International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 12– No. 39, April 2022 – www.ijais.org

Overview of Big Data in Cloud Computing Environment

Sara Kaleem
Asia Pacific University of Technology and Innovation (APU), Malaysia

ABSTRACT
Cloud computing is a game-changing technology that helps organizations innovate and compete in the digital economy by lowering upfront infrastructure expenses, and democratizing access to computing and storage capacity. The cloud computing technology provides hardware, software and platforms as a service. It is the explicit solution that enters into the picture in order to process, store, and exchange a massive amount of data generated every second. This paper aims to focus on the use of cloud in big data.

General Terms
Cloud, SaaS, PaaS, IaaS, Cloud Providers

Keywords
Big data, Cloud computing, Security, Privacy, Services

1. INTRODUCTION
Data is one of the most valuable assets for businesses in any industry. The increasing importance and volume of data has generated a new problem which the standard analysis approaches can no longer handle. The solution to this large volume and variety of data is big data. The term big data relates to a wide range of data kinds being produced by a variety of system, that are typically too vast to browse or query on a standard computer. Since the big data cannot be handled or processed locally on a single computer system, the concept of cloud computing was established.

Cloud computing simply refers to the use of a computer from a remote location. The National Institute of Standards and Technology describes cloud computing as a model for enabling 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 (Mell and Grance, 2011). In other words, a cloud is a networked collection of IT resources made up of shared pool of hardware and software resources situated in data centers. The top 4 cloud providers as mentioned by the Gartner magic quadrant are Amazon Web Services, Microsoft Azure,Google Cloud Provider and Alibaba cloud.

In company-run data centres, storing massive and voluminous amounts of data is inconvenient. Furthermore, extracting insights from this data is a time-consuming process that takes a long time to perform and produce results. All these problems can be solved with the help of cloud computing by pushing all the data to the cloud.

Cloud computing has many features, some of the major ones are:

1. Provides a virtualized technological platform that enables for the virtualization of physical resources, computing, and network capabilities. Whatever the deployment patterns, cloud computing attempts to maximize the use of existing technology among a large number of users (Alam, 2020).
2. Provider requirements based on current demand criteria are possible also known as dynamic procurement.
3. Allows access to the Internet via a variety of devices, including PCs, laptops, and mobile phones.
4. Processors are used to regulate and automate the infrastructure, as well as to provide monitoring and billing information (Alam, 2020).

The goal of this study is to delve into the concepts of big data and cloud computing and the relationship between them and also discuss the top 3 cloud providers; Amazon Web Services, Microsoft Azure and Google Cloud Provider.

2. BIG DATA
The term "Big Data" refers to large amounts of data collected from a variety of sources, including social media, sensors, and cell phones. Big data, according to IBM, is a term used to describe new data sets that are too large or complex for standard relational databases to acquire, manage, and analyse with low latency. One or more of the following qualities characterize big data: high volume, high velocity, or high variety."

Big Data underwent a series of evolutionary phases after entering the twenty-first century, and software in a suitable environment was built (Hong et al., 2018). Big Data has reached a certain magnitude, not only in terms of size, but also in terms of data technology, as a result of the rise of information exchanges.Servers, storage systems, cloud services, and networking equipment can be included in Big Data infrastructure.

2.1 Harnessing Data through Cloud Computing Platform
The data can be harnessed through the Cloud Computing platform and utilized in a variety of ways. One of the major issues is ensuring the privacy and security of data stored, communicated, processed, and distributed in the cloud, as well as the people who use that data.

- **Search:** Before being sent to the cloud, data is normally encrypted to protect security. Due to the generation of massive amounts of data, for e.g., in healthcare such as patient records, tests, and other data has given rise to the term "big data," and the necessity to store and move this data to the cloud has also become critical. Data encryption on the cloud is also promoted in order to improve data security and privacy. However, getting access to these encrypted files, such as healthcare records, is difficult. Looking and sharing encrypted information is more troublesome than looking and sharing plain information. In any case, it may be a key duty for the cloud benefit supplier since clients anticipate the cloud to perform a rapid search and create the results without jeopardizing information security. A ciphertext-policy attribute-based mechanism with a keyword search and data sharing (CPAB-KSDS) was introduced for quick data search and sharing of encrypted data on cloud (Ge et al., 2020).

