ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

Climate Change Detection using Hadoop with MapReduce

Ms. Priyanka More[1], Ms. Sunita Nandgave[2], Ms. Megha Kadam[3]

M.Tech Student, Department of Computer Engineering, G. H. Raisoni College of Engineering and Management, Wagholi, Pune, India[1]

Asst. Professor, Department of Computer Engineering, G. H. Raisoni College of Engineering and Management, Wagholi, Pune, India[23]

ABSTRACT: Big Data is a term refers to a collection of large amount of data which requires new technologies to make potential to get value from it by analysis and capturing method. In every aspect of human life, weather has a lot of importance. It has direct impact on each part of human society or human beings. Accurate analytics of weather collecting, storing and processing a large amount of weather data is necessary. So a scalable data storage platform and efficient or effective change detection algorithms are required to monitor the changes in the environment. An existing or traditional data storage techniques and algorithms are not applicable to process the large amount of weather data. In the proposed system, a scalable data processing framework that is Map-Reduce is used with a climate change detection algorithms which is Spatial Cumulative Sum algorithm and Bootstrap Analysis algorithm. This paper presents, the large volume of weather data is stored on Hadoop Distributed File System (HDFS) and Map-Reduce algorithm is applied to calculate the minimum and maximum of climate parameters. Spatial Autocorrelation based climate change detection algorithm is proposed to monitor the changes in the climate of a particular city of india.

KEYWORDS: Big Data, HADOOP, Map-Reduce, Temperature

I. INTRODUCTION

Analyzing a huge data sets is the method of big data which comprises a classes of data types. The big data maintain a significant amount of data and process that data. It is traditional data analysis which can handle only the structured data, but not unstructured data. In big data, it can process both structured as well as unstructured data. Big data contains various datasets which are in different dimensions having the ability to generally employed on software tools which manages, captures, processes the data accurately. Big data size may varies from terabytes to several petabytes of data.

Weather analytics is the employment of technology to analyze the behavior of the environment for a given area or city. It is essential for farmers, disasters, business agriculturist and in sports etc. weather analytics is one of the most exciting and fascinating domain and plays a very important role in aerography. There are numerous conditions in an excellent implementation of weather analytics for example in data mining methods, it cannot analyze weather in short-term efficiently[1].

The most popular Big Data handling and processing technique is Hadoop Map-Reduce which is currently used. Map-Reduce is a technique which executes parallel and distributed algorithm across large data using number of clusters. In the proposed system, Map-Reduce algorithm is used to calculate minimum and maximum temperature of a particular city and Spatial Cumulative Sum(CUSUM) based algorithm is proposed to detect the changes in the climate which produces the results in the form of graphs with temperature values[5].

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

II. REVIEW OF LITERATURE

Mrs. Shraddha V. Shingne [2] developed a weather prediction system by using a data mining. The data mining steps like data collection, pre-processing, data cleaning, transformation and smoothing are used. Different techniques of data mining like Classification, K-means clustering and Outlier analysis were applied on weather data to analyze minimum, maximum and average temperature. A method of determination of the occurrence of rare patterns in weather is to be proposed.

Basvanth Reddy and Prof B. A. Patil [9] worked on a system for prediction of maximum and minimum temperature of particular area of particular year. Some basic details of YARN and Map-Reduce were explained to implement or proposed a system.

Ye Ding, Yanhua Li [10], worked on detecting and analyzing urban regions with high impact of weather change on transport using the weather traffic correlation detection method. This proposed work fill up the gap in the analysis of the impact of weather to traffic from some locations to all road networks overall city.

Mr. Sunil Navadia [12] worked on a weather analytics a novel technique for measuring and analyzing weather data. The aim of this work is to analyze the chances of rainfall by using various hadoop tools. Here, the proposed system used as a tool and Naive Bayesian algorithm is used to collect data from large volume of data as input and analyzed future rainfall.

