Security Issues with Big Data in Cloud Computing

Reeta Mishra
Department of Information Technology
K. J. Institute of Engineering & Technology, Savli, Vadodara, Gujarat

Abstract— This paper have detailed analysis between big data and cloud computing security issues and challenges focusing on the cloud computing types and the service delivery types. Big data is a data analysis methodology enabled by recent advances in technologies and architecture. 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. Cloud computing is a set of it 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. It advantages includes scalability, resilience, flexibility, efficiency and outsourcing non-core activities .As security is one of the main challenges that hinder the growth of cloud computing. At the same time, service providers strive to reduce the risks over the clouds and increase their reliability in order to build mutual trust between them and the cloud customers.

Key words: Cloud Computing; Big Data, Data Security, Cloud Types, Spheres

I. INTRODUCTION

A. *Big Data*

Big data is the realization of greater business intelligence by storing, processing, and analyzing data that was previously ignored due to the limitations of traditional data management technologies. Big data is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools. The challenges include capture, curation, storage, search, sharing, analysis, and visualization. The trend to larger data sets is due to the additional information derivable from analysis of a single large set of related data, as compared to separate smaller sets with the same total amount of data, allowing correlations to be found to "spot business trends, determine quality of research, prevent diseases, link legal citations, combat crime, and determine real-time roadway traffic conditions.

The four dimensions of Big Data

- Volume: Large volumes of data
- Velocity: Quickly moving data
- Variety: structured, unstructured, images, etc.

Veracity: Trust and integrity is a challenge and a must and is important for big data just as for traditional relational DBs

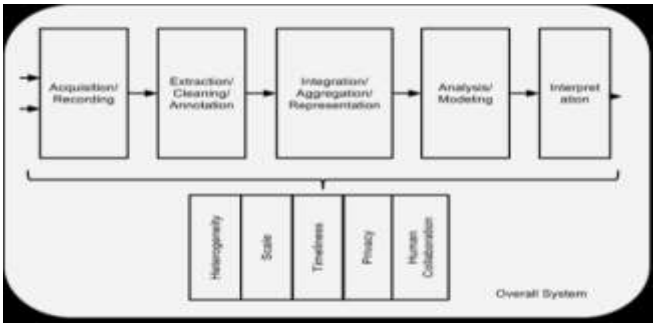

Fig. 1: Big Data

B. *Cloud Computing*

Cloud computing is a set of IT 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. Usually Cloud Computing services are delivered by a third party provider who owns the infrastructure. Cloud Computing holds the potential to eliminate the requirements for setting up of high-cost computing infrastructure for IT-based solutions and services that the industry uses. It promises to provide a flexible IT architecture, accessible through internet from lightweight portable devices. This would allow multi-fold increase in the capacity and capabilities of the existing and new software. Properties of cloud computing are as follows-

1) 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.

2) 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.

3) Reduced Cost:

Cloud computing puts a stop to the illegal reproduction and distribution of software. Some software on the Cloud is free. 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.

II. CLOUD COMPUTING SERVICE DELIVERY MODELS

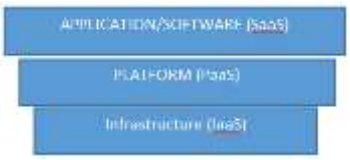

Fig. 2: Cloud Computing Service delivery models

A. *Infrastructure as a Service (IaaS):*

In the infrastructure as a Service (IaaS) model, consumers are given full freedom to manage their data on the server. Here the service provider is only responsible for raw storage, computing power, networks, firewalls, and load balancers and this is often manifested as a virtual machine. A client business pays on a per-use basis whenever the equipment is used to support computing operations such as: storage, hardware, servers, and networking equipment. [2]. Infrastructure as a service is a cloud computing model that

has received most attention from the market, with an expectation of 25% of enterprises planning to adopt a service provider for IaaS [3].