- **Update:** Another important data related task on cloud is the updating of data. Ramesh et al., 2018, proposed the Merkle B+ Hash Tree (MBHT), which was utilized to address security and integrity issues as well as dynamic data updates on the cloud.

- **Storage and Processing:** Another data harnessing method which is commonly being used is the implementation of Hadoop on cloud. Hadoop can be used for storing huge amount of structured/unstructured data by splitting data files into blocks and distributing across multiple nodes in the clusters. Due to its flexibility, accessibility, scalability, parallel processing, ease of programming, and fault tolerance capabilities, several companies are implementing Hadoop with cloud computing. To support spatial big data applications, a cloud enabled Hadoop framework was developed, which coupled cloud technologies and high computational resources with the traditional Hadoop framework (Tripathi et al., 2018). Amazon's Elastic MapReduce and Rackspace's Elastic MapReduce are two common solutions to moving Hadoop on cloud (Khan et al., 2017).

- **Analyse:** Big data entails handling petabytes (and, perhaps, exabytes and zettabytes) of data, and the cloud's scalable environment enables the deployment of data-intensive applications that support business analytics. Analytics in cloud computing, such as measuring and tracking social media engagement, is essentially applying analytics principles to data stored on cloud drives rather than on individual servers or drives. The cloud also makes internal communication and cooperation easier, allowing more staff to have access to important analytics and streamlining data exchange (Chan, 2018). A cloud-based platform, WebMeV, was introduced for analysing and visualizing cancer genomic data (Wang et al.,2017). It allows users to analyse and visualize large cancer related datasets.

3. CLOUD MODELS FOR BIG DATA ANALYTICS

There are two types of cloud models: [1] Cloud Service Model, [2] Cloud Deployment Model. Figure 1 below is the NIST Cloud Reference Model.It is a high-level conceptual model that serves as a useful tool for describing the Cloud's requirements, structure, and functioning. The cloud reference model allows for the examination of security, interoperability, and portability standards. It establishes a set of actors, processes, and functions that aid in the creation of Cloud architectures.

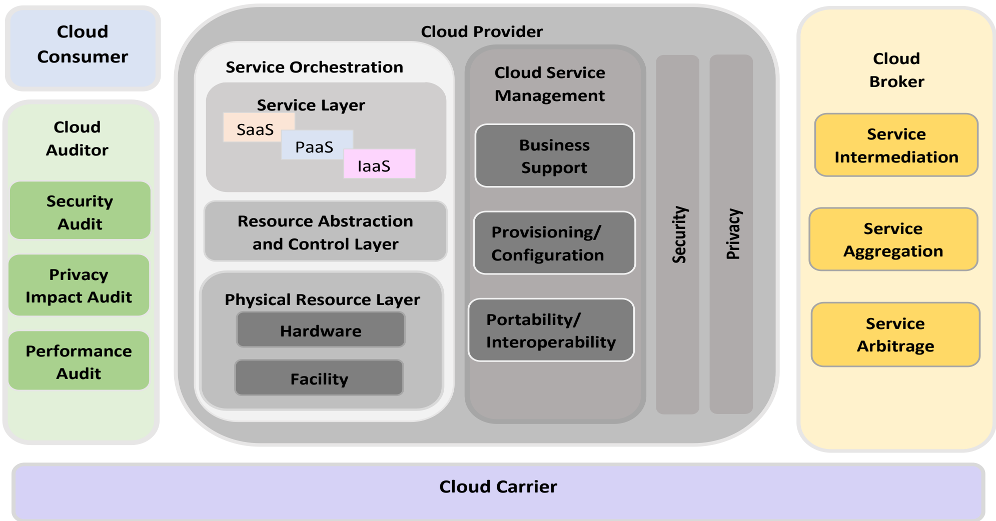

Fig 1: Cloud Reference Model

3.1 Cloud Service Models

The services and capabilities given to customers are defined by a cloud service model. As per NIST, the cloud service is mainly divided into three types:

- **Software as a Service (SaaS):** SaaS enables users to access programs that run on the cloud provider's infrastructure and are controlled by the cloud provider (Kokila et al., 2016). It is the most common service model. It allows users to access cloud services via a network connection, often the Internet and a Web browser (Liu et al., 2011). Some examples of SaaS are Salesforce.com, Google Apps, Office 365, etc.