Ms. D. Jayanthi [13] worked on weather data analysis with the help of spark which is an in memory Computing framework. Here, a spark instance and i-python notebook were created. After that weather data set is collected from weather sites and loaded into notebook. The spark streaming is done by creating data into number of partitions also distributed the partitions into a clusters in which the highest average precipitation and temperature values were executed for highest ten weather stations and displayed results. If the volume of data and iterations will increase, then the work can be extended.

A. Gayathri [17] worked on using data mining methods to survey on weather data. Various kinds of forecasting that is Now casting, Short range, Medium range and Long range forecasting were disscuss. weather parameters also a data mining methods were described along with the classification algorithm like Bayesian classification, Back propagation to analyze weather.

III. MAPREDUCE PROGRAMMING

Map-Reduce is a process which will be work in three steps, namely map, shuffle, and reduce. The mapper's job is to process the input data in map stage. In Hadoop file system, the input data is in sort of file and is collected from various weather sites. And the reducer will take the output from the Map as an input and combined that data into a set of tuples.

IV. PROPOSED METHODOLOGY FOR WEATHER ANALYTICS

4.1 Driver Operation

The driver which set up job, submit it, and wait for process to complete. It is taken from a configuration file to specify the input or output directories. Also it can accept script based on mapper and reducer without re-compilation [7].

4.2 Mapper Operation

The mapping is a simple process in that the variables which matched certain will be sent to the reducer. It consider mappers is acts like a distributed search capability and pull (key, value) pairs of a file. The input file format reader of hadoop opens files and which starts to read file for (key, value) pairs. Once it determine (key, value) pair, it reads both key and values which passes to the mapper and mapping operator which is used to filter out (key, value) pairs which not match the criteria. Since mapper is not a part of Hadoop which read data. The mapper is collecting data from input file format reader and input file format reader is modified to read sequenced files. This routine opens a file and performs a simple loop to read each (key, value) within file. The desired key if filter matched, then values which are

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

read into memory and passed to the mapper. If filter is not match, then values were skipped. In the mapper, null values is to filter for calculation of a place, id is used as a key and combination of date and place is used as Key [7].

4.3 Reduce Operation

The resulting (key, value) pairs which matched the criteria is analyzed and forward to reducer with sequencing and complete mapping process, Once a (key, value) object has created, a comparator is needed to order keys. If data is combined, a group comparator is also needed. In a partitioner must be created in order to handle partitioning data into groups of sorted keys.With all these components in place, Hadoop takes the (key, value) pairs which is created by using mappers and group and sort them as specified way. Hadoop assumes that all values share a key will sent to same reducer and a single operation over a large data set will employ on one reducer, This gives us result in number of output files. [7]

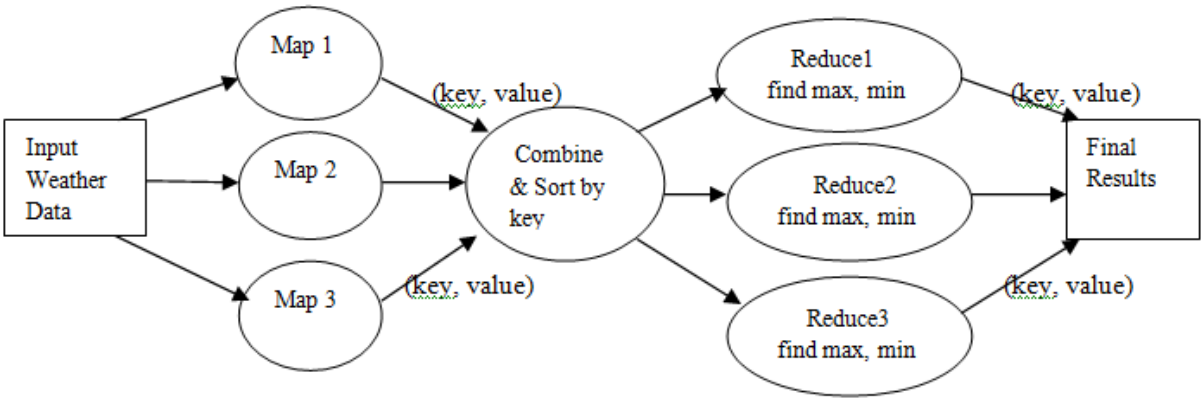

Figure 1: Map-Reduce Model

4.4 Map-Reduce Model

The Map-Reduce process which is executed for minimum and maximum operation on the National climatic data is given as follows :

