ANALYSIS OF TEMPERATURE, PRESSURE AND WIND SPEED USING HADOOP

Dr. N. Usha Rani[1] , V Sai Vivek[2] , P. Tejo Sahithi[3] , P.Rishika Reddy[4]
[1] Assistant Professor,Department of CSE, SVU College of Engineering, Tirupati.
[2, 3, 4] Department of computer science, SVU College of Engineering, Tirupati

Abstract:

Big data is the term used for large and complex data which is difficult to process using traditional data processing applications. Hadoop is an open source framework for storing data and running applications on clusters of commodity hardware. Large amount of atmospheric data related to temperature, pressure and wind speed is collected and analyzed using Hadoop and the results are correlated to know how the wind speed, pressure and temperature are related to each another. In the current work Hadoop frame work (single node cluster) and Hive QL queries are used to analyze the change in wind speed and pressure with temperature. Relation between temperature wind speed and pressure is observed in the present work. Pressure and temperature are inversely proportional to eachother and directly proportional to windspeed

Key words: **Big Data Hadoop, Hive, single node cluster**

Introduction:

In general big data is defined in terms of volume, velocity and variety (VVV). Volume or size of the data is very important in big data analysis. Large amount of in comprehensive data is used in big data analysis. Velocity is the measure of how fast the data is accessing. The date should be processed, updated, modified and be retrieved with great speed. Large variety of data either structured or unstructured for variety of applications is used in big data analysis. Large amount of data is collected for analyzing weather conditions [1]. Pressure and temperature are inversely proportional to eachother and directly proportional to windspeed . As wind speed decreases the temperature increases. Pressure increases with decreasing wind speed. The present work is used to study the variations of temperature, pressure and wind speed which influences atmospheric change. These parameters are interrelated and influence climatic change in any country. By using Hadoop frame work, the data related to temperature, pressure and wind speed is analyzed, the relation between these parameters is shown. Hive is a data ware house solution for Hadoop that is flexible to work with large scale data volumes.

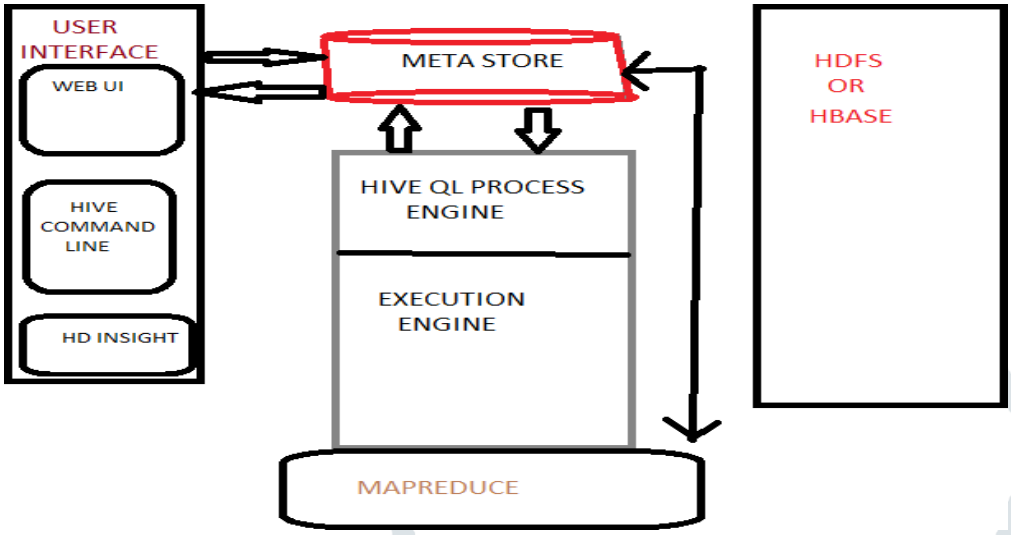
Hadoop:

Hadoop is used to deal with these huge sets of data and analyze the data for future prediction or for decision making. Hadoop Distributed File System (HDFS) is a distributed file system designed to run on commodity hardware. It is similar to existing file distributing systems with more fault-tolerant at a reasonable cost. This is access to wide variety of application date and is suitable for large amount of data. We make use of one of the tools or packages that helps in analyzing and determining data sets with some specific output

Hadoop environment is collected from cloud era domain. Cloud era is a virtual machine package that contains all the environment that need to get set to make use of Hadoop packages, Cloud era has preinstalled with jdk, Hadoop, impala, spark and hive.

