International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 01, January-2022

Review Paper Weather Data Management using hadoop in Zimbabwe

Muchaneta Mayepi[1], Mainford Mutandavari[2]
Department of Information Technology, Department of Software Engineering
Harare Institute of Technology University
Harare Institute of Technology, P.O Box BE277, Belvedere Harare

***Abstract*: Meteorological Services Department (ZMSD) in Zimbabwe is experiencing high costs in transmitting daily data due to different formats of data from its diverse models of automatic weather stations to the centralized database at Head Office (HO) Belvedere in Harare. The current configuration makes MSD less competitive and efficient its weather data management posture. However, surveys indicate that emerging technologies of big data analytics such as Hadoop MapReduce can be extended into weather data management offering multiplicity of functions that includes: prediction, management of transmission of data and also managing the range of weather elements to monitor at anytime and anywhere. This paper reviews on how weather data is transmitted, processed and stored.**

Keywords: Hadoop, MapReduce, Meteorological Services Department (MSD), automatic weather station, Application Programme Interface (API).

I. INTRODUCTION

This is a review of related work conducted by diverse range of authors and the associated outcomes and recommendations are accounted for. This is meant to provide insights into the relevance of Hadoop MapReduce tool in weather data management be it in analytics or in data processing capabilities. Surveyed literature indicates a bias of use of MapReduce for weather data analytics than for other data management challenges such as attempts to manage the costs of transferring data from each automatic weather station to the central database [10], [11], [13]. Most of these authors concentrated on weather analysis capabilities without focusing attention of challenges in transmitting large data files from weather stations to the central database.

II. OVERVIEW OF HADOOP

Hadoop is a software framework for storing and analysis of large volumes of data during processing. It splits an input file into fixed size pieces of data (input splits) and contains many records in the form of key-value pairs. Data is divided or splitted into smaller chunks to run map functions on every key value pair generated. Hadoop runs the MapReduce tasks on the same worker nodes where the data is located, the reducer task shuffles, grouped key- value pairs are downloaded to the local machine where the reducer is running and produces new key value pairs as an output and that's a new record has been created. The key-value pairs are sorted into a larger data list and groups the equivalent keys so that the values can be integrated easily in the reducer task [3].

III. RELATED WORK

Mohammed acknowledges that modern weather systems have numerous automatic weather stations that provide wide variety of parameters of the weather such as temperature, wind speed, humidity, and pressure among others. Mohammed sought to use the MapReduce and Hadoop techniques to analyse historical weather data of a region in Malaysia. The author did setup an experiment in a physical cluster environment in which three desktop computers were deployed. Linux Ubuntu 14.04 was used for this experiment where one computer ran the Name Node and Resource Manager while two computers ran the DataNode and DataManager. Each of the computers had a minimum of 4GB of memory, 1TB of disk space running on Core i7 processor. Furthermore, Hadoop 2.7.1 version was used. Existing weather datasets files showing different parameters of weather was input in HDFS. Once the data was stored in HDFS, the next step was to use the Map function to extract something of interest and then create the output. The output was then optimised simply by organising the output. The organised output is sent to the reduce function to compute a set of results from which the final output was produced. The results indicated that instead of using traditional systems that consumes much time, Hadoop/MapReduce technique is an appropriate framework to process huge distributed amount of weather data. The research discovered that MapReduce is elastic and scalable to accommodate the growing amount of weather data [10].