1. Weather data set files is inserted into sequence files on head node of Hadoop Distributed File System.
2. The sequence files which are loaded into Hadoop file system with replica factor three.
3. The job provides Map-Reduce operation which is submitted to head node to run. The head node schedules with job tracker and on cluster jobs which is to be run. Hadoop distributes mappers to all data nodes which contains data to analyze.
4. For reading the input format reader opens up each sequence file which passes all the (key, value) pairs to map function on each node.
5. The mapper determines, if key matches the criteria for query. If mapper keeps (key, value) pairs for delivery to the reducer. If not, The pair is discarded. The keys, values in a file are read and also analyze by using the mapper.
6. The (key, value) pairs will match query are to sent to reducer and reducer performing the average function on the sorted pairs to create final (key, value) pair results.
7. In a Hadoop Distributed File System as a sequence file, the final output is stored.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

V. PROPOSED ALGORITHM

The proposed framework for detection of climate change and raw weather station data are shown in Fig.2. A big climatic data is reduced with Hadoop MapReduce framework. Proposed Spatial Cumulative Sum algorithm is used to monitor the day wise changes in the climate from many years. MapReduce algorithm is used to create a table also.[3]

A. Cumulative Sum Method(CUSUM) :

Cumulative Sum method [3] is use to find drastic changes in the mean value of quantity of interest. Here cumulative sum method is used to monitor the changes in the climate. For 'n' data points X1, X2,…, Xn, calculate average by equation as follows:

$$\bar{X} = X1 + X2 + \ldots + Xn$$

The Cumulative Sum value SI is calculated by the equation:

$$SI = Si{-}1 + Xi - \bar{X} \text{ for } i = 1, 2, \ldots, n$$

where consider, S0 = 0

Calculate maximum and minimum, by equation given in below :

$$S\_max = \max i=0, 1,..., n\ Si$$

$$S\_min = \min i=0,1, ..., n\ Si$$

Find S_diff values to detect the changes in cumulative sum value SI, by equation as :

