Big-data Management using Map Reduce on Cloud: Case study, EEG Images' Data

Sahar Mahdie Klim

Department of Computer Engineering / College of Engineering / Misan University
Email: sahar_mahdi@uomisan.edu.iq

Abstract

Database is characterized as an arrangement of data that is sorted out and disseminated in a way that allows the client to get to the data being put away in a simple and more helpful way. However, in the era of big-data the traditional methods of data analytics may not be able to manage and process the large amount of data. In order to develop an efficient way of handling big-data, this work studies the use of Map-Reduce technique to handle big-data distributed on the cloud. This approach was evaluated using Hadoop server and applied on EEG Big-data as a case study. The proposed approach showed clear enhancement for managing and processing the EEG Big-data with average of 50% reduction on response time. The obtained results provide EEG researchers and specialist with an easy and fast method of handling the EEG big data.

Keywords: Big-data, Cloud Computing, Electroencephalogram, MapReduce, Hadoop,.

1. Introduction

Database is characterized as an arrangement of data that is sorted out and disseminated in a way that allows the client to get to the data being put away in a simple and more advantageous way. By utilizing this strategy, the client will be permitted to direct any sort of adjustments over a specific arrangement of data. There are diverse sorts of databases that are right now being utilized by kind of data that should be put away. Among one the most generally utilized databases that is being utilized to get to element data is the value-based database [1]. The utilization of this database to recover dynamic data, for example, stock data, has been extraordinarily examined as of late. The other typical kind of database is known as illustrative database [2], which is used for social event and organizing static data, for instance, legitimate results or land comes about that don't much of the time change on a steady base.

The utilization of a database takes into account arranging a huge substance of data that are recovered from different assets in an expansive data framework known as data warehousing [3]. This framework is actualized for a simple recovery of inquiry and investigating forms in a productive way, rather than getting to exchange forms. The data put away in the warehousing framework comprises of authentic data, which are for the most part recovered from verifiable data, other than different sources as well.

This framework works by isolating examination workload from exchange workload, with the goal that this would take into consideration arranging the data selected from different assets effectively. What's more, this framework capacities by enrolling data through Extraction, Transportation, Transforming and Loading (ETL) arrangement [4]. Moreover, other application procedures, for example, online diagnostic procedures and customer examination apparatuses take into consideration gathering the

data and sending the accumulated data effortlessly to the business clients in a proficient way.

Subsequently, the usage of data warehousing has hoisted the calculation inside the business sector and which to be sure has been created further by the execution of systems administration framework. This has added to updating the execution of an incredible measure of data in an auspicious productive way. The primary reason for data warehousing framework is to make a long haul stockpiling framework so as to be utilized by the individuals who require getting documented data for future reference [5].

In spite of the way that the foundation of databases takes into account a proficient and simple system of recovering data adequately, conventional databases have their own particular shortcoming focuses [6]. For instance, standard database and programming techniques were not prepared for being used to handle, mastermind and have control over an extensive volume of data, for instance, sorted out, semi-composed or unstructured data [7]. This opening prompts the headway of a wonder known as Big Data [8]. This term is utilized to delineate enormous dataset comprising of 4-V definitions: Volume , Velocity , Variety and Value , (for example, electronic restorative records, electrocardiogram and biometrics information) [9].

In any case, these datasets were appeared to force an issue with capacity, representation and investigation [9]. Consequently, keeping in mind the end goal to settle this issue, new programming frameworks have been made to experience the issues connected with these datasets. The recently created programming frameworks are manufactured especially to arrive parallelism from huge accumulations of processing groups, instead of acquiring them from super PC. These figuring bunches comprise of ordinary procedures, it can connected to Ethernet connections or some other unobtrusive switches. These redesigned programming structures are frequently called "Distributed File System" (DFS) [10].

This framework presents bigger joins than the circle squares that is found in customary working frameworks. What's more, DFS gives an additional element of data replication to ensure against data disappointment, which for the most part happens because of the expansive dataset being entered. Subsequently, this has risen the need to develop the utilization of Map Reduce keeping in mind the end goal to precisely prepare Big-Data and concentrate the potential prescriptive and prescient Big-Data [11] [12].

