Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 07, July-2019

Role of Big Data in Cloud Computing: A Review

Dr. Harjinder Kaur
Principal ,
Akal Group of Technical & Management Institutions,
Mastuana Sahib,

Dr. Major Singh Goraya
Asso. Prof.,
Dept. of Computer Science & Engineering,
SLIET , Longowal

***Abstract* - Big data is a data analysis methodology that contributes in the rapid growth of various applications using in daily life like social network analysis, semantic web analysis and bioinformatics network analysis. Cloud computing is a model that is spreading everywhere in order to deliver big data services. This paper presents a brief introduction about the prototype of cloud computing. Cloud service models and deployment models explored in the prototype. Overview of big data along with its technologies is discussed. The basic issues and challenges in delivering big data are mentioned in rest of the paper.**

Index Terms - Cloud computing, cloud service model, cloud deployment model, big data, big data technologies.

I. INTRODUCTION

Cloud is a server collection that is distributed across the internet to store, process and manage the data. Cloud computing provides the hardware and software services through internet. It enables big data to manage and allocate the stored data in proper manner[9]. Cloud computing provide the security to big data with hadoop, environment like file and network, encryption, logging, nods authentication, layered frame work for assuring cloud etc. Different types of models are using for mining the big data. Cloud computing approach is valuable as it has the adopted technologies to handle the large amount of data. It provides the interconnectivity between the devices and data that is further assist to exchange the data and connected to other devices. In 2014, the connected devices were 3.7 billions and it will reach at estimated 25 billions till 2020. Big data term is basically used to convert the data into information. The major purpose of big data is to store and manage, visualize and analyze huge amount of data per day.

II. CLOUD COMPUTING

Cloud computing provides enormous computing resources to the user applications through internet in the large scale distributed computing environment. Cloud computing is characterized by multiagency, fast deployment, low cost, scalability, rapid provisioning, elasticity and ubiquitous network access [1]. Therefore, users can access the resources as per requirement from the cloud on pay-as-use basis. They are not required to buy and install the required resources locally[2]. For example, if we need a particular resource occasionally then it is the better option to use the resource from cloud instead of investing to own the resource. In the modern world in the presence of cloud, we can store the large data on cloud and may access it whenever and wherever required. Cloud computing is inherently flexible architecture as we can access the resources on our demand and need basis [1,3]. Resources are always available on the cloud and are made available to different user domains on demand basis. Furthermore, through virtualization the high capacity servers in cloud are provisioned to different users in parallel. Virtual machines which are fabricated using virtualization software are allocated to the users instead of the physical machines [4]. Cloud computing reduces the resource cost by pay-as-per usages and also offers high storage capacity.

III. CLOUD SERVICE MODELS AND DEPLOYMENT MODEL

A. Cloud Service Models

a. ***Software as a Service (SaaS)*** **-** In this service model user can access the software applications and database existing on the cloud. It provides the environment to run a particular application. If a user does not have the required software or hardware resource locally, he can access it from the cloud [5].

b. ***Platform as a Service (PaaS)*** - In this service model user can configure and install the software on the cloud. PaaS provides operating system, programming language and web server to design the software. User can save the time and cost to buy all these resources [5].

c. ***Infrastructure as a Service (IaaS)*** - In this service model cloud provides infrastructure like virtual machines, storage, network, IP addresses and other specialized software or hardware resources through high speed networks to the users [5]. In IaaS, an environment is provided to deploy and run the infrastructure (hardware/software) in the distributed cloud environment.

B. Cloud Deployment Models

a. ***Public Cloud*** - In public cloud, cloud service is accessible to all the users in the public domain. Any organization and person can access the resource and data from the cloud directly without the involvement of any third party [5]. Cloud users pay for the resources (hardware and software) as per usage.

b. ***Private Cloud*** - In this deployment model different organizations built and maintain their local cloud. Therefore it is known as private cloud. Private cloud is primarily controlled and maintained by the owning organization itself [5]. Private cloud serves the needs internal needs of an organization. Due to local control,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 07, July-2019

privacy and security can be maintained at a higher level in the private cloud.

c. ***Hybrid Cloud*** - Private and public cloud can be combined to maintain a hybrid cloud. An organization may decide to access a few services which involve high costs from the public cloud. Sometimes the critical services may be confined to private cloud whereas other services may be accessed from the public cloud. Thus two clouds are integrated together which forms a hybrid cloud. User of an organization can access the public cloud through the private cloud.

d. ***Community Cloud*** - Many organizations may decide to implement and maintain a cloud together. This cloud is called a community cloud. Community clouds are preferred when organizations want to share their resources together. Through community clouds the common resources may be easily shared and distributed among the participant organizations.

