

## Big Data and Current Cloud Computing Issues and Challenges

*M Sowmya Reddy*

Lecture in computer science

[Msowmya1235@gmail.com](mailto:Msowmya1235@gmail.com)

### Abstract

Big data applications are a great benefit to organizations, business, companies and many large scale and small scale industries. We also discuss various possible solutions for the issues in cloud computing security and Hadoop. Cloud computing security is developing at a rapid pace which includes computer security, network security, information security, and data privacy. Cloud computing plays a very vital role in protecting data, applications and the related infrastructure with the help of policies, technologies, controls, and big data tools. Big data is a data analysis methodology enabled by recent advances in technologies and architecture. Cloud computing is a set of its services that are provided to a customer over a network on a leased basis and with the ability to scale up or down their service requirements. Its advantages include scalability, resilience, flexibility, efficiency and outsourcing non-core activities.

This paper introduces a detailed analysis between big data and cloud computing security issues and challenges focusing on the cloud computing types and the service delivery types.

However, big data entails a huge commitment of hardware and processing resources, making adoption costs of big data technology prohibitive to small and medium sized businesses. It offers an innovative business model for organizations to adopt its services without upfront investment irrespective of the potential gains achieved from the cloud computing, the organizations are slow in accepting it due to the security issues and associated challenges. Security is one of the major issues which hamper the growth of cloud. The use of big data could provide sufficient benefit to a small to medium sized company to the extent that the business would be willing to commit resources to implement big data technology in-house. However, the level of benefit is difficult to determine without some experience. The main focus is on security issues in cloud computing that are associated with big data. Moreover, cloud computing, big data and its applications, advantages are likely to represent the most promising new frontiers in science

Keywords – Big Data, Cloud Computing, Scalability, Infrastructure, IT

### INTRODUCTION

The concept of cloud computing originated in early 1960s, when John McCarthy an American

computer scientist predicted the eventual convergence of computing infrastructure, which would allow a great degree of versatility in the

distribution of IT resources in order to meet fluctuating and unpredictable business demands. It is an extremely successful paradigm of service oriented computing which has revolutionized the way computing infrastructure is abstracted and used. The popularity and versatility of the internet enabled companies like Amazon to modernized their data centers and adopt the cloud architecture which allowed them to outsource computing power to external customers. This service was later launched in 2006, after which numerous companies took the initiative to create cloud-based infrastructure. Cloud Computing is a term used to describe a new class of network based computing that takes place over the Internet or a model that relies on a large, centralized data center to store and process a great wealth of information.

It can be defined as a collection of integrated and networked hardware, software and Internet infrastructure called a platform i.e. using the Internet for communication and transporting hardware, software and networking services to clients. This platform hides the complexity and details of the underlying infrastructure from users and applications by providing very simple graphical interface or API (Applications Programming Interface) and also provides on-demand services that are always on, anywhere, anytime and anyplace. It is a model for enabling 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 intervention. Computing power and storage space is provided on-demand to companies that outsource their IT management to the cloud service provider.

Cloud computing is a way to increase the capacity or add capabilities dynamically without investing in new infrastructure, training new personnel, or licensing new software. But as more and more information are placed in the cloud, concerns begin to grow about the security of the cloud environment. Security issues in cloud computing has played a major role in slowing down its acceptance. This work is a survey more specific to the different security issues and the associated challenges that has emanated in the cloud computing system.

### **Cloud Computing**

Cloud Computing is a technology which depends on sharing of computing resources than having local servers or personal devices to handle the applications. In Cloud Computing, the word “Cloud” means “The Internet”, so Cloud Computing means a type of computing in which services are delivered through the Internet. The goal of Cloud Computing is to make use of increasing computing power to execute millions of instructions per second. Cloud Computing uses networks of a large group of servers with specialized connections to distribute data processing among the servers. Instead of installing a software suite for each computer, this technology requires to install single software in

each computer that allows users to log into a Web-based service and which also hosts all the programs required by the user. There's a significant workload shift, in a cloud computing system.

