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 – II

Course Code: CSE3025

Couse Name : Large Scale Data Processing

Instructor Name: Ramesh Ragala

Slot : B2

Problem:

The QR factorization and the SVD are two fundamental matrix decompositions with applications throughout scientific computing and data analysis. For matrices with many more rows than columns, so-called "tall-and-skinny matrices," there is a numerically stable, efficient, communication-avoiding algorithm for computing the QR factorization. It has been used in traditional high performance computing and grid computing environments. For MapReduce environments, existing methods to compute the QR decomposition use a numerically unstable approach that relies on in-directly computing the Q factor. In the best case, these methods require only two passes over the data. In this digital assignment , you need describe how to compute a stable tall-and-skinny QR factorization on a **MapReduce architecture with implementation** in only slightly more than 2 passes over the data. Then you can compute the SVD with only a small change and no difference in performance.

Material: A new direct TSQR method, indirect TSQR methods that use the communication avoiding TSQR algorithm, and a standard unstable implementation for MapReduce (Cholesky QR).

Reference: https://www.cs.cornell.edu/~arb/papers/mrtsqr-bigdata2013.pdf