Riyaz [11] proposed the collection, processing and storing of huge amounts of weather data for accurate prediction of the weather by way of leveraging on the Hadoop MapReduce technologies. The author successfully implemented the MapReduce on Hadoop to analyse temperature data. The author collected input datasets from the various automatic weather stations and made use of the Input Split method of MapReduce created mapper jobs in the Mapper process, created lists was shuffled. Basically, the MapReduce algorithm managed to average temperature using the mapper function and reduce these in the results. The study managed to find the average, minimum and maximum temperatures necessary in weather prediction in the reducing phase and stored such output in cloud database. This study concluded that Hadoop MapReduce technology can be utilized to optimize weather prediction as more and more data is building up for many meteorological services around the world.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 01, January-2022

In another study [2] [3]. The author attempted to describe how data is recovered, retrieved and stored using cloud computing. General Packet Radio Service (GPRS) was used as a method for pushing files from Automatic Weather Stations to the database for easy monitoring and weather prediction. Meteorological weather stations runs a Conventional Weather Station where data is collected basing on the equipment on the site by a Meteorological Observer, records the information into the register and send it to the Meteorological Head Office for data capturing by the Post Office. Automatic Weather Station (AWS) that consists of sensors that records and store data on the data logger and internal memory. This data can be collected directly using a laptop in case the GPRS transceiver is down.

Furthermore, AGNET system was developed in order to understand how automated weather data network for Agriculture can be optimized. The author collected hourly weather data from a remote agriculture weather station using a system developed in Nebraska that interrogate, the incoming data and send it to the mainframe computer. Data availability to users is through the AGNET system which is said to have lower costs, simple site requirements and maintenance [4].

Another author developed an embedded weather monitoring system to record and transmit weather data to the cloud database using sensors. This cloud-based system have many advantages such as data access to users from anywhere, data replication to reliable database as to allow online user easy access of weather data [14]. This system provided an increased data collection capacity from distributed weather stations [12] .

LoRa wireless infrastructure is a prototype for a weather station developed as an alternative technology to existing wireless connectivity modules that are already in use. LoRa-based weather station has a gateway and two nodes that reads the values from sensors which are then displayed on the interface Things Speak and a suggestion the use LoRa of when designing sensor networks that would require several nodes [6].

CLIMSOFT is a climate database management tool developed for the improvement of data management in the Southern African countries. This tool transmit data from weather stations (Manual and Automatic) into the main database. Automatic Weather Station is probed to check for new data after every two hours and transmitted to the main database[8].

IV. SUMMARY OF THE REVIEW

The above discourse indicates that adoption of newer weather technologies only took place in the 21st century. Majority of the developments in weather data management systems are driven by developments in computing and Internet platforms. The emergence of automation in the electronic equipment and computing capabilities enabled the development of automatic weather stations that automatically senses, store and even transmit data over networks. Data is being transmitted at greater speed (velocity) from different sources (variety), thereby increasing in volume. Therefore these developments coupled by advancement in analytics capabilities catalysed the development of technologies that can efficiently process, store and transmit data such as the Hadoop MapReduce tools [3] [8]. The current study therefore sought to use MapReduce to process the weather data and store the data in a cloud environment to not only improve efficiency in processing, and transmission of data but to do so in scalable cloud environment that allows users to access such critical information anywhere at any time.

V. RESEARCH GAP

Related work surveyed indicates a bias of most studies to concentrate on data analytics capabilities using MapReduce [10] . These past studies did not pay attention to the broader capabilities of MapReduce in terms of offering clear opportunity to reduce file sizes from huge weather data files to small files. Transmitting of huge files in an era of data bursts and huge bandwidth may operationally be costly for weather institutions who depend on dispersed weather stations and weather data sources. Ability to manage the transmission and storage of weather data is thus a significant investment opportunity for weather institutions that want to adapt to emerging technologies. This study is thus focused on using Hadoop/MapReduce to optimise on data transfer from weather stations to the Head Quarters of MSD.

In addition, past studies used weather datasets from remote environments outside Zimbabwe. Attempting to adopt previous researches' prescriptions may thus be out of context. It was therefore pertinent to develop a solution system that is responsive to the Zimbabwean environment [2].

One of the important gaps that was hardly explored by most of the previous studies is the integration of MapReduce with cloud-based database system. Mohammed and Posada[10] illustrate Map Reduce solutions for standalone systems. The relevance of such systems to growing Internet and networked weather computing systems is unclear. This study is novel in integrating and embedding MapReduce functionalities within a cloud-based database system. An architecture of this nature provides not only scalability of the database and efficiency in processing of the data but also improves weather data visibility and accessibility anywhere and anytime.

