Research For Efficient Water Resource Visualization Using Big Data Analytics

[1]Bharatwaj Krishnan, [2]Ajinkya Nimhan, [3]Akshay Atole, [4]Shravan Vikhe
[1,2,3,4]Department of Computer Engineering,
[1]Padmabhooshan Vasantdada Patil Institute of Technology, Pune, India

Abstract : Water scarcity is a serious problem that the state of Maharashtra is facing today. Scarcity of water leads to long term drought when addressed incorrectly. Thus a strong demand for building a real-time system to support water resources analysis, modeling and prediction of drought arises. Models and approaches currently existing lack the desirable accuracy and speed needed in predicting and analysing drought. Here, this paper proposes a big data analytics based approach to support drought analysis and prediction based on widely diverse data sets including climate sensors, satellite and weather data, drought conditions and water usage reports. It also reports an implemented system supporting big data analytics for drought. It also uses an easy to understand Drought Index and presents data based on existing models and approaches, as well as proposed models.

***IndexTerms* - Data Mining, Knowledge Engineering, Cloud Computing, Data Analytics, Visualization.**

I. Introduction

Drought conditions prevail in over 29,000 villages in Maharashtra. Statistics say 3,228 farmers committed suicide in Maharashtra in 2015. Aurangabad district in Marathwada, which is witnessing the worst drought of the century, has only one percent water left of its 2.17 billion cubic meter capacity. Water trains have been sent to the worst affected areas, but the government is said to be implementing long-term projects as well. This is only a temporary solution. If water could be managed as well as electricity is, drought can be fought well. The website of The Ministry of Power, India contains real time information regarding the consumption, storage and price of electricity throughout the country. This helps in the proper management and distribution of electricity. A similar system for water would be very beneficial and can help in foreseeing drought conditions and acting accordingly.

II. PROPOSED SYSTEM

In this approach we present a mechanism to forecast and visualize water availability with respect to population and usage. This approach will be fast and accurate when compared to the current system.

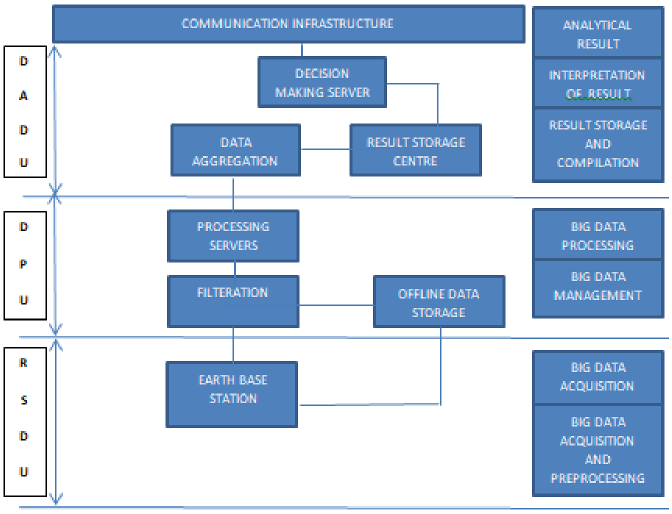
III. ALGORITHM

The algorithm used in this proposed system is **KNN**. K nearest neighbors (KNN) is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition already in the beginning of the '70s as a non-parametric technique.

IV. SYSTEM OVERVIEW

The three main modules of the proposed architecture are-

1. Remote Sensing Big Data Acquisition Unit (RSDU)
2. Data Processing Unit (DPU)
3. Data Analysis and Decision Unit (DADU).



4.1 Remote Sensing Big Data Acquisition Unit (RSDU)

Remote sensing encourages the enlargement of earth beacon system as cost-effective identical data acquisition method to satisfy specific computational requirements. Therefore, the need for parallel processing of the massive volume of data was required, which could efficiently analyse the Big Data. For that reason, the proposed module is introduced in the remote sensing Big Data architecture that gathers the data from various reservoirs around the globe. It is possible that the received raw data are distorted by scattering and absorption by various parameters. We assume that the intermediate can correct the erroneous data. However, to make the raw data into the data format. Some relational data pre-processing techniques are data integration, data cleaning, and redundancy elimination. After pre-processing phase, the collected data are transmitted to a processing server. The data must be corrected in different methods to remove distortions caused due to the motion of the platform relative to the reservoirs. We divided the data processing procedure into two steps, such as real-time Big Data processing and offline Big Data processing. In the case of offline data processing, the Earth Base Station transmits the data to the data centre for storage. This data is then used for future analyses. However, in real-time data processing, the data are directly transmitted to the filtration and load balancer server (FLBS), since storing of incoming real-time data degrades the performance of real-time processing.[1]

