14th International Conference on Recent trends in Engineering, Applied Science and Management
Vedant College of Engineering & Technology, Bundi, Rajasthan (India) (ICRTESM-19)
16th February 2019 www.conferenceworld.in ISBN : 978-93-87793-71-2

ADVANCE COMPUTING ENVIRONMENT OF DATA PROCESSING IN TO CLOUD WITH BIG DATA ROTATION

Vijay Malav[1] , Dr. Amit Sharma[2]

[1]*M. Tech. Research Scholar, Department of Computer Science & Engineering , Vedant College of Engineering & Technology ,Bundi ,Rajasthan, India*
[2] *Professor, Computer Science & Engineering Department, Vedant College of Engineering and Technology, Bundi, Rajasthan, India*

Abstract: *Monstrous development in the size of information or huge information created through distributed computing have been watched. Tending to huge information is a testing and time-requesting assignment that requires a substantial computational foundation to guarantee fruitful information handling and information investigation. Gigantic capacity and figuring for information curation are requests of huge information registering and furthermore, it requests preparing that could be conveyed from cloud frameworks. Distributed computing is a model that is spreading wherever so as to convey huge information administrations. Little to extensive organizations would to a great extent profit by enormous information. Incorporating huge information registering and cloud would mean making best utilization of the striking highlights in both. The ascent of huge information in distributed computing is explored in this examination.*

Keyword: Big data computing, Cloud computing, data analysis, data processing

I INTRODUCTION

Cloud administration models regularly comprise of PaaS, SaaS, and IaaS. PaaS, for example, Google's Apps Engine,Salesforce.com, Force stage, and Microsoft Azure, alludes to various assets working on a cloud to give stage registering to end clients. SaaS, for example, Google Docs, Gmail, Salesforce.com, and Online Payroll, alludes to applications working on a remote cloud foundation offered by the cloud supplier as administrations that can be gotten to through the Internet . SaaS has certain benefits, for example, less demanding programming organization, programming similarity over the business, less demanding coordinated effort, and worldwide openness . IaaS, for example, Flexi scale and Amazon's EC2, alludes to equipment gear working on a cloud given by specialist organizations and utilized by end clients upon interest . This is accepting most consideration from market because of arrangement of administrations, for example, fiasco recuperation, PC as an administration, stockpiling as an administration, work area foundation, cloud blasting and so on., along these lines giving pinnacle load ability to variable procedures.

II CLOUD STORAGE & SERVICES

Cloud storage is a model of computer data storage in which the digital data is stored in logical pools. The physical storage spans multiple servers (sometimes in multiple locations), and the physical environment is typically owned and managed by a hosting company. These cloud storage providers are responsible for keeping

the data available and accessible, and the physical environment protected and running. People and organizations buy or lease storage capacity from the providers to store user, organization, or application data.

Cloud storage services may be accessed through a collocated cloud computing service, a web service application programming interface (API) or by applications that utilize the API, such as cloud desktop storage, a cloud storage gateway or Web-based content management systems.

