Name - Vasu Kalariya

Roll - PE29

Sub - BDA

Lab Assignment - 4

Aim: Write Hadoop Mapreduce Program using Java

Objective:

→ To learn Hadoop concepts

→ To learn Mapreduce Processing

Problems: Install core Hadoop components of write a Map-reduce Program to count no. of occurence of words in text file

Theory:

→ Introduction to Hadoop

Hadoop is an open source software programing framework for storing larage amount of data & performing the computation. Its framework is based on Java programming with some code in C & shell script.

→ Introduction to Mapreduce processing framework

Mapreduce is a software framework & programing model used for processing huge amounts of data MapReduce program works in two phases. Map tasks deal with splitting & mapping of data while reducer tasks Shuffle & Reduce data

→ Introduction to Hadoop Streaming

Hadoop Streaming is a generic API which allows writing Mappers & Reducers in any language. But the basic Concept remains same. Mappers & Reducer recieve their input & output as <key, value> pairs. Apache Hadoop uses streams as per UNIX standard between your application & Hadoop system.

Streaming is best fit for text processing. Data view is line oriented & processed as a <key, value> pair separated by 'tab' character

→ Study of HDFS Master/slave Architecture

HDFS is a block-structured file system where each file is divided into blocks of pre-determined size These blocks are stored across a cluster of one or several machines. In architecture single Namenode is master & all other nodes (DataNodes) as a slave nodes

Input: Text file with data in HDFS

Output: Output file created in HDFS

Platform: Windows

Conclusion: Succesfully implemented & studied Map-reduce programming model

FAQs

1 Explain the DFS, YARN shell commands and services in Hadoop.

Commands: start-dfs.sh: Fire Datanode & namenode
start.yarn.sh: Fire yarn

Services: DataNode
Namenode
NodeManager
Resource Manager
JPS
} Hadoop Daemons

Vasu Kalariya(PE29)

2 Write the different shell commands used in HDFS for following.

1) Creating a folder
→ hdfs dfs -mkdir /dirname

2) Copying file from local file system to HDFS
→ hdfs dfs -copyfromlocal /inputfile path /desired path

3) Deleting an existing folder
→ hdfs dfs -rm -r /path to folder

4) Displaying contents of a file at command line
→ hdfs dfs -cat /path to file