- **Platform as a Service (PaaS):** PaaS allows cloud users to deploy user-created or acquired applications onto the infrastructure of the cloud provider. An operating system, a programming language execution environment, databases, and web servers are the services included in PaaS. Some examples of PaaS are Microsoft Azure, Google App Engine, Pivotal Cloud Foundry, IBM Cloud, etc.

- **Infrastructure as a Service (IaaS):** IaaS allows cloud users to manage the cloud provider's underlying infrastructure from an abstract or virtual perspective (Liu et al., 2011). IaaS enables the cloud customer to manage and deploy their own software, including the underlying operating system. Some examples of IaaS are Amazon EC2, S3, Google Compute Engine, OpenStack, etc.

From the big data point of view these service models provide the same services but for slightly different functionalities. Figure 2 below, illustrates a summary of Big Data as a Service (BDaaS) Model.

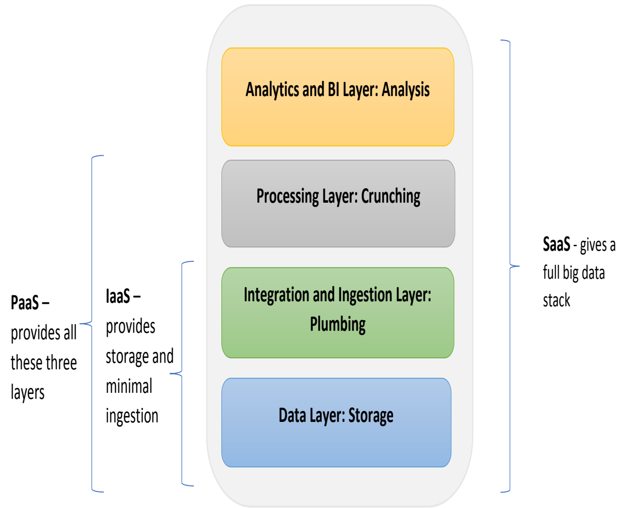

Fig 2: Big Data as a Service Model

The Big Data Infrastructure as a Service simply provides the cloud service provider's fundamental functions. Amazon's AWS IaaS Architecture, which mixes S3 and EC2, is an example of a Big Data IaaS paradigm. The Big Data Platform as a Service provides a whole Big Data Stack similar to Amazon S3 or RedShift. This does not include ETL and BI. Big Data Software as a Service encapsulates the complete big data stack into a single application. The prominent SaaS solutions for a complete data warehouse solution in the cloud are Azure Data Warehouse and AWS Redshift.

A study was conducted which proposed a Big Data-as-a-Service Framework, that included three main planes, i.e., sensing plane, cloud plane and application plane (Wang et al., 2017). A thorough study was done on the concepts of cloud and the characteristics of the three main service model were combined and implemented in the framework along with the storage and management of data on cloud.

3.2 Cloud Deployment Models

The three essential characteristics of the cloud environment are size, ownership, and access. The deployment model for the cloud environment is defined based on these three features. As per NIST, there are 4 deployment models:

- **Private cloud:** A private cloud is utilized by a single corporation, which may have multiple user groups. The organization, a third party, or a combination of the two may maintain, own, and manage it. Physically, it could be located at the organization's headquarters or elsewhere (Simon, 2017).

- **Public cloud:** A public cloud is one that is accessible to the general public and is owned, operated, and delivered by a company, academic institution, government agency, or a combination of the three (Simon, 2017; Kokila, 2016; Liu et al., 2011). In most cases, the Public Cloud will be located at the provider's physical site. Cloud users get access to the Public Cloud's resources as a service.

- **Community cloud:** As per the NIST (Liu et al., 2011),a Community Cloud is utilized by a specific group or community of users from a variety of organizations that are united by a shared objective or issue. It maybe claimed, overseen, and managed by one or more community organizations, a third party, or a blend of the three, and it may take place on or off premises (Simon, 2017).