Investigative is generally used to represent diverse purposes and strategies used to direct different procedure in a dataset. There are three sorts of examination. [11]:

A) Descriptive investigation: A method that is utilized to subtract the dataset of premium and make typical reports that can be utilized to respond inquiries, for example, "what was the deal? What is the issue? What activities are required? ".

B) Predictive investigation: As a reason for future data being given by illustrative systematic methodology, a prescient logical methodology has been created. This methodology utilizes the need of utilizing measurable models of the chronicled datasets with a specific end goal to figure more data about what's to come. Prescient logical methodology is a valuable methodology of noting inquiries like "why is this event? What will happen next?" this sort of prescient methodology depends on the how shut the data acquired as contrasted and other factual models.

C) Prescriptive investigation: this kind of expository methodology includes the utilization of different data model, for example, multi variables incitement and recognizing the connection between different variables. This methodology is proficient for noting certain inquiries, for example, "What could happen if a specific situation is utilized? What is the most fitting situation to be actualized? "

Elite registering approach [13] is a technique for organizing parallel pathways for running anomalous state application programs in a compelling, more tried and true and less repetitive way. This strategy applies for structures that work at a skimming motivation behind 1012 consistently. Prevalent enrolling is utilized for handling convoluted issues and reasonably fulfilling research practices using bleeding edge PC workplaces, affectation and by executing distinctive preparing resources.

The distinctions among lattice processing and disseminated registering frameworks are as per the following [14]:

a)   A disseminated figuring framework can deal with hundreds or a large number of PC framework, which are described of having a constrained access of preparing assets, for example, CPU, memory and capacity. Then again, the lattice processing framework concentrates more on the proficiency of coordinating heterogeneous frameworks with administration servers, stockpiling, ideal workload and system.

b)   A matrix registering framework is all the more particularly intended to improve calculation inside different authoritative spaces, which to a great degree varies from the customary conveyed figuring framework.

For the greater part of figuring frameworks, a solitary processor is generally utilized alongside its principle memory, store and nearby plate. Beforehand, applications that called for parallel PCs with numerous processors and certain equipment projects were utilized. Be that as it may, the interest of utilizing extensive scale web administrations required the interest of all the more registering being done in light of establishment. What's more, registering hubs were appeared to significantly diminish the expense when contrasted with exceptional reason parallel machine. Along these lines, recently created registering offices have added to the improvement of more upgraded programming structures. The upside of using such structures is that it keeps the unwavering quality issues that may emerge when figuring equipment framework is made out of a great many free sections that could neglect to work appropriately [14].

Moreover, these systems can likewise handle parallelism, which may happen accidently. The group creates errands that can be observed by the expert hub, which is usually known as the NameNode [15], and it is accountable for piecing the data, duplicating it, bringing in the data to the circulated processing hubs, checking the status of the bunch and gathering/assembling the outcomes gained. Notwithstanding the advancement of DFS, other abnormal state programming has been composed.

The most famous structure created is the MapReduce system. MapReduce is a typical programming system that is used for data serious application that are displayed by Google [16]. MapReduce assembles thoughts from different useful programming frameworks the especially characterizes Map and Reduce assignments so as to run colossal arrangements of appropriated data all the more proficiently. The usage of MapReduce permits a wide range of figurings on the expansive scale to be done on registering groups all the more viably and in a more capable manner that could diminish equipment disappointments amid the calculation procedure [17].

In this manner, the Hadoop made simpler to process substantial measure of data sequentially where it could read the entire envelope each one in turn [18]. While, in ordinary programming every document is perused and dissected independently. This will therefor takes time that longer and requires greater memory and limit. The "Hadoop" moreover saved attempt that is required in obtaining the code, as run of the mill programming requires to perform programming on the records. On the other hand the aide code is speedier and can be consolidated successfully.

1.1. Problem Declaration

As a result of the extension the data volume in the figuring space, a magnificent enthusiasm of realizing more beneficial technique that go for securing, separating and planning the data signal is principal. Electrophysiological data is noteworthy for investigation, management and clinical investigation in epilepsy and other essential issue. "EEG" data is a kind of biomedical sign dataset and clinical "Big-Data" that have "4 V" (volume, combination, speed and regard), which including more than 100 "multi-channel" signal. Without a doubt, Which imperative to actualize successful methodology so as to oversee huge EEG datasets. One of the ordinarily utilized strategies as a part of dissecting data is Ensample Empirical Mode Decomposition (EEMD) strategy which help in breaking down EEG flags yet this technique takes longer time if data are broke down successively.