Local computers no longer have to take the entire burden when it comes to running applications. Cloud computing technology is being used to minimize the usage cost of computing resources. The cloud network, consisting of a network of computers, handles the load instead. The cost of software and hardware on the user end decreases.

The only thing that must be done at the user's end is to run the cloud interface software to connect to the cloud. Cloud Computing consists of a front end and back end. The front end includes the user's computer and software required to access the cloud network. Back end consists of various computers, servers and database systems that create the cloud. The user can access applications in the cloud network from anywhere by connecting to the cloud using the Internet. Some of the real time applications which use Cloud Computing are Gmail ,Google Calendar, Google Docs and Drop box etc.,

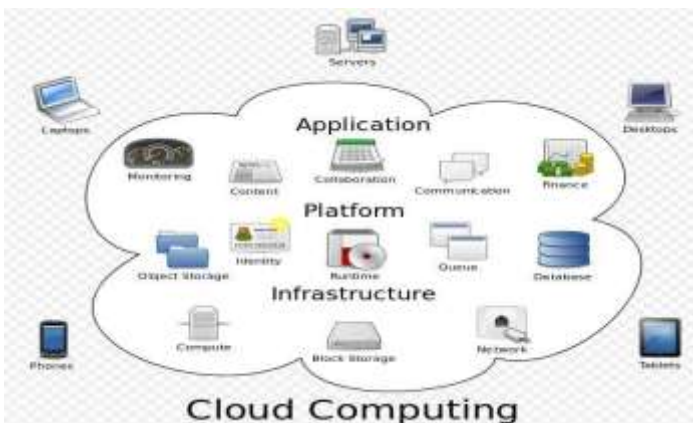


Fig1. Cloud Computing

### Types of Clouds

The Cloud Computing model has three types of clouds model which are

- 1) Public cloud
- 2) Private cloud
- 3) Hybrid cloud
- 4) Community cloud

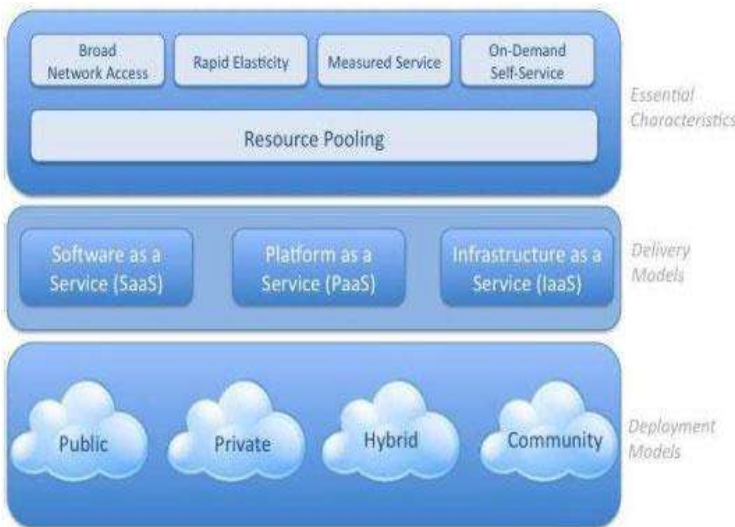
### Public cloud

A public cloud is the pay- as-you-go services available to the general public. In this configuration, a business does not own the core technology resources and services but outsource these to service providers. Public cloud is also considered to be an external cloud. Public cloud describes cloud computing in the traditional mainstream sense, whereby resources are dynamically provisioned on a fine-grained, self-service basis over the Internet, via web applications/web services, from an off-site third-party provider who shares resources and bills on a fine-grained utility computing basis. It is typically based on a pay-per-use model, similar to a prepaid electricity metering system, whose flexibility caters for spikes in demand for cloud optimization. Public clouds are less secure than the other cloud models because it places an additional burden of ensuring all applications and data accessed on the public cloud are not subjected to malicious attacks.