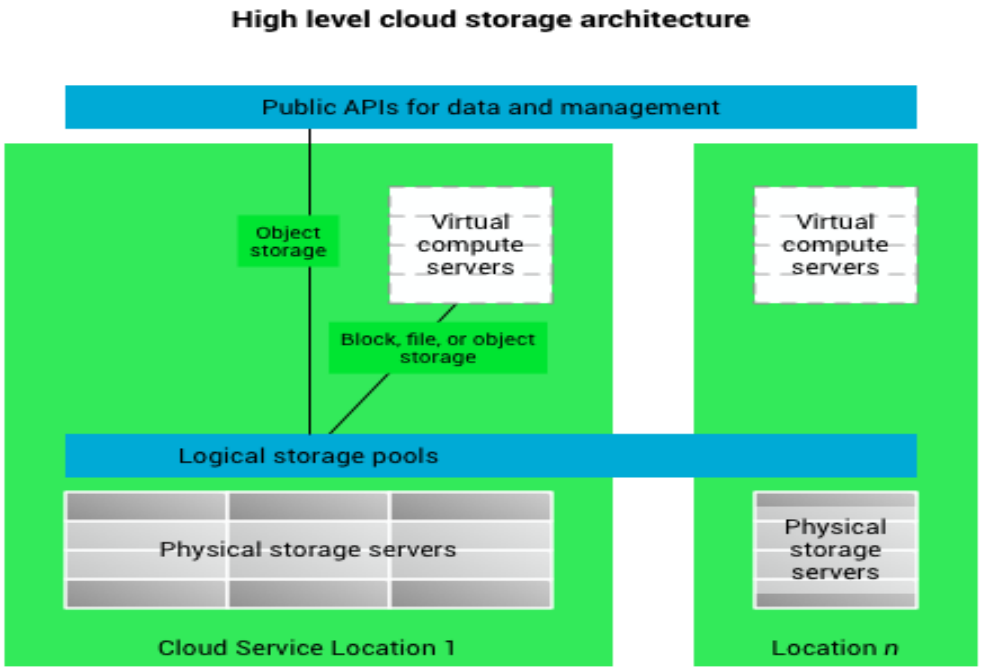

Figure 1 Cloud Storage Architecture

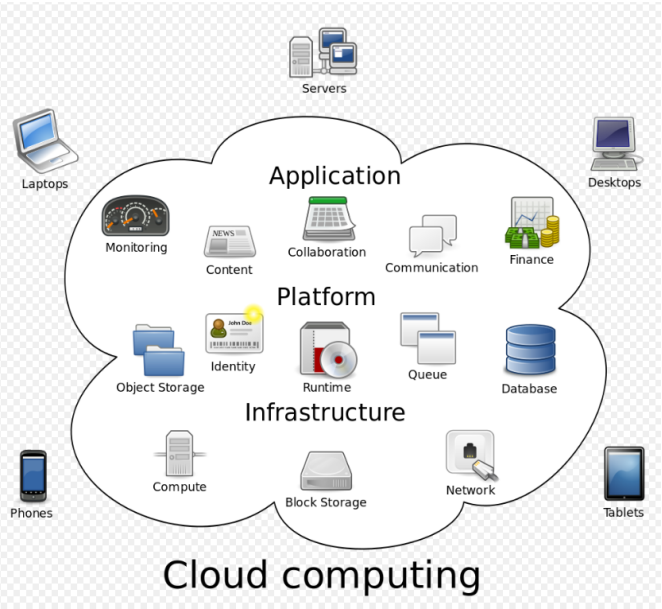

Figure 2 Cloud Services

2.1 Cloud Computing

Genesis of the term „cloud computing" is ambiguous, although it sounds to be borrowed from the habit of employing sketches of clouds to symbolize networks. The custom of catering remote connection to computing activity through networks contributed to prevalent usage of this caption. Cloud computing cite to an exemplary of network computing where a program or utility executes on a connected servers instead of confined computing apparatus. Corresponding to the conventional client-server or mainframe model, a node associates with a server to accomplish a job.

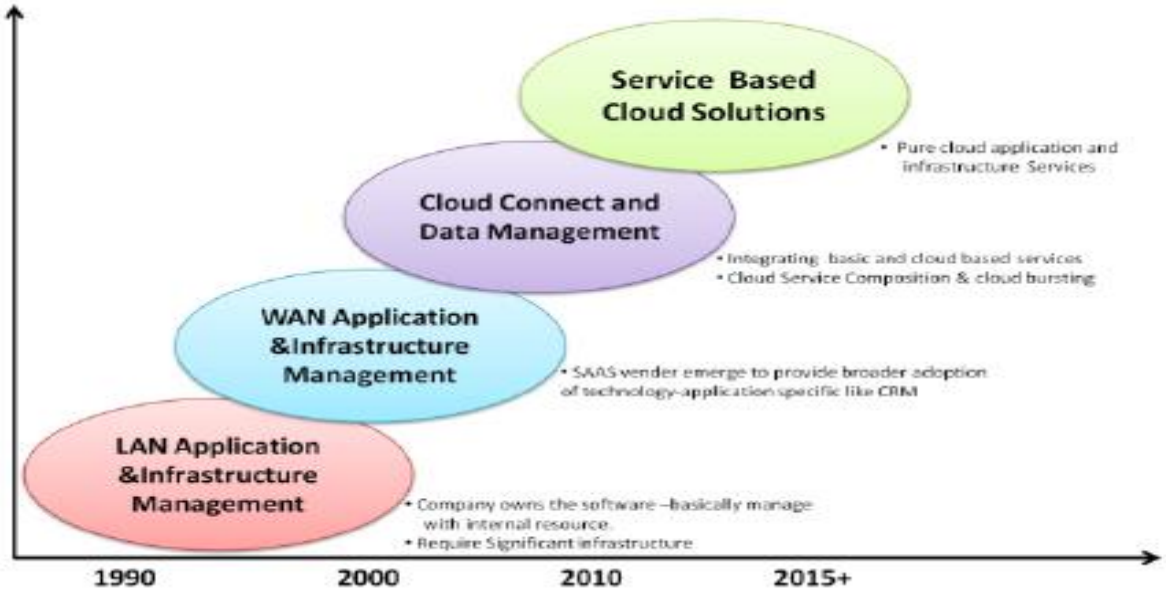

Figure 3: Year wise Expansion of Cloud

The digression with cloud computing is that the computation may be executed on a single or many linked nodes at the same instance, applying the notion of virtualization. Virtualization allows multiple servers to be designed and distributed among several autonomous „virtual" servers, operating separately seeming to the node to be a single device. These virtual servers are core, extensile, mountable and un-mountable, un-influencing the nodes.

