International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 11 | Nov -2016 www.irjet.net p-ISSN: 2395-0072

The Rising of Big Data in Cloud Computing: A Study

Ms. Tanvi Desai[1], Ms. Rikita Patel[2]

[1]Asst. Professor, MCA, Anand Institute of Information Science, Gujarat, India
[2]Asst. Professor, MCA, Anand Institute of Information Science, Gujarat, India

---

Abstract - *Big data is an inborn feature of the cloud and with the rapid growth of rising applications like social networking, semantic Web analysis and bioinformatics networking, a variety of data are spawned. In fact, it is expected that half of the total data will be on the cloud. Big Data and cloud computing are two important issues in the recent years; Cloud Computing reduces the need to maintain expensive computing hardware, dedicated space, and software. Massive escalation in the degree of data or big data generated through cloud computing has been observed. This paper includes the hike of big data in cloud computing. This paper also presents the relationship between big data and cloud computing, characteristics of big data, and Big Data technologies.*

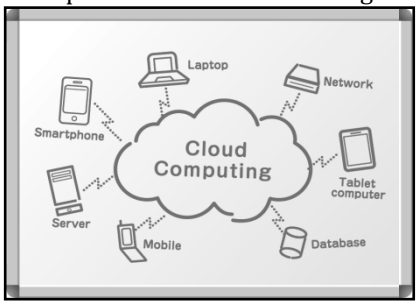
Key Words: **Big Data, Cloud Computing, Semantic Web Analytics, Bioinformatics Networking.**

1. INTRODUCTION

Cloud Computing

"A Cloud is a type of parallel and distributed system consisting of a collection of inter-connected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreement established through negotiation between the service provider and consumers."[1]

As the successful approach of service oriented, cloud Computing provides different types of services to individuals and Business for on demand basis. The elasticity, pay per use, low upfront investments are the major characteristics that make the cloud computing is everywhere platform for deploying inexpensively reasonable enterprise infrastructure settings.



Cloud computing services are categorized into three different levels [3]:

Software as a Service (SaaS):

These types of application are generally designed for end-users, delivered over the web. SaaS works so much better for students because it provides access to applications anytime, anywhere, for any type of devices like laptop, smart phone, tablet, or other web-enabled device. Adding more users or scaling the software to more classrooms or campuses is becomes very easy task with SaaS. As an example, a college can scale its SaaS solution from 50 students to 5,000 in a matter of hours – unimaginable in the conventional IT scenario.

Platform as a Service (PaaS):

PaaS is the collection of development tools and services which is used for coding and deploying the applications quick and efficient. With PaaS, Students, teachers or other academicians can develop new applications or services in the cloud which is platform independent, and also make them widely available to users through the Internet. It also provides services for testing, deploying, collaborating on, hosting, and maintaining applications.

Infrastructure as a Service (IaaS):

IaaS is the combination of hardware and software that powers it all – servers, storage, networks, operating systems. These are also known as On demand data centers which provide compute power, memory, and storage, typically priced per hour according to resource consumption. It can be used to satisfy the infrastructure needs of students, staff or any other academia's.

Big Data

Big data is a huge amount of data which are complex and which are not handled by traditional data processing application. Big data is being generated by everything around us at all times. Every digital process and social media exchange produces it. Systems, sensors and mobile devices transmit it. Big data is arriving from multiple sources at an alarming velocity, volume and variety. [2]

Characteristics of Big Data

1) Data Volume: The Big data word itself defines the volume. Currently the data measures in peta bytes and is supposed to increase to zeta bytes in nearby future

2) Data Velocity: Velocity in Big data is a concept which involves the speed of the data coming from various sources. Velocity concerns the different rates at which data streams may get in or out the system.

3) Data Variety: Data variety is an evaluation of the richness of the data demonstration – text, images, video, audio, etc. Data being generated either structured or unstructured from various resources like web Pages, social media sites, e-mail, documents.

2. THE SOURCE OF BIG DATA

Big data is the result of advances in mobile technology that now include digital video, photography, audio, and advanced email and text features. Users are collecting data in numbers that were never seen a decade ago; likewise, applications like Face book, You Tube, Instagram, etc. We can imagine what's going on every minute on the web

- On Netflix 80,000 hours of video streamed
- 1.7 million photos likes per minute on Instagram
- 300,000 Shares on Snap chat
- Interest pinners pin nearly 10,000 images

3. COMPANY USES BIG DATA IN CLOUD

Amazon

As Amazon is the king of e Commerce Company uses big data to improve its performance. Amazon takes advantages of its big data resources to give more relevant product recommendations and improve its customer care quality. The customer profiles developed using big data resources enabled Amazon to create highly personalized marketing messages for its customers. It uses historical data on purchases made by customers to give them highly customized product suggestions. It also analyze the past customer data which helped Amazon in giving suggestions to new customers who were buying from its portal for the first time. This forecasting model of Amazon uses these big data from customers Amazon activity, including time on site, duration of views, links clicked and hovered over, shopping cart activity and wish lists for predicting what the customer wants.

