ISSN: 2278-4632
Vol-09 Issue-10 No. 1 October 2019

AN ANALYTICAL RESEARCH ON EFFICIENCY, LOAD BALANCING AND SCHEDULING ASSESSMENT FAILURE PREDICTION IN COMPUTING USING BIG DATA IN THE CLOUD NETWORK

Mr. Sudhir

Dr. Ramesh Kumar Associate Prof. CSE **OPJS University, Churu (Rajasthan) – India**

Abstract- Cloud computing refers to accessing, configuring and manipulating resources (such as software and hardware) in a remote location. It is defined cloud computing in terms of distributed computing "A cloud is a type of parallel and distributed system that contains a set of interconnected and virtualized computers that are dynamically provided and presented as one or more computing resources. Agreements service-based - level agreements established through negotiation between the service provider and consumers." Cloud computing is a model to allow convenient on-demand network access to a shared group of configurable IT resources (e.g. servers, networks, storage, services and applications) that can be delivered and released quickly with minimal management effort or interaction from service providers. "Cloud computing offers various resources in the form of on-demand end-user services. This article assesses loud computing and Big Data enable science discoveries and application developments.

Keywords: Software, Hardware, Servers, Networks, Storage, Services, Application Developments

1. Introduction

Modern progress is digitalizing our lives more and more, which has led to rapid data growth. Such multidimensional data sets are valuable because of the potential to discover new knowledge and develop ideas for making decisions from it. Analyzing this large amount of data from multiple sources can help organizations plan the future and anticipate changing market trends and customer needs. Although the Hadoop framework is a popular platform for processing larger datasets, there are several other computing infrastructures available for use in various application domains.

We live in the information age and an important measure of today's times is the amount of data generated everywhere around us. Data is becoming more valuable. Companies try to unlock the hidden potential of data and offer a competitive advantage. MRC statistics predicted that data analysis and the Hadoop market, which represented $ 8.48 billion in 2015, are expected

ISSN: 2278-4632
Vol-09 Issue-10 No. 1 October 2019

to reach one billion by 2022. The global big data market has estimated that it will go from one billion in 2013 to one thousand million in 2018.

Gartner predicted that data will grow 800 percent over the next five years and 80 percent of the data will not be structured (email, documents, audio, video and social media content) and 20 percent will be structured (e-commerce transactions and contact information).

Today's largest scientific institute, CERN, produces over 200 PB of data per year in the Large Hadron Collider project (as of 2017). The amount of data generated on the Internet has already exceeded 2.5 exabytes per day. 400 hours of YouTube videos are uploaded in one minute, 3.6 million Google searches are performed worldwide every minute of every day, more than 656 million tweets are shared on Twitter and more than 6 are shared, 5 millions of pictures on Instagram every day. When a dataset becomes so large that its storage and processing become difficult due to the limitations of existing tools and resources, the dataset is known as big data. It is the first part of the journey towards delivering ideas for decision making. But instead of focusing on people, this process uses much more powerful and evolving technology, given the latest advances in this field, to quickly analyze large data streams, from a variety of sources, and to produce a single data stream useful knowledge.

2. The Difference Between Big Data & Cloud Computing

Before discussing how they go together, it is important to make a clear distinction between "Big Data" and "Cloud Computing". Although they are technically different terms, they are often seen together in the literature because they interact synergistically with each other.

- **Big Data**: This simply refers to the very large sets of data that are output by a variety of programs. It can refer to any of a large variety of types of data, and the data sets are usually far too large to peruse or query on a regular computer.
- **Cloud Computing**: This refers to the *processing* of anything, including Big Data Analytics, on the “cloud”. The “cloud” is just a set of high-powered servers from one of many providers. They can often view and query large data sets much more quickly than a standard computer could.

Essentially, “Big Data” refers to the large sets of data collected, while “Cloud Computing” refers to the mechanism that remotely takes this data

ISSN: 2278-4632
Vol-09 Issue-10 No. 1 October 2019

in and performs any operations specified on that data.

