

INVESTIGATION OF BIG DATA ARCHITECTURE TO INTEGRATE CLIMATE CHANGE TO INFECTIOUS DISEASE

Poonam Sharma¹ Pawan Kumar²

¹Research Scholar, NIILM University, Kaithal

²Assoc. Professor, NIILM University, Kaithal

Abstract:

In this paper we have thoroughly Investigated and Found Some rare Results with the Help of Cloudera, HDFS, Hadoop, Computational Method a model has been developed for the calculation of Big Data Architecture to Integrate climate change to infectious disease Dengue, Malaria and Chikengunia. By Considering Proposed Big Data Architecture and some experimental data has been Compared with Present Study four type of Model Technique has been discussed in Details.

Keywords: Cloudera, HDFS.

1.Introduction:

Descriptive analytics is the kind of analysis that is performed to describe an organization's current circumstances. The data used in this instance can include customer feedback, sales numbers, website traffic--essentially any data that is a record of past events that can be used to analyze business up to the present. Predictive analytics uses the same type of data, and sometimes the descriptive outcomes, to predict what have happen given the current circumstances. Businesses often employ machine learning and various forms of predictive modeling to make predictions. Think of predictive analytics as what have happen if current organizational practices and habits remain the same.Prescriptive analytics is less fortune teller and more medical doctor. Instead of simply predicting what have happen, prescriptive analysis tweaks certain variables to achieve the best possible outcome, and then prescribes that course of action. Diagnostic: Diagnostic study so far considered in the pPresent Study the main part of study is model equation is Specially Descriptive it is fundamental types and Minimum Big Data Diagnostics Current Value change in the apply in source this analysis can be used for countries. The No. of Light Followers, Shares, Fans past an facebook.

2.Tools and Techniques

In this paper we have also used some tools and techniques such as CDH (Cloud era) Hadoop Hive Map reduce and R etc. All these tools have explained below:

1. CDH (Cloudera): Hadoop is an ecosystem of open source components that fundamentally changes the way enterprises store, process, and analyze data. ... CDH, Cloudera's open source platform, is the most popular distribution of Hadoop and related projects in the world Hadoop is an ecosystem of open source components that fundamentally changes the way enterprises store, process, and analyze data.
2. Hadoop: Hadoop does distributed processing for huge data sets across the cluster of commodity servers and works on multiple machines simultaneously. To process any data, the client submits data and program to Hadoop. HDFS stores the data while MapReduce process the data and Yarn divide the tasks. Hadoop is an Apache open source framework written in java that allows distributed processing of large datasets across clusters of computers using simple programming models.
3. HDFS: The way HDFS works is by having a main « Name Node » and multiple « data nodes » on a commodity hardware cluster. All the nodes are usually organized within the same physical rack in the data center. Data is then broken down into separate « blocks » that are distributed among the various data nodes for storage. HDFS exposes a file system namespace and allows user data to be stored in files

4.Spatial Commutative Algorithm

In this study four “Spatial Commutative Algorithm” I have study of the “Big Data” is defined by volume, velocity, and variety of data. Big data is very complex to process by traditional data processing techniques and tools. Nowadays, data generation sources like telescopes, satellite, sensor networks, social networks, wearable devices, mobile devices, streaming machines and high throughput instruments are continuously generating a large volume of data. Recently, big data analytics has been applied in various domains, such as healthcare, business process, scientific research, natural resource management, share marketing, social networking, community administration and climate modeling. Climate data is observed from various advanced sensor technologies and is used to represent the seasonal changes. Weather data collected from different climate laboratory and advanced computing technologies are used to give valuable information to the

world. Meteorological data is most often used to predict the weather and other climate-related phenomena. In addition, climate data is also used for various purposes that lead to a significant development in weather forecasting, rocket launching, and public health. World Weather Records (WWR) is originally developed by the world climate organization in 2009. The primary goal of World Weather Records (WWR) is to maintain the huge size of records such as monthly temperature, wind speed, rainfall, precipitation and pressure data that are collected from thousands of weather stations around the globe. In recent years, the number of stations is increased noticeably. Especially, many weather stations have been collecting metrological data in a continuous manner. World Metrological Organization (WMO) has been collecting the day-to-day weather data in the form of digital since 1920. Metrological data collected from WMO are digitally published by nine issues. It include 2010-2011, 2011-2012, 2012-2013, 2013-2014, 2014-2015, 2015-2016, 2016-2017, 2017-2018, 2018-2019 and 2019-2020. World Meteorological Organization (WMO) Commission for Climatology (CCI) maintains NoSQL based database to store the massive amount of data related to world weather ex- tremes and various abnormal conditions. The Commission for Climatology (CCI) published the world weather extreme data that are available online to users. The essential role of this database is to maintain the huge extreme of various climate parameters respect to space and time. The database consists of following climate parameters it includes maximum/minimum observed temperature, wind speed and most precipitation on earth respect to space and time. In addition, the CGI database also maintains the most destructive earthquake, hurricanes, floods, storms, and tornadoes.