### Private cloud

A private cloud is internal data center of a business that is not available to the general public but uses cloud structure. In this configuration, resources and services are owned by the business, with the services accessible within the business through the intranet and since the technology is owned and operated by the business, this type of cloud is more expensive than a public cloud. It is also more secure and because of its specified internal exposure, only the organization and designated stakeholders may have access to operate on a specific Private cloud. A private cloud is an internal cloud residing inside the company's firewall and managed by the company.

cloud; it uses a public cloud for some tasks and a private cloud for other tasks. In this model, a company uses the public cloud to expedite extra tasks that cannot be easily run in the company's data center or on its private cloud. A hybrid cloud allows a company to maintain critical, confidential data and information within its firewall while leveraging the public cloud for non-confidential data. The private cloud portion of the hybrid cloud is accessed by company employees, both in the company and on the go, and is maintained by the internal technology group. The private cloud part of the hybrid cloud is also accessed by the company employees but is maintained by external service providers. Each portion of the hybrid cloud can connect to the other portion.



NIST Visual Model of Cloud Computing Definition

**Hybrid cloud**

Hybrid cloud is a combination of both public and private cloud, when a company uses a hybrid

**Community cloud**

Community cloud is a private cloud that is shared by several customers with similar security concerns and the same data and applications sensitivity.



**Big Data**

Big Data is the word used to describe massive volumes of structured and unstructured data that are so large that it is very difficult to process this data using traditional databases and software technologies. The term “Big Data” is believed to be originated from the Web search companies who had to query loosely structured very large distributed data. The three main terms that signify Big Data have the following properties:

**a) Volume:** Many factors contribute towards increasing Volume - storing transaction data, live streaming data and data collect from sensors etc.,

**b) Variety:** Today data comes in all types of formats – from traditional databases, text documents, emails, video, audio, transaction s etc.,

**c) Velocity:** This means how fast the data is being produced and how fast the data needs to be processed to meet the demand.

The other two dimensions that need to consider with respect to Big Data are Variability and Complexity.

**d) Variability:** Along with the Velocity, the data flows can be highly inconsistent with periodic peaks.

**e) Complexity:** Complexity of the data also needs to be considered when the data is coming from multiple sources. The data must be linked,

matched, cleansed and transformed into required formats before actual processing.

Technologies today not only support the collection of large amounts of data, but also help in utilizing such data effectively. Some of the real time examples of Big Data are Credit card transactions made all over the world with respect to a Bank, Walmart customer transactions, and Face book users generating social interaction data.

When making an attempt to understand the concept of Big Data, the words such as “Map Reduce” and “Hadoop” cannot be avoided.

### **Hadoop**

Hadoop, which is a free, Java-based programming frame work, supports the processing of large sets of data in a distributed computing environment. It is a part of the Apache project sponsored by the Apache Software Foundation. Hadoop cluster uses a Master/Slave structure. Using Hadoop, large data sets can be processed across a cluster of servers and applications can be run on systems with thousands of nodes involving thousands offer bytes. Distributed file system in Hadoop helps in rapid data transfer rates and allows the system to continue its normal operation even in the case of some node failures. This approach lowers the risk of an entire system failure, even in the case of a significant number of node failures. Hadoop enables a computing solution that is scalable, cost effective, flexible and fault tolerant. Hadoop Framework is used by popular companies like



Google, Yahoo, Amazon and IBM etc., to support their applications involving huge amounts of data. Hadoop has two main sub projects – Map Reduce and Hadoop Distributed File System (HDFS).

### Map Reduce

Hadoop Map Reduce is a framework used to write applications that process large amounts of data in parallel on clusters of commodity hardware resources in a reliable, fault-tolerant manner. A Map Reduce job first divides the data into individual chunks 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 job are both stored in a file-system. Scheduling, Monitoring and re-executing failed tasks are taken care by the framework.

### Hadoop Distributed File System (HDFS)

HDFS is a file system that spans 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.

### Big data advantages

In Big data, the software packages provide a rich set of tools and options where an individual could map the entire data landscape across the company, thus allowing the individual to analyze the threats

he/she faces internally. This is considered as one of the main advantages as big data keeps the data safe. With this an individual can be able to detect the potentially sensitive information that is not protected in an appropriate manner and makes sure it is stored according to the regulatory requirements.