- **Hybrid cloud:** The hybrid cloud framework is made up of two or more distinct cloud foundations (private, community, or open) that stay partitioned entities but are connected by standardized or restrictive innovation that permits data and application compactness (Simon, 2017).

When it comes to the suitability of deployment models for big data analytics, Hadoop on cloud is the best example.

Hadoop on Public Cloud: Hadoop distributions like Hortonworks, Cloudera, and BigInsights may be deployed and run-on public clouds like Rackspace, Microsoft Azure, and Amazon Web Services (Khan et al., 2017). The term "Hadoop-as-a-Service" is commonly used to describe such a setup. The problem with such solutions is that they rely on cloud providers' Infrastructure-as-a-Service (IaaS). Except for Amazon EMR, the user must install and configure Hadoop on all of the accessible alternatives. Amazon EMR offers a service known as "MapReduce-as-a-Service" (Khan et al., 2017). Hadoop cluster can also be utilized with S3 if the developer does not want to use HDFS as the default storage solution. Some applications may necessitate the use of these characteristics, making S3 an indispensable storage solution. Aside from that, Hadoop with S3 may be the best storage solution for businesses who currently have data on S3.

Hadoop on Private Cloud: The user has more control over Hadoop settings on the cloud with the private cloud (Khan et al., 2017). Some of the benefits of using Hadoop on a private cloud are improved cluster control and visibility, improved data privacy and security issues mitigating.

4. DATA INTENSIVE CLOUD COMPUTING

Large-scale data computing is referred to as data intensive computing. When data-intensive applications are placed on clouds, a variety of requirements and challenges occur. The cloud must be scalable and accessible at all times. It should also make large-scale data processing and input/output activities easier. A data-intensive cloud could be deployed as a private cloud, serving users of a single business, or as a public cloud, giving shared resources to a large number of users, depending on its intended application. Some of the characteristics of data intensive computing are:

- It is vital to reduce data transfer in order to obtain high performance in data intensive computing. By allowing the algorithms to run on the node where the data is stored, the system overhead is reduced and performance is improved.

- The data intensive computing system employs a machine-independent method in which the run-time system is in charge of scheduling, execution, load balancing, communications, and program mobility.

- The dependability and availability of data are extremely important in data intensive computing. Data intensive computing is aimed to overcome the issues that traditional large-scale systems face, such as hardware failures, communication errors, and software defects.

- Data intensive computing is built for scalability, meaning it can handle any quantity of data and meet time-sensitive requirements. One of the most significant advantages of data intensive computing is the scalability of both the hardware and software architecture.

To efficiently cope with the vast amount of Web data for applications such as search engines and social networking, Internet companies such as Google, Yahoo, Microsoft, Facebook, and others needed a new processing approach (Furht & Vallinustre, 2016). Several alternatives have evolved, including Google's Map Reduce architecture, which is currently accessible as an open-source Hadoop implementation utilized by Yahoo, Facebook, and others.

Google's MapReduce architecture and programming paradigm is an example of a modern systems architecture built for processing and analysing enormous datasets, and it is now being utilized by Google in many applications to analyse massive volumes of raw Web data with great success. Because the system handles features like splitting input data, scheduling and executing jobs across a processing cluster, and handling communications between nodes, even programmers with no prior experience with parallel programming may use a large distributed processing environment with ease.The Google File System (GFS) sits beneath and on top of the MapReduce architecture. GFS was created to be a fault-tolerant, scalable distributed file system for very big data files and data-intensive applications that could run on commodity hardware clusters. GFS has proven to be quite successful for data-intensive computation on very big files, but it is less effective for tiny files, which might lead to hot spots if multiple MapReduce operations visit the same file (Furht & Vallinustre, 2016).

Apache Hadoop, an open-source Java-based software system, is used in a wide range of data-intensive applications such as marketing analytics, image processing, machine learning, and web crawling. Hadoop features a distributed file system called HDFS, which is similar to Google MapReduce's GFS implementation.