4. BIG DATA ANALYSIS SOLUTIONS

This is an open source java based programming framework supporting for the processing of large sets of data in a distributed computing environment. Using Hadoop, big data sets can be processed over cluster of servers and apps may be run on system with thousands of nodes involving terabytes of information [4]. Hadoop Framework is used by many well known companies like Google, Yahoo, Amazon and IBM etc., to manage big data.. Hadoop has two main sub parts – Map Reduce and Hadoop Distributed File System (HDFS)

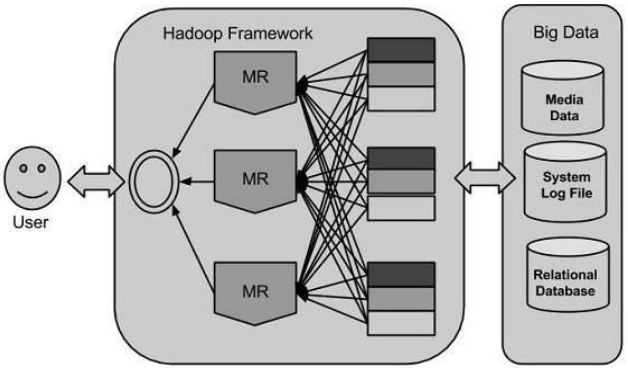


Fig: 2 Hadoop Frameworks [5]

Map Reduce

Hadoop Map Reduce is a framework used to write applications that process big data in parallel on clusters of commodity hardware resources in a reliable, fault-tolerant manner. A Map Reduce initially separates the data into individual break apart which are processed by Map jobs in parallel. The outputs of the maps sorted by the framework are then input to the reduce tasks. Generally the input and the output of the process are both stored in a file-system. Scheduling, Monitoring and re-executing failed tasks are taken care by the framework.

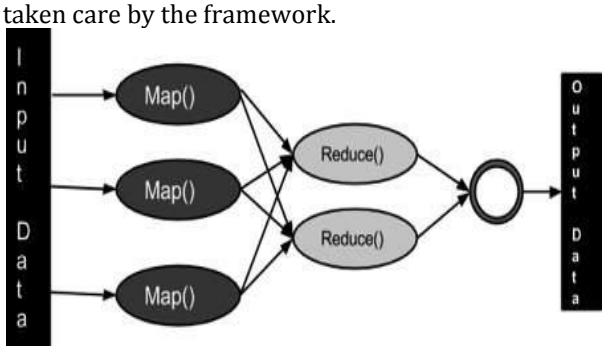


Fig:3 Map reduce[5]

Hadoop Distributed File System (HDFS)

HDFS is a file system that extent all the nodes in a Hadoop cluster for data storage. It links together file systems on local nodes to make it into one large file system. HDFS improves reliability by replicating data across multiple sources to overcome node failures.

5. BIG DATA ISSUES

Large volume of data being generated at a rapid speed and to store these data is a big deal as a consequences company must define the policies for access mechanism and for the expiration of data these data. As data being stored at cloud security, privacy and data management become a major issue.

Security

1. Distributed Framework

Big data implementation distributes on parallel computation through multiple clients which creates opportunities for violation of security. Identifying a malicious or unreliable client and protecting the data from these unreliable processors is a complex task to involving big data.

2. Non Relational Data Stores
As traditional database doesn't supports unstructured data,
Organization migrates to NoSQl Databases to manage unstructured data. NoSQL databases, which are themselves lack of security.

Privacy
Big data can enable "invasions of privacy, invasive marketing, decreased civil liberties, and increased state and corporate control". The amount of information collected on each individual can be processed to provide a surprisingly complete picture. As a result, organizations that own data are legally responsible for the security and the usage policies they apply to their data. [7]

Data Management
To handle petabytes of data current technologies are not satisfied and speed of increasing storage capacity is very slow compared to growing the data. Thus revolution for modernization of information framework is essential.

5. BIG DATA APPLICATIONS

Here is the list of interesting application [6]:

Stock market sentiment analysis by means of Google Trends has been revealed to correlate well to chronological index declines and rises, which is perhaps not surprising but interesting in terms of significance as a big data application. This research no doubt requires more analysis, but is compelling. An interesting consideration, though, is what will happen as these machine-based trading systems come online by the side of presented programmed trading.

Picasa photo sorting by Google is a practical tool that allows a user to sort, query, and automatically identify faces using CV techniques collective with machine learning. This is a grand way to get a sense for the value of big data services and applications. It makes it clear that big data analytics will require highly developed analytics such as CV and methods like machine vision.

Recommendation systems like Pandora (music), Netflix (movies) and Amazon (books and products) use customer data and multiple agents in an approach known as *collaborative filtering.*

Booking systems for tourism are being enhanced by integrating customer preference, logistics, and prior history to make helpful suggestions for that always-arduous task of planning travel.

6. CONCLUSIONS

This paper gives a detail of Cloud Computing and Big Data. It also gives brief overview of Big Data Solution and Its application. In future, make way for the even more efficient use of the big data by the user on a cloud computing environment. It is needed that the computer scholars and IT professionals to work together and make a successful and long term use of cloud computing and explore new ideas for the usage of the big data over cloud environment.

REFERENCES

[1] Cloud Computing: Vision, Tools, Technologies for Delivering Computing as the 5th Utility by Rajkumar Buvya
[2] https://www.ibm.com/big-data/us/en/
[3] Ms. Tanvi Desai, Cloud Computing in Education Sector, International Journal for Innovative Research in Science & Technology Volume 2 Issues – 10 Year of Publication: 2016
[4] Venkatesh H# A Study on Use of Big Data in Cloud Computing Environment, International Journal of Computer Science and Information Technologies, Vol. 6 (3) , 2015, 2076-2078
[5]https://www.tutorialspoint.com/hadoop/hadoop_big_data_solutions.htm
[6] Sam B. Siewert, Assistant Professor, University of Alaska Anchorage IBM Developers Work
[7] http://www.business.com/technology/privacy-and-security-issues-in-the-age-of-big-data/

BIOGRAPHIES

Ms.Tanvi Desai
Assistant Professor
Anand Institute of Information Science, Anand

Ms.Rikita Patel
Assistant Professor
Anand Institute of Information Science, Anand