IPASJ International Journal of Computer Science (IIJCS)
Web Site: http://www.ipasj.org/IIJCS/IIJCS.htm
A Publisher for Research Motivation ..... Email: editoriijcs@ipasj.org
Volume 5, Issue 1, January 2017 ISSN 2321-5992

BIG DATA-AS-A- SERVICE (BDAAS) IN CLOUD COMPUTING ENVIRONMENTS

Dr Minimol Anil Job

Assistant Professor, Faculty of Computer Studies, Arab Open University, Kingdom of Bahrain

Abstract

Evolution of Information Technology moves all business organizations to digital business. Organizations treat data as an asset because it can be defined as a collection of stored truths that are open to interpretation and manipulation for organizational processes. Transactional, social, mobile, cloud and sensor data available now a days offer huge potentials in organizational processing. Database, big data and business intelligence technologies interact to create a new business technology. The technologies are increasingly becoming data-centric. The availability of data acquisition, collection and storing platforms are becoming a necessity because central storage of data in a database reduces data redundancy, data isolation, and data inconsistency and allows for data to be shared among users of the data. These large datasets, popularly known as Big Data, are difficult to manage using traditional computing technologies. Cloud computing Cloud computing eliminates the need to maintain expensive computing hardware, dedicated space, and software. In this research paper the researcher analyses the implementation of Big Data as a service in cloud environments and identified the factors to be considered implementing Big Data-as-a- Service on the cloud

Keywords: Data Processing, Big Data, Data Analytics, Cloud Computing

1. INTRODUCTION

Most recent innovations in the IT industry are related to cloud computing and Big Data. Data are analyzed to study the operation of each industry in order to utilize them for maximum efficiency. With the prevalence of service computing and cloud computing, more and more services are emerging on the internet, generating huge volume of data, there is a demand for combining Big Data and cloud computing to create new technologies that can solve the Big Data problem. Big Data based on cloud computing is an idea used by many vendors. Data as a service (DaaS) enables data to be shared among clouds, systems, applications. DaaS makes it easier for data architects to select data from different pools, filter out sensitive data, and make the remaining data available on-demand. A key benefit of DaaS is the elimination of the risks and burdens of data management to a third-party cloud provider. The combination of Big Data technologies and cloud computing platforms introduced of a new category of technology called Big Data as a Service or BDaaS. In this paper the researcher has conducted a literature review of BDaaSand implementation of BDaaS on cloud infrastructure and summarized the findings.

2. RELATED LITERATURE

The National Institute for Standards and Technology (NIST) defines cloud computing as “Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction”. [1][13].Cloud computing is built around a series of hardware and software that can be remotely accessed through any web browser. Once a cloud is established, how its cloud computing services are deployed in terms of business models can differ depending on requirements. The primary service models being deployed are commonly known as:

- Software as a Service (SaaS) — Consumers purchase the ability to access and use an application or service that is hosted in the cloud.
- Platform as a Service (PaaS) — Consumers purchase access to the platforms, enabling them to deploy their own software and applications in the cloud. The operating systems and network access are not managed by the consumer, and there might be constraints as to which applications can be deployed.
- Infrastructure as a Service (IaaS) — Consumers control and manage the systems in terms of the operating systems, applications, storage, and network connectivity, but do not themselves control the cloud infrastructure.

IPASJ International Journal of Computer Science (IIJCS)
Web Site: http://www.ipasj.org/IIJCS/IIJCS.htm
A Publisher for Research Motivation ..... Email: editoriijcs@ipasj.org
Volume 5, Issue 1, January 2017 ISSN 2321-5992

