

# Lab 06

## Programming Map-Reduce

IT563 Data Mining, Winter'2019, DAIICT, Gandhinagar; pm\_jat

Do following-

1. Setup Hadoop. Hopefully, following link helps  
<https://dzone.com/articles/install-a-hadoop-cluster-on-ubuntu-18041>

Implement following algorithms using map-reduce-

2. Perform a natural join using map-reduce as discussed in lectures.
3. Compute TF-IDF matrix for given set of documents.  
May refer the paper: <https://download.atlantis-press.com/article/2562.pdf>

Deliverables:

1. Source Code
2. Data Files and Outputs

Submit all deliverables in single compressed file named as <your\_id>\_lab06.zip