5. REAL TIME PROCESSING OF BIG DATA

Real-time processing necessitates a constant stream of data intake, processing, and output. Various research has been conducted on the service models of the cloud for real-time processing. Each model consists of several layers which needs to be altered according to the requirement or to achieve a predictable behaviour of a cloud system (Danielsson, 2018). An example of real-time data is Internet of Things (IoT). IoT devices are used in a variety of systems, including smart cities, smart homes, and environmental monitoring (e.g., smart grid) (Zobaed& Salehi, 2019). Many of these systems (for example, monitoring systems) necessitate low latency (real-time) data analytics (Zobaed& Salehi, 2019). The relationship between IoT, big data and cloud computing can be described as IoT being the data source, Big Data as data analysis platform, and Cloud Computing as the place for data storage, scale, and access speed. IoT is the source for data that needs to be extracted for analysis purposes in a corporation. The appropriate data can then be analysed and extracted using Big Data to generate the required information (McKenna, 2021). Big Data not only processes massive volumes of data in real time, but it also saves the information using various storage technologies, making it crucial for utilizing IoT functions and data extraction. IoT and Cloud Computing, when used together, enable systems to be automated in a cost-effective manner that allows for real-time control and data monitoring (McKenna, 2021). Big Data and Cloud can be used together to store enormous amounts of data and enable scalable processing and real-time data analysis. To manage big data in health service applications, a hybrid model of IoT and Cloud Computing was introduced, with the goal of improving the performance of healthcare systems by reducing stakeholders' request execution time, optimizing the required storage of patients' big data, and providing a real-time data retrieval mechanism for those applications. (Elhoseny, 2018).

The results showed that the proposed model surpassed state-of-the-art models by 50 percent in overall execution time and the system's efficiency in terms of real-time data retrieval was improved by 5.2 percent. The lack of physical infrastructure required to bring Big Data, IoT, and the Cloud together lowers costs and allows you to concentrate on improving analytical capacity rather than worrying about maintenance and support. Regular updates can be delivered and knowledge of any infrastructure breaches can be reported up instantly by combining the Cloud with IoT and also helps to improve security. Some of the tools for real time big data processing are Apache Spark, Apache Storm (used by companies like Spotify, Yelp, and WebMD).

6. CRITICAL REVIEW OF POPULAR CLOUD PROVIDERS

This section critically reviews the big data platforms by top 3 cloud providers: AWS, Azure and GCP. Table 1 gives the big data platforms provided by these cloud providers.

Table 1: Cloud Providers

Review	Amazon Web Services (AWS)	Microsoft Azure	Google Cloud Platform (GCP)
About	AWS (Amazon Web Services) is Amazon's complete cloud computing platform.	Microsoft Azure is a public cloud computing platform originally known as Windows Azure.	Google Cloud Platform is a set of Google's public cloud services.
Service Models	IaaS PaaS SaaS	IaaS PaaS SaaS Serverless	IaaS PaaS SaaS

Big Data Platform	Amazon EMR Amazon Redshift Amazon Kinesis	Azure Data Lake Analytics Azure Stream Analytics Azure Databricks	Google cloud BigQuery Google cloud Data Fusion Google cloud Bigtable
Companies	Expedia Yelp Zillow Duolingo McDonalds	eBay Samsung BMW Boeing	PayPal Twitter Target Toyota Nintendo

6.1 Amazon Web Services (AWS)

AWS (Amazon Web Services) is an extensive and evolving Amazon Cloud Computing Platform that contains a mix of Service Infrastructure (IaaS), Service Platforms (PaaS) and Service packaged Software (SaaS). AWS services can provide tools for organizations including computer power, storage of databases and distribution of content. AWS is divided into numerous services; each of them can be modified according to the demands of the user. Users can view the AWS Service's setup and server maps. It provides services across Availability Zones in areas world-wide from dozens of data centres. Amazon Simple Store (S3), for backup data, collecting and analytics, provides scalable object storage. Amazon EMR, Amazon Redshift, Amazon Kinesis and Amazon DynamoDB are some of the big data platforms provided by AWS.

