

Name - Vasu Kalariya

Roll - PE29

Sub - BDA

Lab Assignment - 5

Aim: Implement MapReduce in Hadoop and perform file handling in HDFS for any domain such as Agriculture / Healthcare using JAVA.

Objective:

- To learn and implement Map-reduce program
- To learn of use of Map-reduce processing in various Big Data Application

Problem statement

Write a Java program to perform map reduce processing for input files related to Weather Monitoring and Sales Datasets

Theory:

→ Working of Map reduce with suitable example

1. Record Records, 2) Map, 3) Combine, 4) Partition
5) Shuffle & sort / reduce 6) Output format.

→ Job tracker: acts like a Master

→ Task tracker: acts like a slaves.

- A job is divided into multiple tasks which are then run onto multiple data nodes in a cluster.
- It is the responsibility of job tracker to coordinates the activity by rescheduling tasks to run on different data nodes.
- Task tracker's responsibility is to send the progress report to the job tracker.
- In case of failure, the job tracker can reschedule it a different task tracker.

→ Advantages of Map Reduce ~~with~~ ~~suitable~~ Processing in Big Data application.

- Highly scalable
- Very self effective situation for business
- Highly flexible
- ~~Tracker~~ Takes very less time to process very large data
- Provides fault tolerance
- Based on simple programming model.

Input: Text / csv file with data

Output: Year-wise Maximum Temp / Display / Payment type wise Maximum Amount spent.

Platform: Windows.

Conclusion: Thus, performed Mapreduce

FAQs

- 1 How the key and values are extracted from input files in the Mapper?

The Map reduce framework operates on $\langle \text{key}, \text{value} \rangle$ pairs i.e framework views the output to the job as a set of $\langle \text{key}, \text{value} \rangle$ pairs and produces a set of $\langle \text{key}, \text{value} \rangle$ pairs as the output of the job conceivably of different types.

- 2 Explain the command ~~word~~ to run a Map Reduce program from command line with example

Command:

hadoop jar path-of-jar-file name of-driver
input-file-path-hdfs output-Holder-name

Eg:

hadoop jar D:\MRjar Driver /BDA input.txt MRout

- Q 3 Discuss the ~~implement~~ importance of Job class in the MapReduce program.

The job class, allows the user to configure the job, submit it control into execution and query the state.