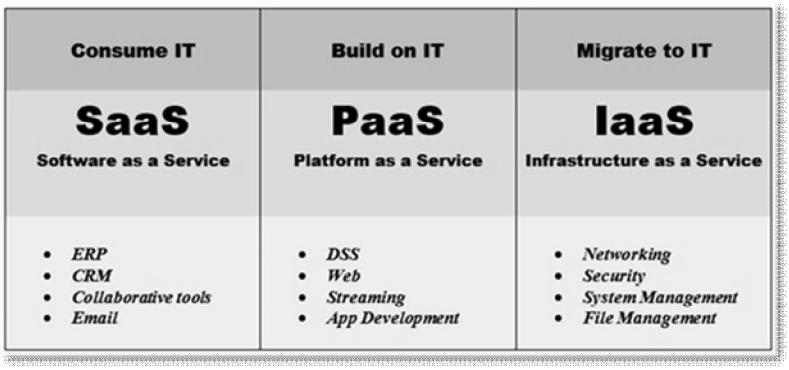

Figure 1: Cloud Computing Stack

3. DATA AS A SERVICE (DAAS)

Data as a Service (DaaS) is an information providing and distributing model in which needed data such as text, images, sounds, and videos are made available to customers over the internet [2][3]. DaaS is based on the concept that the needed data can be provided on demand to the user regardless of geographic separation of receiver and provider. The advancement of IT introduced the new concept of DaaS. Especially, due to the emergence of Service Oriented Architecture (SOA) the actual platform on which the data resides is less important. The technological developments of mobile data, sensors, and near-field communications have led to more data being more easily collected [4][5]. Sources of data include mobile devices and sensors embedded in products and physical things, social content from texts, tweets, posts, blogs, click stream data from the Web and Internet searches and videos and photos from retail and user-generated content., M2M sensors embedded in everything from airport runways to casino chips for example the Internet of Things [6][7]. Enterprise data deals with the data related to financial, medical, research, and customertransactions and also from B2B transactions.

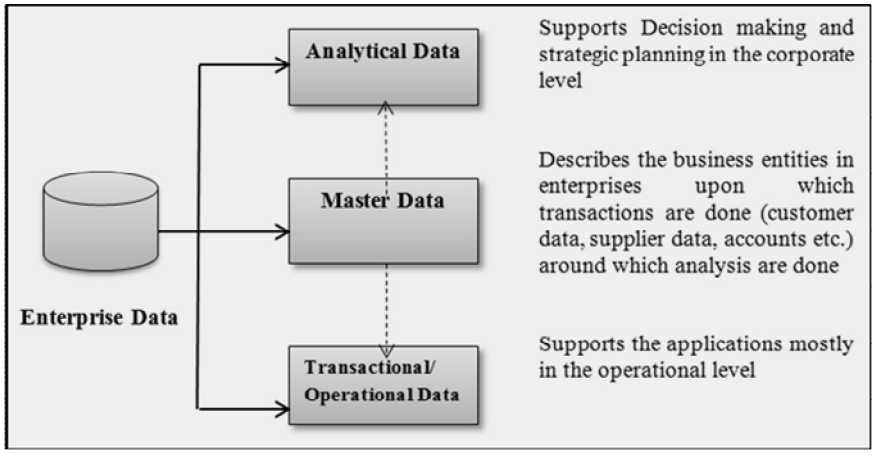

Figure 2: Enterprise data types: transactional, master, and analytical data.

DaaS can be considered as the solution of the growing volume and variety of data as well as the competencies requires managing this huge amount of data[12]. Organizations use data as a means to generate competitive advantage. Data are worthless if it cannot be analyzed, interpreted, understood, and apply the results in context. Poor quality data cannot be trusted and may result in the inability to make intelligent business decisions. The capability to capture, process, format, and distribute data in near real-time or faster requires a huge investment in data management architecture and infrastructure. End users need to see data in a meaningful format and context if the data are to guide their decisions and plans and the managers need context in order to understand how to interpret collected organizational data. The output would not be useful to the organization, if the wrong analysis or datasets are used. Cloud computing providescomputing resources such as servers, storage, operating systems, and networks; and enables the sharing of data on a selective basis. DaaS makes it easier for data selection from different pools, filter out sensitive data, and make the remaining data available on-demand. In 2013, Gartner conducted a survey of over 720 IT executives from organizations across various sectors to identify the key drivers in the implementation of Big Data technologies [10]. According to the survey, Big Data is not replacing the traditional systems but is supplementing it with additional capability to handle large datasets. Along with increasing demand for infrastructure

IPASJ International Journal of Computer Science (IIJCS)
Web Site: http://www.ipasj.org/IIJCS/IIJCS.htm
A Publisher for Research Motivation ..... Email: editoriijcs@ipasj.org
Volume 5, Issue 1, January 2017 ISSN 2321-5992

capability, IT decision makers are facing a problem of understanding which technologies are required and how they fit into the technology landscape.