- **Amazon EMR** is the top large data platform for the industry to use open-source tools such as Apache Spark, Apache Hive, Apache HBase, Apache Flink, Apache Hudi and Presto to process enormous quantities of information. Low cost, elasticity, reliability, ease of use, security, and customization are all features of Amazon EMR. It can be used for machine learning, ETL, Clickstream analysis, Real-time streaming, interactive analysis and Genomics. Companies using Amazon EMR include Expedia, Yelp, Zillow, etc.

- **Amazon Redshift** can be used for querying and combining exabytes of structured and semi-structured data across data-warehouses, operational database, and data lake using standard SQL. Redshift can simply export query results from open formats like as Apache Parquet back onto S3 data lake, allowing for more analysis from other analytics services such as Amazon EMR, Amazon Athena and Amazon SageMaker. It can be used for Business intelligence, operational analytics on events and predictive analytics. Yelp, Duolingo, COMCAST, McDonalds, and several other companies use Amazon Redshift.

- Amazon Kinesis enables the collection, processing and analysis of data in real time and streaming to gain immediate insights and respond fast to new information. In addition to providing flexibility for selecting the tools which best meet application requirements, Amazon Kinesis delivers important cost-effectiveness of processing Streaming data. Amazon Kinesis is able to ingest real-time data for machine learning, analytics, and other applications, including video, audio, application logs, web-clickstreams, and IoT telemetry information. Netflix, Zillow, Veritone and SONOS use Amazon Kinesis.

6.2 Microsoft Azure

Microsoft Azure is a public cloud computing platform originally known as Windows Azure. It delivers a variety of cloud services like computing, analysis. Users can select and choose new applications from these providers or execute current applications in public. Due of the different services available to Microsoft Azure, its case studies are very diversified. Cloud Virtual machines or cloud containers are one of Microsoft Azure's most popular applications. Azure is a fast, flexible, and affordable platform.

- **Azure Data Lake Analytics** supports the design of data transformation algorithms in a range of languages, including U-SQL (a proprietary Microsoft language, which brings together the advantages of SQL and C#), Python, .NET and R (Maayan, 2020). It is able to process data petabytes.

- **Azure Stream Analytics -** An end-to-end stream can be constructed for streaming events via Azure Stream Analytics. The technology is serverless. Stream Analytics allows the determination of a data streaming analytics pipeline with SQL syntax data processing and production within minutes. Depending on the amount and the output of streaming data, it escalates elastically. Because streaming data frequently requires extremely high efficiency and real-time replies, Azure Stream Analytics enables sub-second latency with a guaranteed event processing "exactly once" (Maayan, 2020). It also provides access to 99.9 percent.

- **Azure Databricks** is an Apache Spark analytics service. Apache Spark is a veteran tool used to process large quantities of high-speed unstructured data. Databricks supports languages like TensorFlow and PyTorch, such as Python, Scala, Java, SQL, and R, which enable Spark data to interact with any of these languages and frameworks (Maayan, 2020). Databricks also incorporates Azure Machine Learning, allowing you to access a wide range of prepared algorithms.

6.3 Google Cloud Platform (GCP)

Multiple services for large data storage and analysis are provided by the Google Cloud Platform. The most important is perhaps BigQuery, a SQL-compatible high-performance engine that can analyse very huge volumes of information in seconds. GCP provides various additional services to build a complete Cloud-based Big Data infrastructure, including Dataflow, Dataproc and Data Fusion.

- **Google Cloud BigQuery** enables data sets containing huge volumes of data to be stored and queried. The service employs a table structure, supports SQL and connects all GCP services easily. For batch processing and streaming, BigQuery can be used. This is a perfect service for offline analytics and interactive queries.

- **Google Cloud Data Fusion** is an integration solution fully managed that allows stakeholders of varying levels of expertise to prepare, transmit and transform information. With a point-and-click, Data Fusion facilitates the creation of code-free ETL/ELT data pipelines. Data Fusion, a portability project for working with hybrid and multicloud connections, is an open-source project.

- **Google Cloud Bigtable** is a fully managed NoSQL database service that offers exceptional performance for large-scale data loads. Bigtable works on an open source HBase API with a low latency storage stack, is available worldwide. The service is perfectly suited for time, financial, commercial, graphical and IoT applications. It enables central Google services such as analysis, search, maps and Gmail.

7. SECURITY AND PRIVACY CONCERNS

While providers of cloud services maintain their customers a degree of security and privacy, this is not expressly set and users can face security and privacy problems together with others. Other typical concerns facing users in the Big Data cloud include storage, data quality, interoperability and portability, and cost-related problems (Balachandran, 2017). As discussed earlier in this research, storage and analysis of the vast cloud data has its favourable effects. However, the security and privacy aspects of the cloud are one of the main issues regarding the use of the cloud for large data analysis. The idea of online data storage on the Cloud emphasizes the problem of data leakage and the loss of data. If not secured correctly, sensitive information stored in the cloud may be leaked. Data leakage or loss of data saved on your cloud may be caused by several reasons. These might be caused by data violations, misconfiguration and poor changes, lack of a cloud security architecture, problems related to identification credentials or key management, account hacking and deprivation, malicious insiders and unsecure user interfaces (Riaz et al., 2020). These problems may, however, be dealt with using security apps, file systems encryption, data loss countering tools, etc (Balachandran, 2017). Interoperability and portability that can be affected by the supplier lock-in are another worry of using the cloud for Big Data. Customers who have used a cloud service provider's service cannot move freely with another cloud service provider and are therefore subject to the cloud service provider leaving room for large cost and incompatibility with other providers (Martins et al., 2016). This situation is particularly serious. This problem is because the cloud service providers are not standardized and can be combated by completing research before using the services of a given cloud service provider. The cloud provides customers with enormous storage capabilities, in which users can store big volumes of data that are essential to their enterprises. However, it might cost businesses, since cloud services offer the different service providers have no suitable standardized pricing strategy. Big data are also becoming difficult for a corporation, as the company may have to store increasingly important data while competing with other companies. Data quality problems may also occur as firms hold increasingly important data to affect the accuracy and timely accessibility of insights (Balachandran, 2017). The cost problem is the growing amount of saved data, bandwidth and computer power necessary to process the data, as well as the vast number of data saved. The costs may vary with the variance in the demand for resources, which might affect an organization's budgeting.

8. CONCLUSION AND FUTURE SCOPE

This paper presented an overview of the cloud computing technology in detail and also gave insight on how cloud computing can help industries to manage big data. The cloud reference model was discussed in detail which included discussion about deployment and service models.This paper also included critical overview of the top 3 cloud providers as per the Gartner magic quadrant; Amazon Web Services, Microsoft Azure and Google Cloud Platform. The future work can be focused more towards the security and privacy of the big data over cloud as it is still one of the major concerns.

9. REFERENCES

[1] Mell, P. and Grance, T., 2011. The NIST definition of cloud computing.

[2] Liu, F., Tong, J., Mao, J., Bohn, R., Messina, J., Badger, L. and Leaf, D., 2011. NIST cloud computing reference architecture. NIST special publication, 500(2011), pp.1-28.

[3] Liang, H, et al (2018) Big Data in Health Care: Applications and Challenges. Data and Information Management; Warsaw Vol. 2, Issue 3. pp. 175-197.

[4] Logo, S., Introduction to Cloud Computing.

[5] Alam, T., 2021. Cloud Computing and its role in the Information Technology. IAIC Transactions on Sustainable Digital Innovation (ITSDI), 1, pp.108-115.

[6] Ge, C., Susilo, W., Liu, Z., Xia, J., Szalachowski, P. and Liming, F., 2020. Secure keyword search and data sharing mechanism for cloud computing. IEEE Transactions on Dependable and Secure Computing.

[7] Tripathi, A.K., Agrawal, S. and Gupta, R.D., 2018. A COMPARATIVE ANALYSIS OF CONVENTIONAL HADOOP WITH PROPOSED CLOUD ENABLED HADOOP FRAMEWORK FOR SPATIAL BIG DATA PROCESSING. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 4(5).

[8] Chan, M., 2018. Cloud and Big Data: Why Cloud Computing is the Answer to Your Big Data Initiatives. [online] Thorn Technologies. Available at: <https://www.thorntech.com/big-data-in-the-cloud/> [Accessed 5 September 2021].

