International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST173430

Weather Data Management Using Hadoop MapReduce in Zimbabwe

Muchaneta Mayepi[1], Eng. Mainford Mutandavari[2], Tinashe Blessing Chuwe[3]

[1]Department of Information Technology, Harare Institute of Technology, P.O Box BE277, Belvedere, Harare, Zimbabwe
[2]Department of Software Engineering, Harare Institute of Technology, P.O Box BE277, Belvedere, Harare, Zimbabwe
[3]Department of Information Technology, Harare Institute of Technology, P.O Box BE277, Belvedere, Harare, Zimbabwe

Article Info
Volume 9, Issue 6
Page Number : 305-310

Publication Issue
November-December-2022

Article History
Accepted : 15 Nov 2022
Published : 04 Dec 2022

ABSTRACT

The Meteorological Services Department (MSD) in Zimbabwe is experiencing huge costs in transferring data of over Gigabytes from its different models of automatic weather stations into centralized database at Head Quarters (HQ) Belvedere in Harare. 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 study therefore sought to design and develop an experimental set-up to understand how Hadoop MapReduce can be used to manage data files from various weather stations simulated as Application Programming Interface (APIs). The developed system was conducted entirely in a cloud environment and the results suggest that Hadoop MapReduce function is an appropriate big data tool to manage weather data and offers better options especially for cash strapped organizations such as Meteorological Services Department (MSD). Adoption of MapReduce is economic and elastic to accord changing weather data management requirements. The research used a small dataset of 100 records which might offer comprehensive model for use for practical purposes. Further studies with more realistic huge datasets might be pertinent.

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

I. INTRODUCTION

According to Kumari [14], the procedural steps taken to complete a project constitute the study methodology. The methodology has been broken down into the different stages of the experimental stages: designing, development, debugging and testing, and finally the validation stage. These are accounted for in this chapter. Each step had sub-tasks to be completed before commencing on the next step in a waterfall approach since the project was small. Chapter 3 begins with listing of the tools,

apparatus and materials used to design, develop, test, and validate the developed solution. The study intends to map and reduce automatic weather station data such that it reduces the large weather station data of an average estimate of more than 1GB of data being functional weather data management systems. Rather a simulated weather station data obtained from the Internet and the virtual machine with the Application Programming Interface (API) that was running the algorithm plus the virtualised database were hosted on a subscribed private cloud: Digital Ocean Private cloud. To simulate the MapReduce of the weather data, the researcher personal data was used to programme the API and used to sign on to the virtualised system. The following paragraphs provides more detailed information of work done on each stage of the project.

II. METHODS AND MATERIAL

This section of the paper contains a list of the hardware, software and dataset requirements for the design and development of a Hadoop MapReduce algorithm to manage the weather station data from large file sizes of 1GB to small files of 10MB that will be transmitted from each weather station to the central database hosted on the cloud. The local machine being the laptop of the researcher were used for two purposes: working as one of the weather stations and for logging on to the cloud based virtualised weather stations (API) and centralised database representing the HQ weather station. The weather data was processed in the cloud. Users can access the weather station and HQ weather data anywhere at any time and select the format of data representative of their weather station (CSV, text or API specific form). This has been done after realising that MSD automatic weather stations have different formats. The design was scalable in that it also allowed any future data formats to be customised and accommodated. Any future expansion only involves adding more virtualised space and storage space. The designed system is elastic and scalable. Table 1.1 transmitted to HQ to small file sizes of 10MB. In this set up, live weather station data could not be used to avoid disruption of provides detailed description of the tools used in their broad categories: hardware, software, and dataset.

Table1. List of the required hardware, software, and dataset for the experiment

Software	**Purpose**	**Source from which it was obtained**
Subscribed Private cloud - Digital Ocean	To host the database and virtual machines for processing (map reduce) the data from 3GB to as small as 10MB	The Department of Meteorological Services.
Django	Developing the web interface for remote log to the virtualised cloud environment	Internet free download software
Python	Programming the API to perform the map reduce function	Internet free download software
Windows 10	Operating system (OS) for the personal laptop to host the local applications and log on to the Digital Ocean environment.	Internet free download software
Hadoop Mapreduce	Running weather dataset	Internet free download software
Hardware		
Laptop – HP EliteBook	Accessing the platform	Departmental Resource
Dataset		
Total records and fields -	3 records and 113 fields	Department of Meteorological Services.
Number of records for Testing data	3 Record	Department of Meteorological Services.
Total number of records for validation	3 Records	Department of Meteorological Services.