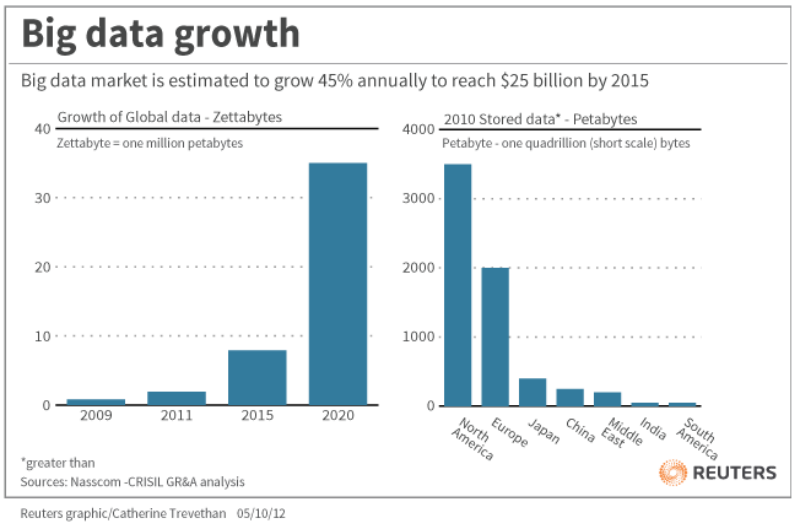

Figure 3: Big Data growth from 2009 to 2020 [14]

4. BIG DATA-AS-A- SERVICE (BDAAS)

The amount of data that produced and consume is growing rapidly in this digital world. Increasing use of social media and use of smart devices generate large amounts of data. If this data is not managed properly it will reach to invaluable information. These large datasets are popularly known as Big Data. This big data is difficult to manage using traditional computing technologies [15]. New technologies are emerging in the market to address the problem of managing and analyzing Big Data. Organizations are finding it difficult to implement these Big Data technologies effectively due to problems such as lack of available expertise [16][17]. Big Data-as-a-Service provides common big data related services to users to enhance efficiency and reduce cost. Table 1 shows the Big Data acquisition methods.

Table 1 Big Data acquisition methods

Type	Sources
Documents	Doc, XLS, PDF, HTML etc.
Sensor Data	Smart electric readers, biometric devices, medical devices, car sensors, road cameras etc.
Media	Images, Video, Audio etc
Social Networks	Facebook, LinkedIn, Twitter, Google+ etc
Public Web Pages	News channels, weather portals, Wikipedia etc
Business Applications	CRM systems, ERP solution systems, HR systems, Project Management software etc.
Data Storages	RDBMS, NoSQL, Hadoop, file systems etc.
Machine Log Data	Attendance logs using machines, Application logs, event logs, server data etc.

BIG DATA CHARACTERISTICS (THE FOUR VS)

The specific attributes that define big data are called the four V's: volume, variety, velocity, and veracity. Figure 3 shows the four Vs and the date types

IPASJ International Journal of Computer Science (IIJCS)
Web Site: http://www.ipasj.org/IIJCS/IIJCS.htm
A Publisher for Research Motivation ..... Email: editoriijcs@ipasj.org
Volume 5, Issue 1, January 2017 ISSN 2321-5992

Figure 4: The four Vs of Big Data

Volume: Large volumes of structured and unstructured data are analyzed.
Variety: The analytic environment has expanded from pulling data from enterprise systems to include big data and unstructured sources.
Velocity: Speed of access to reports that are drawn from data defines the difference between effective and ineffective analytics.
Veracity: Validating data and extracting insights that managers and workers can trust are key factors of successful analytics. Trust in analytics has grown more difficult with the explosion of data sources.
Big Data analytics is a challenging task that requires expensive software and large computational infrastructure, as well as effort. The value of data has been recognized by organizations and realized that data can be analyzed for various purposes. For example organizational data can be used foroperational, tactical and strategic decision making, assessing risk and introducing mechanisms to increase sales. Organizations are moving to a single infrastructure which provides common functionality of big data management, and flexible enough to handle different types of big data and big data analysis tasks [36]. The various infrastructures are Big Data Infrastructure-as-a-Service,Big Data Platform-as-a-Service and Big Data Analytics Software-as-a-Service

