ISSN XXXX XXXX © 2016 IJESC

Research Article Volume 6 Issue No. 12

Weather Forecasting Using Hadoop

Rutuj Deshpande[1], Kamlesh Kamble[2], Rajesh Deshmukh[3], Yogesh Wagh[4]
Computer Department MESCOE, PUNE, India

Abstract:
Weather Forecasting is an important concept since the weather is changing drastically now days. These techniques of weather prediction are used for various purposes like farming, fishing, traveling plan etc. This paper gives the brief idea about the techniques used for mining and processing the weather data and display the result to required user. These techniques include Map Reduce, Naïve Bayes etc. For handling large amount of data Hadoop System is used. This paper provides the knowledge for developing the web application which obtains the data including location and personal details of user and by using the Map reduce and Naïve Bayes techniques for training the data. The required information about the weather and its prediction is sent to user.

Keywords: Big Data, Hadoop, Map Reduce, Naïve Bayes, HDFS, HBase, Weather, and Prediction.

I. INTRODUCTION

Now-a-days, everyone is getting familiar with the concept of Big Data. It is well known to all IT industries as they have to handle fast growth rates data coming from scientific research or business simulations. The need for storing the data is increasing. Handling the large volume of data is a hectic task. Google has also developed Google File System and also Map reduce for handling large volume of data. The storing and processing of such large data, hadoop system is used. Hadoop is an open source system developed by Apache in Java. It is developed to scale to very large clusters. The clusters are automatically fragmented from the storage. Hadoop uses Map reduce technique for scaling and storing the data. This paper includes the details for developing a web application which interacts with the subscriber and providing the required information to solve the queries of the subscriber. The web application will obtain all the personal information of the subscriber like Name, occupation, contact information, etc. while registering this application. Also the location of the subscriber is obtained while registration phase. This Web application will extract the weather data from various sources like Google Weather service then arrange and process the data in hadoop system using the techniques like Map Reduce and Naive Bayes techniques. The weather data here is processed further and the changes in the weather are predicted from the past and real time weather information. This prediction is then sent to the user in form of text message.

II. SYSTEM ARCHITECTURE

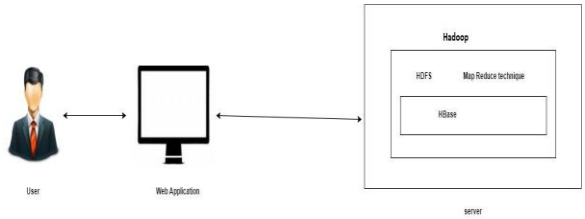


Figure.1. System Architecture

Important points of the system architecture are: -

1] User: - The user is any person who login in the web application like farmers, fishermen, tourist etc.

2] Web Application: - The application developed for obtaining the login details of user and location. This application also sends the required weather information to the user.
3] Hadoop: - It is used where there is need to handle large amount of data.
4] HDFS: - It stands for Hadoop Distributed File System. This is file system used by hadoop to store the data.
5] Map Reduce: - It is technique is used for arranging the data in appropriate order such that it can be used according to need.
6] HBase: - It is data base used in Hadoop which is used for storing the data.

III. Hadoop

There are many industries which use Hadoop System for handling large data. Various applications in such industries are spam filtering, network searching, click-stream analysis, and social recommendation. Many researches are based on Hadoop. The industries using Hadoop are Yahoo, Facebook, etc. Hadoop is described in 3 v's. First v is volume; it is the amount of data it stores. Second v consists of velocity which is transfer speed of data which is high in Hadoop. Third v is Variety; it has various types of data stored in system.