2.3. CLOUD ENVIRONMENT LAYERS Cloud computing attracts many managers and organizations. There are many similar terminologies that are usually utilized for describing cloud computing, these terms such as: distributed, grid, cluster, virtualization, on-demand, utility, and software-as-a-service. In other words, cloud computing refers to end -users connecting with applications running on sets of shared servers, often hosted and virtualized, instead of a traditional dedicated server.

A. Deployment Models The deployment models can be categorized into four categories namely Public cloud, Private cloud, Hybrid cloud and Community cloud [6]. These categories will be described in details as follows:-

- **Public Cloud,** this model is owned by an organization for selling the cloud services and the design of infrastructure is made in order to be available for industries, organizations and businesses.

- **Private Cloud,** this model is managed by the organization itself or by a third party. Private cloud may be either off or on premises. The major characteristic of this model is that the infrastructure of the cloud is private, in addition to its availability to a single organization.
- **Hybrid Cloud,** this model is similar to the private cloud as it is managed by third party or by organization itself and may exist off or on premises. But the cloud infrastructure may combine two or more clouds (public, private or community).
- **Community Cloud,** this model is similar to the previously mentioned private and hybrid cloud as the organization or third party are allowed to manage it and also exists off or on premises. But in community cloud, multiple organizations with common interests, requirements, or considerations share the infrastructure.

III BIG DATA

Big Data is a data analysis methodology enabled by some new generation of technologies and architecture which support high-velocity data capture, storage, and analysis. This refers to dealing with data that may be structured or unstructured or semi-structured, textual or non-textual and the amount of data is voluminous. Big data and traditional data warehousing systems have similar goals to deliver business value through the analysis of data, the difference lies in the analytics methods and the organization of the data . The large volume and the variety in data will demand some pre-processing tasks in order to integrate the data, clean it, and filter it . Big data is characterized by five Vs namely, Volume, Velocity, Variety, Veracity, and Value .

1. Volume: Big data comes in one size: XXL ranges from 30-50 terabytes through enterprises. The available storage cannot handle structured and unstructured data thus causing a big problem for enterprises.

2. Velocity: Velocity defines the speed of data that enters the enterprise which is then analyzed to increase the profit of business before the value of the information lost.

3. Variety: Data can be structured, unstructured, semi structured or a mix of three. It comes in a variety of forms like logs files, tweets, images, videos, audio, text, PDF files, click streams etc.

4. Veracity: Veracity means "compliance with truth or fact". Data sources are of different qualities with differences accuracy, coverage and timeliness.

5. Value: It refers to the processing of the data and producing it during analysis. Value of data is not for single time use and will be reused in the future in combination with other data sets.

Big data computing is a budding theory or concept in data science that focuses on multidimensional information mining for scientific discovery and business analytics over large-scale infrastructure. The data collected/produced from several scientific explorations and business transactions often require tools to achieve proficient data management, analysis, validation, visualization, and dissemination and at the same time preserving the inherent value of the data . Cloud is a collection of a number of terrestrial servers spread across the internet and these are used to collectively store, manage and process data. Cloud computing is the use of hardware and software services that are delivered as a service over the internet. Platform as a Service (PaaS), Infrastructure as a Service (IaaS), and Software as a Service (SaaS) are the three cloud computing models .

Unlike the traditional Cloud services, analytics deals with high-level capabilities which demand highly specialized resources such as data and analysis skills of domain experts. For this reason, under certain business models – especially those where data and models reside on the provider's premises – not only ordinary Cloud services, but also the skills of data experts need to be managed. To achieve economies of scale and elasticity, Cloud-enabled Big Data analytics needs to explore means to allocate and utilized specialized resources in a proper manner . There are some challenges in migration of big data into cloud, such as: i) Scalable data management ii) Data management for large applications iii) Large multitenant databases iv) Security issues for cloud computing, big data, map reduce and Hadoop environment.

