



Big Data in Cloud Computing Enhancing the Future Tehhnologies

Neha Rajendra Thakare

B.E. 4th Yr. CSE

Department of Computer Science
and Engineering

Mauli Group of Institution College of Engineering and Technology Shegaon

Abstract

Big Data is collection of data which is in a huge size. It refers to diverse set of information that grow at increasing data. As data is growing rapidly with time, addressing such a big data is highly challenging and time demanding task. It needs a large computational infrastructure for successful data processing & analysis. The term cloud computing means storing and accessing the data and programs on remote servers that are hosted on the internet. Big data solves most of the present problems but still need improvements. At present levels of services required to improve execution efficiency. In present era Cloud is using big data processing technology to enhance application aggregation, data aggregation and data utilization. Cloud computing is used to eliminate expensive computing hardware, software and space. Hence it is best technology for complex computing. This paper deals data processing in cloud computing environments using Big Data applications and an overview of both technologies and cases of success when integrating big data and cloud technology.

KEYWORDS:- Cloud Computing, Big Data, data aggregation, computational infrastructure, complex computing .

I. INTRODUCTION

In recent years, there has been an increasing demand to store and process more and more data, in domains such as finance, science, and government. The continuous increase in the volume of data captured by organizations, such as the increase of social media, Internet of Things (IoT) & multimedia, has produced an enormous flow of data in either structured or unstructured format. Systems that support big data, and host them using cloud computing, have been. Our contributions to the current state-of-the-art is done by providing an overview over the issues to improve or have yet to be addressed in both technologies.

Systems that support big data, and host them using cloud computing, have been. Our contribution to the current state-of-the-art is done by providing an overview over the issues to improve or have yet to be addressed in both technologies.

One of the core challenges in the context of Cloud computing is the management of very huge volumes of data. This is totally independent of the resource type which is shared in the Cloud – data bases are either directly visible or accessible to clients as part of the Infrastructure or are hidden behind service interfaces.

It means that data required be partitioning and replicating across dissimilar data centers on internet. Best search engines such as Amazon, Google, have begun to establish new data centers for providing Cloud computing applications. Society is becoming increasingly more instrumented and storing vast amounts of data.

Analytics solutions that structured and unstructured data are important as they can help organizations to gain insights not only from their privately acquired data, but also from huge amounts of data publicly available on the Web.



Fig. A - Big Data-Cloud Computing

This paradigm is being well liked termed as Big Data. Big data are distinguished by three aspects: (a) numerous data, (b) data cannot be classified into regular relational databases, and (c) data are generated, captured, and processed speedily. The progress in data storage and mining technologies allow for the preservation of increasing large amount of data described by a change in the nature of data held by organizations.

II. BIG DATA

Big data is a term that describes large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information.

Big Data Resources –

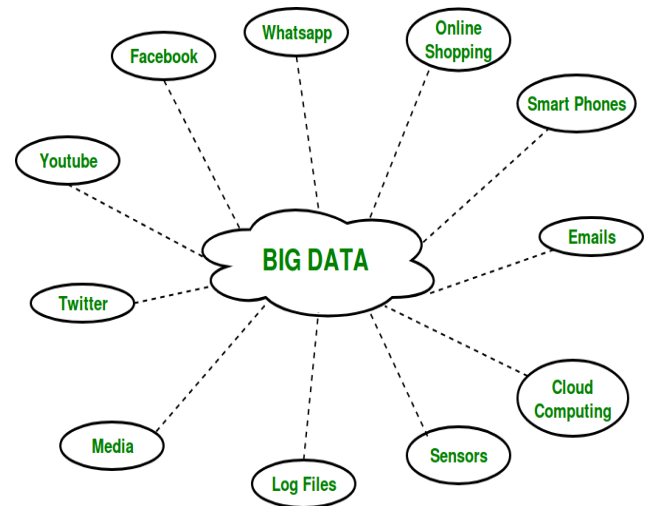


Fig. B - Big Data Resources

The prime objective of big data analysis is to process data of high volume, velocity, variety, value and veracity using various traditional and computational intelligent techniques.

Big data concept four major dimensions or properties named (7Vs).

1. Volume

According to the amount of data located in the storage medium. The volume of the data is its size. It's a massive amount of data in data stores

2. Velocity

Velocity represents the speed at which data is processed and becomes accessible. How fast the data is generated and processed as per client demands, determines real potential in the data

3. Variety

Variety describes one of the biggest challenges of big data. The insights may come without structure. Organizing the data in a meaningful way is no simple task when the data itself changes rapidly.

4. Variability

Variability is different from variety. It means constantly changes that is inconsistency of data.

5. Veracity

Veracity ensures the data is accurate, which requires processes to keep the insufficient data from accumulating in your systems.

6. Visualization

Visualization is critical in today's world. Using charts and graphs to visualize large amounts of

complex data is much more effective in conveying meaning than spreadsheets and reports and formulas.

7. Value

Value is the end game. After addressing the above 6V's . You want to be sure your organization is getting value from the data.

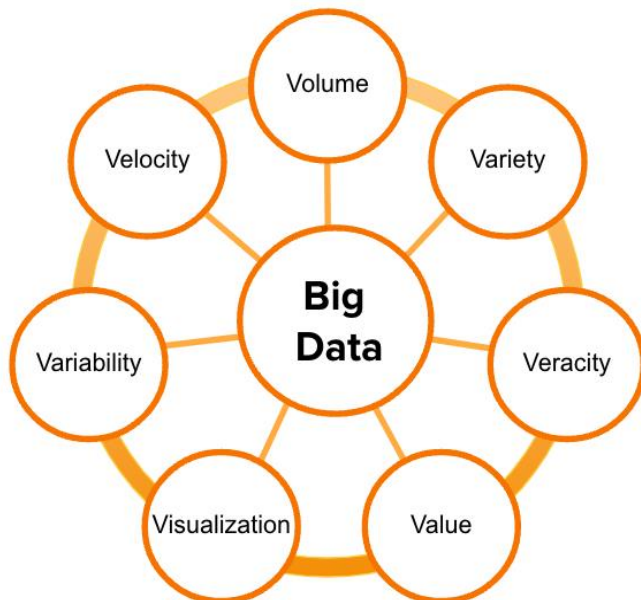


Fig. C - 7 V's of Big Data

Big Data platforms :

Big Data platform is integrated IT solution for Big Data management which combines several software systems, software tools and hardware to provide easy to use tools system to enterprises.

Ex. Hadoop, Colouera, Amazon Web Services, IBM Open Platform etc.

Importance & Advantages

Big Data importance lies in the fact that how the company utilizes the gathered data. It is in the form of -

1. Cost Savings
2. Time-Saving
3. Understand the market conditions
4. Social Media Listening
5. Boost Customer Acquisition and Retention
6. Solve Advertisers Problem/ Offer Marketing
7. Innovations and Product Development

Big data makes companies capable to innovate and redevelop their products. The use of Big Data in a wide range of industries including Finance

and Banking, Healthcare, Education, Government, Retail, Manufacturing, and many more.

There are many companies like Amazon, Netflix, Spotify, LinkedIn, Swiggy .etc which use big data analytics. Banking sectors make the maximum use of Big Data Analytics. Education sector is also using data analytics to enhance students' performance as well as making teaching easier for instructors.

Advantages for Organizations using big data to target customer-centric outcomes, tap into internal data and build a better information ecosystem.

Improved business processes: Probably the biggest advantage of big data is it helps businesses to gain a huge competitive advantage. Apart from being able to understand, as well as, target customers better, analyzing big data can result in the improvement and optimization of certain facets of business operations

Fraud detection: This advantage of using big data comes from the implementation of machine learning technologies.

Improved customer service: One of the most common goals among big data analytics programs is improving customer service.

III. CLOUD COMPUTING

Cloud - The cloud is a metaphor for the network of data centers which store and compute information available through the internet.

Cloud Computing - Cloud computing refers to an efficient method of managing lots of computer servers, data storage and networking

The cloud computing means storing and accessing the data and programs on remote servers that are hosted on the internet.

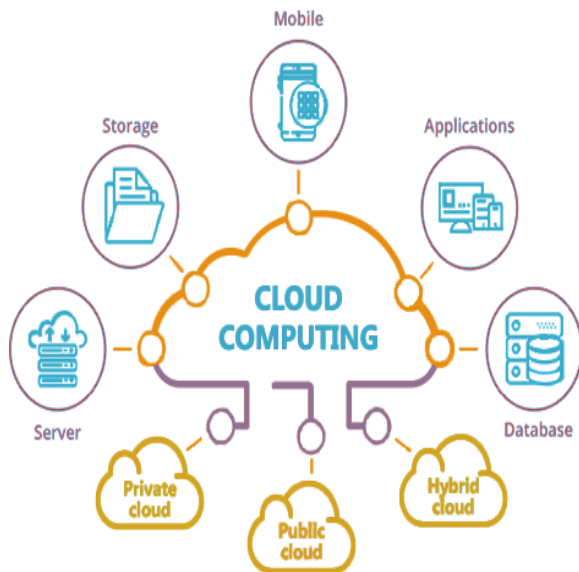


Fig. D - Cloud Computing system

Following are the features or characteristics of Cloud Computing:

- a. Great Availability of Resources
- b. On-demand Self-service
- c. Easy Maintenance
- d. Large Network Access
- e. Availability
- f. Automatic System
- g. Economical
- h. Security
- i. Pay as you go

IV. BIG DATA IN THE CLOUD

Storing and processing big volumes of data requires scalability, fault tolerance and availability. Cloud computing delivers all these through hardware virtualization.