5. BIG DATA ANALYTICS IN CLOUD COMPUTING

Big data analytics is the process where advanced analytic techniques operate on big data. Big Data analysis Steps are Acquisition, Recording, Extracting/data cleaning, Integration/Representation, Analysis and Visualization. Cloud is known as nature architecture for Big Data-as-a-Service. There are several solutions for Big Data analysis related to Cloud computing. The various solutions have been created because of the wide range of analytics requirements. Bid data Analytics can be descriptive, predictive or prescriptive. Also Big Data can have various levels of its four characteristics, variety, velocity, volume, and veracity. It is important for organizations to understand the requirements in order to choose appropriate Big Data Analytics tools. Big Data Analysis Requirements are significant requirements for conducting these inquiries in an expedient way: Minimize data movement, Use existing skills and Attend to data security. Following are some of the tools for analyzing Big Data.

- Discovery tools are useful throughout the information lifecycle for rapid, intuitive exploration and analysis of information from any combination of structured and unstructured sources.
- Business Intelligence tools are important for reporting, analysis and performance management, primarily with transactional data from data warehouses and production information systems.
- In-Database Analytics include a variety of techniques for finding patterns and relationships in your data.
- Hadoop is useful for pre-processing data to identity macro trends or find nuggets of information, such as out of-range values.
- Decision Management includes predictive modeling, business rules, and self-learning to take informed action based on the current context.

The elements of big data cloud are: *Big data infrastructure services*-offers core services such as compute, storage, and data services for big data computing; *Big data platform services* - offers schedulers, query mechanisms for data retrieval and data-intensive programming models to address several big data analytic problems; *Big data analytics services* - big data analytics as services over big data cloud infrastructure. Figure 4 shows the elements of big data infrastructure services on a cloud.

IPASJ International Journal of Computer Science (IIJCS)
Web Site: http://www.ipasj.org/IIJCS/IIJCS.htm
A Publisher for Research Motivation ..... Email: editoriijcs@ipasj.org
Volume 5, Issue 1, January 2017 ISSN 2321-5992

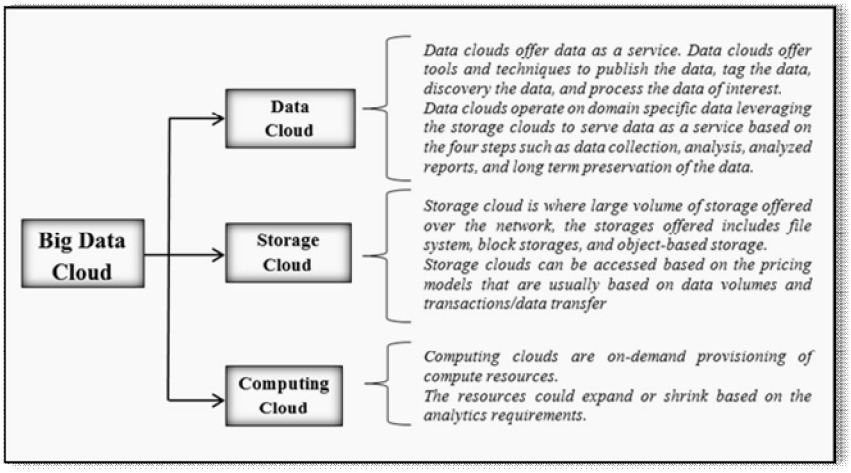

Figure 5: Elements of big data infrastructure services

6. FACTORS TO BE CONSIDERED IMPLEMENTING BIG DATA-AS-A- SERVICE ON THE CLOUD

Cloud computing will help in improving the problems of big data analytics by providing resources on-demand with costs proportional to the actual usage of the service. Following are some of the factors to be considered while implementing big data and analytics on the cloud.

