Emerging Load Testing Approach of Big Data & Cloud Computing

Ms. Namita Sharma, Ms. Roshan Kumari, Mr. Vijay Kumar Tiwari, Devanshu Dube, Poonam Bhargav

Assistant Professor, MCA Noida Institute of Engineering & Technology, Greater Noida
Assistant Professor, MCA Noida Institute of Engineering & Technology, Greater Noida
Assistant Professor, MCA Noida Institute of Engineering & Technology, Greater Noida
Assistant Professor, Dept . of Computer Science, JIMS Engineering Management Technical Campus, Greater Noida
College of Computer Science and IT, Department of Information Technology and Security Jazan University, Kingdom of Saudi Arabia

Abstract:
Switch over a few utterances by using information technology in a variety of ways assembles big amounts of data. Such data necessitates selling out and storage. The cloud is an online storage space model where data is stock up on numerous virtual servers. Big data processing represents a new face up to in computing, particularly in cloud computing. Data Science dispensation engross data acquirement, storage and analysis. In this respect, there are many questions including, what the connection between big data and Data Science is computing. The answer to these difficulties will be talk about in this paper, where the big data and cloud computing will be studied, in adding up to receiving acquainted with connection between them in the terms of security and confronts. We have suggested a period for big data, and a model that illustrates the relationship between big data and cloud computing.

Keywords: Big data, Hadoop, Map Reduce, Data Science, Cloud Computing.
DOI: 10.24297/j.cims.2023.2.5

1. Introduction

Data is the underdone objects for information prior to sorting, accumulate and processing. It cannot be used in its chief outline prior to processing. Information represents data after processing and analysis [1]. The technology has been inhabited and used in all aspects of life, increasing the demand for storing and processing more data. As a result, more than a few systems have been developed including cloud computing that support big data. While big data is answerable for data storeroom and processing, the cloud provides a trustworthy, easy to get to, and scalable ambiance for big data systems to function [2]. Big data is defined as the

capability of digital data created from dissimilar sources of technology for example, antennas, digitizers, scanners, mathematical modeling, mobile phones, Internet, videos, e- mails and social networks. The data types embrace texts, geometries, images, videos, sounds and combinations of each. Such data can be in a straight line or indirectly related to geospatial information [3]. Cloud computing refers to on-demand computer supplies and systems accessible across the set of connections that can endow with a number of incorporated computing services without local resources to make easy user access. These resources contain data storage competence, backup and self-bringing together [4]. Most IT transportation computing consists of services that are make obtainable and delivered through public centers and servers based on them. Here, clouds come into view as personage access points for the estimate needs of the end user. It is normally projected for money-spinning recommend to meet the QoS requirements of customers or consumers, and as you would look forward to include service level conformity (SLAs) [5]. They are an online storage space model where data are stored on manifold virtual servers, rather than being hosted on a specific server, and are more often than not provided by a third party. The hosting companies, which have tremendously developed data centers, rent spaces that are stockpile in a cloud to their consumers in line with their needs

2. Why Big Data

Big data come close to and is fashioned through electronics operations from multiple sources. It necessitates appropriate processing influence and high capabilities on behalf of analysis [9]. The significance of big data lies in the systematic use which can facilitate manufacture an informed judgment to provide better and more rapidly services [10]. The term big data is categorize on the huge amount of high-speed big data of different types; this data cannot be processed and store in accepted computers.

A. Volume

It symbolizes the measure of data produced from plentiful sources which demonstrate the huge data in numbers by zeta bytes. The volume is most perceptible measurement in what concerns to big data.

B. Variety

It symbolizes data types, with, increasing the number of Internet users all over the place, smart phones and groups of populace networks users, the decipherable form of data has changed

from structured data in databases to unstructured data that includes a large number of set-up such as images, audio and video clips, SMS, and GPS data [11].

C. Velocity

It exemplifies the speed of data regularity from dissimilar sources, that is, the speed of data manufacture such as Twitter and Facebook. The huge swell in data volume and their timekeeping read aloud the need for a system that ensures super-speed data analysis.

D. Veracity

It symbolize the excellence of the data, it demonstrates the accurateness of the data and the self-assurance in the data content. The excellence of the data incarcerate can vary greatly, which influence the accuracy of analysis. Although there is wide conformity on the forthcoming value of big data, the data is almost worthless if it is not accurate [12].