B. Platform as a service (PaaS):

Platform as a Service is a level above Infrastructure as a service (IaaS). In the PaaS model, consumers are provided with an operating system, programming language execution environment, database, and web server. They are not concern with the cost and management in the hardware and software layers. PaaS is the use of cloud computing to provide platforms for the development and use of custom applications [4]. The PaaS solutions include application design and development tools, application testing, versioning, integration, deployment and hosting, state management, and other related development tools [5]. Businesses attain cost savings using PaaS through standardization and high utilization of the cloud-based platform across a number of applications [6]. Other advantages of using PaaS includes: lowering risks by using pretested technologies, promoting shared services, improving software security, and lowering skill requirements needed for new systems development [7].

C. Software as a Service:

Software as a service (SaaS) is the level above Platform as a service (PaaS). In this model, consumers are given access only to the application software, which can be run remotely from the data centers of the cloud service provider. The provider is responsible for the maintenance and support of the infrastructure and operating platforms i.e. it provides businesses with applications that are stored and run on virtual servers in the cloud [8], Since cloud service providers specialize in one area, they can provide reliable service at a fraction of the cost. The business is not charged for hardware, only for the bandwidth for the time and number of users necessary. SaaS applications are accessed using web browsers over the Internet therefore web browser security is vitally important. Information security officers will need to consider various methods of securing SaaS applications. Web Services (WS) security, Extendable Markup Language (XML) encryption, Secure Socket Layer (SSL) and available options which are used in enforcing data protection transmitted over the Internet.

D. Hardware as a service (HaaS):

HaaS offers only the hardware. Time sharing developed into the practice of managed services [11]. In a managed service situation, the managed service provider (MSP) remotely monitors and administers hardware located at a client's site as contracted [12]. A problem with managed services was the necessity for some MSPs to provide hardware on-site for clients, the cost of which needed to be built into the MSP's cost [12]. The HaaS model allows the customer to license the hardware directly from the service provider which alleviates the associated costs [12]. Vendors in the HaaS arena include Google with its Chromebooks for Business, CharTec, and Equus [13].

III. BIG DATA AND THE CLOUD

Big Data is a data analysis methodology enabled by recent advances in technologies and architecture which support high-velocity data capture, storage, and analysis. 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. Cloud computing offers the promise of big data implementation to small and medium sized businesses. Data sources extend beyond the traditional corporate database to include email, mobile device output, sensor-generated data, and social media output Big Data requires huge amounts of storage space. While the price of storage continues to decline, the resource needed to leverage big data still poses financial difficulties for small to medium sized businesses. A typical big data storage and analysis infrastructure will be based on clustered network-attached storage (NAS) [14]. Clustered NAS infrastructure requires configuration of several NAS "pods" with each NAS "pod" comprising of several storage devices connected to an NAS device [14]. The series of NAS devices are then interconnected to allow massive sharing and searching of data [14]. Data storage using cloud computing is a viable option for small to medium sized businesses considering the use of Big Data analytic techniques. Cloud computing is on-demand network access to computing resources which are often provided by an outside entity and require little management effort by the business. A number of architectures and deployment models exists for cloud computing, and these architectures and models can be used with other technologies and design approaches [15]. Owners of small to medium sized businesses who are unable to afford adoption of clustered NAS technology can consider a number of cloud computing models to meet their big data needs. Small to medium sized business owners need to consider the correct cloud computing in order to remain both competitive and profitable.

A. Types of Clouds

The Cloud Computing model has three types of clouds model which are – the public cloud, the private cloud, and the hybrid cloud.

1) 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. 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.

2) 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.

3) Hybrid Cloud:

Hybrid cloud is a combination of both public and private cloud, when a company uses a hybrid 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 [1]. A hybrid cloud allows a company to maintain critical, confidential data and information within it 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.

4) 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.

