Reg. No.:

Name :

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computing Science and Engineering

Digital Assignment – I

Course Code: CSE3025

Couse Name : Large Scale Data Processing

Instructor Name: Ramesh Ragala

Slot : B2

Problem:

Stochastic SVD is a stochastic technique for computing large dimensional approximate low rank SVDs (Singular Value Decomposition)with rank reaching potentially into hundreds of singular values with just very few passes over data.

This particular project is focused on developing techniques allowing for parallelization of SSVD computation on top of MapReduce framework and Mahout's vectorization framework.

Most of SSVD applications are expected to be revolving around PCA (Principal Component Analysis) as well as LSI. Mahout also uses SVD computations for recommender work, although quality of such recomendations is often (IMO) questionable as in practice such recommendation input is often too sparse for SVD to make good predictions.

Please material: https://github.com/dlyubimov/ssvd-lsi , http://cobweb.cs.uga.edu/~squinn/mmd_s15/lectures/lecture18_apr2.pdf

Develop a MapReduce application using JAVA to compute Stochastic SVD.