E . Value

It represents the value of big data, i.e. it shows the magnitude of data after analysis. This is voluntary to the fact that the data on its own is approximately insignificant. The value lies in careful analysis of the exact data, the in sequence and ideas it make available. The value is the final phase that comes after dispensation volume, velocity, variety, contrast, validity and visualization [13]

There have been much reconsideration to the big data until they reached [14]. In this paper, based on the connection between cloud computing and big data, will suggest a new term, virtualization, which practically represents. The data structure is by default. The virtualization of big data is a headway that focuses on generates virtual structures for big data systems. Virtualization technology is the key apparatus used to be of assistance cloud computing hold large amounts of data flexibly and make easy the process of managing big data.

3. The Type And Nature Of The Data

Data in spacious variety is a set of values that are in the form of numbers, letters, and symbols and supplementary forms where they are alarmed with a scrupulous idea and subject. The data does not make intelligence without analysis, and is, therefore, bring mutually for use. It represents input, although in sequence is output after processing, i.e. data is entered into the system first, and then progression until it comes out in the form of useful in order that has a

comprehensible meaning and against which decisions are made. Big data come from multiple establishment including sensors and free texts such as social media, unstructured data, metadata and other geospatial data collected from web logs, GPS, medical devices, etc. [15]. The big data is congregation from different sources, so it is in several forms, including:

A. Structured data

It is the controlled data in the form of tables or databases to be processed.

B. Unstructured data

It characterizes the leading proportion of data; it is the data that people assemble daily as texts, images, videos, messages, log records, click-streams etc.

C. Semi-structured data

It is look upon a variety of structured data but not deliberate in tables or databases, for example XML documents or JSON [16]

4. Segregation between Conventional Data and Big Data

In broad-range, the data in the sphere of technology is a set of letters, words, numbers, symbols or images, but with the enhancement of multitasking equipment tools the data has become different in substance and establishment [17]. In illumination of this, big data become able to be seen which differs from conventional data. Dissimilarities among customary data and big data are shown in Table1:

	Traditional Data	**Big Data**
Volume	MB and GB	PBs and ZBs
Data production Rate	Extended Period of Time	More speedy
Data Type	Structure	Semi Structure, Unstructured
Data Sources	Centralized	Multiple sources, and distributed
Data Store	RDBMS	HDFS, No SQL

Table 1 Comparison between traditional and big data

5. Cloud Computing Era

It is a section that refers to on-demand computer supplies and systems that can make available a number of incorporated computer services without being hurdle by local resources to facilitate user access. These resources include data storage, backup and self-bringing together, as well as software handing out and scheduling tasks [19]. Cloud computing is a shared resource system that can present a assortment of online services such as virtual server storage, and applications and licensing for desktop applications. By leveraging regular resources, cloud computing is able to accomplish growth and provide volume [20].

A. Characteristics of cloud computing

That cloud computing is one of the scattered systems that correspond to a convoluted model. NIST has identified important facet of the cloud, as it condensed the concept of cloud computing in five characteristics as follows:

B. On-demand self-service

Cloud services endow with computer resources such as storage and processing as needed and without any human interference.

C. Broad network access

Cloud computing resources are easy to get to over the network, mobile and pleasing to the eye devices even sensors can access computing resources on the cloud.

D. Resource Pooling

Cloud proposal users allocate a vast array of computing resources; users can patch up on the nature of resources and the geographic spot they have a preference but cannot determine the exact physical location of these resources.

E. Rapid Elasticity

Resources from storage media, network, relaxation of the rules units and applications are always available and can be increased or decreased in an exactly immediate fashion, allowing for high scalability to guarantee most favorable use of resources.

F. Measured service

Cloud systems can review the progression and spending of resources as well as examination, have power over and treatment in a completely translucent manner [21].

6. Cloud Computing Service Models

Cloud computing types are top secret on the basis of two models: cloud computing service models and cloud computing consumption models as in Figure 1

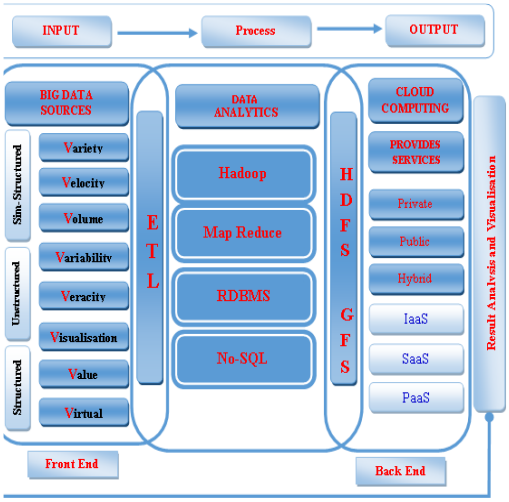

Figure 1. Cloud computing model

A. Software as a service (SAAS

Cloud examination providers make available a variety of software applications to users who can use them without install them on their computer. The user is not responsible for whatever thing other than fiddle with the settings and customizing the service as correct to his needs. SAAS helps big-data clients to perform data.