There are some common characteristics of big data, such as big data integrates both structured and unstructured data. Addresses speed and scalability, mobility and security, flexibility and stability. In big data the realization time to information is critical to extract value from various data sources, including mobile devices, radio frequency identification, the web and a growing list of automated sensory technologies.

All the organizations and business would benefit from speed, capacity, and scalability of cloud storage. Moreover, end users can visualize the data and companies can find new business opportunities. Another notable advantage with big-data is, data analytics, which allow the individual to personalize the content or look and feel of the website in real time so that it suits the each customer entering the website. If big data are combined with predictive analytics, it produces a challenge for many industries. The combination results in the exploration of these four areas:

- 1) Calculate the risks on large portfolios
- 2) Detect, prevent, and re-audit financial fraud
- 3) Improve delinquent collections

- 4) Execute high value marketing campaigns

should be a balance between data privacy and national security.

### **Benefits of big data in security**

For marketing and research, many of the businesses use big data, but may not have the fundamental assets particularly from a security perspective. If a security breach occurs to big data, it would result in even more serious legal repercussions and reputation damage than at present. In this new era, many companies are using the technology to store and analyze peta bytes of data about their company, business and their customers. As a result, information classification becomes even more critical. For making big data secure, techniques such as encryption, logging, honey pot detection must be necessary. In many organizations, the deployment of big data for fraud detection is very attractive and useful.

The challenge of detecting and preventing advanced threats and malicious intruders must be solved using big data style analysis. These techniques help in detecting the threats in the early stages using more sophisticated pattern analysis and analyzing multiple data sources. Not only security but also data privacy challenges existing industries and federal organizations. With the increase in the use of big data in business, many companies are wrestling with privacy issues. Data privacy is a liability, thus companies must be on privacy defensive. But unlike security, privacy should be considered as an asset, therefore it becomes a selling point for both customers and other stakeholders. There

### **Benefits of Cloud Computing**

There are lots of benefits in using Cloud computing to render or access computing resources. Presently a lot of people use

Cloud computing without even knowing what it means. For example, Gmail, Yahoo mail, You Tube, and Skype users...are all in the Cloud. Increasingly companies and organizations are becoming aware of the huge benefits that Cloud computing provides. Some of these benefits include:

#### **Flexibility and storage**

With Cloud computing Files are stored in the "Cloud". This allows for development in the organization because workers no longer have to worry about the storage of documents. Also, workers can access office files from wherever and whenever. Workers can also work together virtually even when they are not at the same place at the same time. Various documents can be viewed simultaneously provided Internet connection is available.

#### **Time saving**

Alongside easy collaboration, Cloud computing also aids the easy access to information. Easy access in this context could be seen in how fast it is to access Gmail, Yahoo mail, mailboxes in

general. It is fast and easy in contrast to the time it would take to download and install software.

### Reduced Cost

Cloud computing puts a stop to the illegal reproduction and distribution of software. Some software on the Cloud is free. For example, most SaaS solutions have a pay-as-you-go pricing model instead of a large up-front investment. Such pricing models allow end users to pay only for what they use thus freeing up resources such as time and money for other more important (core) business activities. Cloud computing is therefore cheaper and less labor intensive for companies. There is no need to buy and install expensive software. There is no need to acquire, track and manage software licenses.

### Issues and Challenges

Cloud computing comes with numerous security issues because it encompasses many technologies including networks, databases, operating systems, virtualization, resource scheduling, transaction management, load balancing, concurrency control and memory management. Hence, security issues of these systems and technologies are applicable to cloud computing. For example, it is very important for the network which interconnects the systems in a cloud to be secure. Also, virtualization paradigm in cloud computing results in several security concerns. For example,

mapping of the virtual machines to the physical machines has to be performed very securely.