This work should answer the underneath inquiries:

•   What does the MapReduce and Hadoop can offer as far as capacity and examination of EEG Big-Data?

•   How does the Hadoop vary from typical programming methods in preparing the EEG Big-Data?

1.2. The Solution Proposed

In order to improve the adequacy of separating "EEG Big-data" for all the additionally sympathetic and straightforwardness of considering patient cases, "EEG Big-data" ought to be completed close by "Hadoop" by using MapReduce technique. The use of MapReduce and "Hadoop" on scattered structures, for instance, Cloud Computing can add to the basic advancement in clinical Big-Data planning and utilize. Likewise, it will present new open entryways in the creating time of "Big-Data" examination and enhance the aftereffect of clinical "EEG Big-Data" explanatory gadgets.

The change of the "Hadoop" made it a great deal less requesting to consider and dismember immense measure the data in a favorable profitable way. The "Hadoop" executes the use of the "Map" and "Reduce" shapes, that let the designer to control the data that they are enthused about examination. Besides, Hadoop plays out the work speedier, where it takes each one of the archives and explore every one of them meanwhile, however in regular programming the records ought to be poor down freely. Additionally, "Hadoop" obtain the code in a speedier and compelling course as differentiated and the standard programming technique.

Thusly, it is basic to outline the upsides of realizing the "Hadoop" in the figuring space of taking care of "EEG Big-data", This procedure will agree to the clinical and EEG Signal get ready examiners, to procure their required information and accomplish their destinations swiftly and definitely, with slightest tries. In addition, this proposed system will make it straightforward to manage them "EEG" huge data in improved way.

This search shed the light on the points of interest that the "Hadoop" and MapReduce had added to the preparing and the programming space. Furthermore, it used an example of "EEG Big-data" to be taken care of using "Hadoop" server in light of the MapReduce method and depict an overhauled strategy for data use and examination in a capable, brisk, and correct way. This accomplishment, evaluated the planned approach of using the MapReduce methodology on "Hadoop" server in light of flowed figuring structure, to set up the "EEG Big-data" to make it basic for clinical individuals and "EEG" signals investigators to recuperate the required data for their work and analyzes, absolutely with brief era. This planned system and they got contributed occurs, which will be showed up in later ranges, could be considered as a distress in the Clinical field specially the "EEG" signals taking care of.

1.3. Paper Organization

Next to this section, the second section will specify foundation focused on the issue understudy. The third section will the talk about the instruments, approach and methodology that are utilized as a part of this study and their choice criteria. The proposed approach and the assessment test will be examined in the fourth area. In addition, the fifth area will contain the outcomes got from the proposed approach and their examination notwithstanding correlations amongst the and the nearest related work, with a specific end goal to assess the commitment and upgrades that this study has advertised. At last yet essentially, the 6thsection will abridge the work demonstrate the conclusions got from this study.

2. Evolution of Big-data

In the course of recent years, the growing usage of using web organizing, new improvements, for instance, the possibility of "Internet of Things" (IoT) make huge measure of information that can yield extremely valuable data if fittingly managed. It is called as "Big Data". These datasets portray voluminous measure of composed, semi sorted out and unstructured information.

Along these lines, "Big Data" has created to restrain the value, arrangement and velocity of datasets; in view of that reality, the conventional advances of planning information were not able. This is by virtue of it divided, hard to reach, excessively immense, move too fast, and unnecessarily trapped. Pros evaluate that 80 to 90 of affiliations' information are unstructured and the measure of unstructured information is turning out to be rapidly and much of the time usually than composed database. Therefore, routine programming methods, for example, "Relational Database Management System (RDBMS)", and "SQL" were ended up being too progressively to set up the volume of information in favorable path since they cannot oversee Petabyte and Exabyte of information.

Dataset is continually updating and making new information that they ought to be readied, the more quickly of information is being created, the more basic to have a base that can scale to process information as quick as it is changed and "Big Data" consolidate unstructured information that it doesn't stay in line segment of database.