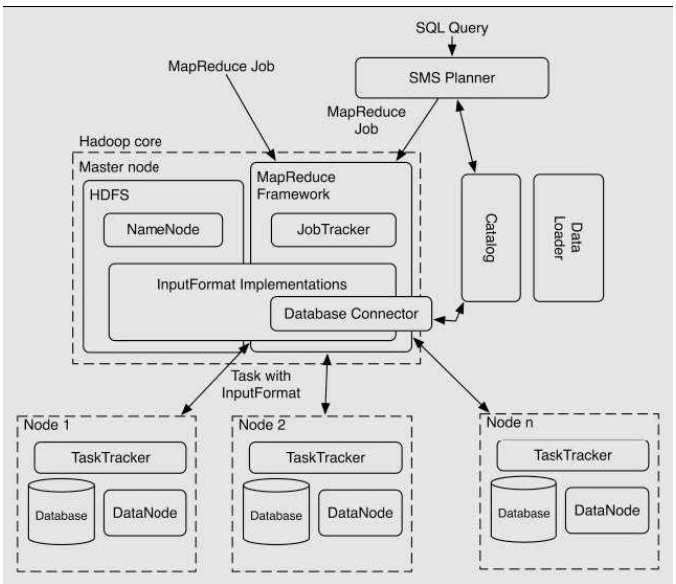


Figure.2. Hadoop Environment

IV. MAP REDUCE

The data collected in this application is stored in raw format in files. These files are given input to the Hadoop system where map reduce technique is used. This technique is used for arranging the data in required format. An input file contains weather data sets for example value of temperature, time, place etc. These data sets are split into data chunks. These data chunks are known as records. Map reduce technique can be explained in three functions. These are driver function, mapper function, reducer function.

Driver function: - This function actually sets up job, submits it and waits for completion of that job.

Mapper function: - Hadoop system uses input file format reader which opens every input file sequentially. It reads the file for finding key-value pairs. Input file format reader reads the file and perform simple loop to extract every key-value pair. Filter is used to match desired key. Values are read into memory and pass to mapper if match with desired key, otherwise skipped.

Reducer function: - The problem is reduced in data chunks. These data chunks are distributed over many data nodes. Each data node process the sub problem given to it and output is given to reducer node. These data node forms tree like structure where each data node send its output to previous data node and finally produce the required output.

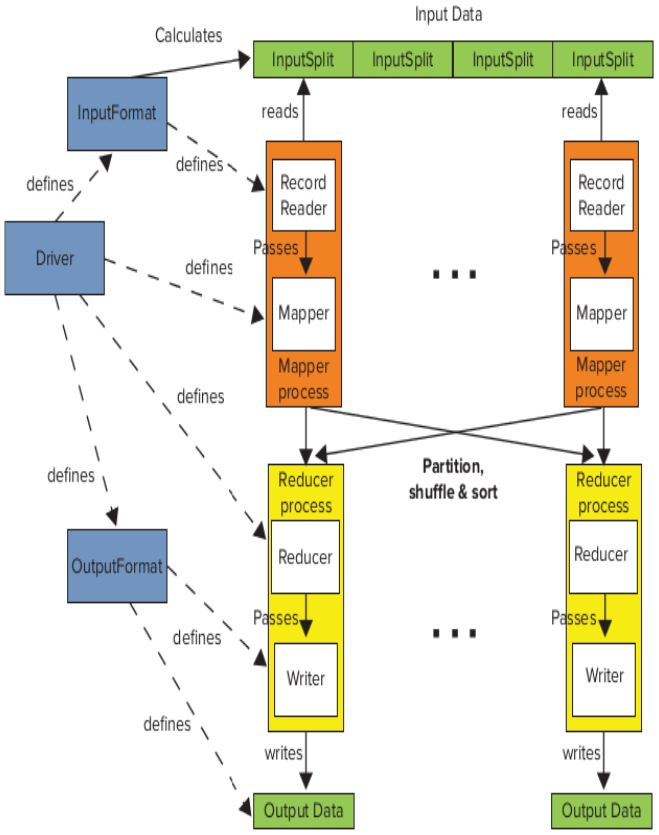
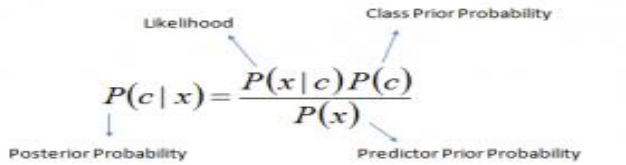


Figure.3. Map Reduce Technique

V. NAÏVE BAYES

Naïve Bayes is a general classification algorithm. It is widely used for classification of text, numerical value, etc. This classification algorithm uses some naïve assumption about class conditional independence for each feature. It uses probability for finding output target class with maximum posterior probability.

The Naïve Bayes algorithm is as follows: -



$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

P (c|x) is the posterior probability of class (c, target) given predictor (x, attributes).
P(c) is the prior probability of class.
P (x|c) is the likelihood which is the probability of predictor given class.
P(x) is the prior probability of predictor.

