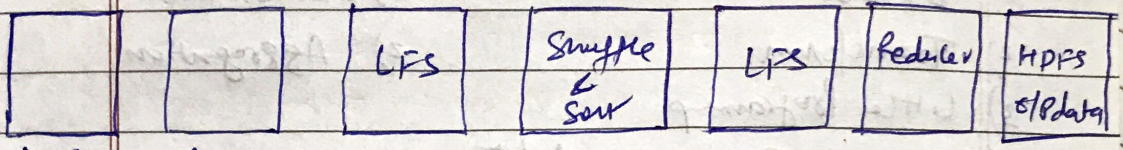
Date:_____ Page:_____

Map Reduce Stages



HDFS i/p data — Mapper

Mapper

1) Data locality
2) Split optional size (equal to block size)
3) Map o/p written to LFS
4) Each map task operate on single HDFS block

Reduce

1) No data locality
2) Reducer o/p is written to HDFS (Reliability)
3) final o/p
4) operate on intermediate data (S/S) - Map task o/p

Shuffle and Sort

1) Sort and consolidate data from all mapper
2) Happens after all map task, before Reduce task start
3) Partitioning (divide based on key value)
4) Movement of intermediate record from mapper to reducer (Shuffling)
5) All value of same key should land on same reducer.

Date: Page:

Mapper

1) Extract
2) Transform
3) Little Programing
4) Deal with small amount of data
5) Work in parallel.

Reducer

1) Statistical
2) Aggregation

- If map fail — Task retry – 4 times
  Job retry – 2 times

MR Flow