3. Proposed Methodology

The used Electrophysiological information for this examination was gotten from, which record for patients inside the five days time of their specialist's office stay and release multi-channel dataset. Remembering the true objective to record electrophysiological information, 30-40-channels are regularly included. Each of these channels is

made out of 20 "EEG", 4 "EKG", 1 channel for oxygen checking, 2 occupies are incorporated into respiratory hailing and 1 is used for watching circulatory strain. Around, 20 EDF records are exposed for each patient inside the 5 days cross period. Each EDF report is made out of information record and each information record contains seconds of sign demonstrated by a particular number of tests are showed up in record header and chronicled as Cloud Wave Metadata [19].

In the aide, the information is changed over into an once-over and that in truth could save time while differentiating qualities and foresee delay in the run time as it has a record. In the wake of changing over the information to an once-over, the product designer would have the option of picking the information that are required. Along these lines, the aide has diverse components, where it has a record, change over the substance to once-over and make the measure of the information more diminutive concurring the information which required. When the customer needs to intensify the range of the information, the cursor ought to be moved. Regardless, this movement is not any more required as "JAVA" can use to oversee content records that might be segregated by commas and space and a while later these information can carry on to the rundown by changing over the substance archive to an aide. By then, when moving to the diminishing or count, the customer can make the code that is required on the once-over that was made. The summary does not have a cursor since it is considered as a document that starts with zero and the once-over could be gathered into a key or a value.

The transport of the information could be on more than one server so that if we rent from the "cloud" various servers, this information ought to be added to the cloud. By then, a request ought to be directed to accumulate the information inside the information coordinator. In this way, while selecting the aide decrease, the engineer needs to give the name of the particular yield and the way could be an IP deliver which is conferred to various diverse servers. The implies that required while coordinating an aide decrease are the underneath:

1. Created database building
2. Database inclusion.
3. Dataset processing and looping.
4. Building the guide work, which reduces that we did.

There were particular classes. The principal class, that is the overall public class. The static class, which is fanned and rest mapper that reduces stage from "JAVA". The product design needs to choose the kind of the information and execute is to the mapper.

4. The Experiment of Evaluation

"Hadoop" is an engine that considers running techniques simply in a brief period. As a product engineer or an architect, it is fundamental to know the establishment of the "Hadoop"; which suggests the spot of where each one of the records will be secured at. Each one of the information can be secured in H-work coordinator and in case that the designer needs to change the name of the record, the earth route need to change. In order to change the earth way, the course of the "Hadoop" should be changed from windows settings. By then, to turn the path on a database, the database that the architect need to manage should have the same structure to create an authentic test on it. By then, remembering the ultimate objective to make the aide, the product build needs to consider the structure of the database, which could be a substance record, picture or some other course of action.

The following stride that should be done is determining the course that I have for putting away every one of the data inside an data organizer. While managing the Hadoop, every one of the data that we have as a content record will be gathered and assembled bit by bit inside the info envelope. It could be conceivable that we would have more than one organizer however the greater part of the envelopes ought to be as a content and ought to have the same structure. This is significant as though the software engineer needs to build up a little program; the calculation that we need to create to break down the data needs steady structure. Along these lines, when we reach to the guide structure, every one of the data that enters that guide ought to have a consistent structure for more proficient investigation. In this manner, every one of the data in an organizer ought to have the same number of sections and the same data sort, while these datashould be composed in a specific arrangement. Case in point, if the software engineer takes a bit of data as a number and after that it shows up in one document as a content, this would bring about a noteworthy issue. At that point, the "Hadoop" will build the guide on every one of the documents that are found inside the

info envelope, regardless of what number of organizers we have.

Presently, we should assemble a guide with a specific end goal to do a file (rundown) of the data. To develop this progression, the content document should be changed over to information or list, as this progression will add to minimizing the span of the data and dispose of the considerable number of data that are not required. For our situation, we just need to know the patient's ID and the time that light was killed. In this manner, we are no more keen on the various data that are found in the rundown. Notwithstanding the way that the guide is steady, we will have the capacity to extricate the data that

we require from it as indicated by the kind of data that we need to separate from the guide. The work on the last stride is to create the yield record, which has every one of the qualities that have been determined by the developer. For our situation, we picked the estimations of which time has been killed after 12 pm as it were. Inside the "Hadoop", there is an order that gives the data as per its accessibility, we simply call attention to the data and the yield organizer that are found inside the "Hadoop". It has been accounted for that the season of "CPU" for our data was zero because of the little size of the documents that we have.

The beneath square graph compresses the strategy took after amid the execution procedure.
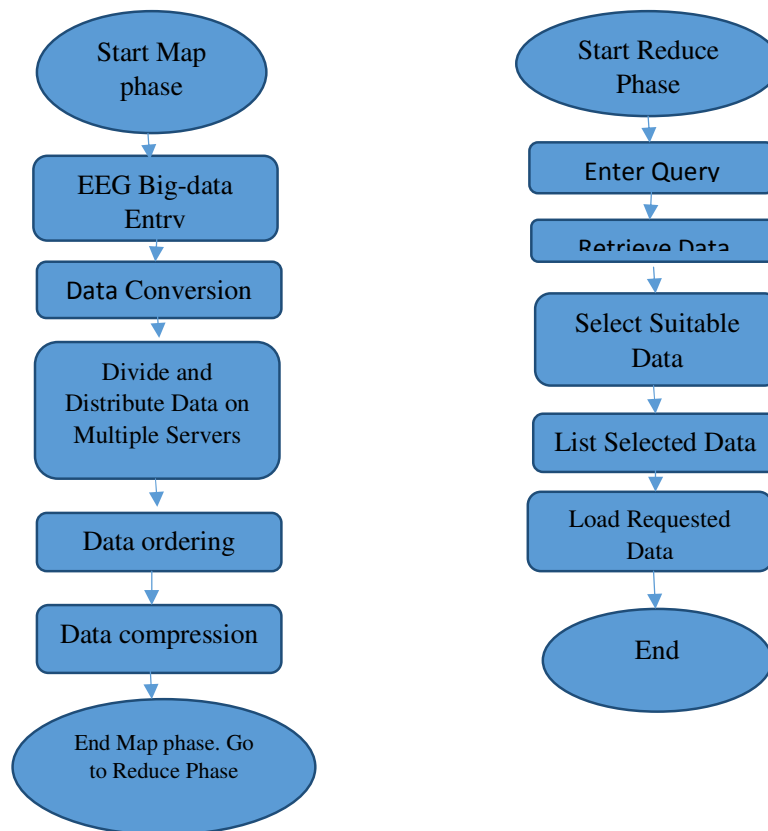

Fig. 1. Proposed Approach Block Diagram

The above block diagram contains the steps of the two phases, "Map" and "Reduce". Then the next section will show the obtained outcome and comparison between the proposed approach and the previous one.

5. Experimental Results

In this search, we utilized the inputfiletemazepam_ effects_on_sleep.txt [24]

keeping in mind the end goal to acquire the results.txt record. At that point, we utilized the Map-Reduce-EEG.java to lead the guide procedure and afterward create the diminish procedure keeping in mind the end goal to get the information that are of our fundamental attention. We were occupied with acquiring the information, where the time of lights turn off after 05:00AM. The data underneath demonstrates the Java file,

Map-Reduce-EEG.java. The following is the "JAVA" assessment code test for the trial.

5.1. The Function of The Map

That show the evaluation code for the map phase.

```
public void MapEEG (LongWritable key, Text
value, OutputCollector<Text, Text> output,
Reporter reporter)
throwsIOException { String roweeg =
value.toString(); String curser = null; String
Tokenizereeg = new StringTokenizer(line,"\t");
String var_agent= s.nextToken();
While (eeg.hasMoreTokens())
{curser= eeg.nextToken();}
 String timeOfSleep =curser; output.collect(new
Text(timeOfSleep), new Text(timeOfSleep)); }
```

5.2. Reducer Class

That show the "JAVA" code for the reduce phase.

```
public static class ReduceEEG extends
MapReduceBase implements Reducer< Text,
Text, Text, Text>
{ public void reduceRRG( Text key, Iterator
<Text> values, OutputCollector<Text,Text>
output, Reporter reporter)
throwsIOException
{while (values.hasNext())
{ TextSec=values.next();LocalTimeLocTim= new
LocalTime(sec );
if(LocTim.isAfter(LocalTime.MORNING)) {
output.collect(key, new Text(LocTim));
        }}}
```

