UE20CS581 – Advanced Big Data Analytics

Assignment-1

Title: Generating Inverted Index for Mahabharata

Professor: Dr. K.V. Subramaniam

Submitted by: Suraj S (PES1PG20CS046)

Divya M S (PES1PG20CS053)

GitHub Repository: https://github.com/surajs757/Inverted_index_mapreduce

Summary:

- splitting files using csplit command. ⸚SEP⸚
- Removing of stop words and cleaning data ⸚SEP⸚
- Understanding how the MapReduce model works. ⸚SEP⸚
- Learning how to set the mapper class and reducer class and ⸚SEP⸚updating context for each job. ⸚SEP⸚
- Learning about the input and key value pairs. ⸚SEP⸚

Output: