Compendious study of Big Data as a Service (BDaaS)

Chethan S.[1]

[1] *Department of Information and Communication Technology MIT, Manipal*

Abstract: Cloud computing is actually a broad term that describes various range of services. There is a growing diversity of services that can be made available over the internet via cloud computing rather than providing services locally. Big data as a service (BDaaS) actually means to offer analysis services of large data sets which are usually complex in nature over the internet via cloud services. Big data can be both structured and unstructured which can be analyzed to get an insight which can be used to take better decisions. In this paper we see how the big data can be provided as a service over the internet and the different parameters to be considered for this purpose.

Keywords: Cloud computing, big data, BDaaS, analytics, Hadoop

1.   INTRODUCTION

Cloud computing is actually broad term that describes various range of services. There is a growing diversity of services that can be made available over the internet via cloud computing rather than providing services locally. In layman's term cloud computing means nothing but storing and accessing resources, data and/or applications over the internet instead of our system's hard disk. Cloud computing enables users to access resources and applications anywhere in the world on demand, which is motivating business houses to develop software applications which can be used by many users by consuming them as a service rather than to use them on their individual systems [1]. Big data is a term which usually describes any voluminous amount of data. Big data can be either structured (example: Relational data), semi-structured(example: XML data) or unstructured(example: Text, Media etc) which can be analyzed to get a better insight and concrete decisions can be made which results in cost reductions, better operational efficiency and reduced risk for the business. Big data is the data that is too much for a traditional database system to handle i.e., it exceeds processing capacity of a normal DBMS. Big data is made up of 3 V's that is large Volume, high Velocity and huge Variety of data as shown in Fig 1 [2]. Here volume refers to the amount of the data which is usually in terms of GB and more. Velocity refers how fast the data is processed. Since most of the data especially the social media data is fed in real time the speed of processing should be as fast as possible for an efficient system. Finally, Variety means the types of data which can be structured, semi-structured and/or unstructured.

Big data can be made of many data generated by different applications or devices as shown in Fig 2. Big data as a service (BDaaS) actually means to offer analysis services of large data sets which are usually complex
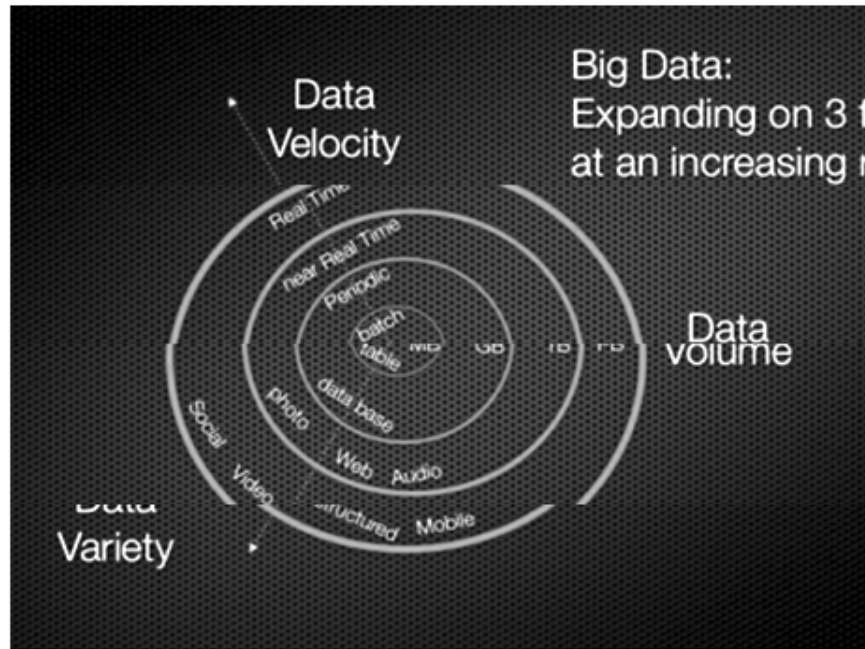
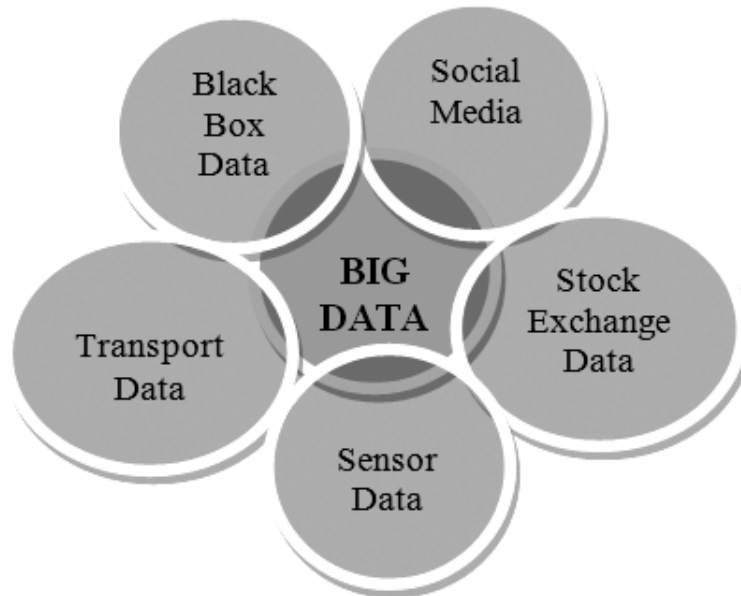**Figure 1: 3 V's which characterizes big data**