4.2 Data Processing Unit (DPU)

In data processing unit (DPU), the filtration and load balancer server have two basic responsibilities, such as filtration of data and load balancing of processing power. Filtration identifies the useful data for analysis since it only allows useful information, whereas the rest of the data are blocked and are discarded. Hence, it results in enhancing the performance of the whole proposed system. Apparently, the load-balancing part of the server provides the facility of dividing the whole filtered data into parts and assign them to various processing servers. Each processing server makes statistical calculations, any measurements, and performs other mathematical or logical tasks to generate intermediate results against each segment of data. Since these servers perform tasks independently and in parallel, the performance proposed system is dramatically enhanced, and the results against each segment are generated in real time. The results generated by each server are then sent to the aggregation server for compilation, organization, and stored for further processing.[1]

4.3 Data Analysis and Decision Unit (DADU)

DADU contains three major portions, such as aggregation and compilation server, results from storage server(s), and decision-making server. When the results are ready for compilation, the processing servers in DPU send the partial results to the aggregation and compilation server, since the aggregated results are not in organized and compiled form. Therefore, there is a need to aggregate the related results and organized them into a proper form for further processing and to store them. In the proposed architecture, aggregation and compilation server is supported by various algorithms that compile, organize, store and transmit the results. The decision part of the architecture is significant since any small error in decision-making can degrade the efficiency of the whole analysis. DADU finally displays or broadcasts the decisions, so that any application can utilize those decisions in real time to make their development. The applications can be any business software, general purpose community software, or other social networks that need those findings (i.e., decision-making).[1]

V. EXPERIMENTAL SETUP

As a case study water reservoir, tap water, population and water usage data of the state of Maharashtra has been analyzed. The state of Maharashtra has been divided into five major regions. The data are collected district wise. They are refined and grouped to form the base for data analytics. After grouping the objects, C# has been used to classify and establish the relationship between the data feed and the objects. On this classified data, a drought index has been devised which visually depicts the data in all the Regions (5 major Regions). ASP.NET and HTML5 is used as front-end to design. The Model View Controller (MVC) Approach has been adopted for implementation. The dataset is strictly normalized and is stored and accessed via MS-SQL Server 2015.

VI. RESULTS

6.1 Screenshot

The above image shows a screenshot of the home page after execution. It also shows the total number of districts, their population and full reservoir capacity of the 5 regions. The current water level and capacity can be updated or changed by signing in with appropriate credentials. The data can be updated as often as required.

VI. FUTURE SCOPE

The developed system will show the obtained result with a high level of accuracy. It is a significant feature of the system, because a wrong or less accurate result can cause water scarcity and lead the authorities into making wrong decisions. The system can analyze and present the obtained results. The program will be ready for use by the administrative authorities.

VII. CONCLUSION

This system provides a real time big data system with user friendly interface which supports and monitors environmental water eco-system. The drought computing and prediction model is based on historic and current data, comprehensive impact factors with data integration and model integration. The application supports prediction based on comprehensive and integrated big datasets such as reservoir, climate change, water usage, population and temperature data. This system provides a realistic visualization of drought.

References

[1] Akshay Atole, Bharatwaj Krishnan, Ajinkya Nimhan, Shravan Vikhe "Efficient Water Resource Visualization using Big Data Analytics", IJSRD - International Journal for Scientific Research & Development| Vol. 6, Issue 02, 2018 | ISSN (online): 2321-0613

[2] Pengcheng Zang, Jerry Gao, A.G.Thomas, "On Building a Big Data Analysis System for California Drought", 2017. Hohai University, Nanjing, China.

[3] Simon Fong, Raymond Wong, Athanasios V. Vasilakos "Accelerated PSO Swarm Search Feature Selection for Data Stream Mining Big Data", IEEE TRANSACTIONS ON SERVICES COMPUTING

[4] J. Heinrich and B. Broeksema, "Big Data Visual Analytics with Parallel Coordinates", 2015. CSRIO, Sydney, Australia

[5] E. E. Moreira, C. A. Coelho, A. A. Paulo, et al., "SPIbased drought category prediction using log-linear models", Journal of Hydrology, 2008, 354(14):116-130.

[6] Swain L. Daniel, M. Tsiang, M. Huagen, D. Singh, A. Charland, B. Rajaratnam, and S. N. Differenbaug, "The Extraordinary California Drought of 2013-2014", American Meteorological Society.
[7] R. Seager, M. Hoerling."Causes of California drought", 2015, American Meteorological Society.
[8] J. McWilliams, 2014, "Meat makes the planet thirsty", Retrieved from http://www.nytimes.com
[9] J. S. Oguntoyinbo, "Drought prediction", Climatic Change, 1986, 9(1):79-90.
[10] A. C. Steinemann, "Using Climate Forecasts for Drought Management", Journal of Applied Meteorology & Climatology, 2009, 45(10):1353-1361
[11] W. Fan and A. Bifet, "Mining big data: Current status, and fore-cast to the future," SIGKDD Explorations, vol. 14, no. 2, pp. 1–5.
[12] M. M. Gaber, A. Zaslavsky, and S. Krishnaswamy, "Mining data streams: A review," ACM SIGMOD Rec., vol. 34, no. 2, pp. 18–26.
[13] P. Domingos, and G. Hulten "Mining high-speed data streams," in Proc. 6th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, New York, NY, USA, 2000, pp. 71–80.