

Big Data –The Change of Datacenter Concept and Challenges

Bindhya Thomas¹, Geevarghese Jacob²

Lecturer, Department of IT and Project Management, College of Administrative and Financial Studies,

AMA International University - Salmabad, Kingdom of Bahrain¹

Network Administrator, Ministry of Interior, Kingdom of Bahrain²

Abstract: The research paper is an introduction to the new change in the IT industry of managing huge amount of data being generated in the new IT revolution. There are different technologies or solutions are introduced in the IT industry to accept the new change of Big Data. Here we explained the structure of Big Data management and its structure process also its shows the features and its use and how it's benefited the IT industry. We have also discussed the business or services which are using Big Data like social Medias. There also suggestions for the betterment for future development.

Keywords: big data, Cloudera Hadoop, Saas, Iaas, Paas.

INTRODUCTION

We are in the era of information explosion. Globally Data are being generated in Exabyte's every day. Digital Data being generated from lots of cheap devices. The concept of Datacenter being changed as the data sizes grows. The legacy systems are not capable to deal with structured and unstructured data being generated. Let's take the example of social media like Twitter, Facebook, Wikipedia etc. These are the infrastructure or platform developed by the companies and people around the world are using to store their data and other communication and socialization activities. As the data size grows the challenges grows as well like managing huge data, especially data mining, data analysis and security. As per Doug Laney, the Gartner Analyst The data growth challenges are three dimensional which are volume, velocity and type of data. Volume denotes increasing volume of data; velocity the processing and the accessing speed of data in and out. Variety denotes the different types of data. Instead of keeping all data in different location Big Data provide consolidated platform with efficiency to provide solution for all needs.

Big Data Structure

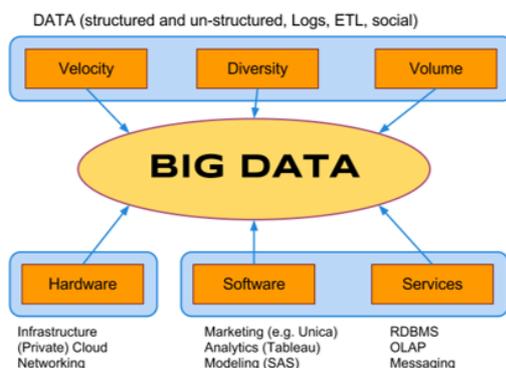


Fig. 1 Big Data Structure

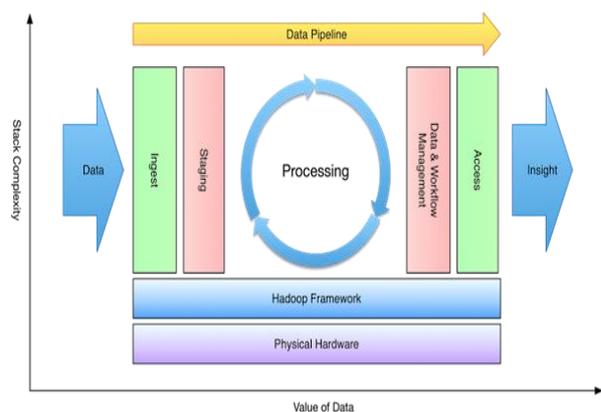


Fig. 2 Big Data Processing

Traditional and Modern Data Processing Comparison

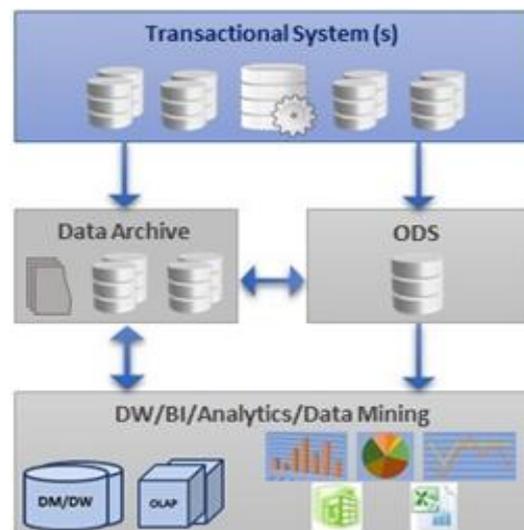


Fig. 3 Traditional data processing & management

The Figure 4 shows the modern level architecture of big data. It is a combination of different concepts and each of them belongs to different levels. Each level has different amount of data, time frame and formats. The figure shown below explain the high-level architecture of big data [3].