VI. CONCLUSION

The above literature indicates that most modern weather data management systems make use of big data analytics and artificial intelligence. However, the trending technologies seem to exclude the data transmission part of the data management chain. Challenges in transmission speeds and capacity seem to be marginalised in most of the reviewed studies, thus provides some pertinent research

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 01, January-2022

gaps identified in the surveyed literature. Integration of the cloud storage and MapReduce, exploration of the same within the context of developing economies, and understanding are the missing links in most researches. Therefore, this study sought to develop an API and use of Hadoop to optimise weather data transfer within a cloud computing setup.

ACKNOWLEDGEMENTS

Firstly, I would like to thank the Lord Almighty God for guiding me throughout this project. My heartfelt gratitude goes to my supervisors Mr Mutandavari, Dr K. Zvarevashe, and Dr.E. Tarambiwa and my friends who gave me academic guidance and inspiration during the conduct of this research.

REFERENCES

[1] K. D. Foote, "A brief history of data management," Dataversity , New York , 2020.

[2] National Aeronautics and Space Administration , "Weather forecasting through ages," National Aeronautics and Space Administration (NASA), New York , 2018.

[3] P. A. Riyaz and S. M. Varghese, "Leveraging MapReduce with Hadoop for weather data analytics," *Journal of computer engineering* , vol. 17, no. 3, pp. 6-12, 2015.

[4] G. Kenneth, I. Hubbard, J. Norman and A. Rosenberg, "Automated weather data network for agriculture," National Aeronautics and Space Agency (NASA), New York, 2015.

[5] D. Stuber, A. Mhanda and C. Lefebvre, "Climate Data Management Systems: Status of implementation in developing countries," *Climate Research,* vol. 3, no. 1, 2011.

[6] E. Murdyantoro, R. Setiawan and I. Rosyadi, "Prototype weather station uses LoRa wireless," *Journal of Physics,* vol. 3, no. 2, p. 9, 2019.

[7] E. E. Filho, A. Albuquerque and M. Nagano, "Identifying SME mortality factors in the life cycle stages: an empirical approach of relevant factors for small business owner-managers in Brazil," *Journal of global entrepreneurship research,* vol. 7, no. 5, 2017.

[8] H. Saini, A. Thakur and S. Ahuja, "Arduino Based Automatic Wireless Weather Station with Remote Graphical Application and Alerts," *2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN),* vol. 3, no. 4, 2016.

[9] K. Lagouvardos, V. Kotroni, I. Bezes and T. Koletsis, "The automatic weather stations NOANN network of the National Observatory of Athens: operation and database," *Geoscience data Journal,* vol. 6, no. 3, 2017.

[10] D. Liberto and J. Berry-Johnson, "Small and Mid-size Enterprise (SME): What is a Small and Mid-size Enterprise (SME)?," Intercompany Solutions, New York, 2020.

[11] M. Aguirre-Munizaga and R. Gomez, "A Cloud Computing Based Framework for Storage and Processing of Meteorological Data," Monash , Washington DC, 2020.

[12] N. Kumari and S. Gosavi, "Real-Time Cloud based Weather Monitoring System," *International Journal of Climate Change,* vol. 4, no. 2, 2019.

[13] R. Posada, D. Nascimento, F. O. Neto, O. Jens, F. Riede and F. Kaspar, "Improving the climate data management in the meteorological service of Angola: experience from SASSCAL," *Advances in science and research,* vol. 13, pp. 97-105, 2016.

[14] S. Abbate, M. Avvenuti, L. Carturan and D. Cesarini, "Deploying a Communicating Automatic Weather Station on an Alpine Glacier by," *Procedia computer science,* vol. 19, pp. 1190-1195, 2013.