3. The Roles & Relationship between Big Data & Cloud Computing

Cloud computing providers often use a "software as a service" model to allow customers to easily process data. A console is usually available which can incorporate specialized commands and parameters, but you can also do everything from the site's user interface. Some products that are generally part of this package include cloud-based virtual containers, machine and database management systems, identity management systems, machine learning capabilities and more. In turn, big data is often generated by large network-based systems. It can be in standard or non-standard format. If the data is in a non-standard format, the AI of the Cloud Computing provider can be used in addition to machine learning to standardize the data.

From there, the data can be exploited through the cloud computing platform and used in various ways. For example, it can be searched, modified and used for future information. This cloud infrastructure enables real-time processing of Big Data.

You can take huge "bursts" of data from intensive systems and interpret them in real time. Another common relationship between Big Data and Cloud Computing is that the power of the cloud allows Big Data analysis to take place in a fraction of the time it used to do.

4. CLOUD COMPUTING AND BIG DATA RELATIONS

Data is the raw material for information before ordering, organizing and processing. It cannot be used in its main form before processing. The information represents the data after processing and analysis. The technology has been developed and used in all aspects of life, increasing the demand for storing and processing more data.

Various systems have been developed, including cloud computing that supports big data. While Big Data is responsible for storing and processing data, the cloud offers a reliable, accessible and scalable environment for Big Data systems to function. Big data are defined as the amount of digital data produced by various technological sources, such as sensors, digitizers, scanners, numerical models, cell phones, the Internet, videos, e-mails and social media. Data types include texts, geometries, images, videos, sounds and combinations of each. Such data can be directly or indirectly related to geospatial information.

ISSN: 2278-4632
Vol-09 Issue-10 No. 1 October 2019

Cloud computing refers to resources and on-demand processing systems available on the network that can provide a series of integrated processing services without local resources to facilitate user access. These resources include data storage, backup and automatic synchronization capabilities. Most of the IT infrastructure computing consists of services provided and provided through public centers and servers based on them. Here, clouds appear as individual access points for consumer computing needs. In general, commercial offers should meet customer or consumer QoS requirements and generally include service level agreements (SLAs).

They are an online storage model where data is stored on multiple virtual servers, rather than being hosted on a specific server, and is generally provided by third parties. Hosting companies, which have advanced data centers, rent space to customers stored in a cloud based on their needs. Expert Erik Brynjolfsson compared big data with a microscope invented in the old days and through which scientists were able to identify and measure things they had never imagined before on a cellular level. This is similar to Big Data, which is a modern microscope with which you can see things and measure data that you didn't expect. With this rapid growth, the question that comes to mind is how these large amounts of data can be stored in cloud environments. We need storage technology that meets fast growing cloud data needs and we need low cost, high reliability and high capacity storage technology.

- **The Models between the cloud and big data**

The most well-known models for providing large data analytics arrangement on clouds are PaaS and SaaS. IaaS usually is not utilized for elevated level data analytics applications; however for the most part, to deal with the storage and computing needs of data, Cloud computing models can help quicken the potential for versatile analytics solutions. Cloud computing is an individual from distributed computing family that provides resources as client services, for example, (SaaS), infrastructure like (IaaS) and a platform as service like (PaaS), however with the coming of large data, the cloud computing model is step by step moving to huge database service including (AaaS, BDaaS) known as (DaaS) database as a service which implies that database services are accessible for applications that are sent in any implementation condition. BDaaS is a type of service like software as a service or infrastructure as a service.

ISSN: 2278-4632
Vol-09 Issue-10 No. 1 October 2019

Enormous data as a service frequently depend on cloud storage to keep up consistent data access to the undertaking that possesses the information and the provider it works with and is viewed as facilitated in the cloud. Comparable kinds of services incorporate (SaaS) or service-based infrastructure (IaaS), where large, explicit data is utilized as service choices to help businesses handle huge data. It provides a great deal of significant worth for organizations today, where a blend of these has been made to make a definitive answer for organizations pushing ahead, DBaaS is as yet a moderately dim term.