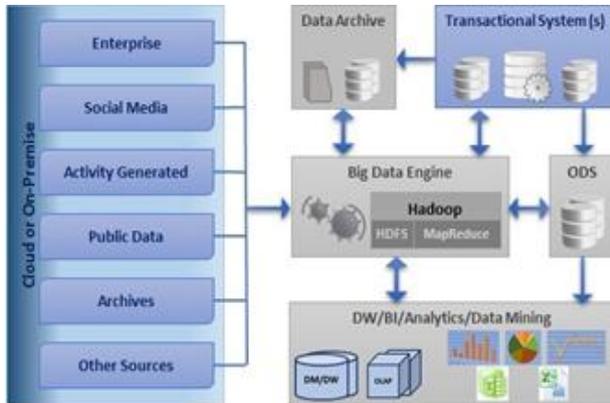


Fig.4. Modern data processing & management

Advantages of Big Data

Communication services are moving from the concept of local data center to cloud. Some institutions like Military, Research Centre’s, Government ministries are still in their own datacenter as the information is critical and confidential. Otherwise all other sectors are moving to cloud to reduce their IT department maintenance cost and availability of information. Here Big Data gives better opportunity, doesn’t matter the volume of data to be processed. As it’s mentioned earlier Big Data systems are capable to prove the information regardless of data growth, without latency. This deals with any format of data.

The below are the examples of cloud services or cloud computing services

1. SaaS: Software as Service.

This service is an example of cloud service. Any application can be hosted in cloud datacenter and what the client required is the communication mostly internet connection to the cloud service. It can be anything like CRM solution, thin client, mobile application etc. The importance of big data is the back in this solution. The advantage is the big data provider can host any number of SaaS services for different customers according to the capacity. The best examples are Amazon and Micros for virtual desktop.



2. IaaS: Infrastructure as Service

Years before we couldn’t imagine how a infrastructure can be given as a cloud service. We can have servers, storages, network switches in the cloud. The customer can control and manage these servers and storages and switches same like it in their local network. The services saves big amount of money and more over this gives the high availability of services without interruption. Another advantage is “Pay as you Grow”. The meaning is customer needs to pay for additional requirement when it required not worrying about if the new system will support. VMWare ESXi, Microsoft Hyper-V, Oracle Virtual Box are the examples for this service

3. PaaS: Platform as Service

This service is widely using now. Companies like Microsoft, Amazon and Google provides these service. Microsoft giving services through Azure and VMWare, Google providing Google App Engine and Amazon Web Services. These companies provides the platform to build datacenters in their cloud datacenter. They provide computing plat forms which includes operating system, database, and webservers. The main advantage is for developers, where they don’t have to buy any additional hardware.

Cloudera Hadoop

Most of the Big Data manufactures uses Cloudera Hadoop software. Hadoop is an open source, framework is java based –programming that supports the processing of big data sets in a distributed computing environment. Hadoop fundamentally changes the way enterprises store, process, and analyze data. Unlike traditional systems, Hadoop enables multiple types of analytic workloads to run on the same data, at the same time, at massive scale on industry-standard hardware. CDH, Cloudera’s open source platform, is the most popular distribution of Hadoop and related projects in the world (with support available via a Cloudera Enterprise subscription).

Growth and acceptability of Big data concept

The major commercial big Data vendors are Dell, IBM, HP, Amazon and Microsoft. All are using Hadoop as the software for Big Data. Major customers are social media companies, scientific research developers, military documentation s and communications, Banks for their millions of transactions in a day.

Recommendations for Big Data

As it mentioned above Big data can accept and process any format of data which is one of the main advantage. But the migration of existing data to Big Data is a challenge, however there is no proper migration method has been introduced which can cause pause of business. As we are in the middle of traditional data storage to big data. Applications and corresponding supports systems are developed for the traditional systems. Before migrating to Big Data there should be a proper testing should happen whether to confirm the application workflow runs

smoothly with Big Data also the migration. For example migration from mainframe to big data is complex. Another recommendation are the backup capacity. Since big data is dealing with large amount of data like Exabyte's, backup is an important concern. An institute can't afford any data loss and all the data and information related to a company located in Big Data, failure of that cause big loss. So we can recommend a Big Data cluster and a disaster recovery plan should be recommended with big data solutions.

CONCLUSION

As data are going big the name of the solution itself is given Big Data. The future data management are getting more complex and sensitive. We can say Big Data is a consolidated solution for the future world's requirement and for sure more development should happen. IT industry is already accepted Big Data solution and major vendors are started their own research and development. This research is the introduction to the new solution and more finding can be done in coming topics.

REFERENCES

- [1]. Elena Geanina ULARU, Florina Camelia PUICAN, Anca APOSTU, Manole VELICANU ; Phd. Students of Institute of Doctoral Studies . Database Systems Journal vol. III, no. 4/2012.
- [2] Nada Elgendy and Ahmed Elragal, "Big Data Analytics: A Literature Review Paper" ICDM 2014, LNAI 8557, pp. 214–227, 2014. © Springer International Publishing Switzerland 2014
- [3] SAS "Five big data challenges And how to overcome them with visual analytics"
- [4] Research Trends " Special issue on Bid data" Issue 30 September 2012
- [5] <https://www.sas.com/resources/asset/five-big-data-challenges-article.pdf>
- [6] <https://www.mssqltips.com/sqlservertip/3133/big-data-basics--part-2--overview-of-big-data-architecture/>

BIOGRAPHIES

Bindhya Thomas is Lecturer in AMA International University-Bahrain. 11 years of experience in teaching and 1.5 years of Industry experience. Completed M.Sc. Computer Science. Published 4 papers.

Geevarghese Jacob is senior system administrator in Ministry of interior. 12 years of IT industry experience as project engineer.