III. RESULTS

This current section presents the experimental results of designing and developing a weather data management based on the mapping and reducing functions of the Hadoop MapReduce system. The purpose of the study was to manage the entire weather data in a cloud environment and use such platforms to reduce the size of files transmitted from one weather station to the central databases – MapReduce function was able to perform this requirement. A recap of the main study objectives may help inform the structure followed in this section. The study had three main objectives:

- To design an application programming interface (API) that will integrate weather data from all automatic weather station in the country.
- To optimize data from stations to the central weather database using Mapreduce
- To design a private cloud for storage of data collected from the automatic weather station.

As outlined in the above research objectives the main purpose of this study was to design and develop a cloud-based weather data management system using Hadoop MapReduce as the main engine behind the data management process.

A. Design of the Hadoop MapReduce algorithm for mapping and reducing MSD weather data

The design of the cloud-based Hadoop MapReduce system for weather data management is depicted in Figure 1 below. The system was entirely hosted on the cloud. Development of the system occurred using cloud-based development tools. The researcher could easily sign on to the system and manipulate how each of the weather station data (API data) could be mapped and reduced to produce a much smaller file size as shall be presented in the proceeding sections of the results chapter. The architecture involved APIs for individual weather station and then an API to integrate the separate files into a consolidated data set ready for mapping and reducing using the mapper and reducer functions of the Hadoop MapReduce cloud-based system. By launching the system on the cloud, it served personal resources of the researcher.

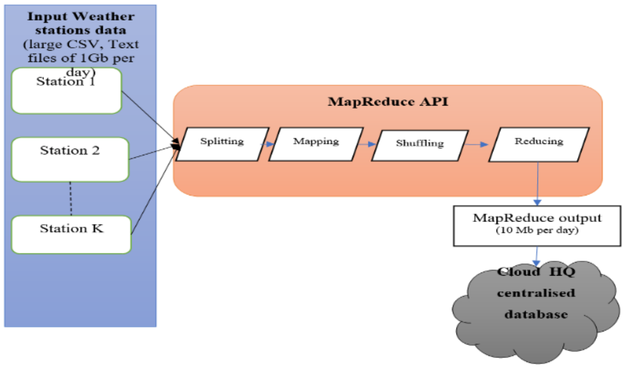
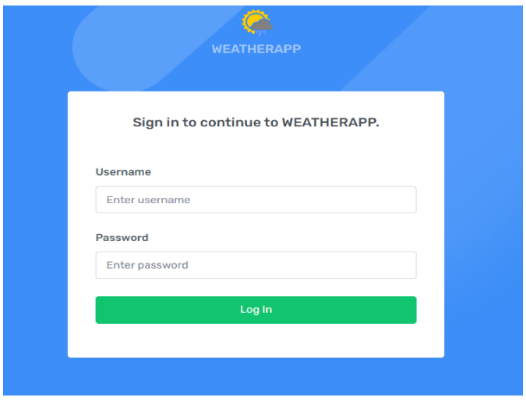


Figure 1. Overview of the system architecture

B. Design an application programming interface (API) to integrate weather data

At this stage it was important to develop an API that would gather data from the distributed files on the weather stations (individual APIs) into a single large weather data file. This was the IP that is used to access the API http://142.93.34.93:5555/ and the following login form will pop up as shown below.

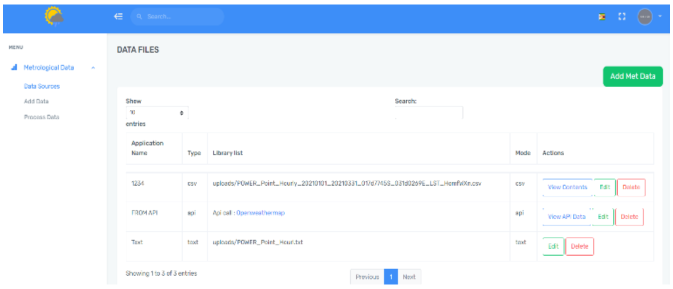
Log on form to the Integrating API



Source: Researcher 2021

The integrating API would feed data from the individual weather stations represented by different data formats: text, CSV and API as depicted below. The integrating API is user friendly as users must select from the presented menus.