HIVE:

HIVE[2] is a query language that is used to retrieve the data from the hive tables that are stored in hadoop distributed file system(HDFS). It is similar to sql queries which is familiar, scalable and extensible. It is mainly used for online analytical processing.



HIVE provides a web user interface which makes easy retrieval of data from that HDFS hive architecture has 3 phases. They are User Interface(UI), Processing Phase(PP), storage**.** Here queries are given from either UI or CLI(Command Line interface) and processed in hive ql (Query Language) process engine and again returns output to UI.

HUE EDITOR:

HUE (hadoop user interface) [3] editor is an user friendly web interface for using hive, impala, spark, pig etc., hue editor can be considered as the web interface of hadoop for retrival and analysis of data with simple instructions. This editor is an inbuilt extension for cloudera [4] virtual machine. File system, list of files that are loaded in to HDFS, query editor and java programming platform, pig platform, map reduce scheduler, work flow of the data can be seen this editor all these features provide user to more ease to understand data and retrieve the data with minimal programming and analyze them.

Proposed work:

Current work is to detect and analyze the data on temperature, pressure and wind speed and how they are interrelated to each other Using Hadoop frame work and Hive queries. In the database [5] pressure is recorded in milli bars, temperature in kelvin*e3, wind speed in Mps.

Implementation:

Hive is a SQL type data base query language that is as similar to SQL queries that would be easy to handle and now upload data bases into Hadoop data base (HDFS) and load them into hive tables, now with appropriate HIVE queries we analyze data

The relation between Temperature and Pressure can be shown in the following table.

Query History Saved Queries Results (17)

	pressure	day10
1	1000	236
2	925	243
3	850	244
4	700	242
5	600	237
6	500	230
7	400	221
8	300	212
9	250	211
10	200	210
11	150	208
12	100	202
13	70	197
14	50	195

Table 1: Relation between pressure and temperature

From the above Table 1, Relation between Pressure and Temperature is analyzed that Pressure is inversely propositional to Temperature.

From the following Figure 2, It is analyzed that pressure increases with Wind Speed.