B. Platform as a service (PAAS)

Cloud service providers endow with platforms, tools and other services to users, where the cloud examination provider administers everything else,

Including the operating system and middleware with assets that make possible you to distribute the whole thing from simple cloud-based complicated.

C. Infrastructure as a service (IAAS)

Cloud overhaul providers make available infrastructure such as storage, computing ability, etc. It is a superficial appearance of cloud computing that provides virtualized computing resources greater than the Internet, In an IaaS model, a third-party contributor hosts hardware, software, servers, storage space and other communications apparatus on behalf of its users [22].

D. Desktop as a Service (DaaS

It is an alternate cloud computing model, as it be at contradiction from conventional models like (SAAS, IAAS, PAAS) in make available that data to users through the network, as data is considered the value of this model [20] in amalgamation with cloud computing based on solving some of the challenges in managing a huge amount of data. For these reasons, DaaS is for myself related to big data whose technologies must be utilized.[21] DaaS provides highly well-organized methods of data distribution and handing out. DaaS is closely related to SaaS (storage as a service) and SaaS (software as a service) which can be reciprocated with one of these models or both of them [20].

7. Testing as a Service (TaaS) on Clouds

In Cloud computing leads an opportunity in offering testing as a service (TaaS) for SaaS, clouds, and cloud-based applications. Testing as a service (TaaS) is flattering a burning research topic in both cloud computing and software engineering research communities. & solutions.

More innovative testing techniques and solutions, and QoS standards are needed to support on-demand testing services in a scalable cloud infrastructure, for example, SaaS testing adequacy and standards for multi-tenancy.

8. Cloud Testing Tool

A*. SOASTA* is aggravated through the essential to the test in construction, quite than in the laboratory environment. Today' s web applications frequently go after agile practices with in frequent builds and lofty change rates. Load testing with legacy tools in the laboratory be able to considerably different from testing in the fabrication environment in terms of scale, pattern, user profiles and network atmosphere.

B) iTKO LISA

iTKO LISA aspires to supply a cloud- based upbringing and virtual services for amalgamated application development, verification and validation. It declares to decrease software release timeline by 40% or additional by means of its innovative come close to support constant combination for development and testing.

C) Cloud Testing

Cloud Testing is instigated via a group of designers and presentation specialists from UK' s largest Website concert Monitoring & Load Testing Company. It intends to maintain traverse browser and functional testing of Web applications.

D) Cloud Sim

Cloud Sim is constructing by CLOUD (Cloud Computing and Distributed Systems) Laboratory at the University of Melbourne in Australia. It endeavors to make available a toolkit for sculpt and replicate the activities of various Cloud components including data canters, virtual machines and supply provisioning services. It can be used for investigate and weigh up cloud strategies inside a controlled simulated environment.

E) D-Cloud

D-Cloud is a devoted simulated test environment construct ahead Eucalyptus, an open-source cloud transportation as long as similar functionalities as Amazon EC2.

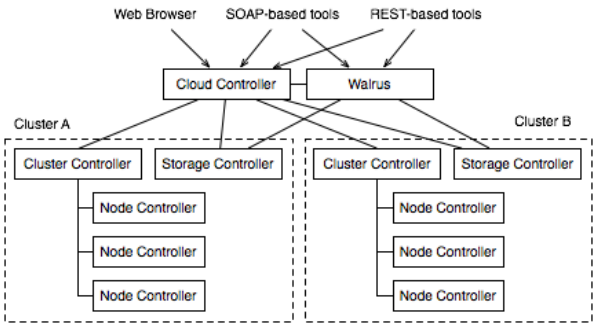

Figure 2: D-Cloud Architecture

f) Cloud9

Cloud9 an academic research project from EPFL in Switzerland, transfers symbolic implementation to the cloud platform Symbolic execution is a significant testing technique introduced in 1970s.

g): Hadoop Unit

Hadoop Unit move arounds JUnit test structure to Hadoop platform. JUnit test cases are created as autonomous Hadoop Map Reduce jobs. The map () function receives test jobs as < test name; test command > pair. At each node, the command is accomplished as a process. The reducer gets < testname; test result > from each map and combines all the results. Experiments shows that a 150-node bunch can fabricate 30 x developments compared with chronological test executions on a local computer.

h) YETI

YETI (York Extensible Testing Infrastructure) also provides a cloud version arbitrary testing tools. It uses Map Reduce to the parallelize the processing of test inputs and results. A beginning assessment was carried using Amazon EC2.

i) Cloud9

Cloud9 an academic research project from EPFL in Switzerland, transfers symbolic implementation to the cloud.