IV BIG DATA

Big data sports by the new technologies and architecture that is basically used for data capture, storage and analysis. Big data is implemented in various ways like e-mail, mobile device output, sensor generated data and social media output. Big data has the need of large storage capacity. Hence, huge data could not be stored at local data base system. Another point is that data is in the form of different structure and use for different source[6]. Big data is the concept of information and communication technology. It is put up the contribution to solve the problem of data warehouses and data mining. Here data mining is a big issue to handle the big database of data warehouses[7]. According to Gartner, " Big data are high-volume, high velocity and high variety information assets that required new forms of processing to enable enhanced decision making, insight discovery and process optimization"[7]. Big data signify six terms- volume, variety, velocity, veracity, value and complexity

a) Volume-it represents to the expanding of data beyond terabytes like transcation data, sensor data.
b) Variety-it refers to the collection of data from various different sources like machine, sensors etc. for example e-mails, audio-visuals, text document etc.
c) Velocity-How data is processed at fast speed. How fast data is accessed. Moreover IOT emerge the new type of data that is collection of sensor data and the control of equators. Big data is implemented in the form of different applications in the smart cities.
d) Veracity- it means collected data as different qualities with different accuracy, coverage and timeliness must be compliance[14].
e) Value-Big data is provided after the processing and analyzing data. Further store data can be used for further uses with the combination of other data sets.
f) Complexity- it manage the complexity of multiple sources data is linked, matches, cleansed and transformed before delivered

Data is distributed on different servers that is stored by the user. Cloud Service Provider (CSP) is delivered the data to the different clients by ensuring the integrity and confidently. It apply the authentication, non-delicacy and data recovery to control the available data.

Third Party Auditor(TPA) – Security is implemented by secure socket layer, point to point tunneling protocol and virtual private networks[8]. Yet various user are accused the data by unauthorization. To overcome from such type of problem third party authentication mechanism is applicable on users as well as cloud service providers. TPA keep track of data transmission and technologies and techniques that apply on data. It is passes through the planning, execution and reporting. Security is checked out at data integrity level.

Encryption Based Storing of Data – To provide such type of security cryptographic techniques are used to encode the server of cloud storage [8]. A key is provided to all the users that store data on storage cloud. Hence, only those users accessed the data from cloud which have access key. Whenever they want to access the data then data is decrypted before retrieved from the cloud. New access key is provided to the user who stored the data on cloud first time.

Privacy Preserving Public Auditing - Homomorphic authenticators technique is used to provide secure data to the users that ensure to the correctly computed of data blocks[8]. It is the combination of four algorithms. Keygen algorithm generates the key for user to access the data. Singen algorithm verifies the metadata along with digital signature. Genproof algorithm generates proof of secured and intact data. Verify algorithm verifies the proof generated by Genproof algorithm. It works in setup phase and audit phase. In setup phase local copy of data is deleted and altered the data files. In audit phase a report is created and delivered to the service provider.

Non Linear Authentication - Through this technique homomorphic non linear authenticator is used randomly[8]. RSA algorithm is used to encrypt and decrypt the data that follows the digital signature for authentication. Extensible authentication protocol is used with RSA especially for handshaking scheme. Client first sends the request to the cloud service provider that is further calculated by hash function. If the value is matched it is authenticated otherwise dismissed.

Secure and Dependable Storage - To eliminate the problems that have occurred in the data storage on cloud error localization technique is used [8]. It verifies the accuracy and provides ensured data security. It uses homomorphic token which verifies the data through erasure coded data. It identifies the error and bad performance of server. Such method monitors errors by only one server at particular time because sometimes it may lead to server failure.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 07, July-2019

V TECHNOLOGIES OF BIG DATA

Hadoop - it is used to form the clusters of data nodes and store data on space utilization. It execute on different environment to handle and transfer data among racks[13]. It is a java-based programming frame work. Hadoop is the part of Apache project sponsored by Apache Software Foundation[15]. Hadoop distributed data on different servers that helps to run the different application. It has lower rate of system failure although various nodes clusters fail. Hadoop is featured by scalable, cost effective, flexible and fault tolerant. Hadoop is used by various popular companies like Google, Yahoo, Amazon, IBM etc[15]. Hadoop architecture is described as task trackers, job trackers, data engine and fetch manager[16]. Task trackers are used for running the tasks. Job trackers manage the cluster resourcing and scheduling the all jobs. Data engine provide information of processing data. Fetch manager fetches the data during the execution of specific task. Hadoop frame is support in various applications[9].

a) **Hadoop Distributed File System (HDFS)** – It able to tolerate the failure of system as well as store huge amount of data. Clusters are created by hadoop to distribute the data among machines. HDFS splits the files into blocks and stored it on the server. It stores three copies of data on different servers [10]. Collection of data files corresponding to both data node and name node. Name node is responsible for accessing of all types of files and data node interacts with itself to perform the operations on file system[11]. Information can be accessed through file system by interacting to name node and data node. Data node provides the data that is queried by the clients which further mention in name node [12].

b) **Map-reduce** – Map-reduce is used to write applications that handles the large amount data processing in a reliable and fault tolerant manner [15]. It splits the data into chunks that are parallel processed through Map jobs. The input and output data during the processing is stored in file system. It also monitoring and re-executing the failed task. The distribution of data is implemented two steps map step and reduce step[7]. It is helpful in solving the large data problems. A query is created and data of related query is mapped to access the related data [6]. The data is further reduced to view the data according to query. Due to the configuration storage requirement, map-reduce is processing through the cloud service providers. It is based on master slave architecture. Master node is basically monitoring, scheduling and re-execution the jobs and slave nodes works according to the direction of master node[9].