$$S\_diff = S\_max - S\_min$$

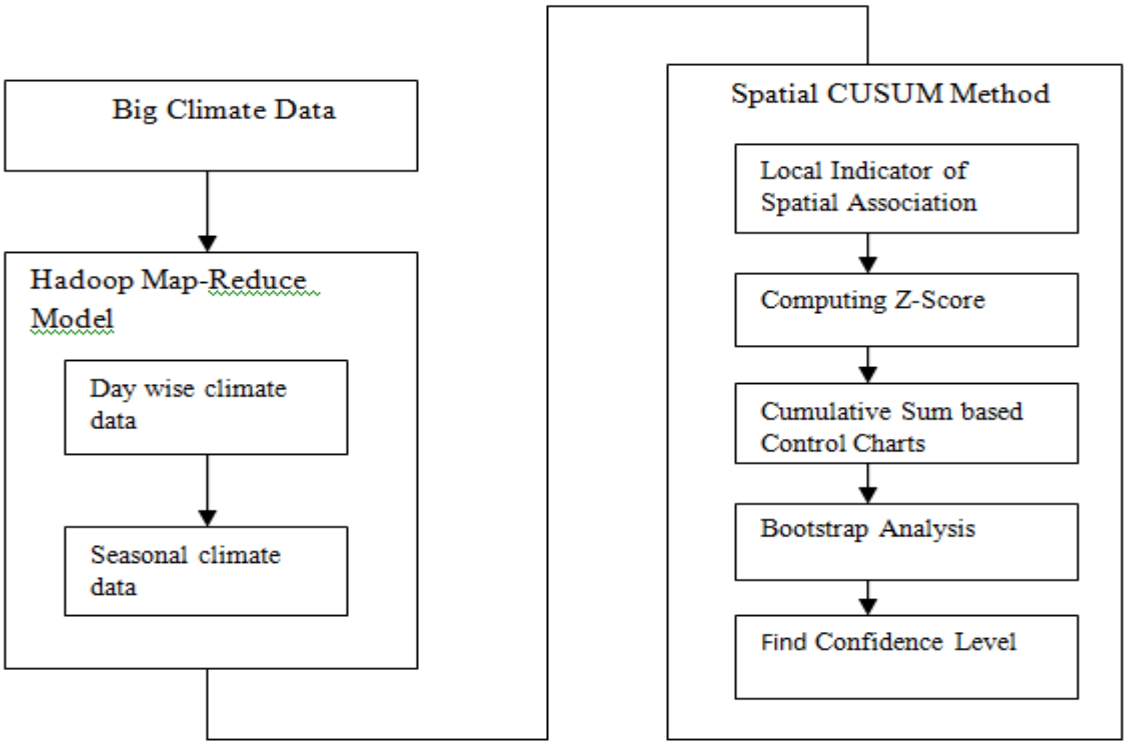

Figure 2: Proposed Framework for Climate Change Detection

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

B. Bootstrap Analysis Method :

This method is used to find change in the average. To calculate significant changes the bootstrap analysis technique is to be used and it is calculated by randomly reordering the original values. A single bootstrap consists of following steps:

1. Re-order original 'n' values X_o 1, X_o 2, …, X_o n without replacement method using sampling.
2. Calculate the cumulative sum values S_o 0, S_o1, …, S_o n.
3. Calculate maximum, minimum and difference that is S_o_max, S_o_min, and S_o_diff.
4. Identify bootstrap difference which is S_o_diff is less than original difference that is S_diff.

Number of bootstrap analyses is performed to determine confidence level of the changes. Confidence level can be calculated by,

Confidence Level (CI) = 100 ∗ X / N %

Where, N = Number of bootstrap samples performed

X = Number of bootstraps for which S_o diff < S diff

C. Spatial Auto-Correlation :

Spatial Autocorrelation which shows correlation of variables. The local Moran's I is used to find correlation between parameters of climate. The Local Moran's I spatial auto correlation which can be calculated by following equation :

Z score zi = xi − $\bar{x}$ SD

Local Moran's Ii = zi wij zj

Where, n = Number of spatial locations indexed by i and j

xi = Variables

$\bar{x}$ = Mean of x

SD = x − x 2 n

Wij = Spatial standardized weight matrix

VI. EXPERIMENTAL RESULTS

National Climatic Data Centers which contains huge historical weather datasets.The Proposed System which uses dataset of National Climatic Data Centers(NCDC) which provides the parameters as station number, station name, date, country, city, temperature, and due point, speed etc. The weather data files are loaded into Hadoop Distributed File System that is HDFS. Then, files are divided and provides to different mappers. The output of mapper is a set of (key, value) pairs, where key is consists of station number, date and value is contains the parameter temperature. After that the output of mappers is combined and sort by using a key. Finally, results which is sent to the reducers. For each reducer calculate maximum(daily, monthly, yearly), and minimum(daily, monthly, yearly), for the parameter temperature in different stations. Then reducers store the final results in HDFS. Due to practical limit, the analysis is executed in Hadoop stand alone mode. Figure 1 shows Map-Reduce framework execution. Figures (3, 4 and 5) presents experimental results which shows daily maximum and minimum temperature change over the years. Also it shows daily maximum and minimum temperature change for a month.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