Yet, it generally alludes to a large group of re-appropriated services and capacities identified with Big Data taking care of in cloud-based condition models for cloud-based huge data analytics , imagine two sorts of services for Cloud analytics, Analytics as a Service (AaaS), where analytics is provided to customers on demand, and they can pick the solutions required for their motivations; and Model as a Service (MaaS) where models are offered as building blocks for analytics solutions, More as of late, terms, for example, Analytics as a Service (AaaS) and Big Data as a Service (BDaaS) are getting well known. They include services for data analysis comparably as IaaS offers computing resources. In any case, these analytics services despite everything need very much characterized agreements since it might be hard to gauge the quality and unwavering quality of results and input data, provide promises on execution times.

5. Challenges And Opportunities

Big data computing requires the use of various techniques and technologies. MapReduce and Hadoop are arguably the most popular and useful frameworks for this purpose. In addition to Cloud Computing, bio-inspired computing was also proposed; Quantum computing and granular computing are potential technologies for computing big data. However, each of these technologies must be adapted for this purpose and is not without potential challenges, which go beyond the scope of this research.

Thanks to the elasticity and scalability of cloud solutions, this technology is one of the pioneers in the big data sector. That said, the feasibility and feasibility of using a synergistic model has yet to be explored. NESSI presented specific challenges for the implementation of existing machine learning techniques for big data computing and the development of analytical solutions, citing the following requirements as fundamental:

ISSN: 2278-4632
Vol-09 Issue-10 No. 1 October 2019

1. There is a need for development of solid scientific foundation, to facilitate selection of method or technique that needs to be chosen.
2. There is a need for development of scalable and efficient algorithms that can be used.
3. The developed algorithms cannot be implemented unless appropriate technological platforms have been selected.
4. Lastly, the solution's business viability must be explored.

Analytics solutions are expected to be uncomplicated and simple regardless of the intricacy of the problem and the challenges in question. Other than this, circulated frameworks and parallel computing is by all accounts, a proper answer for the enormous data problem. Accordingly, the designed arrangement must be slanted towards computational standards that take a shot at these foundations. There are a few distinguished challenges identified with the utilization of Cloud Computing for huge data analytics. Huge data in the cloud experience a few preliminaries, both specialized just as non-specialized.

Specialized challenges associated with cloud-based large data analytics can additionally be isolated into three classifications to be specific, huge data the board, application modeling, and representation. Challenges associated with attributes, storage, and processing of enormous data are remembered for large data the executives. The board of enormous data is a difficult task considering the way that data is constantly expanding in volume. Besides, total and integration of unstructured data, gathered from assorted sources is likewise under research thought. There are two parts of data procurement and integration. Right off the bat, data should be gathered from various sources.

6. Conclusion

Cloud computing has access from anywhere to data resources that extend around the world using a (public) cloud to allow these sources faster access to storage. The nature of Big Data is generated by technologies and locations around the world, which is why the cloud resource service provides and helps in the collection and storage of large amounts of data resulting from the use of technologies.

The cloud computing facility can expand solid equipment to accommodate small and large volumes of data. The cloud can be expanded to manage large amounts of data by dividing it into parts, which happens automatically in IAAS. Expanding the environment is a requirement for big data. Cloud computing has the advantage of

ISSN: 2278-4632
Vol-09 Issue-10 No. 1 October 2019

helping to reduce costs by paying the value of the resources used, which helps develop big data. Flexibility is also considered a requirement for big data.

When we need more storage space for data, the cloud platform can dynamically expand to meet the right storage needs when we want to manage a large number of virtual machines in a single period of time. As for fault tolerance, the cloud helps manage big data in the extraction and archiving process.

Fault tolerance helps SLAs and QOS levels. Service-level agreements specify different rules for regulating cloud service availability. Large companies such as Yahoo, Google, Facebook and others offer web-based services and the amount of data they typically collect through online user interactions has overwhelmed traditional IT capabilities. Therefore, the development of basic infrastructure components should be developed. Apache Hadoop was presented as a realistic benchmark for managing large amounts of unstructured data. Apache Hadoop is an open platform distributed software for data storage and processing. Using Hadoop, large quantities (bytes of pets) can be reliably stored on tens of thousands of servers, increasing performance cost-effectively. MapReduce is based on the distribution of a dataset among multiple servers, the partial results are reassembled. Big data are characterized by diversity, that is, they are of different types and therefore require big data. Big data and cloud computing have been concentrated from a few important angles, and we have reasoned that the connection between them is corresponding. Big data and cloud computing comprise an incorporated model in the realm of distributed system technology. The improvement of big data and their necessities is a factor that persuades service providers in the cloud for consistent advancement because the connection between them is based on the item, the storage, and processing as a typical factor. Big data represents the item, and the cloud represents the compartment.