IV BIG DATA PROCESSING TOOLS

BIG DATA is a term used for a collection of data sets so large and complex that it is difficult to process using traditional applications/tools. It is the data exceeding Terabytes in size. Because of the variety of data that it encompasses, big data always brings a number of challenges relating to its volume and complexity. A recent survey says that 80% of the data created in the world are unstructured. One challenge is how these unstructured data can be structured, before we attempt to understand and capture the most important data. Another challenge is how we can store it. Here are the top tools used to store and analyses Big Data. We can categories them into two (storage and Querying/Analysis).

1. Apache Hadoop

Apache Hadoop is a java based free software framework that can effectively store large amount of data in a cluster. This framework runs in parallel on a cluster and has an ability to allow us to process data across all nodes. Hadoop Distributed File System (HDFS) is the storage system of Hadoop which splits big data and distribute across many nodes in a cluster. This also replicates data in a cluster thus providing high availability.

2. Microsoft HDInsight

It is a Big Data solution from Microsoft powered by Apache Hadoop which is available as a service in the cloud. HDInsight uses Windows Azure Blob storage as the default file system. This also provides high availability with low cost.

3. NoSQL

While the traditional SQL can be effectively used to handle large amount of structured data, we need NoSQL (Not Only SQL) to handle unstructured data. NoSQL databases store unstructured data with no particular schema. Each row can have its own set of column values. NoSQL gives better performance in storing massive amount of data. There are many open-source NoSQL DBs available to analyse big Data.

4. Hive

This is a distributed data management for Hadoop. This supports SQL-like query option HiveSQL (HSQL) to access big data. This can be primarily used for Data mining purpose. This runs on top of Hadoop.

5. Sqoop

This is a tool that connects Hadoop with various relational databases to transfer data. This can be effectively used to transfer structured data to Hadoop or Hive.

6. PolyBase

This works on top of SQL Server 2012 Parallel Data Warehouse (PDW) and is used to access data stored in PDW. PDW is a datawarhousing appliance built for processing any volume of relational data and provides an integration with Hadoop allowing us to access non-relational data as well.

7. Big data in EXCEL

As many people are comfortable in doing analysis in EXCEL, a popular tool from Microsoft, you can also connect data stored in Hadoop using EXCEL 2013. Hortonworks, which is primarily working in providing Enterprise Apache Hadoop, provides an option to access big data stored in their Hadoop platform using EXCEL 2013. You can use Power View feature of EXCEL 2013 to easily summarise the data. (More information).

Similarly, Microsoft's HDInsight allows us to connect to Big data stored in Azure cloud using a power query option.

V Security and Privacy

Cloud computing poses privacy concerns because the service provider can access the data that is in the cloud at any time. It could accidentally or deliberately alter or delete information. Many cloud providers can share information with third parties if necessary for purposes of law and order without a warrant. That is permitted in their privacy policies, which users must agree to before they start using cloud services. Solutions to privacy include policy and legislation as well as end users' choices for how data is stored. Users can encrypt data that is processed or stored within the cloud to prevent unauthorized access.

According to the Cloud Security Alliance, the top three threats in the cloud are Insecure Interfaces and API's, Data Loss & Leakage, and Hardware Failure—which accounted for 29%, 25% and 10% of all cloud security outages respectively. Together, these form shared technology vulnerabilities. In a cloud provider platform being shared by different users there may be a possibility that information belonging to different customers resides on same data server. Additionally, Eugene Schultz, chief technology officer at Emagined Security, said that hackers are spending substantial time and effort looking for ways to penetrate the cloud. "There are some real Achilles' heels in the cloud infrastructure that are making big holes for the bad guys to get into". Because data from hundreds or thousands of companies can be stored on large cloud servers, hackers can theoretically gain control of huge stores of information through a single attack—a process he called "hyperjacking". Some examples of this include the Dropbox security breach, and iCloud 2014 leak. Dropbox had been breached in October 2014, having over 7 million of its users passwords stolen by hackers in an effort to get monetary value from it by Bitcoins (BTC). By having these passwords, they are able to read private data as well as have this data be indexed by search engines (making the information public).

There is the problem of legal ownership of the data (If a user stores some data in the cloud, can the cloud provider profit from it?). Many Terms of Service agreements are silent on the question of ownership.[109] Physical control of the computer equipment (private cloud) is more secure than having the equipment off

site and under someone else's control (public cloud). This delivers great incentive to public cloud computing service providers to prioritize building and maintaining strong management of secure services. Some small businesses that don't have expertise in IT security could find that it's more secure for them to use a public cloud. There is the risk that end users do not understand the issues involved when signing on to a cloud service (persons sometimes don't read the many pages of the terms of service agreement, and just click "Accept" without reading). This is important now that cloud computing is becoming popular and required for some services to work, for example for an intelligent personal assistant (Apple's Siri or Google Now). Fundamentally, private cloud is seen as more secure with higher levels of control for the owner, however public cloud is seen to be more flexible and requires less time and money investment from the user.

