the word distribution Input file.

Input:

* + - * + Data file (docword.nips.txt)

NewClass.java

Generates document-topic distribution for each of the 1500 documents and also generates topic-word distribution (top 10 words per topic) for each topic across all the documents.

Input:

* Document files (/home/gaurang/Personal/Big\_Data/LDA/Doc\_files/\*.txt)