VI. IMPLEMENTATION OF NAÏVE BAYES

The Naïve Bayes algorithm is used for classifying the data on the basis of various weather conditions such as sunny, rainy, overcast, etc. One such example is explained below: -

Table .1. Classifying the data on basis various weather conditions

WEATHER	SOWING RICE
SUNNY	NO
OVERCAST	YES
RAINY	YES
SUNNY	YES
SUNNY	YES
OVERCAST	YES
RAINY	NO
RAINY	NO
SUNNY	YES
RAINY	YES
SUNNY	NO
OVERCAST	YES
OVERCAST	YES
RAINY	NO

FREQUENCY TABLE		
WEATHER	NO	YES
OVERCAST	-	4
RAINY	3	2
SUNNY	2	3
GRAND TOTAL	5	9

LIKELIHOOD TABLE				
WEATHER	NO	YES		
OVERCAST	-	4	4/14	0.29
RAINY	3	2	5/14	0.36
SUNNY	2	3	5/14	0.36
ALL	5	9		
	5/14	9/14		
	0.36	0.64		

Problem: Farmer will sow rice if weather is sunny. Is this statement correct? We can solve it using above discussed method of posterior probability.

P (Yes | Sunny) = P (Sunny | Yes) * P (Yes) / P (Sunny)

Here we have P (Sunny |Yes) = 3/9 = 0.33, P (Sunny) = 5/14 = 0.36, P (Yes) = 9/14 = 0.64

Now, P (Yes | Sunny) = 0.33 * 0.64 / 0.36 = 0.60, which has higher probability.

VII. CONCLUSION

The flow of the project can be summarized in the following ways:-

The flow starts with registration of user on the web application. All the required information are submitted by the user during registration. With this registration the user has created his account on the web application. After users login the web application requests for the required location of which the user wants the weather information. All the weather information is fetched from the Hadoop system where all the weather data is

stored and processed using the techniques such as Map reduce and Naïve Bayes theorem through Google weather services. This information is send to the subscriber through SMS. All the prediction of weather and according to these predictions, advice is given to the subscriber. We conclude that by using this application we get the best result about the weather or climatic condition on user's mobile phones via SMS. Farmers can decide the suitable timing for their farming processes such as sowing, irrigation, etc. Fisherman can decide whether it is a right time to sail their boats or not. Travelers can decide their tour plans according to the weather conditions of that place. So this project is useful various types of users for fulfilling their requirement.

VII. REFERENCES

[1] "Leveraging Map Reduce With Hadoop for Weather Data Analytics" Riyaz P.A.1, Surekha Mariam Varghese2 *1,2. (Dept. of CSE, Mar Athanasius College of Engineering, Kothamangalam, Kerala, India)*

[2] "Weather Prediction Based on Big Data Using Hadoop Map Reduce Technique" Basvanth Reddy1, Prof. B.A Patil2 Department of CSE, KLE Dr. MSS College of Engg., & Tech, Belgaum1 Professor, Computer Science & Engg, KLE DR M S Sheshgiri College of Engg & Tech., Belgaum2

[3] "The Hadoop Distributed File System" Konstantin Shvachko, Hairong Kuang, Sanjay Radia, Robert Chansler, Sunnyvale, California USA{Shv, Hairong, SRadia, Chansler}

[4] "Classification and Forecasting of Weather using ANN, k-NN and Naïve Bayes Algorithms" Nishchala C. Barde1, Mrunalinee Patole2,1RMD Sinhgad School of Engineering, Savitribai Phule Pune University, Pune, India,2Professor, RMD Sinhgad School of Engineering, Savitribai Phule Pune University, Pune, India

[5] "Weather Forecasting using Deep Learning Techniques" Man Galih Salman School of Computer Science Bina Nusantara University Jakarta, Indonesia asalman@binus. edu Bayu Kanigoro School of Computer Science Bina Nusantara University Jakarta, Indonesia Bkanigoro @binus. edu YayaHeryadi School of Computer Science Bina Nusantara University Jakarta, Indonesia YayaHeryadi @binus.edu Abstract- Weat

[6] "Large Scale Text Classication using Map Reduce and Naive Bayes Algorithm for Domain Specied Ontology Building" Joan Santoso, Eko Mulyanto Yuniarno, and Mochamad Hariadi Department of Electrical Engineering, Faculty of Industrial Technology Institut Teknologi Sepuluh Nopember Surabaya, Indonesia Department of Informatics EngineeringSekolah Tinggi Teknik Surabaya Surabaya, Indonesia