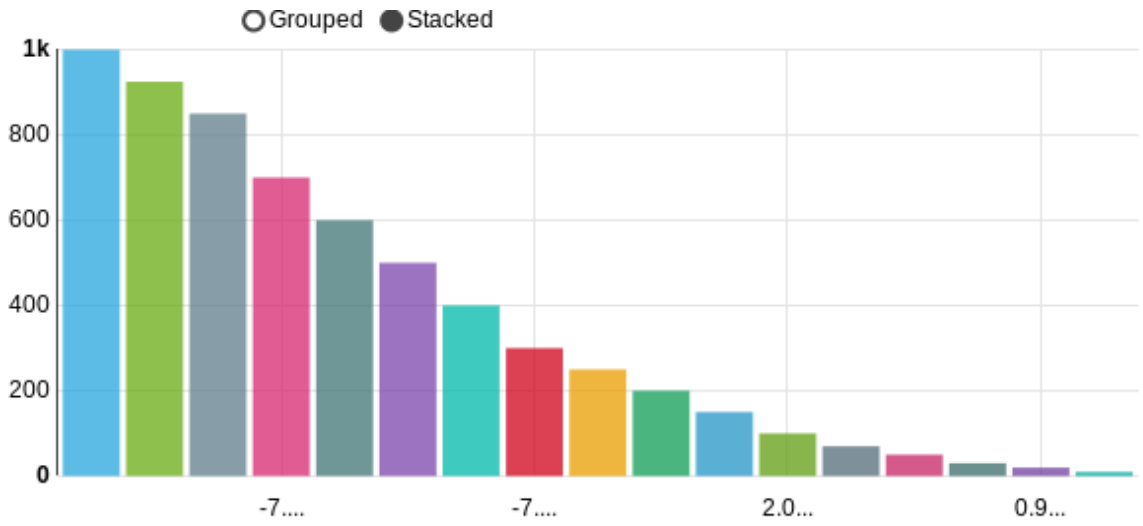


Figure 2: windspeed vs pressure

From the following Figure 3, It is analyzed that, if the Pressure iccreases, then Temperature decreases.

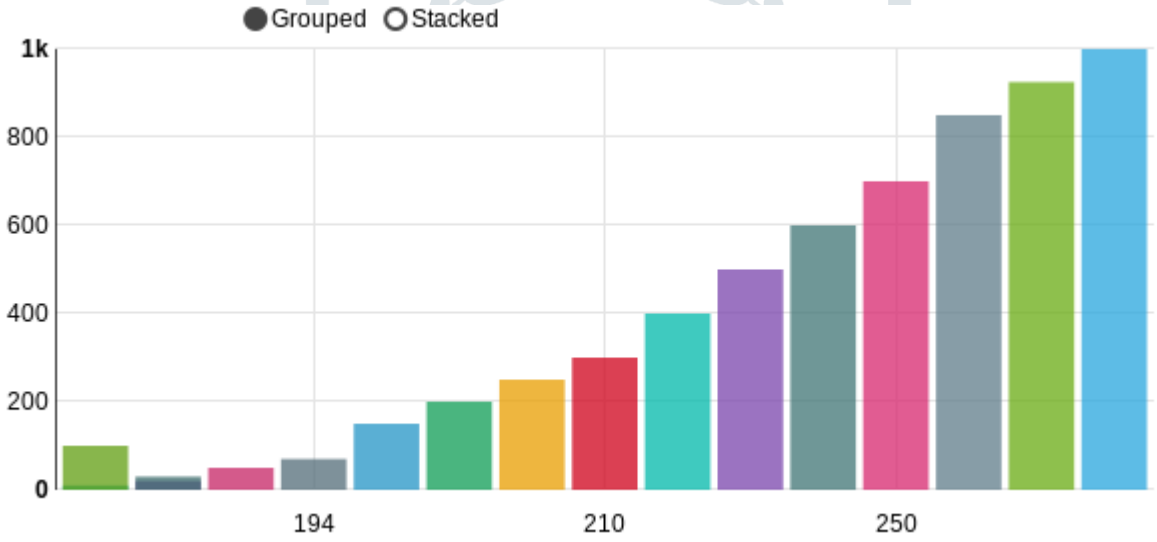


Figure 3: Temperature vs Pressure

Conclusion:

With the increasing amount of daily atmospheric data, It is not possible to process and analyze data on a single system and thus there is a need of Multiple Node HDFS. Once shifted to HDFS System Hive programming proves to be a better tool to analyze data for huge volumes. The relation between the Temperature, Pressure and Wind Speed is correlated for the existing data which is useful for predicting the future atmospheric dateafor climatic changes.

References

1.Veershetty Dagade, Mahesh Lagali, Supriya Avadhan andPriya Kalekar, “Big Data Weather Analytics Using Hadoop”,

International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353 Volume 14 Issue 2 ,APRIL 2015.

2. N pushpalatha and P sudheer, " Data Processing in Big data by using HIVE interface", International journal of advance research in computer science and management studies(IJARCSMS) volume 3, issue4, april2015,

3. M jansi, K pushpa rani, T chaandra shekar reddy, " GLOBAL LAND TEMPERATURE ANALYSIS USING HADOOP USER EXPERIENCE(HUE)". Internaltional journal of civil engineering and technology(IJCTET) volume8, Issue 7, july2017

4. https://www.cloudera.com/downloads/quickstart_vms/5-13.html

5.https://www.esrl.noaa.gov/psd/data/composites/day/