Dashboard for the Integrating API

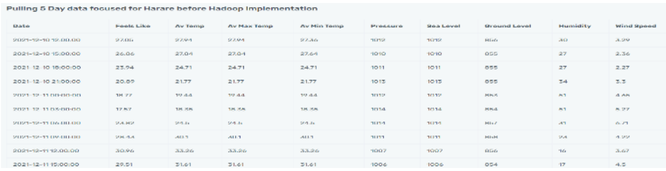


Source: Researcher 2021

As depicted in the Log on form to the Integrating API and Dashboard, upon successful log onto the integrating API, the user is offered with a user-friendly dashboard in which the following parameters of the individual weather station APIs are defined:

- Application name
- The source of data
- Type and format of data

After deciding on the range of weather stations to provide weather prediction, the data from such is integrated producing integrated weather data. However, before that, the user has also to specify the range of days for which weather data is to be integrated as indicated in the output below.

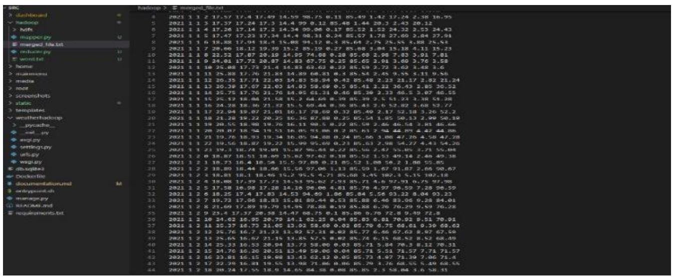


Pulling 5 Day data focused for Harare before Hadoop Implementation

Date	Feels Like	Av Temp	Av Max Temp	Av Min Temp	Pressure	Sea Level	Ground Level	Humidity	Wind Speed
2021-12-10 12:00:00	27.05	27.94	27.94	27.36	1012	1012	856	30	3.99
2021-12-10 15:00:00	26.06	27.04	27.04	27.64	1010	1010	855	27	2.36
2021-12-10 18:00:00	23.94	24.71	24.71	24.71	1011	1011	855	27	2.27
2021-12-10 21:00:00	20.89	21.77	21.77	21.77	1013	1013	855	34	3.3
2021-12-11 00:00:00	18.77	19.44	19.44	19.44	1012	1012	863	51	4.48
2021-12-11 03:00:00	17.87	18.38	18.38	18.38	1014	1014	864	51	5.27
2021-12-11 06:00:00	24.82	24.5	24.5	24.5	1014	1014	867	41	6.71
2021-12-11 09:00:00	28.44	40.1	40.1	40.1	1011	1011	868	24	4.92
2021-12-11 12:00:00	30.96	33.26	33.26	33.26	1007	1007	856	16	3.67
2021-12-11 15:00:00	29.51	31.61	31.61	31.61	1006	1006	854	17	4.5

Source: Integrated API output.

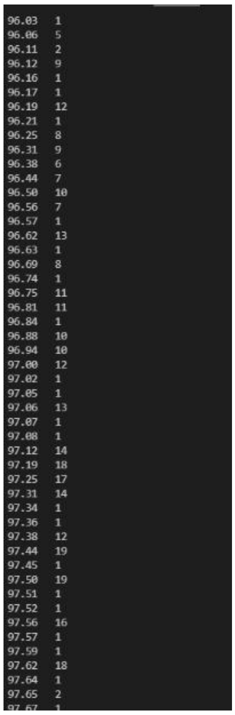
Merging and integrating of weather data before MapReduce functions, the data from csv and text format had to be merged before the Hadoop as depicted below.



Source: Integrating API output file

A. Hadoop MapReduce

The integrated file depicted in table 6 was then inputted to the Hadoop MapReduce system to reduce the weather data into manageable small file sizes. The Hadoop managed to produce the reduce file as depicted in table . This is the IP that is used to access the API http://142.93.34.93:5555/. Digital Ocean is hosting the database. The following results are from the database after hadoop processing.



96.03	1
96.06	5
96.11	2
96.12	9
96.16	1
96.17	1
96.19	12
96.21	1
96.25	8
96.31	9
96.38	6
96.44	7
96.50	10
96.56	7
96.57	1
96.62	13
96.63	1
96.69	8
96.74	1
96.75	11
96.81	11
96.84	1
96.88	10
96.94	10
97.00	12
97.02	1
97.05	1
97.06	13
97.07	1
97.08	1
97.12	14
97.19	18
97.25	17
97.31	14
97.34	1
97.36	1
97.38	12
97.44	19
97.45	1
97.50	19
97.51	1
97.52	1
97.56	16
97.57	1
97.59	1
97.62	18
97.64	1
97.65	2
97.67	1

Source : Hadoop MapReduce functions.

MapReduce managed to extract only the relevant data, in this case highest temperatures were identified though the indexes are different depending with the location of the automatic weather station. The output shows that only needed data was returned minimizing the file sizes to be transmitted as sought in this study.