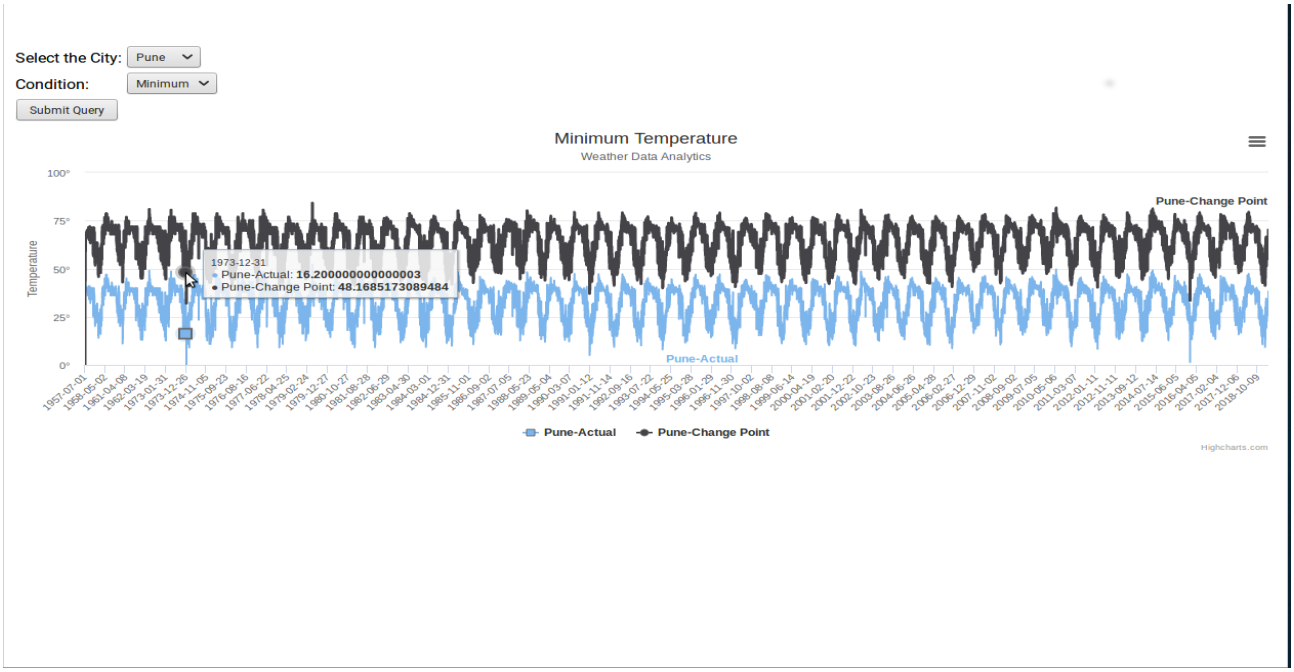

Figure 3 : Results of daily minimum temperature change over the years

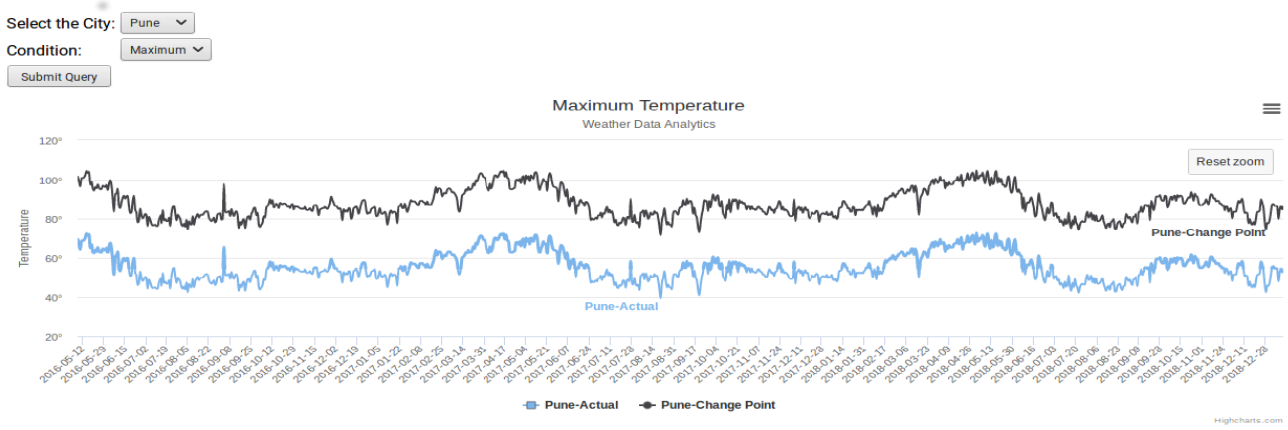

Figure 4: Result of daily maximum temperature change over the years

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

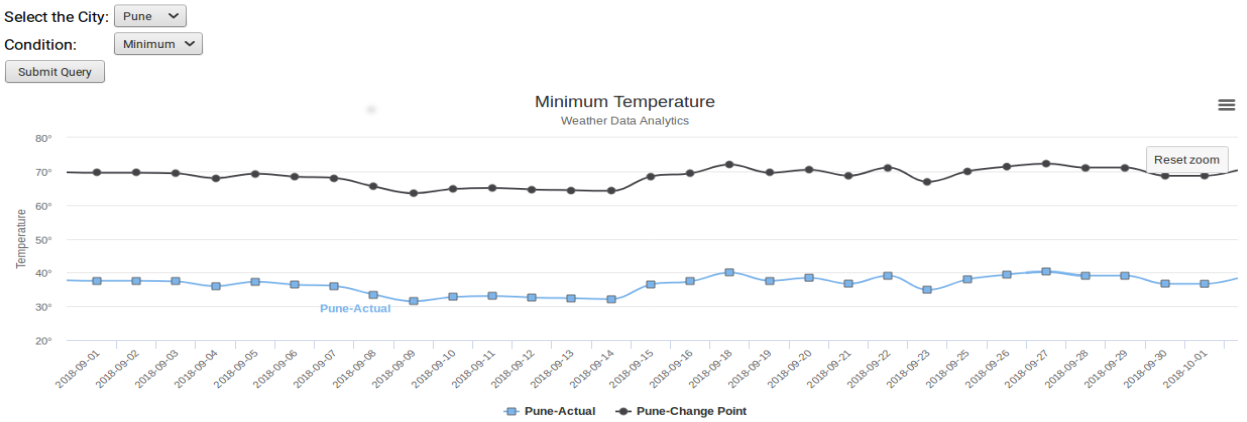

Figure 5: Result of daily minimum temperature change for a month

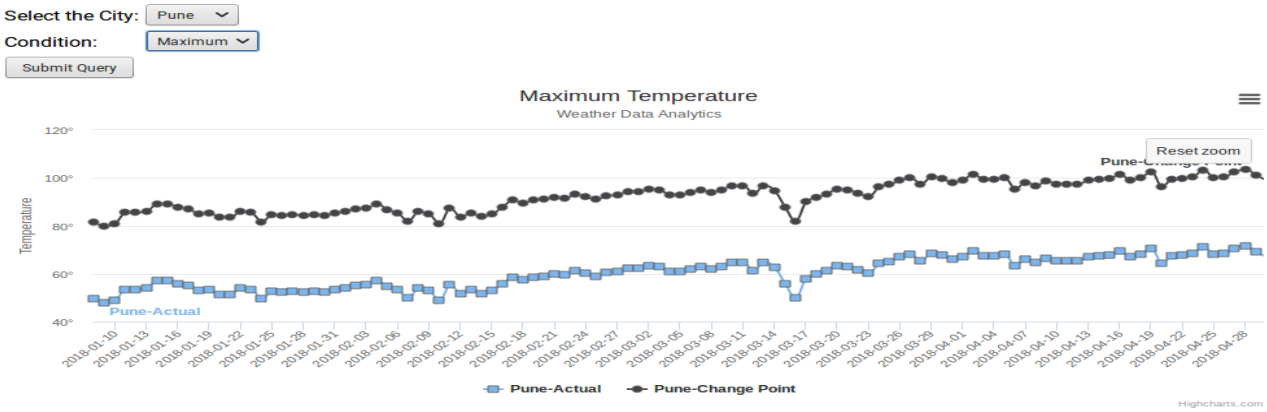

Figure 6: Result of daily maximum temperature change for a month

VII. CONCLUSION AND FUTURE WORK

The traditional or existing systems which processes millions of records is a time consuming process. So here Hadoop with Map-reduce, weather data can be analyzed effectively. Map reduce is a framework which is parallel and distributed systems across large dataset. Using Map-Reduce with Hadoop helps in removing scalability issues. This technology which is used to find huge datasets has the potential for significant enhancement to analyze weather. The major advantage of Map-Reduce with Hadoop framework speeds up the processing of data, where the volume of data is increasing every day. we intend to use the climate change values to predict the diseases is the future work of the proposed system.

REFERENCES