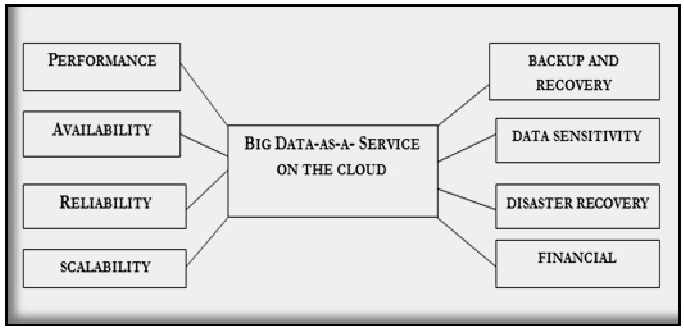

Figure 6: Eight factors to be considered implementing Big Data-as-a- Service on the cloud

Performance –One way of measuring the performance of the services, in this case the data, provided through cloud adoption from user experience. To evaluate this factor there is a need to identify the components to be identified properly which satisfy the performance of services. To evaluate the services there is a need to understand the characteristics of the workload and choose the appropriate technology to support it and then determine whether the performance required of the most demanding workloads can bemet on virtual components or with the use physical ones. The less demanding workloads are usually safe to run on virtual components after validating against the requirements.

Availability- The organizations should determine whether data and solutions must be made highly available to the purposes they are designed for. It is suggested to add flexibility to components where the impact of the necessary data unavailability will poorly affect the service being delivered to the business.

Reliability- The key component of a big data solution is the data so the reliability of data delivery is important. There is a need to investigate the business requirements and deliver reliability for specific business needs. A less expensive capable retry mechanism can be used where guaranteed reliability is not necessary.

Backup and recovery- Ensuring the correct backup and restore systems and procedures should be in place depending on the big data solution used. In business environments, there is a need for providing a backup and recovery solution that protects the data, meets recovery point objective (RPO) and recovery time objective (RTO) thus ensuring business continuity.

Data Sensitivity-The sensitivity of data is an important consideration while implementing big data and analytics on the cloud. Organizations must comply with important regulatory standards based on their domain of execution. A good

suggestion is to keep all sensitive data within the network boundaries of the organization such as on-premises data center or in a private cloud. Organizations can use off-premises big data services for non-sensitive data.

Disaster Recovery (DR)-Organizations should formulate a good DR plan which ensures that there will be a solution available in the case of any disaster. There is an important need to identify that disaster recovery solution is important to the organization. In the case of a disaster, reduced performance is acceptable or not, the minimum application of data and performance requirements are defined or not, how quickly does the business need to recover after a disaster, the RTO and RPO, the acceptability of data loss to be included in a Business Continuity Plan (BCP). The organizations should Design a disaster recovery solution that meets the BCP. The acceptable reduced performance short period of time after a disaster should be defined in a BCP.

Scalability: The need is to determine the best way to provide scalability by adopting proprietary or open source. For big data solutions, scaling can be done horizontally or vertically. Cloud platforms provide an ideal environment for horizontal scaling because of their inherent elastic characteristics and this makes the cloud a good platform to place a component with this requirement.

Financial-The solutions adopted by organizations with the required features to ensure the best outcome can be too expensive and financial considerations are always important in organizations. Organizations should build cloud data solutions based on business need at the most reasonable price that means the solutions should be for *fit for purpose.*

7. BIG DATA LOCATIONS AND PLACEMENT IN CLOUD INFRASTRUCTURE

With the development of technology, a flood of data is created every day by the interactions of billions of people using computers, GPS devices, cell phones, and medical devices [3][4]The placing of big data solution components are important to consider. The data must be placed by meeting the requirements by considering the above mentioned factors such as Performance, Availability, Reliability, Scalability, Financial impact, backup and recovery of data, sensitivity of data and Disaster recovery methods. Usually cost will become the guiding factor if the decision is a choice between on-premises and off-premises cloud locations for a component.

Examples of BDaas solution on clouds are: Microsoft Azure-HD Insight is a Microsoft BDaaS solution that allows instantiation of Hadoop clusters in a cloud environment. HD Insight provides Apache Hadoop as a service, which enables end users to have a more scalable and cost-efficient environment. HDI nsight is the flagship Microsoft solution for cloud-based Big Data analytics, QuBole is a BDaaS company that provides a Hadoop framework on a cloud-based platform. It has features such as auto scaling as well as GUIs for managing the Big Data implementation on a cloud platform. It abstracts the complexity of managing Hadoop clusters and provides real-time scalability.