As indicated by the "Hadoop" capacity above, we have initially extricated the JAVA document that has every one of the records and data in it. At that point, we led the guide procedure to change over every one of the data to rundown that has record and in this way, making the correlation procedure simpler that it is utilized to be in typical programming. At that point, in the lessen procedure; we picked the qualities that the light turned off after 05.00AM. The string line quality will be put away in the content document and it will take the last esteem token. The line will continue moving to the following worth token and after that give the estimation of the last token, where it generally take the last esteem and store it in "java" record. The underneath figures demonstrates the response time about in the wake of making the query utilizing the proposed

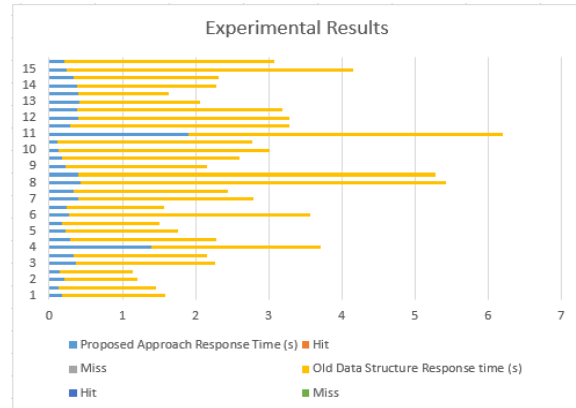strategy and contrasting it and the old data structure technique.

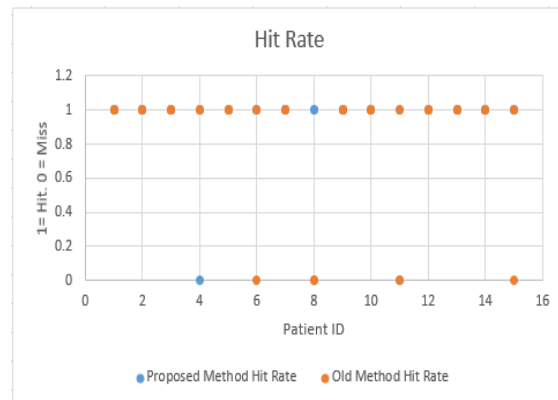**Fig. 2. Experimental Results Request Time.**

Fig. 3. Experimental Results Hit Rates vs Miss Rates

The previous figures demonstrates the pattern of the reply time for the experimental results, demonstrating the improved reply time of the proposed approach in correlation with the conventional strategy for data structure.

The previous results demonstrate clear enhancement on the reply time and precision of the retrieved data, with normal general improvement with average response time of 0.36s, and average hit rate of 93.33%. On the other hand, the old data structure method showed average response time of 2.37s and average hit rate of 83.30%. To have approximately 65% enhancement on the response time and 12% enhancement on the accuracy and hit rate. Which is great contribution, especially when dealing with big data that contains huge about of stored and distributed data.

6. Conclusions

Hadoop is an engine that works alongside the MapReduce with a specific end goal to help the software engineer or designer to acquire the information that they require and bar the various data that they need not bother with. In any case, Hadoop needs a server with excellent attributes to create the work and every one of the procedures that we have to perform. In this study, we utilized the utilization of the MapReduce strategy on Hadoop server to process EEG big data and make the dispersion on cloud. Accordingly, this has made it less demanding for examination and getting that information of concern.The analysis were actualized on Hadoop, clear upgraded results were acquired and contrasted and customary information structure strategy, which demonstrated a normal of half improvement over the conventional technique.

7. References

[1] Londhe, S. R., Mahajan, R. A., &Bhoyar, B. (2013). Overview on Methods for Mining High Utility Itemset from Transactional Database. International Journal of Scientific Engineering and Research (IJSER), 1(4), 12.

[2] Adrian.A. Big Data Challenges. 2013. Database Systems Journal vol. IV, no. 3, 31-40.

[3] Krishnan, K. (2013). Data warehousing in the age of big data. Newnes.

[4] Jin, Q., Liao, H., Srinivasan, S., &Xu, L. (2014). U.S. Patent No. 8,903,762. Washington, DC: U.S. Patent and Trademark Office.

[5] Cloud security alliance. 2013. Big Data Analytics for Security Intelligence. 1-22.