The big data is concerned with the limits of cloud computing. Then again, cloud computing is keen on the sort and wellspring of big data. Cloud computing represents a domain of adaptable distributed resources that utilize high techniques in the processing and the executives of data but decreases the expense. Every one of these attributes shows that cloud computing has an incorporated relationship with big data. Both are moving towards rapid advancement to stay up with progress in technology prerequisites and clients. In the cloud computing condition, fault tolerance with load balancing is one of the primary issues that is required to improve its

ISSN: 2278-4632
Vol-09 Issue-10 No. 1 October 2019

performance. This examination principally centers on fault tolerance and load balancing algorithms in a cloud domain. For this, different existing fault tolerance techniques, load balancing algorithms, and load balancing algorithms, including fault tolerance, are studied. From the review, we have distinguished that there is a need to actualize the autonomic fault tolerance method for various cases of an application running on a few virtual machines, another approach should be developed that incorporate fault tolerance techniques with load balancing algorithm or with existing workflow scheduling algorithms.

With data increasing on a day by day base, big data systems and specifically, logical tools have become a significant power of advancement that provides an approach to store, process, and get information over petabyte datasets. Cloud situations unequivocally influence big data solutions by providing fault-tolerant, versatile, and accessible conditions to big data systems. Albeit big data systems are incredible systems that empower the two ventures and science to get bits of knowledge over data, there are a few worries that need further investigation. Extra exertion must be utilized in creating security mechanisms and institutionalizing data types.

References

1. R. Shimonski. “Windows 2000 & Windows Server 2003 Clustering and Load Balancing”, Emeryville. McGraw-Hill Professional Publishing, CA, USA (2003), p 2, 2003.

2. Ali M. Alakeel, “A Guide to Dynamic Load Balancing in Distributed Computer Systems”, IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.6, June 2010.

3. O. Abu- Rahmeh, P. Johnson and A. Taleb-Bendiab, “A Dynamic Biased Random Sampling Scheme for Scalable and Reliable Grid Networks”, INFOCOMP - Journal of Computer Science, ISSN 1807-4545, 2008, VOL.7, N.4, December, 2008, pp. 01-10.

4. F. Saffre, R. Tateson, J. Halloy, M. Shackleton and J.L. Deneubourg, “Aggregation Dynamics in Overlay Networks and Their Implications for Self-Organized Distributed Applications.” The Computer Journal, March 31st, 2008

5. Randles, M., D. Lamb and A. Taleb-Bendiab, “A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing,” in Proc. IEEE 24th International Conference on

ISSN: 2278-4632
Vol-09 Issue-10 No. 1 October 2019

Advanced Information Networking and Applications Workshops (WAINA), Perth, Australia, April 2010

6. Jameela Al-Jaroodi, Nader Mohamed, and Klaithem Al Nuaimi, "An Efficient Fault-Tolerant Algorithm for Distributed Cloud Services," in proc. 2012 IEEE Second Symposium on Network Cloud Computing and Applications, pp:1-8.

7. Tian-Liang Huang, Tian-An Hsieh, Kuan-Chou Lai, Kuan-Ching Li, Ching-Hsien Hsu, and Hsi-Ya Chang, "Fault Tolerance Policy on Dynamic Load Balancing in P2P Grids", in proc. International Joint Conference of IEEE TrustCom-11/IEEE ICESS-11/FCST-11, 2011, pp:1413-1419.

8. Dhinesh B. L.D , P. V. Krishna, "Honey bee behavior inspired load balancing of tasks in cloud computing environments", in proc. Applied Soft Computing, volume 13, Issue 5, May 2013.