Weather Log Analysis based on Hadoop Technology

M.Kavitha[1], S.Divya[2], K.Keerthi Lakshmi[3], T.N.Anantha Raman[4], E.Arun[5]

[1]ME, Assistant Professor, Computer Science and Engineering, Dhirajlal Gandhi College of Technology, Tamil Nadu, India

[2,3,4,5]B.E. Computer Science and Engineering, Dhirajlal Gandhi College of Technology, Tamil Nadu, India

---***---

Abstract -

Weather data analytics is very important in every aspect of human life. Weather plays a crucial role in every sectors like agriculture, tourism, government planning, industry and many more. Weather has various parameters like temperature, pressure, humidity and wind speed. To process or analyse this huge amount of data or extracting meaningful information is a challenging task now a days. Big data exceeds the processing capability of traditional database to capture, manage, and process the voluminous amount of data. Here in this paper we are working on data analysis using Apache Hadoop and map reduce algorithms.

***Key Words*: Apache Hadoop, Big Data, MapReduce, Weather analysis .**

1. INTRODUCTION

In general, big data shall mean the datasets that could not be perceived, acquired, managed, and processed by traditional IT and software/hardware tools within a tolerable time. Big Data describes any massive volume of structured, semi structured and unstructured data that are difficult to process using traditional database system such as RDBMS . The data is too big, moves too fast and is unstructured and doesn't fit the structures of the architectures. To gain value from this data we need an alternative way to process it. Thus in our project we are dealing with huge amount of unstructured weather data. Our paper focuses on the shifting of processes from single node data processing to Hadoop distributed file system for faster processing and the best technique to process the queries.

1.1 Motivation:

The goal is to analyse and predict temperature which is useful in agriculture, and also for government to take some necessary steps during floods, hurricanes and natural disasters and so on...

1.2 Objective

a. Data are collected from different resources.

b.   These datas are converted into java language for processing.

c.   The hadoop system processes data and displays the predicted output results.

2. SYSTEM IMPLEMENTATION
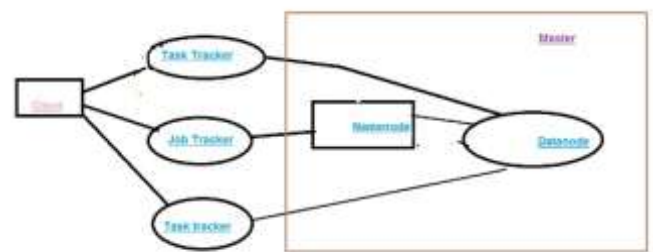
Following diagram shows flow of analysis

FIG-1: Flow Diagram

A. Hadoop:

Hadoop is an open-source framework for processing large amount of data across clusters of computers with the use of high-level data processing easy to use languages, graphical interfaces and administration tools for handling petabytes of data on thousands of computers.
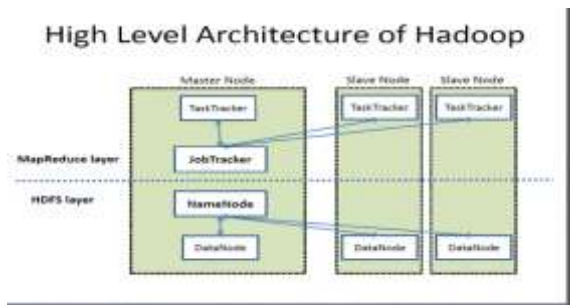
B. Map Reduce:

Map reduce is a processing large datasets in parallel using lots of computer running in a cluster. Map master computer instructs worker computers to process local input data and Hadoop performs shuffle process. Thus master computer collects the results and compilers to answer overall query.

C. Hadoop Distributed FileSystem:

HDFS holds very large amount of data and to store such huge data, the files are stored across multiple machines. These files are stored in redundant fashion and is suitable for the distributed storage and processing.

Hadoop Architecture

3. COMMUNICATION INTERFACES

A. Hardware Requirements:

- **Processor:** Pentium P4 or higher version
- **Motherboard :** Genuine Intel
- **RAM :** Min 512MB
- **Hard Disk :** 20 GB HDD and higher
- Monitor
- Mouse and Keyboard as Input Devices

B. Software Requirements:

- **Operating system:** Ubuntu Operating System
- **Technology Used:** Java and Bigdata Hadoop
- **IDE:** Netbeans IDE 8.2
- **Tools used:** cloudera
- Cloud as a storage

4. MODULES OF THE PROJECT

1. Input Dataset

 The data fields collected from various resources and the parameters are like humidity, temperature, pressure and wind speed etc...

2. Filtering Dataset

The huge amount of data is the loaded onto a hadoop distributed file system. This file system consists of number of clusters .Once the data is loaded onto the hdfs file system the data is balanced across the clusters. The unwanted and irrelevant datas against constraint parameters are filtered.

3. Mapping and Reducing

Map job scales takes data sets as input and processes them to produce key value pairs. Reduce job takes the output of the Map job i.e. the key value pairs and aggregates them to produce desired result.

4. Displaying Result Module

 The datas are processed using data node and namenode of hdfs and the output is also displayed in form of graphs for better knownledge about the predicted result.

5. CONCLUSION

With the increasing amount of daily data its impossible to process and analyze data on a single system and thus there's a need of Multiple Node HDFS system. Once shifted to HDFS System Hadoop proves to be better tool to analyze data for huge volumes. Thus huge weather data can be easily processed with high end systems using Hadoop distributed file system in a very efficient manner.

REFERENCES

[1] Aditya B. Patel, Manashvi Birla, Ushma Nair, (6-8-Dec. 2012),"Addressing Big Data Problem Using Hadoop and Map Reduce".

[2]D.Rajasekar,C.Dhanamani,S.K.SandhyaASurveyon"BigData Concepts and Tools"   Volume 5(February.2015) IJETAE-2250-2459.

[3]Kosha Kothari, Survey Ompriya of various Clustering KaleTechniques"   for   BigData   in   DataMining" Volume1,Issue7,2014 IJIRT-2349-6002.