[1] Gunasekaran Manogaran, Daphne Lopez, "Spatial cumulative sum algorithm with big data analytics for climate change detection", Computer and Electrical Engineering, 2017 Elsevier ltd.
[2] Priyanka Chouksey, Abhishek Singh Chauhan, "A Review of Weather Data Analytics using Big Data" International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 6, Issue 1,PP 130-152,January 2017

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

[3] Ms. Himanshi Jain, Ms. Raksha Jain, "Big Data in Weather forecasting: Applications and Challenges", International Conference On Big Data Analytics and computational Intelligence (ICBDACI),2017
[4] Khalid Adam Ismail, Mazlina Abdul Majid,"Big Data Prediction Framework for Weather Temperature Based on MapReduce Algorithm" IEEE Conference on Open Systems (ICOS), Vol.7, Issue 1,PP 244-304,October 2016
[5] Riyaz P.A., Surekha Mariam Varghese, "Leveraging Map Reduce With Hadoop for Weather Data Analytics" IOSR Journal of Computer Engineering (IOSR-JCE), Volume 17, Issue 3, Ver. II (May – Jun. 2015), PP. 06-12
[6] Meena Agrawal, A.K.Pandey, C. P Agrawal, "A Hadoop based Weather Prediction Model for Classification of Weather Data",IEEE 2017