IV. CURRENT CHALLENGES

List of current challenges with big data in cloud computing are given below-

- Cloud Data Management & Security
- Data Encryption
- Migration of virtual Machines
- Interoperability
- Access Controls
- Energy Management
- Multi-tenancy
- Server Consolidation
- Reliability & Availability of Service
- Common Cloud Standards
- Platform Management
- Service Level Agreements (SLA's)

In this paper our main focus is on Data Security & Management which is one of the most challenging issues. 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.

A. 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.

B. 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.

C. Data Level:

The challenges that can be categorized under data level deals with data integrity and availability such as data protection and distributed data.

1) Generic Types:

The challenges that can be categorized under general level are traditional security tools, and use of different technologies

Cloud computing is a set of IT services that are provided to a customer over a network so it the service provider/organizations should take care about the privacy scenario in terms of "user spheres". In cloud computing have three spheres.

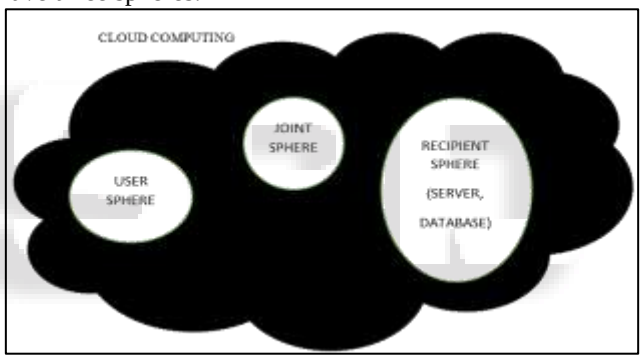

Fig. 3: Cloud computing along with its sphere

a) User sphere

In this data is stored on user's desktop ,pc ,laptops ,mobile phones ,RFID ETC. Organizations responsibility is to provide access to users and monitor that access to ensure misuses does not happen. They should have monitor on what data is transferred from client to data recipient? Is the user explicitly involved in the transfer? Is data storage transient or persistent?

b) Recipient sphere

In this data lies with recipients: server and database of network provider, service provider or other parties with whom data recipient shares data. .Here Organization's responsibility is to minimize user's privacy risk by ensuring that unwanted exposure of personal data of users does not happen. Here alerts like -What data is being shared by data recipient with other parties? Can the user expect / anticipate a transfer of his data by the recipient? Is personal data is secure? Is data storage transient or persistent?

c) Joint sphere

In this data lies with web service provider's server and database .This is in between sphere where it is not clear to whom does the data belong :to the users or data recipient. Here Organization's responsibility is to provide to users some control over access to themselves and to minimize

user's future privacy risk. Here alerts like – Is the user fully aware of how his/her data is used and can control this ?

V. CONCLUSION

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 .If proper and regular monitoring is performed then only we can secure the data in clouds and the big data is smooth enough for normal usage.

REFERENCES

[1] Hao, Chen, and Ying Qiao. "Research of Cloud Computing based on the Hadoop platform.".Chengdu, China: 2011, pp. 181 – 184, 21-23 Oct 2011.[4] Y, Amanatullah, Ipung H.P., Juliandri A, and Lim C. "Towa
[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
[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_EGuide_061212.pdf
[9] Carraro, G., & Chong, F. (2006, October). Software as a service: An enterprise perspective. Retrieved from
[10] ComputureWeekly.com. (2009, March). Hardware as a service. Retrieved from http://www.computerweekly.com/feature/Hardware-as-a-Service
[11] Rouse, M. (2007, December). Hardware as a service. Retrieved from http://searchitchannel .techtarget.com/definition/Hardware-as-a-Service-in-managed-services
[12] 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/
[13] White, C. (2011). Data Communications and Computer Networks: A business user's approach, (6th ed.). Boston: Cengage Learning.
[14] IOS Press. (2011). Guidelines on security and privacy in public cloud computing. Journal of EGovernance, 34 149-151. DOI: 10.3233/GOV-2011-0271