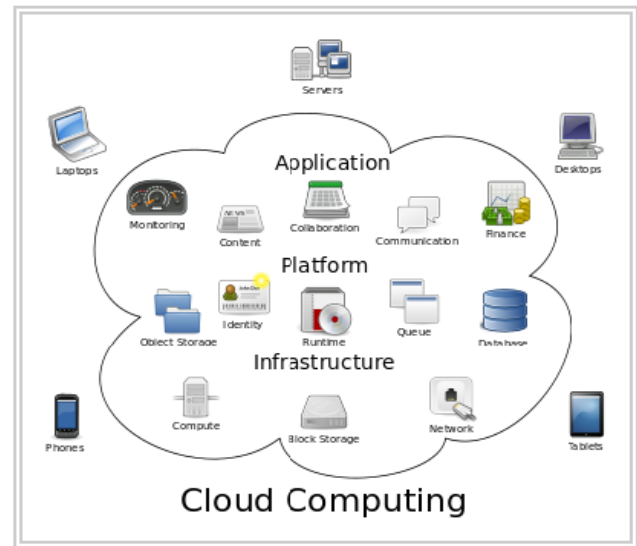


Fig. E - Big Data & Cloud Computing Ecosystem

Thus, big data and cloud computing are two compatible concepts as cloud enables big data to be available, scalable and fault tolerant. Business regard big data as a valuable business opportunity. Companies such as Google, IBM, Amazon and Microsoft also provide ways for consumers to consume big data on demand. Next, we present two examples Nokia and RedBus, which discuss the successful use of big data within cloud environments.

A) Cloud deployment models

Cloud deployment models are grouped broadly into four models: private cloud, public cloud, community cloud and hybrid cloud.

Private cloud is the most secure way to utilize cloud computing. The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers. It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.

Public cloud is provisioned for open use by the public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.

Hybrid cloud is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability.

Community cloud is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns. It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or open use by the public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them.

Types of Cloud Computing Deployment Models



Fig. E - Cloud deployment models

B) Cloud Service Delivery Models

Cloud provides usually three different basic services.

SaaS – Software as a Service In simple this is a service which leverages business to roll over the internet. SaaS is also called as “On demand software” and is priced on pay-per-use basis. SaaS allows a business to reduce IT operational costs by outsourcing hardware and software maintenance and support to the cloud provider.

PaaS – Platform as a Service PaaS is quiet similar to SaaS rather than SaaS been offered through web the PaaS creates software, delivered over the web. PaaS provides a computing platform and solution stack as a service. In this model user or consumers creates software using tools or libraries from the providers.

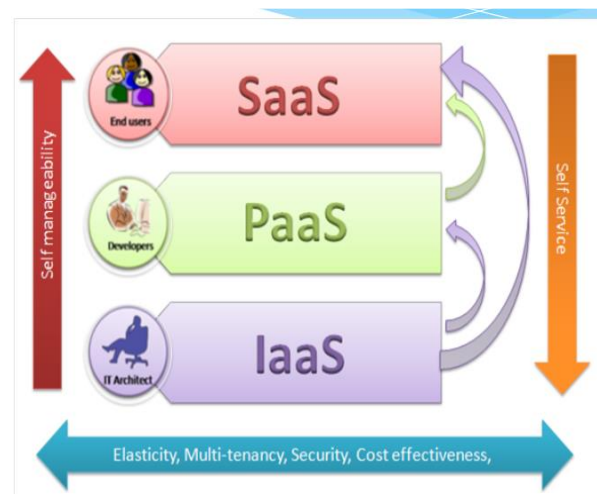


Fig. F -Service Delivery Models

IaaS – Infrastructure as a Service. Infrastructure is the foundation of cloud computing. It provides delivery of computing as a shared service reducing the investment cost, operational and maintenance of hardware. Infrastructure as a Service (IaaS) is a way of delivering Cloud Computing infrastructure – servers, storage, network and operating systems – as an on-demand service.