[6] Hernandez, M. J. (2013). Database design for mere mortals: a hands-on guide to relational database design. Pearson Education.

[7] Barbulescu, M., Grigoriu, R., Halcu, I., Neculoiu, G., Sandulescu, V. C., Marinescu, M., &Marinescu, V. (2013, January). Integrating of structured, semi-structured and unstructured data in natural and build environmental engineering. In Roedunet International Conference (RoEduNet), 2013 11th (pp. 1-4). IEEE.

[8] Mayer-Schönberger, V., &Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

[9] Birke, R., Bjoerkqvist, M., Chen, L. Y., Smirni, E., &Engbersen, T. (2014). (Big) data in a virtualized world: volume, velocity, and variety in cloud datacenters. In Proceedings of the 12th USENIX Conference on File and Storage Technologies (FAST 14) (pp. 177-189).

[10] Vashist, S., & Gupta, A. (2014). A Review on Distributed File System and Its Applications. International Journal of Advanced Research in Computer Science, 5 (7).

[11] Das. T, Kumar. M. BIG Data Analytics: A Framework for Unstructured Data Analysis. 2013. Vol 5, 0975-4024.

[12] Lee, K. H., Lee, Y. J., Choi, H., Chung, Y. D., & Moon, B. (2012). Parallel data processing with MapReduce: a survey. AcMsIGMoD Record, 40(4), 11-20.

[13] Bryan, B. A. (2013). High-performance computing tools for the integrated assessment and modelling of social–ecological systems. Environmental Modelling & Software, 39, 295-303.

[14] Pourqasem. J, Karimi. S and Edalatpanah. S. Comparison of Cloud and Grid Computing. (2014). American Journal of Software Engineering, Vol. 2(1), 8-12.

[15] Mukhopadhyay, D., Agrawal, C., Maru, D., Yedale, P., &Gadekar, P. (2014, December). Addressing Name Node Scalability Issue in Hadoop Distributed File System Using Cache Approach. In Information Technology (ICIT), 2014 International Conference on (pp. 321-326). IEEE.

[16] Maniar, K. B., & Khatri, C. B. (2014). Data Science: Bigtable, Mapreduce and Google File System. International Journal of Computer Trends and Technology (IJCTT), 16(03), 115-118.

[17] Vadivel.M&Raghunath.V. 2014. Enhancing Map-Reduce Framework for Bigdata with Hierarchical Clustering. International Journal of Innovative Research in Computer and Communication Engineering, 2 (1), 2320-9801.

[18] Raghupathi. W &Raghupathi. V. 2014. Big data analytics in healthcare: promise and potential. Health information science and systems, 2 (3), 1-10.

[19] Tsai. C, Lai. C, Chao. H, Vasilakos. 2015. A. Big data analytics: a survey, Journal of Big Data, 1-32.

إدارة البيانات الضخمة باستخدام تقنية المابريديوس على الحوسبة السحابية: دراسة حالة، بيانات صور المخططات التحليل العصبي

سحر مهدي غليم

قسم هندسة الحاسبات/ كلية الهندسة/ جامعة ميسان

البريد الالكتروني:sahar_mahdi@uomisan.edu.iq

الخلاصة

قاعدة البيانات، تعرف بمجموعة من البيانات المرتبة والموزعة بطريقة تسمح للمستخدم بالوصول إليها بطريقة سهلة ومناسبة. لكن في عصر البيانات الضخمة، الطرق القديمة في إدارة البيانات لا تستطيع إدارة هذا الكم الهائل من البيانات. لكي يتم تطوير طريقة فعالة في إدارة هذه البيانات الضخمة. هذا البحث يدرس استخدام تقنية المابريديوس في معالجة البيانات الضخمة الموزعة على السحابة المحوسبة. هذه التقنية تم تقييمها باستخدام خادم الهادوب وتم تطبيقها على بيانات مأخوذة من صور التخطيط الدماغي، كدراسة حالة. الطريقة المقدمة في هذا العمل أظهرت تحسن واضح على إدارة ومعالجة البيانات الضخمة. بمعدل تحسين يساوي ٥٠% مقارنة بالطرق الاعتيادية في معالجة وإدارة نفس البيانات. وهو ما يوفر طريقة سهلة وسريعة في التعامل مع البيانات الضخمة.