Figure 2: Possible data under Big Data

in nature over the internet via cloud services. Predicting the requirements for a big data project would be consternating. It's not easy of nearly impossible to decide the kind of data that would be stored, the storage space required and the power required to process the big data [3]. BDaaS offers business enterprises environments with the fleetness, malleability and near limitless power to process the data [3].

2. THE LATEST FAD IN BIG DATA: BDAAS

The most common concept in cloud computing are software as a service, platform as a service and data as a service. By mixing all these technologies and by scaling up the data involved as shown in Fig 3, the concept of
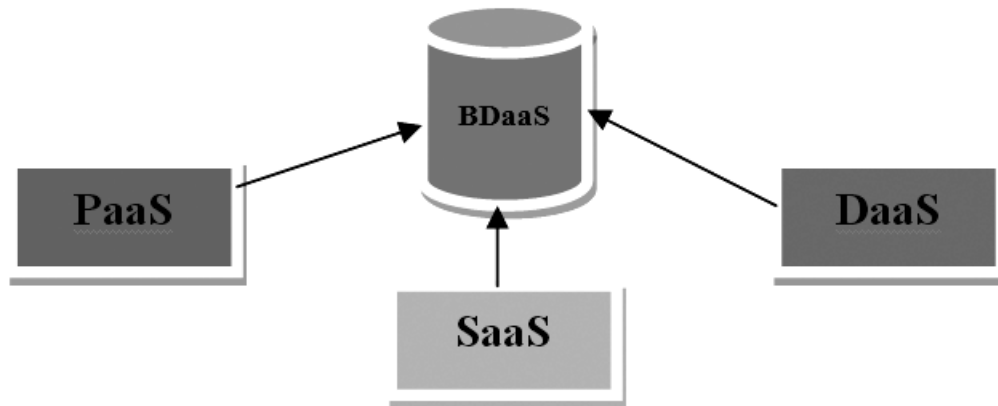
Figure 3: BDaaS: Best of PaaS, SaaS and DaaS

BDaaS was arrived at [4]. Big data as discussed is the term used to refer the ever growing and huge amount of data that are generated, stored and analyzed which gives us a better insight which in turn drives business growth. Organizations can also use hadoop to analyze data by using open source softwares but the cost to launch a big data initiative could be still superabundant, on the other hand BDaaS permits enterprises to make use of cloud computing facilities and they need to pay only for the amount of computing capacity needed which is also economical for the organization [5]. The concept of BDaaS eliminates cost associated with deployment of hadoop and instead concentrates on getting more insight about the data.

BDaaS should necessarily provide following four functions [6] [9]:

1. High Performing Service Oriented Architecture (SOA) – BDaaS should necessarily provide an architecture which provides fast data processing components and analysis tools which improves business.

2. Capability of Virtualization in Cloud – BDaaS is based on cloud computing which is scalable and allows multiple entities to work as single unit and new ones can be added if amount of data increases.

3. Event-Driving Processing of complex data – BDaaS allows data management in EDP(Explanatory, Descriptive and Predictive) modules and due to its real time processing BDaaS is accurate, timely and affordable.

4. Analytical tools to support business intelligence - Since BDaaS consists of huge amount complex data which requires data mining, querying , reporting etc it should be able to convert unstructured data into informative for business intelligence.

BDaaS enables business organizations to provide the users pervasive access to various interfaces and big data tools such as MapReduce, Sqoop etc. BDaaS is also auto-scaling and self-managing technology so organizations can achieve access without much operational change. BDaaS provides both hardware and software services which can be made use by the both big and small organizations to gain most from the collected data and use the knowledge gained to improve the business. Small organizations might not generate and use as much data when compared to a big business but BDaaS facilitates this so that even smallest business can get benefit of big data through BDaaS [7]. Most compelling reason for organizations to consider BDaaS is its ability to provide quick data analysis services without additional infrastructure and investment since services are provided on the cloud. BDaaS now allows organizations access to data that was previously possible for only for large organizations [7].