[7] Basvanth Reddy and Prof B. A. Patil, "Weather Prediction on Big Data Using Hadoop Map Reduce Technique", IJARCCE, ISSN: 2278-1021 Volume-05, Issue-06, Page No (643-647), June, 2016
[8] Ye Ding, Yanhua Li, "Detecting and Analyzing Urban Regions with High Impact of Weather Change on Transport", IEEE Transactions on Big Data -2016
[9] E Sreehari, J. Velmurugan and Dr. M. Venkatesan, "A Survey Paper on Climate Changes Prediction Using Data Mining", IJARCCE, ISSN: 2278-1021 Volume-05, Issue-02, Page No (294296), February, 2016
[10] Mr. Sunil Navadia, "Weather Prediction: A novel approach for measuring and analyzing weather data", IEEE International conference on I-SMAC-2017
[11] K.Anusha, K.Usha Rani, "Big Data Techniques for Efficient Storage and Processing of Weather Data", International Journal for Research in Applied Science&EngineeringTechnology(IJRASET), Volume 5 Issue VII, July 2017
[12] Veershetty Dagade, Mahesh Lagali,Supriya Avadhani and Priya Kalekar, "Big Data Weather Analytics Using Hadoop", International Journal of Emerging Technology in Computer Science & Electronics, Volume-14, Issue-02, April 2015
[13] Dr. Doreswamy, Ibrahim Gad, "BIG DATA TECHNIQUES: HADOOP AND MAP REDUCE FOR WEATHER FORECASTING", International Journal of Latest Trends in Engineering and Technology Issue SACAIM 2016, pp. 194-199