IV. DISCUSSION

The findings from this study are indicative of the fact that Hadoop MapReduce can be effectively used to manage weather data especially configured in a distributed file system – different weather station data files. Though several surveyed literatures did not exactly use the Hadoop MapReduce only for reducing the file sizes, the underlying application is appropriate to address a number of weather data management functions [1]. The novel approach to use Hadoop MapReduce in managing weather data is an emerging concept as illustrated by some authors [2] [9]. Reddy [9) demonstrated that MapReduce can easily be used to predict weather. Its predictive power is more accurate that current techniques such as simple forecasting models. Aguirre-Munizaga [2] extended the Hadoop MapReduce into a cloud working environment to understand the performance of the system with scalable and dynamic working platforms. However, these past studies focused on use of Hadoop MapReduce for predictive purposes similar to other common applications of the application [9]. The current study findings are unique in that there is little known about use of MapReduce to manage and optimise weather data to help firms reduce operational costs and interact seamless anywhere at any time within a cloud working setup. The results are promising to the weather data management entities especially given that huge amounts of data in diverse formats and sources is characterising modern weather data management systems [8]. The big data handling capability of Hadoop and its compatibility with cloud environment makes management of data much easier and visible. Current weather data management wait for regular FTP file transfers from each weather station to a centralised repository and then use of rudimentary predictive models to forecast weather. Adoption of Hadoop as demonstrated in this environment and works by Jyothi [13] shows that cloud-based solutions are emerging as the prominent weather data management solutions. Nonetheless, the study findings were conducted using a small size dataset of 100 records limiting the validity of the current model to manage weather data. More data sets would have been necessary. Using such small data sets underutilised the capabilities of MapReduce and also the elasticity opportunity of a cloud environment.

V. CONCLUSION

Hadoop MapReduce has wide range of possibilities to manage and optimise the transmission and prediction of weather events. However, as already alluded to, its application in novel fields such as weather data management is still emerging and open to debate. Despite these limitations, the mere fact that Hadoop MapReduce is adaptive to huge volumes and diverse set of data and seamlessly interface with components in a cloud-based working environment, it remains a potential source of weather management capabilities for the weather management institutes. Care should however, be practised to ensure that the cloud system is secured using separate security frameworks as the platform tends to centralise operations creating a significant single point of failure.

VI. Recommendations

Against the research results and growing adoption of Hadoop MapReduce in various fields of operations, it can be recommended that:

- MSD explores the use of Hadoop MapReduce software in a cloud-based environment to manage its data, in order to reap the optimization and predictive power of the tool while introducing cloud-based approach to extraction, accessing, manipulation and transmission of the data.
- Further research might be necessary to be conducted with large datasets typical of a weather data management environment.

Results obtained in this study were all based on simulation and thus might be quite varied to the actual weather data management environment.

VII. REFERENCES

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

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

[3]. 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.

[4]. National Climate Data Center, “Managing legacy climate data in the 20th century,” National Climate Data Center (NCDC), Washington DC, 2020.

[5]. 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.

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

[7]. L. E. Wood, “Automatic weather stations,” Friez Instrument, Baltimore, 2016.

[8]. V. Dagade, M. Lagali, S. Avadhani and P. Kalekar, “Big data weather analytics using Hadoop,” International journal of emerging technology in computer science and electronics, vol. 14, no. 2, 2015.

[9]. B. Reddy and B. A. Patil, “Weather prediction based on big data using Hadoop Map Reduce technique,” International Journal of Advanced Research in computer and communication engineering , vol. 5, no. 6, 2016.

[10]. A. Mohammed, K. Adam, N. Abu and M. Majid, “Weather data analyis using Hadoop: Applications and challenges,” Materials science and engineering , vol. 551, 2019.

[11]. 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.

[12]. National Aeronautics and Space Administration , “Weather forecasting through ages,” National Aeronautics and Space Administration (NASA), New York , 2018.

[13]. S. N. Jyothi, B. Divija, S. K. Tharun and S. Reddy, “Weather data analysis using Hadoop MapReduce and Spark,” Journal of Xian University of Architecture and Technology, vol. 12, no. 4, 2020.

[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.

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

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

[17]. 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.

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

[19]. 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.

Cite this article as :

Sh Muchaneta Mayepi, Eng. Mainford Mutandavari, Tinashe Blessing Chuwe, "Weather Data Management Using Hadoop MapReduce in Zimbabwe", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 9 Issue 6, pp. 305-310, November-December 2022. Available at doi : https://doi.org/10.32628/IJSRST173430

Journal URL : https://ijsrst.com/IJSRST173430