5.Result and Discussion

This paper uses Cumulative sum control charts to find the differences of each sample value from the target value. CUSUM control charts are also called as time-weighted control chart used to monitor the small shifts in the mean of a process. The traditional CUSUM control chart is used to monitor the changes in rainfall, precipitation, maximum temperature, minimum temperature, humidity, wind speed and solar and the results are shown in figures respectively. The cumulative sum is not the cumulative sum of the values. Instead, it is the cumulative sum of differences between the values and the average (target value).

Because the mean is subtracted from each value, the cumulative sum also ends at zero. A CUSUM control chart is used to identify the cumulative sums (CUSUMs) of the deviations of each sample value from the target value. Moreover, small drifting in the mean value will lead to steadily increasing or decreasing cumulative deviation values. However, interpreting the changes identified by the control charts can still be difficult. In order to analyze these changes, a change- point analysis can be performed. The CUSUMs are also named as change points. CUSUM control chart for identifying the CUSUMs of the deviations of rainfall from the target value.

Visual Layer

Long-time period evaluation of the discovered temperature sorts of Haryana the utilization of temperatures recorded on a day time scales indicates that there could also be no substantial fashion in excessive and occasional temperatures displaying an boom of about 1.00C to 1.20C. Haryana emits approximately 2% of worldwide warming gases (GHG) .This layer incorporates topical detection and occasion detection tools. Data Visualization Layer plays a significant role to predict and Represent the results in Pictorial Form. We have many visualization Tools such as kafka, Orange, ArCGis, R, Python etc.

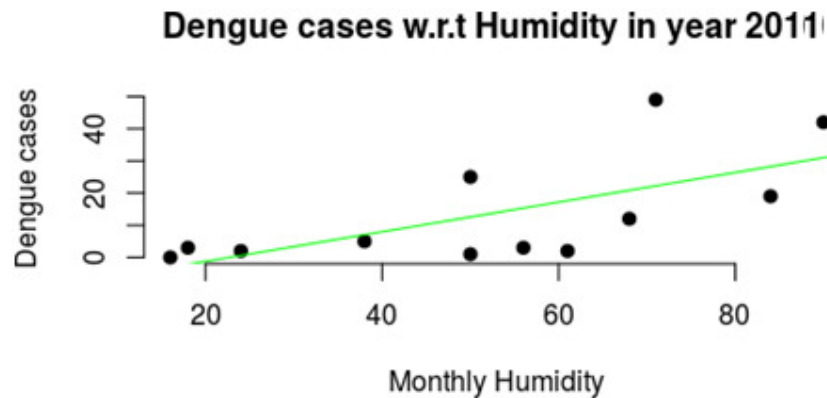


Figure 1: Dengue Case w.r.t Humidity in year 2011

The most common visualization technique is using to transform the information into a , bar graph or pie chart. Line Graph, Scatter Graph and Box Plot etc In my Research I used R Technique to Represent the Results. I adjusted the reversal impact the usage of plot function () in R.

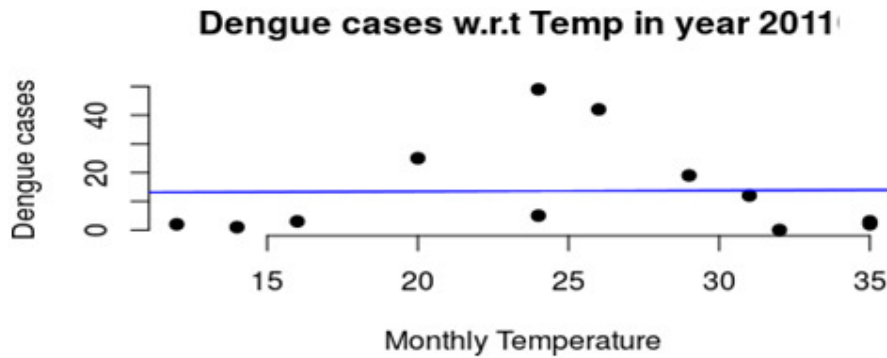
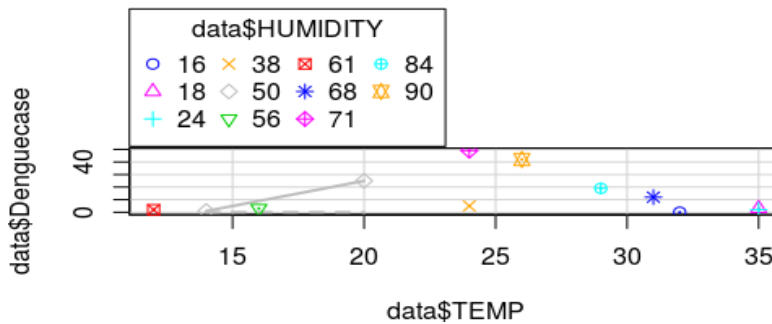


Figure 2: Dengue case w.r.t Temperature in year 2011



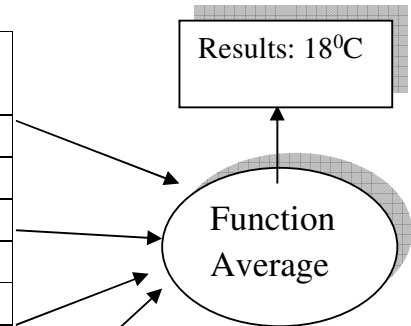
(Figure 3: Results)

In my Research, we are used some parameters such as latitude, longitude, temperature, humidity, rain, cloud, and wind. By prediction, I found temperature, Humid and Rain are important factors that Impact on Infectious disease. Above Results Shows that the same as I have written above and this Results for Dengue Cases by the impact of Climate change.