VI CONCLUSION

Diagram of bigdata approach, birthplace, application, difficulties and processing are given in this paper.Challenges and issues, asset the board, execution and enhancement of bigdata organizing in distributed computing are likewise examined in this paper. Bigdata is a developing stage to address the largescale information for basic leadership and learning. Huge information changed the current conventional database methods with compelling registering and information remaining task at hand preparing with new inventive systems and machine learning strategies. The Hadoop MapReduce is the best answer for headway of cloud highlights. Relocation calculation and open source methods will change the cloud.

VII Reference

[1] Abdullah, A. M., & Aziz, R. H. H. (2016, June). New Approaches to Encrypt and Decrypt Data in Image using Cryptography and Steganography Algorithm., International Journal of Computer Applications, Vol. 143, No.4 (pp. 11-17).

[2] Singh, G. (2013). A study of encryption algorithms (RSA, DES, 3DES and AES) for information security. International Journal of Computer Applications, 67(19).

[3] Gaj, K., & Chodowiec, P. (2001, April). Fast implementation and fair comparison of the final candidates for Advanced Encryption Standard using Field Programmable Gate Arrays. In Cryptographers' Track at the RSA Conference (pp. 84-99). Springer Berlin Heidelberg.

[4] Stallings, W. (2006). Cryptography and network security: principles and practices. Pearson Education India.

[5] Yenuguvanilanka, J., & Elkeelany, O. (2008, April). Performance evaluation of hardware models of Advanced Encryption Standard (AES) algorithm. In Southeastcon, 2008. IEEE (pp. 222-225).

[6] Lu, C. C., & Tseng, S. Y. (2002). Integrated design of AES (Advanced Encryption Standard) encrypter and decrypter. In Application-Specific Systems, Architectures and Processors, 2002. Proceedings. The IEEE International Conference on (pp. 277-285).

[7] Mohamed, A. A., & Madian, A. H. (2010, December). A Modified Rijndael Algorithm and it's Implementation using FPGA. In Electronics, Circuits, and Systems (ICECS), 2010 17th IEEE International Conference on (pp. 335-338).

[8] Pramstaller, N., Gurkaynak, F. K., Haene, S., Kaeslin, H., Felber, N., & Fichtner, W. (2004, September). Towards an AES crypto-chip resistant to differential power analysis. In Solid-State Circuits Conference, 2004. ESSCIRC 2004. Proceeding of the 30th European IEEE (pp. 307-310).

[9] Deshpande, H. S., Karande, K. J., & Mulani, A. O. (2014, April). Efficient implementation of AES algorithm on FPGA. In Communications and Signal Processing (ICCSP), 2014 IEEE International Conference on (pp. 1895-1899).

[10] Nadeem, H (2006). A performance comparison of data encryption algorithms," IEEE Information and Communication Technologies, (pp. 84-89).

[11] Diaa, S., E, Hatem M. A. K., & Mohiy M. H. (2010, May) Evaluating the Performance of Symmetric Encryption Algorithms. International Journal of Network Security, Vol.10, No.3, (pp.213-219).

[12] Jain, R., Jejurkar, R., Chopade, S., Vaidya, S., & Sanap, M. (2014). AES Algorithm Using 512 Bit Key Implementation for Secure Communication. International journal of innovative Research in Computer and Communication Engineering, 2(3).

[13] Selmane, N., Guilley, S., & Danger, J. L. (2008, May). Practical setup time violation attacks on AES. In Dependable Computing Conference, 2008. EDCC 2008. Seventh European (pp. 91-96). IEEE.

[14] Berent, A. (2013). Advanced Encryption Standard by Example. Document available at URL http://www. networkdls. com/Articles/AESbyExample. pdf (April 1 2007) Accessed: June.

[15] Benvenuto, C. J. (2012). Galois field in cryptography. University of Washington.

[16] Lee, H., Lee, K., & Shin, Y. (2009). Aes implementation and performance evaluation on 8-bit microcontrollers. arXiv preprint arXiv:0911.0482.

[17] Padate, R., & Patel, A. (2014). Encryption and decryption of text using AES algorithm. International Journal of Emerging Technology and Advanced Engineering, 4(5), 54-9.

[18] Reddy, M. S., & Babu, Y. A. (2013). Evaluation of Microblaze and Implementation of AES Algorithm using Spartan-3E. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 2(7), 3341-3347.