[9] Wang, Y.E., Kutnetsov, L., Partensky, A., Farid, J. and Quackenbush, J., 2017. WebMeV: a cloud platform for analyzing and visualizing cancer genomic data. Cancer research, 77(21), pp.e11-e14.

[10] Kokila, P.M., Saravanan, P., Jagadhesan, B. and Sharmila, R., 2016. Big Data and Cloud Computing Service Models and Nosql Deployment. International Journal of Science and Engineering Invention, 2(07), pp.192-to.

[11] Wang, X., Yang, L.T., Liu, H. and Deen, M.J., 2017. A big data-as-a-service framework: State-of-the-art and perspectives. IEEE Transactions on Big Data, 4(3), pp.325-340.

[12] Simmon, E., 2017. DRAFT-evaluation of cloud computing services based on NIST 800-145. NIST Special Publication (SP), pp.500-322.

[13] Zhang, Q., Cheng, L. and Boutaba, R., 2010. Cloud computing: state-of-the-art and research challenges. Journal of internet services and applications, 1(1), pp.7-18.

[14] Khan, S., Shakil, K.A. and Alam, M., 2017. Big data computing using cloud-based technologies, challenges and future perspectives. arXiv preprint arXiv:1712.05233.

[15] McKenna, N., 2021. The Relationship Between IoT, Big Data and Cloud Computing. [online] McKenna Consultants. Available at: <https://www.mckennaconsultants.com/relationship-between-iot-big-data-and-cloud-computing/> [Accessed 7 September 2021].

[16] Danielsson, J., Tsog, N. and Kunnappilly, A., 2018, December. A Systematic Mapping Study on Real-Time Cloud Services. In 2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion) (pp. 245-251). IEEE.

[17] Elhoseny, M., Abdelaziz, A., Salama, A.S., Riad, A.M., Muhammad, K. and Sangaiah, A.K., 2018. A hybrid model of internet of things and cloud computing to manage big data in health services applications. Future generation computer systems, 86, pp.1383-1394.

[18] Zobaed, S.M., Salehi, M.A., Zomaya, A. and Sakr, S., 2019. Big Data in the Cloud.

[19] Balachandran, B.M. and Prasad, S., 2017. Challenges and benefits of deploying big data analytics in the cloud for business intelligence. Procedia Computer Science, 112, pp.1112-1122.

[20] Riaz, S., Khan, A.H., Haroon, M., Latif, S. and Bhatti, S., 2020, August. Big data security and privacy: Current challenges and future research perspective in cloud environment. In 2020 International Conference on Information Management and Technology (ICIMTech) (pp. 977-982). IEEE.

[21] Opara-Martins, J., Sahandi, R. and Tian, F., 2016. Critical analysis of vendor lock-in and its impact on cloud computing migration: a business perspective. Journal of Cloud Computing, 5(1), pp.1-18.

[22] McCoy, L., n.d. Microsoft Azure Explained: what it is and how to use it | CCB Tecnology. [online] CCB Technology. Available at: <https://ccbtechnology.com/what-microsoft-azure-is-and-why-it-matters/> [Accessed 7 September 2021].

[23] Maayan, G., 2020. Eight Big Data Analytics Options on Microsoft Azure - DATAVERSITY. [online] DATAVERSITY. Available at: <https://www.dataversity.net/eight-big-data-analytics-options-on-microsoft-azure/#> [Accessed 7 September 2021].

[24] Yifat Perry, P., 2021. Google Cloud Big Data: Build a Big Data Architecture on GCP. [online] Cloud.netapp.com. Available at: <https://cloud.netapp.com/blog/gcp-cvo-blg-google-cloud-big-data-build-a-big-data-architecture-on-gcp> [Accessed 7 September 2021].

[25] Meier, R., 2017. What are the Google Cloud Platform (GCP) Services?. [online] Medium. Available at: <https://medium.com/google-cloud/what-are-the-google-cloud-platform-gcp-services-285f1988957a> [Accessed 7 September 2021].

[26] Link.springer.com. 2016. Big Data Technologies and Applications | SpringerLink. [online] Available at: <https://link.springer.com/book/10.1007%2F978-3-319-44550-2> [Accessed 7 September 2021].