Data security not only involves the encryption of the data, but also ensures that appropriate policies are enforced for data sharing. In addition, resource allocation and memory management algorithms also have to be secure. The big data issues are most acutely felt in certain industries, such as telecoms, web marketing and advertising, retail and financial services, and certain government activities. The data explosion is going to make life difficult in many industries, and the companies will gain considerable advantage which is capable to adapt well and gain the ability to analyze such data explosions over those other companies. Finally, data mining techniques can be used in the malware detection in clouds. The challenges of security in cloud computing environments can be categorized into network level, user authentication level, data level, and generic issues. Cloud computing is associated with numerous challenges and the major challenges that prevent Cloud Computing from being adopted are as follows:

**1) Network level:** The challenges that can be categorized under a network level deal with network protocols and network security, such as distributed nodes, distributed data, Internode communication.

**2) Authentication level:** The challenges that can be categorized under user authentication level



deals with encryption/decryption techniques, authentication methods such as administrative rights for nodes, authentication of applications and nodes, and logging.

**3) Data level:** The challenges that can be categorized under data level deals with data integrity and availability such as data protection and distributed data.

**4) Generic types:** The challenges that can be categorized under general level are traditional security tools, and use of different technologies

**5) Security:** Security issue plays the most important role in hindering Cloud computing acceptance. Security issues such as data loss, phishing, botnet poses serious threats to organization's data and software. For example, hackers can use Cloud to organize botnet as Cloud often provides more reliable infrastructure services at a relatively cheaper price for them to start an attack.

**6) Costing Model:** Cloud consumers must consider the tradeoffs amongst computation, communication, and integration. While migrating to the Cloud can significantly reduce the infrastructure cost, it does raise the cost of data communication, i.e. the cost of transferring an organization's data to and from the public and community Cloud and the cost per unit of computing resource used is likely to be higher. This problem is particularly prominent if the consumer uses the hybrid cloud deployment

model where the organization's data is distributed amongst a number of public/private (in-house IT infrastructure)/community clouds.

**Charging Model:** The elastic resource pool has made the cost analysis a lot more complicated than regular data centers, which often calculates their cost based on consumptions of static computing. Moreover, an instantiated virtual machine has become the unit of cost analysis rather than the underlying physical server. For SaaS cloud providers, the cost of developing multitenancy within their offering can be very substantial. These include: re-design and redevelopment of the software that was originally used for single-tenancy, cost of providing new features that allow for intensive customization, performance and security enhancement for concurrent user access, and dealing with complexities induced by the above changes.

**Service Level Agreement (SLA):** Since cloud consumers do not have control over the underlying computing resources, they do need to ensure the quality, availability, reliability, and performance of these resources when consumers have migrated their core business functions onto their entrusted cloud. It is therefore vital for consumers to obtain guarantees from providers on service delivery. Typically, these are provided through Service Level Agreements (SLAs) negotiated between the providers and consumers. The very first issue is the definition of SLA specifications in such a way that has an appropriate level of granularity, namely the

tradeoffs between expressiveness and complicatedness, so that they can cover most of the consumer expectations and is relatively simple to be weighted, verified, evaluated, and enforced by the resource allocation mechanism on the cloud.

### Conclusion

Despite the benefits enumerated, it is surprising that not many companies and organizations are rushing to leverage the advantages of Cloud computing, especially in developing countries because the benefits of cloud computing are tempered by two major concerns –

- 1) Security
- 2) Loss of control.

Although Big data and Cloud computing is a new phenomenon which is set to revolutionize caution must be exercised in the way we use the Internet. There are many new technologies emerging at a rapid rate, each with improvements in making living much easier for users. However, there is a need for a cost-performance trade off while deliberating on what type of cloud service to adopt. If the data being processed is considered mission critical to the company, the more expensive private cloud, implemented in-house, would provide a more secured environment with the company keeping the mission critical data in-house.

The use of big data could provide sufficient benefit to a small to medium sized company to the extent that the business would be willing to

commit resources to implement big data technology in-house. However, the level of benefit is difficult to determine without some experience. Cloud computing implementation of big data has the potential to become a frontrunner in promoting a secure, virtual and economically viable IT solution in the future.