XI PERFORMANCE OF LOAD TESTING APPROACH ON CLOUD

There are various parameter which are applied by the user beyond the overload on the server to check the performance and capacity using bids.

Test Scenarios	Performed by % Users	Total Load of 40 Users	Total Load of 80 Users	Total Load of 120 Users	Total Load of 160 Users	Total Load of 200 Users
Review Pay Cheques	40%	15	32	48	64	80
Add Time Report	20%	8	16	24	32	40
View Cancel Time Report	10%	4	8	12	16	20
View Calendar Others	10%	4	8	12	16	20
View Calendar Team – 10 Direct Reports View Calendar Team – 28 Direct Reports View Calendar Team – 62 to 74 Direct Reports	4%	1 1 1	1 1 1	1 2 2	2 2 3	2 3 3
Approve Time	4%	1	3	5	7	8
Approve and Cancel Time	4%	1	3	5	6	8
Manage Direct Reports	4%	2	3	5	6	8
Successful Candidates	2%	1	2	2	3	4
Job Requisition Summary	2%	1	2	2	3	4

Table 2 Testing Parameter

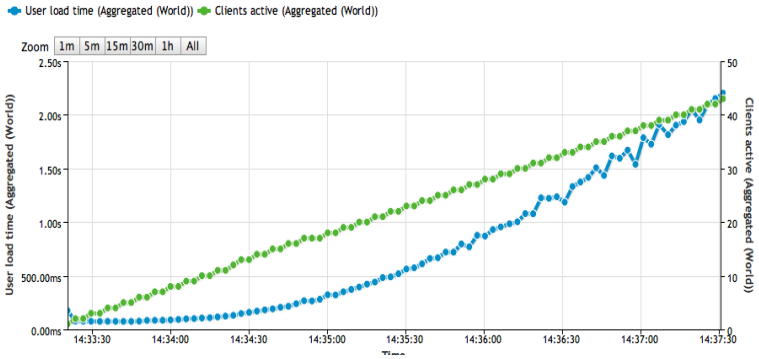

Figure 3: Performance Analysis of Load Testing on cloud.

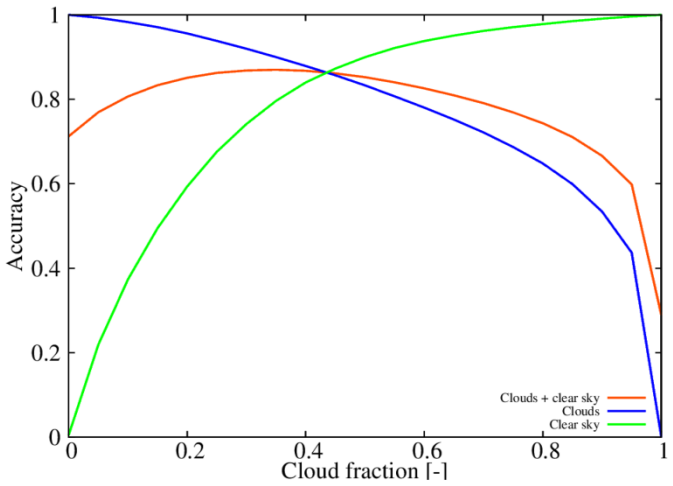

Figure 4: Cloud Fraction during Load

References

1. Charmaz, K., and A. Bryant. "The SAGE Handbook of Grounded Theory: Paperback Edition." (2010).
2. Neves, Pedro Caldeira, Bradley Schmerl, Jorge Bernardino, and Javier Cámara. "Big Data in Cloud Computing: features and issues."
3. Klous, Sander, and Nart Wielaard. We are Big Data: The Future of the Information Society. Springer, 2016.
4. https://www.ibm.com/big-data/us/en/ Bello-Orgaz G, Jung JJ, Camacho D. Social big data: Recent achievements and new challenges. Information Fusion. 2016 Mar 31;28:45-59.
5. Boyd, D., & Crawford, K. (2011, September). Six provocations for big data. In A decade in internet time: Symposium on the dynamics of the internet and society (Vol. [6]. Oxford: Oxford Internet Institute.