Examples - Nokia and RedBus, which discuss the successful use of big data within cloud environments.

a) IBM Cloud

IBM Cloud offers the most open and secure public cloud platform for business, a next-generation hybrid multicloud platform, advanced data and AI capabilities, and deep enterprise expertise across the industries. IBM Cloud hybrid cloud solutions deliver flexibility and portability for both applications and data. IBM Cloud solutions can help your organization with modernize existing application, Build and scale native applications, Integrate applications and data across multiple clouds.

b) NOKIA

Nokia was one of the first companies to understand the advantage of big data in cloud environments. Several years ago, the company used individual DBMSs to accommodate each application requirement. However, realizing the advantages of integrating data into one application, the company decided to migrate to Hadoop-based systems.

c) RedBus

RedBus is the largest company in India specialized in online bus ticket and hotel

booking. This company wanted to implement a powerful data analysis tool to gain insights over its bus booking service. Its datasets could easily stretch up to two terabytes in size. The application would have to be able to analyse booking and inventory data across hundreds of bus operators serving more routes.

C) Big Data and Cloud Computing Challenges

Cloud and big data technologies work very well together. Even though the partnership between these two technologies have been established, both still pose some challenges.

i) *Data Storage* - Storing and analysing large volumes of data that is crucial for a company to work requires a vast and complex hardware infrastructure. With the continuous growth of data, data storage device is becoming increasingly more important, and many cloud companies pursue big capacity of storage to be competitive.

ii) *Data Quality* - Accuracy and timely availability of data is crucial for decision-making. Big data is only helpful when an information management process is implemented to guarantee data quality.

iii) *Security and Privacy* - Security is one of the major concerns with big data. To make more sense from the big data, organizations would need to start integrating parts of their sensitive data into the bigger data.

To do this, companies would need to start establishing security policies which are self-configurable:

iv) *Hacking* and various attacks to cloud infrastructure would affect multiple clients even if only one site is attacked. These risks can be mitigated by using security applications, encrypted file systems, data loss software, and buying security hardware to track unusual behaviour across servers.

v) *Service Delivery and Billing* - Budgeting and assessment of the cost will be very difficult unless the provider has some good and comparable benchmarks to offer.

vi) *Performance and Bandwidth* Cost Businesses can save money on hardware but they must spend more for the bandwidth.

All these challenges should not be considered as road blocks in the pursuit of cloud computing. It is

rather important to consider these issues and the possible ways out before adopting the technology.

V. CONCLUSIONS

With data increasing on a daily base, big data systems Big Data is responsible for data storage and processing, the cloud provides a reliable, accessible, and scalable environment for Big Data systems to function well.

Cloud environments strongly leverage big data solutions by providing fault-tolerant, scalable and available environments to big data systems.

The Data processing in cloud computing environments using Big data application and an overview of both the technologies and cases of success found when integrating Big data and Cloud Technology.

This paper also introduced the characteristics, trends and challenges of big data. In addition to that, it investigates the benefits and the risks that may rise out of the integration between big data and cloud computing.

VI. FUTURE SCOPE OF STUDY

The further goal is to investigate the mechanisms that adaptable software can use to trigger scalability at different levels in the cloud stack. Thus, accommodating data peaks in an automatic and reactive way.

Privacy standards will play an important role in fostering the adoption of cloud services by promoting social responsibility and addressing privacy challenges.

Large amount of WASTE data in clouds. How to remove or utilize it to be considered for/in future.

VII. REFERENCES

- [1] Padgavankar M.H and Gupta, S.R. (2014). Big Data Storage and Challenges, International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2218-2223.
- [2] Sijin; L. Guo, Y. Guo, C. Wu, M. Ghanem, R. Han. Elastic Application Container: A Lightweight Approach for Cloud Resource Provisioning (http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6184989). 2012 IEEE 26th

International Conference on Advanced Information Networking and Applications (AINA). pp. 15–22.

[3] Puneet Singh Duggal, Sanchita Paul, “Big Data Analysis : Challenges and Solutions”, International Conference on Cloud, Big Data and Trust 2013, Nov 13-15, RGPV

[4] <https://www.dqindia.com/>

[5] <https://ekumbh.aicte-india.org/allbook.php>

[6] IBM InfoSphere Streams, <http://www.ibm.com/software/products/en/infospherestreams>.

[7] IBM SmartCloud Enterprise,
<http://www-935.ibm.com/services/us/en/cloud-enterprise/> (2012).

[8] [https://timesofindia.indiatimes.com/blogs/voices/trends-in-cloud-computing-for-the-year-2023/](https://timesofindia.indiatimes.com/blogs/voices/trends-in-cloud-computing-for-the-year-2023/Trends-in-Cloud-Computing-for-the-year-2023) Trends in Cloud Computing for the year 2023

[9] Web-pages from Amazon AWS, Microsoft Azure and Google CDP

[10] F. Shirazi and A. Qbal, “Community clouds within M-commerce: a privacy by design perspective,” *Journal of Cloud Computing*, vol. 6, no. 1, 2017.view at: Google Scholar

[11] www.ignou.ac.in

[12] <https://www.wikipedia.org/>