Cloud computing may prove to be a better option for most businesses down the line. Indicators show that more people are beginning to understand what it entails, how it works and its usefulness. But still, the fear of surrendering absolute control of core business operations and processes to the Cloud still remains a nagging issue that will only go away in the future.

Although it will be extremely difficult to provide a complete solution to securing the Cloud in relative terms thereby completely erasing the anxiety and disquiet which the idea of outsourcing key and crucial business operations, and technology may create. Security measures such as data encryption is advised and should be employed by end-users to reduce the fear and risk of data invasion. If the fear of Cloud computing are addressed, Cloud computing will revolutionize the world of information technology, even in developing countries Cloud computing promises real benefits to companies seeking a competitive edge in today's economy.

### REFERENCES

- 1) Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Zaharia, M. (2010, April). A view of cloud computing. *Communications of the ACM*, 53(4), 50-58. DOI: 10.1145/1721654.1721672.
- 2) Rouse, M. (2010b, August). Infrastructure as a Service. Retrieved from <http://searchcloudcomputing.techtarget.com/definition/Infrastructure-as-a-Service-IaaS>
- 3) Cisco. (2009). Infrastructure as a Service: Accelerating time to profitable new revenue streams. Retrieved from [http://www.cisco.com/en/US/solutions/collateral/ns341/ns991/ns995/IaaS\\_BDM\\_WP.pdf](http://www.cisco.com/en/US/solutions/collateral/ns341/ns991/ns995/IaaS_BDM_WP.pdf)
- 4) Salesforce.com. (2012). The end of software: Building and running applications in the cloud. Retrieved from <http://www.salesforce.com/paas/>
- 5) Géczy, P., Izumi, N., & Hasida, K. (2012). Cloudsourcing: Managing cloud adoption. *Global Journal of Business Research*, 6(2), 57-70.
- 6) Oracle. (2012). Oracle platform as a service. Retrieved from <http://www.oracle.com/us/technologies/cloud/oracle-platform-as-a-service-408171.html>
- 7) Jackson, K. L. (2012). Platform-as-a-service: The game changer. Retrieved from <http://www.forbes.com/sites/kevinjackson/2012/01/25/platform-as-a-service-the-gamechanger/>
- 8) Cole, B. (2012). Looking at business size, budget when choosing between SaaS and hosted ERP. E-guide: Evaluating SaaS vs. on premise for ERP systems. Retrieved from [http://docs.media.bitpipe.com/io\\_10x/io\\_104515/item\\_548729/SAP\\_sManERP\\_IO%23104515\\_EG\\_uide\\_061212.pdf](http://docs.media.bitpipe.com/io_10x/io_104515/item_548729/SAP_sManERP_IO%23104515_EG_uide_061212.pdf)
- 9) Carraro, G., & Chong, F. (2006, October). Software as a service: An enterprise perspective. Retrieved from [http://msdn.microsoft.com/en-us/library/aa905332.aspx#enterprisertw\\_topic3](http://msdn.microsoft.com/en-us/library/aa905332.aspx#enterprisertw_topic3)
- 10) Rouse, M. (2010a, August). Software as a service. Retrieved from <http://searchcloudcomputing.techtarget.com/definition/Software-as-a-Service>
- 11) .techtarget.com/definition/Software-as-a-Service
- 12) ComputerWeekly.com. (2009, March). Hardware as a service. Retrieved from <http://www.computerweekly.com/feature/Hardware-as-a-Service>
- 13) Rouse, M. (2007, December). Hardware as a service. Retrieved from <http://searchchannel>
- 14) Panettieri, J. (2011, June 13). Can Google take hardware as a service (HaaS) mainstream? MSPMentor. Retrieved from <http://www.mspmentor.net/2011/06/13/can->

- google-takehardware- as-a-service-haas- mainstream/ 16) IOS Press. (2011). Guidelines on security and privacy in public cloud computing. Journal of EGovernance, 34 149-151. DOI: 10.3233/GOV-2011-0271
- 15) White, C. (2011). Data Communications and Computer Networks: A business user's approach, (6th ed.). Boston: Cengage Learning.