Traditional Big Data	Big Data as a Service
Structured and unstructured data	Structured and unstructured dataon cloud environment
Scalability in processing and storage achieved through distributed architecture	Scalability on demand through a combination of cloud computing and distributed architecture
Advanced analytics functions	Advanced analytics functions with on-demand computing power
Limited accessibility of data	Global accessibility of data

The differences when shifting Traditional Big Data to BDaaS

8. CONCLUSION

The technological developments of mobile data, sensors, and near-field communications have led to more data being more easily collected. The amounts of data generated by various activities in organizations are increasing data by day. Organizations considers this huge generated volume of data which is known as Big Data as a way of making better decisions and also as obtaining advantage over their competitors. The bid data management and analysis is a challenging

and time consuming task. Every day newer BDaaS technologies are appearing in the market. Cloud computing is seen by organizations as a viable way to reduce costs and increase implementation efficiency. Cloud computing helps organizations by providing resources on-demand with costs proportional to the actual usage and the area of Big Data Computing using Cloud resources is moving fast. Cloud computing plays a key role for Big Data by providing infrastructure and tools as well as it is a business model that Big Data analytics can follow. Vendors are increasingly introducing cloud-enabled versions of their transitional Big Data technologies. This trend gives an indication that most of the Big Data technologies used in various industries today will be cloud-enabled in the near future.

References

[1]. Big data: science in the petabyte era," Nature 455 (7209):1, 2008.
[2]. Buyya R, Vecchiola C, Selvi T. Mastering in Cloud Computing – Foundations and Applications Programming.Morgan Kaufman: USA, 2013.
[3]. Cisco Cloud Computing - Data Center Strategy, architecture, and Solutions, Cisco Systems, (2009)
[4]. D. Logothetis and K. Yocum, "Ad-hoc data processing in the cloud," Proceedings of the VLDB Endowment, vol. 1, no. 2,pp. 1472–1475, 2008
[5]. Dialogic Corporation (2010) 'Introduction to cloud computing: white paper',available at http://www.dialogic.com/products/docs/whitepapers/12023-cloud-computing-wp.pdf (accessed 12 December 2011).
[6]. Dialogic Corporation (2010) 'Introduction to cloud computing: white paper',availableat http://www.dialogic.com/products/docs/whitepapers/12023-cloud-computing-wp.pdf
[7]. Douglas and Laney, "The importance of 'big data': A definition,"2008.
[8]. Implementation of Cloud Computing in Education – A Revolution, Saju Mathew, International Journal of Computer Theory and Engineering, Vol. 4, No. 3, June 2012
[9]. James Manyika et al., "Big data: The next frontier for innovation, competition, and productivity," McKinsey Global Institute, Tech. rep. May 2011.
[10]. Kart, Heudecker and Buytendijk 2013, Gartner Survey
[11]. Mark Raskino, Jackie Fenn, and Alexander Linden, "Extracting Value From the Massively Connected World of 2015," Gartner Research, Tech. rep. 2005.
[12]. Mell, P., & Grance, T. (September 2011). The NIST Definition of Cloud Computing. Retrieved January 28, 2014, from http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.
[13]. Mell, P., Grance, T. (2011). The NIST definition of cloud computing. Gaithersburg, MD: National Institution of Standards and Technology (NIST).
[14]. Reuters , Stastitics Graph-Big Data growth 2099 to 2020
[15]. S. Lohr, "The age of big data," New York Times, vol. 11, 2012
[16]. V. W. Consulting, "Big data, big impact: New possibilities for international development," The World Economic Forum, Tech. Rep., 2012.
[17]. Venkata Narasimha Inukollu, Sailaja Arsi and Srinivasa Rao Ravuri ―Security Issues Associated With Big Data In Cloud Computing‖ International Journal of Network Security & Its Applications (IJNSA), Vol.6, No.3, May 2014.

Author

Minimol Anil Job is working as an Assistant Professor in ITC Department, in AOU Bahrain Branch. She has more than ten tears of academic experience in the field of computer science and information technology. Her current research interests include Information Technology Management in the specialist area of Database management. She has published several papers in the area of information technology, e-learning, cloud computing and green computing in international journals. Her further area of interest is software development.