6. SHAN, Y. C., Chao, L. V., ZHANG, Q. Y., & TIAN, X. Y. (2017). Research on Mechanism of Early Warning of Health Management Based on Cloud Computing and Big Data. In Proceedings of the 23rd International Conference on Industrial Engineering and Engineering Management 2016 (pp. 291-294). Atlantis Press, Paris.
7. Parvin Ahmadi Doval Amiri and Mina Rahbari Gavgani, 2016. A Review on Relationship and Challenges of Cloud Computing And Big Data: Methods of Analysis and Data Transfer. *Asian Journal of Information Technology, 15: 2516-2525*
8. Chen, Min, et al. Big data: related technologies, challenges and future prospects. Heidelberg: Springer, 2014.
9. Demchenko, Yuri, et al. "Big security for big data: Addressing security challenges for the big data infrastructure." Workshop on Secure Data Management. Springer, Cham, 2013.
10. McAfee, Andrew, and Erik Brynjolfsson. "Big data: the management revolution." Harvard business review 90.10 (2012): 60-68.
11. Liebowitz, J. (Ed.). (2014). *Bursting the big data bubble: The case for intuition-based decision making*. CRC Press.
12. Sremack, Joe. Big Data Forensics–Learning Hadoop Investigations. Packt Publishing Ltd, 2015.
13. Franks, Bill. Taming the big data tidal wave: Finding opportunities in huge data streams with advanced analytics. Vol. 49. John Wiley & Sons, 2012.
14. Furht, Borko, and Flavio Villanustre. Big Data Technologies and Applications, Chapter 1, Springer, 2016.
15. Calheiros, Rodrigo N., et al. "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms." Software: Practice and experience 41.1 (2011): 23-50.
16. R.Subhulakshmi , S.Suryagandhi , R.Mathubala , P.Sumathi, An evaluation on Cloud Computing Research Challenges and Its Novel Tools, International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST) Volume 2, Special Issue 19, October 2016.
17. Fonseca, N., & Boutaba, R. (2015). Cloud services, networking, and management. John Wiley & Sons.
18. https://www.ibm.com/blogs/cloud-computing/2014/01/cloud-computing-defined-characteristics-service-levels/
19. Zhang, Q., Cheng, L., & Boutaba, R. (2010). Cloud computing: state-of-the-art and research challenges. Journal of internet services and applications, 1(1), 7-18

20. Amed, F. F. (2015). Comparative Analysis for Cloud Based e-learning. Procedia Computer Science, 65, 368-376.
21. Vacca, J. R. (Ed.). (2016). Cloud Computing Security: Foundations and Challenges. CRC Press. ch-15.
22. https://support.rackspace.com/how-to/understanding-the-cloud-computing-stack-saas-paas-iaas
23. Terzo, O., Ruiu, P., Bucci, E., & Xhafa, F. (2013, July). Data as a service (DaaS) for sharing and processing of large data collections in the cloud. In Complex, Intelligent, and Software Intensive Systems (CISIS), 2013 Seventh International Conference on (pp. 475-480). IEEE.
24. R. Calheiros, R. Ranjan, A. Beloglazov, C. De Rose, and R. Buyya, "CloudSim: a Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms," Software: Practice and Experience, vol. 41, no. 1, pp. 23–50, 2011
25. T. Banzai, H. Koizumi, R. Kanbayashi, T. Imada, T. Hanawa, andM. Sato, "D-Cloud: Design of a Software Testing Environment for Reliable Distributed Systems using Cloud Computing Technology," in Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, 2010, pp. 631–636
26. Jerry Gao, Xiaoying Bai and Wei-Tek Tsai "Cloud Testing- Issues, Challenges, Needs and Practice" Software Engineering : An International Journal (SEIJ), Vol. 1, No. 1, SEPTEMBER 201
27. M. Oriol and F. Ullah, "YETI on the cloud," in Third International Conference on Software Testing, Verification, and Validation Workshops, 2010, pp. 434–437
28. Xiaoying Bai, Muyang Li, Bin Chen, Wei-Tek Tsai, Jerry Gao "Cloud Testing Tools" Proceedings of The 6th IEEE International Symposium on Service Oriented System Engineering (SOSE 2011
29. Cloud Testing. [Online]. Available: http://www.CloudTesting.com