BDaaS provides more flexible and efficient way to access data as we see in Table I which gives a comparison of BDaaS with traditional big data [8].

Table 1
Comparison of BDaaS with traditional big data

Traditional Big Data	*Big Data as a Service*
Scalability in processing and storage achieved through distributed architecture	Scalability on demand through a combination of cloud computing and distributed architecture
Data storage on HDFS or distributed platform	Virtualized data storage on a distributed platform.
Structured and unstructured data	Structured and unstructured data on cloud environment
Advanced analytics functions	Advanced analytics functions with on-demand computing power
Limited accessibility	Ubiquitous accessibility
Analytical capability derived through custom coding	Analytical capability derived from out-of-box domain-specific algorithms along with custom coding.

Table 2
Benefits of BDaaS over Hadoop

Parameters	*Hadoop*	*BDaaS*
Setup Time	Building in-house Hadoop is time consuming and would take several months	Since BDaaS is based on cloud computing access to infrastructure necessary tools within minutes
Pricing	Hadoop is a hardware which might need frequent maintenance and upgrades which is not economical	BDaaS uses cloud which is based on pay per-use basis and is economical
Expertise Required	Person who has good knowledge on Hadoop is required to carry out a project on Bigdata	Cloud eliminates need to find a data scientist to help implement a bigdata project and provider provides with necessary tools to troubleshoot problems if any.
Storage Space	Storage space is necessary for resources and softwares to run Hadoop	Since everything is on cloud no extra storage space is required

4. BDAAS: OPENING WORLD OF LIMITLESS POSSIBILITIES

Data now a days is generated at a very fast pace, and most of this generated data is unstructured and the pace with which these data are generated makes it impossible for a traditional database to handle the volume of the data and more since the data is mostly unstructured it's not possible for normal traditional database to handle data unless its converted into a structured form. Hadoop was created to address all the mentioned issues related with big data. Even though Hadoop was a major stepping stone, it still had limitations for small and midsized businesses since maintaining Hadoop was not economical option. BDaaS addresses resources and economic issues related to Hadoop since it is based on cloud computing which is flexible, scalable and less costly [10]. Advantages of BDaaS over Hadoop are summarized in Table II [10].

5. CONCLUSION

Big data is a term which usually describes any voluminous amount of data. Big data can be either structured, semi-structured or unstructured which can be analyzed to get a better insight and concrete decisions can be made which results in cost reductions, better operational efficiency and reduced risk for the business. Big data as a service (BDaaS) actually means to offer analysis services of large data sets which are usually complex in nature over the internet via cloud services. The most common concept in cloud computing are software as a service, platform as a service and data as a service. By mixing all these technologies and by scaling up the data involved, the concept of BDaaS was arrived at. Four important functions that the BDaaS should necessarily provide was discussed. BDaaS was found to have more benefits when compared to traditional BigData and also Hadoop. Undoubtedly big data as a service is one the most useful and interesting concept which is of tremendous use for business organizations especially SMD's

REFERENCES

[1] Rajkumar Buyya , Chee Shin Yeoa , Srikumar Venugopal , James Broberg , Ivona Brandic, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility", Future Generation Computer Systems 25 (2009) 599–616

[2] Simmi Bagga,Satinder Kaur "Big Data and Hadoop", International Journal of Advanced Engineering Research and Applications (IJAERA) Vol. – 1, Issue – 6, October – 2015

[3] Nathan Aeder, "Are You Ready For Big Data As A Service (BDaaS)" 2015

[4] Bernard Marr, "The Next Big Thing In Big Data: BDaaS" Jul 20, 2015

[5] Jonathan Buckley, "How to Choose a Big Data-as-a-Service Company" October 29, 2015

[6] Ahmed Banafa, "Big Data as a Service: the Next Big Thing?" June 2014

[7] Ricky M. & Monique L. Magalhaes, "Big Data-As-A-Service", 24 Sept. 2015

[8] Benoy Bhagattjee," Emergence and Taxonomy of Big Data as a Service" ,CISL, Massachusetts Institute of Technology, May 2014

[9] Jon Capistrano " The Rise of Big Data-As-A-Service (BDaaS)" November 25, 2015

[10] Gil Allouche, "Big Data as a Service has Arrived" 2013 Available at https://cmsreport.com/articles/big-data-as-a-service-has-arrived-5805