Experiment 6

Shashwat Shah
60004220126
TY BTECH COMPSB

Aim: Implement matrix multiplication and word frequency count using MapReduce.

Theory: MapReduce is one of the main features of Hadoop. It is a type of programming paradigm used in hadoop. The major features of map reduce is to perform the distributed processing in parallel in hadoop cluster which makes the hadoop working so fast

Mapreduce has 2 functions

Map() - Takes input from disk as key, value pairs processes them and produces another set of intermediate <key, values> pair as output.

Reduce()- Also takes input as key value pairs. Output of mapping function.

⇒ Matrix multiplication

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \qquad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

Result (K) = $i \times k = 2 \times 2$

Formula for mapping

Matrix A (K, V) = (i, k) (A_{ij}, A_{ij}) for all k

Matrix B (K, V) = (i, k) (A, j, A_{ij}) for all i

Word count

		Dear, 1	Dear, 1	
		Room, 1	Dear, 1	
	Dear Room	River, 1		
	River		Room, 1	Dear, 2
		Car, 1		River, 3
Dear Room River	Car River	River 1	River 1	Car, 2
Car River River	River	River 1	River 1	Room, 1
Dear Car Bear			River 1	Bear, 1
	Dear Car	Dear, 1		
	Bear	Car, 1	Car 1	
		Bear, 1	Car 1.	
			Bear 1	

Input Splitting Mapping Shuffling Reducing

Conclusion: Thus, we have implemented matrix multiplication & word frequency count using map reduce.