Forty percentage of the world's population, approximately three million citizenry , stay in regions susceptible to dengue. Dengue is often the most purpose of contamination in risky regions. Each 12 months, a couple of four hundred million citizenry be afflicted by dengue. a few hundred million citizenry fall ill from the infection, and 22,000 die from dengue. Most citizenry with dementia recover with out ongoing problems. the prospect of dying amongst people with excessive dengue is 0.8% to 2.5%, and with ok remedy that's much but 1%.

Table No. 1 Results of Climate 2011-2020

Longitude	Latitude	Year	Max. Temperature
76.0856 ⁰ E	29.0588 ⁰ N	2011	17 ⁰ C
76.0856 ⁰ E	29.0588 ⁰ N	2012	17 ⁰ C
76.0856 ⁰ E	29.0588 ⁰ N	2013	17 ⁰ C
76.0856 ⁰ E	29.0588 ⁰ N	2014	16 ⁰ C
76.0856 ⁰ E	29.0588 ⁰ N	2015	18 ⁰ C
76.0856 ⁰ E	29.0588 ⁰ N	2016	19 ⁰ C
76.0856 ⁰ E	29.0588 ⁰ N	2017	19 ⁰ C
76.0856 ⁰ E	29.0588 ⁰ N	2018	17 ⁰ C
76.0856 ⁰ E	29.0588 ⁰ N	2019	17 ⁰ C
76.0856 ⁰ E	29.0588 ⁰ N	2020	19 ⁰ C



HoIver, people with very low blood strain could have a mortality price of the maximum amount as 26%. After Gurgaon, forty seven instances from Panchkula district of Panchkula and different instances from different components of the province had been pronounced, a fitness branch spokesman stated today.

Conclusion

As the Popularity of Cloud in Development of Big Data is Growing Day by Day Number of Clod users are increasing, Citify data is being Stored in Clud and Cloud server Provider for my Research Work Related with Dengue Viruses, Malaria and Chikengunia Disease deals with Responsibility of Handling Data and Information the most critical effects for users in Data and that Data has to been safe Hence, what due to Increasing Security Problem Nowadays the with Challenges is to Search and Implement new ways to Secure critical Data of different Organization People. Better Method are Coming up for development of Big Data Architecture to Integrate of effects of Climate Change on Infectious Disease.

Here I Discussed the Privacy maintained issued of development of Big Data Architecture to Integrate of effects of Climate Change on Infectious Disease as well as different Risk in this Research work I analyzed here different security features of three Disease Dengue, Malaria and Chinkengunia

Refernces

1. Aggarwal, C. C. and Zhai, C. (2012), *Mining text data*, Springer Science and Business Media.
2. Al-kahtani, M. S. and Karim, L. (2017), 'An efficient distributed algorithm for big data processing', *Arabian Journal for Science and Engineering* pp. 1–9.
3. Anuradha, M., Dandekar, R. and Banoo, S. (2014), 'Laboratory diagnosis and incidence of dengue virus infection: A hospital based study, perambalur', *International Journal of Biomedical Research* **5**(3), 207–210.
4. Aydin, G., Hallac, I. R. and Karakus, B. (2015), 'Architecture and implementation of a scalable sensor data storage and analysis system using cloud computing and big data technologies', *Journal of Sensors* **2015**.
5. Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A. and Escobar, G. (2014), 'Big data in health care: using analytics to identify and manage high-risk and high-cost patients', *Health Affairs* **33**(7), 1123–1131.
6. Ben-Haim, Y. and Tom-Tov, E. (2010), 'A streaming parallel decision tree algorithm', *Journal of Machine Learning Research* **11**(Feb), 849–872.
7. Faghmous, J. H. and Kumar, V. (2014), 'A big data guide to understanding climate change: The case for theory-guided data science', *Big data* **2**(3), 155–163.
8. Fatima, M. and Pasha, M. (2017), 'Survey of machine learning algorithms for disease diagnostic', *Journal of Intelligent Learning Systems and Applications* **9**(01), 1.
9. Ferreira Cordeiro, R. L., Traina Junior, C., Machado Traina, A. J., López, J., Kang, U. and Faloutsos, C. (2011), Clustering very large multi-dimensional datasets with mapreduce, in 'Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining', ACM, pp. 690–698.
10. Fiore, S., Mancini, M., Elia, D., Nassisi, P., Brasileiro, F. V. and Blanquer, I. (2015), Big data analytics for climate change and biodiversity in the eubrazilcc federated cloud infrastructure, in 'Proceedings of the 12th ACM International Conference on Computing Frontiers', ACM, p. 52. for Disease Control, C. (2016), 'Laboratory guidance and diagnostic testing', *Dengue. CDC*.