

Big Data Analytics Framework in Weather Forecasting Analysis

1. **S. Zahir Hussain,**
*Assistant Professor,
Department of Computer Science
PERI College of Arts and Science
Chennai*

2. **P Usha**
*Assistant Professor
Department of Computer Science,
PERI College of Arts and Science
Chennai*

3. **P. Dinesh**
*Assistant Professor
Department of Computer Science and Computer Applications
PERI College of Arts and science
Chennai*

Abstract:

Weather Forecasting plays a primary role of our daily routine lives. Weather Forecasting plays a dominant role of Agriculture Sector, Tourism Sector and Government Body Agencies. Prior knowledge of weather can be very useful for human to prepare themselves for any undesirable climatic conditions. Various Weather parameters like temperature, pressure, humidity, wind speed etc, plays important role in the analysis of weather condition. Big Data Analytics is the process used to Analyzing the Data Properly Systematically to generate an Accuracy results. Now a day's several parts of society are interested in Big Data Analytics will give accuracy in results. In order to achieve that we can use Data Management Techniques to solve the problem to produce Feasible and Optimistic Results This paper deals a systematic literature review for Big Data Analytics and it's approaches in Weather Forecasting Analysis using MapReduce and Spark.

Keywords: *Big Data Analytics, Weather Forecasting, MapReduce and Hadoop.*

I. Introduction:

Weather Forecasting plays an important role of our daily routine life. Systematic and Accuracy Prediction of weather is a key element of various sectors like Agriculture, Tourism and other Government sectors, and plan for natural calamities like flood and rain. Human mind and mood can change Optimistic and Pessimistic or tiredness based on the weather and climatic conditions. Prediction of weather climatic condition is a primary challenge for every human being to sustain for survival. To study the climate and Meteorology, we can do a detail study of interdisciplinary and Systematic Scientific study of atmosphere, temperature, pressure, humidity and wind etc .

Normally Temperature pressure and wind are measured by different parameters like Thermometer barometer and anemometer. Observing these parameters are collected by various sensors and deployed by different geographical locations. Data is accumulated by meteorological department of various countries.

Big Data Technologies like Storm, NoSQL are useful to store and process the huge amount of data. it is important to study their the performance and efficiencies in various domain. In this project we can study the impact of Big Data Analytics and Map reduce spark to analyze weather data Analytics .

II. LITERATURE REVIEW

The Research of Shreddha [1] proposed a method for weather forecasting by adaptive techniques patterns of weather are identified, by various steps' of data mining like data collection, data pre-processing, data cleaning, data transformation and smoothing are explained. For knowledge discovery various methods of mining like Classification, Prediction, Clustering and Outlier Analysis are discussed. K-means clustering algorithm was discussed in detail for weather data. Veershetty et al. [2] worked on building a platform using Hadoop to analyze the weather data. Temperature and yearly precipitation were chosen as weather parameter for

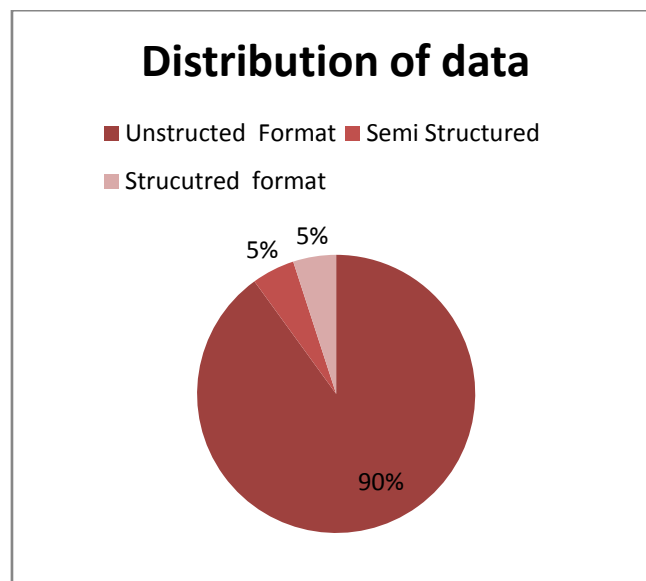
extraction and analysis. The performance comparison of weather data using Pig and Hive is shown. The performance of HIVE is proven to be better in results.

J Denissen et al. [3] has studied the effect of weather on daily mood using multilevel approach. Six weather parameters (temperature, wind power, sunlight, precipitation, air pressure, and photoperiod) were examined to predict the daily mood (positive effect, negative effect, and tiredness). This study showed the importance of routine weather to a daily life of human being.

A Zaslavsky et al. [4] explains the Sensing as a Service and Big Data. Billions of sensing devices are connected to a computer networks and leading to generate huge amount of data on daily basis. Storage and processing of this enormous data is becoming a challenge. To process this data Hadoop, Spark and NoSQL [4] technology can be used. A Katal, M Wazid and R Goudar [5] talked about the issues, challenges, various tools and good practices about handling a Big Data. Various technical challenges to a computer scientist like fault tolerance, scalability, quality of data and processing of heterogeneous data are mentioned. Parallel programming model like MapReduce, Spark and Distributed File System are proposed as a good tool for Big Data.