VI CHALLENGES IN BIG DATA

a) **Availability** - Cloud computing allowed to access the data by individually authorized. Due to the distribution of data on various clouds, it decreases the performance of accessing data[13]. Another issue is to tackle the transforming of data into suitable forms.

b) **Security -** The security becomes at high risk when users accessed personal and other sensitive information through credit/debit cards[9]. Moreover, different organizations have their own rules and regulations regarding the secure information. Hence, multi-level security is required for all types of data [7]. It needs a privacy preserved data model to provide security to the sensitive data of any organization or personal information. Sometimes hackers store or lose the data due to the poor security.

c) **Scalability** - There is a mismatching of data speed and CPU speed. Due to the large volume of data, it is not delivered to the processor at proper time. Many applications require parallel computing like navigation, social networks, finance, internet search, timeliness etc[7]. There is a need of the cloud service to provide the services of infrastructure, platform and applications at required time to maintain the scalability[9].

d) **Big Storage** - Data can be in different forms it may be text, images, audio/video etc. Such type of data is use by different mediums like mobile devices, aerial sensory technologies, remote sensing, radio frequency identification readers etc. Such type of data require large storage device with huge space and higher input/output speed[7]. It is most difficult to access the information from unstructured data[13]. Large amount of data could not be retrieved at proper time. Hence, it becomes more tuff in file systems.

VII CONCLUSION

In the recent time cloud computing has emerged as a paradigm in computing science. The primary reason for its large scale adaptability is cost saving by using the remote computing resources as per demand as well as flexibility. Big data is an emerging platform to manage and distribute the large scale data. It converts the traditional data base techniques into effective innovative and machine learning techniques. HDFS and Mapreduce techniques are use to deal with huge amount of data. Although, big data is serve with numerous facilities still it has many issues and challenges.

VIII REFRENCES

[1] Sumit Jaiswal, Subhash Chandra Patel and Ravi Shankar Singh, “Secured Outsourcing Data & Computation to the Untrusted Cloud – New Trend”, CSI Communications, Vol. 38 (12), 2015.
[2] http://csrc.nist.gov/publications/nistpubs/ 800-145/SP800-145.pdf
[3] Torry Harris, “Cloud Computing - An Overview”
[4] J. Srinivas, K.Venkata Subba Reddy and Dr. A. Moiz Qyser, “Cloud Computing Basics”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 1(5), 2012.
[5] www.Wikipedia.com
[6] Bernice M. Purcell, “Big data using cloud computing”, Journal of Technology Research, http://www.aabri.com
[7] D.P. Acharjya, Kauser Ahmed P, “ A Survey on Big Data Analytics: Challenges, Open Research Issue and Tools”, International Journal of Advanced Computer Science and Applications, Vol. 7(2), 2016
[8] J. Raja, M. Ramakrishan, “A Comprehensive Study on Big Data Security and Integrity Over Cloud Storage”, Indian Journal of Science and Technology, Vol. 9(40), October 2016
[9] Sheetal Singh, Vipin Kumar Rathi, Bhawna Chaudhary, “Big Data and Cloud Computing: Challeges and Opportunities”, International Journal of Innovations in Engineering and technology, Vol. 5(4), August 2015.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 07, July-2019

[10] Iqbaldeep Kaur, Navneet kaur, Amandeep Ummat, Jaspreet Kaur, Navjot Kaur, " Research Paper on Big Data and Hadoop", International Journal of Computer Science and Technology, Vol. 7940, Oct-Dec. 2016.

[11] Rehana Hassan, Rifat Manzoor, Mir Shahnawaz Ahmad, " Big Data and Hadoop : A Survey", International Journal of Computer Science and Mobile Computing, Vol. 6 (7), July 2017

[12] Sumit Kumari, " A Review paper on Big Data and Hadoop"

[13] S. Harini, K. Jothika and K. Jayashree, "A review of big data computing and cloud", International Journal of Pure and Applied Mathematics, Vol. 118(18), 2018

[14] Rohit Chandrashekar, Maya Kala, Dashrath Mane, "Integration og Big Data in Cloud computing environment for enhanced data processing capabilities", International Journal of Engineering Research and General Science, Vol. 3(2), 2015.

[15] V Akhila Reddy, G Rakesh Reddy, " Study and Analysis of Big Data in Cloud Computing" , International Journal of Advance Research in Computer Science and Management Studies, Vol. 3(6), 2015.

[16] Venkatesh H, Shrivasta D Perur, Nivedita Jalihal, " A Study on Use og Big Data in Cloud Computing Environment", International Journal of Computer Science and Information Technologies, Vol. 6(3), 2015.