Huge amount of data is being generated in various domains like Social Networks, Weather Forecasting, Scientific Experiment, Stock Exchange Market, Bioinformatics etc. This huge amount of data is growing at an exponential speed. 90% of data which is available today is in unstructured format while rest 10% comes under the category of structured and semi structured format.

Processing unstructured data is challenge because all the available tools require data to be in structured format. Hence, we need technology and solutions which can process this huge unstructured data

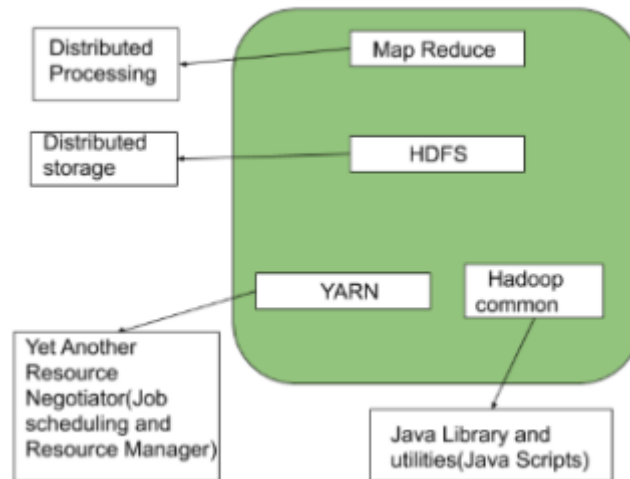


Big Data:

Big Data can be defined as a large amount of data which requires advanced tools and technologies so that it becomes possible to extract value from it by capturing and analysis process. Big data is more than just size. There are three characteristics of Big Data

- a. Volume - It refers to the huge size of data which is being accumulated over the period of time. In case of weather data thousands of sensors are generating data for various weather parameters.
- b. Velocity - It refers to the speed at which data is arriving. If the tool cannot process the data in a time equal to or less than at which it is arriving then it will lead to huge accumulation of unprocessed data.
- c. Variety - It refers to various formats in which data is being generated. Like textual data, binary data, images and video data. In ASCII format XML,, CSV are the different format in which data may come. Separate parser is a need to process each type of format.
- d. Hadoop- Hadoop is the open source software library framework implemented in java for processing huge data on a cluster of commodity computers in a parallel and distributed manner. Hadoop is good for batch processing job. Hadoop has three major components viz. Hadoop Distributed File System (HDFS), Map Reduce Processing Engine and YARN. . HDFS - It is the open source implementation of the Google file system published in white paper it is block oriented, distributed, fault tolerant, reliable, scalable and robust file system supporting huge amount of storage.

Spark Apache spark is fast and general purpose engine for large scale data processing [14-15]. Architecture of spark has spark core at it bottom and on top of which Spark SQL, MLlib, Spark streaming and GraphX libraries are provided for data processing. Apache Spark is very good in memory computing. Spark has its own cluster management but it can work with Hadoop also. There are three core building blocks of Spark programming. Resilient Distributed Datasets (RDD), Transformations and Action.RDD is an immutable data structure on which various transformations can be applied. After transformation any action on RDD can lead to complete lineage execution of transformation before result is produced.



Hadoop Architecture

YARN - It is Yet another Resource Negotiator. It takes the responsibility of resource Management from Map Reduce engine which was there in Hadoop 1.x. It gives the separation between job execution and resource management. It also enables the platform to run another type of programming model like MPI to be executed on top of it.

III. OBSERVATION

Weather Analytics is very useful to every sector of human society. For Weather Data Analysis people has mostly used the temperature as weather parameter Forecasting Analysis. Big Data technology like Hadoop and Spark are being used as a solution to address the challenges caused by Big Data generation.

HDFS has provided the solution for the storage challenge of Big Data.

Hadoop MapReduce is good for batch processing and is giving good result on Hadoop cluster.

IV. CONCLUSION:

Weather Forecasting Analytics has great influence on human society be it agriculture, tourism, sport event, government planning, news agency, industrial farming etc, Weather data is generated by different sensor and produce a storage and processing the results. Big Data Technology like Hadoop Spark can use effectively to handled the weather forecasting data. Various studies are analyzed by different parameters.

REFERENCES

- [1]. 1.Miss. Shraddha V. Shingne, Prof. Anil D.Warbhe and Prof. Shyam Dubey, "Weather Forecasting using Adaptive technique in Data Mining", International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), ISSN: 2321-8169, PP: 091 - 095
- [2]. Veershetty Dagade, Mahesh Lagali, Supriya Avadhani and Priya Kalekar, "Big Data Weather Analytics Using Hadoop", IJETCSE, ISSN: 0976-1353 Volume-14 Issue-02, April 2015
- [3]. Basvanth Reddy and Prof B. A. Patil, "Weather Prediction on Big Data Using Hadoop Map Reduce Technique", IJARCC, ISSN: 2278-1021 Volume-05, Issue-06, Page No (643-647), June, 2016.
- [4]. Zaharia, Matei, et al. "Spark: cluster computing with working sets." HotCloud 10 (2010): 10-10.
- [5]. Riyaz P.A., Surekha M.V., "Leveraging Map Reduce With Hadoop for Weather Data Analytics" IOSR Journal of Computer Engineering, Volume 17, Issue 03, May-June 2015