

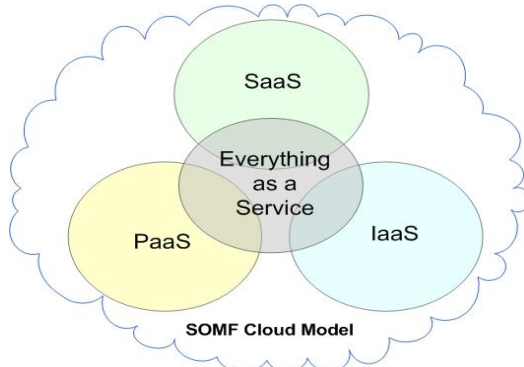
APPLICATION OF BIG DATA ON CLOUD COMPUTING

Arshita Madan, Ishika, Yash Bhardwaj
Manav Rachna International Institute of Research and Studies

Abstract— Cloud computing is set of resources and services offered with the Internet. Cloud services are delivered from data centres throughout the world. With the help of cloud computing user can easily share, store and retrieve their data from anywhere. Cloud computing provides hardware, software and infrastructural storage to many users at a time. However, this technology is still in its initial stages of development, as it suffers from threats and vulnerabilities that prevent the users from trusting it.

In this paper we identify the most vulnerable security threats/attacks in cloud computing. Service model of cloud computing is also explained. To maintain data security and data integrity is also provided. We have also explained the big data.

I. INTRODUCTION



In this age of continuous increment in any sort voluminous information is known as "BIG DATA". Huge information is a procedure empowered by other innovations and engineering which bolster high speed information investigation. Huge information will be data accumulated from the web empowering administrations, online networking and different sources. Essentially "BIG DATA" alludes to the expansion in the volume of information that is tough, puzzling, problematic and entangled to store, to process and dissect through conventional database advancements. Information is not anymore insignificant to organised database, documentation or conventional database advancements.

The advancements of the whole mechanics have made it easy for institutions to put in the findings, as well as while strategizing but in nearly all aspect of the functioning of many institutions, firms. As a result, using this type of technology all resolution reached after consideration of institutions or organisation can be made in structured way.

Big data requires huge amount of storage space. The resource needed to leverage big data can face financial difficulties for small to medium sized businesses.

In this manner information stockpiling sees "Cloud computing" a best alternative for little to medium measured administrations. Cloud computing when all is done is a facilitated administration spread over web. Distributed computing authorises different associations or firms to utilise a figure for example, virtual machine for loads of information, instead of having registering foundation. Accordingly, in distributed computing is a web-based computing, where different administrations, for example, stockpiling of use, substantial measure of information put away through the internet. Cloud registering is a massive call by the system access to figuring assets, given by an outside firm. Organisation models for distributed computing inculcates stage as an administration (PaaS), programming as an administration (IaaS), and equipment as an administration (HaaS).

The 3 types of cloud computing are public cloud, the private cloud, and last one is the hybrid cloud. A public cloud is the pay as you utilise the services. A private cloud is inner data centre of an organisation or business not given to the common public but based on cloud body. The hybrid cloud is a combination of the public cloud computing and the private cloud computing.

Three noteworthy explanations behind little to medium measured organisations to utilise distributed computing for enormous information method usage are equipment cost cut, preparing cost decrease, and the ability to test the estimation of huge information. The significant stress over distributed computing are security and loss of control. Cloud figuring and huge information go as an inseparable unit or we can state they are co-joined. Huge information gives clients the capacity to utilise ware figuring to process circulated inquiries over various datasets and return resultant sets in an opportune way. Huge information sources from the cloud and web are put away in an appropriated composed database and prepared through a

programming model for extensive datasets with a parallel conveyed calculation in a bunch.

Big data make effective use of distributed storage technology based on cloud computing rather than local storage attached to a computer or electronic device.

II. LITERATURE SURVEY

Big data is the big amount of data that is difficult to store, process and analyse. Big data is a new ongoing trend in the IT industries for new demands of the technologies.[4] Big data is classified in the following 4v's: -

1. Volume
2. Variety
3. Velocity
4. Value

Volume- volume defines the extraction of large amount of data that is further expandable. It includes formation of hidden information, patterns from the data through data analysis. The example of volume is the collection of the longitudinal data from smart devices.[4]

Variety- variety defines the different types of data that are being assemble from different sources like social networking, applications, sensors etc. These data types usually include texts, videos, photographs, GIFS etc. The data extracted can be in structured or unstructured forms. Mostly the mobile data assemble is in unstructured format.[4]

Velocity- velocity defines the speed at which the data is transferring. The content of the data can change as every data is being retrieved from a different source.[4]

Value- value is the most important and valuable feature of the characteristics of data. Value is the method of assembling and searching the hidden values from massive datasets.[4]

The big data is important as the presence of huge data storing in cloud. Big data is widespread among IT companies to help them save and analyse the large amount of data easily. Cloud computing is the delivery of different services through the internet. These resources include tools and applications like data storage, servers, databases, networking and software. Rather than keeping files on a proprietary hard drive or local storage device, cloud-based storage makes it possible to save them to a remote database. If an electronic device has access to the web, it has access to the data and the software programs to run it. Cloud computing is a popular option for people and businesses for several reasons including cost savings, increased productivity, speed and efficiency, performance, and security. Thus, need of storing, processing and analysing large amount of data sets has been taken by many organisations and businesses to adopt cloud computing technology. And due to its

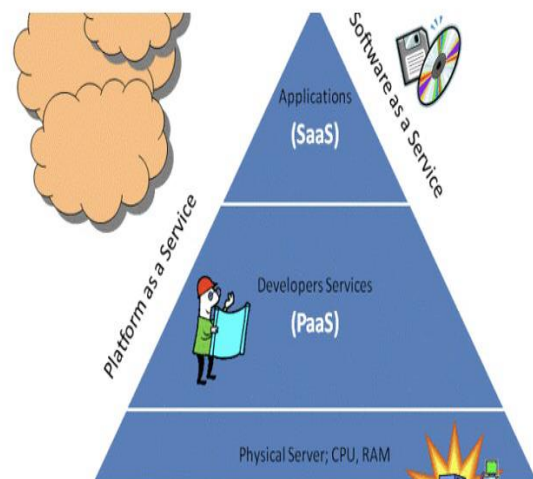
flexibility, of storing data in large volume, it is wireless networking, security, ease of sharing the data, etc increases its popularity and taking "CLOUD COMPUTING" to new heights.

III. BIG DATA AND THE CLOUD

Big data and cloud mean that data are so big and huge that database systems cannot analyze data sets. Data sets are not traditional body structured data, rather are the data from new sources including emails and sensors that can be accessed over the Internet. Big data can serve challenges for archiving and analytics.[4] Model for storing huge amounts of data internally is network storage with clustered storage. It begins with a storage that is connected to the network, consists of many computers connected to one single computer which is used as device. Clustered network attached storage is expensive for small and medium sized companies. For a lower cost service cloud provider can provide it at cheaper rates. Big data processes analytics using a programming pattern called MapReduce. In MapReduce request is made and data is mapped so that we can find key value associated with the query and then result is reduced to a set of data which will respond to the query.[4] Huge amounts of data are needed by the MapReduce pattern. Simultaneously mapping is done by each individual network attached to the storage device and mapping requires a parallel processing. Parallel processing of MapReduce is demanding and requires specific configuration.

IV. CLOUD COMPUTING SERVICE MODELS

Models for cloud computing include PAAS- platform as a service, SAAS-software as a service, HAAS- hardware as a service, IAAS-infrastructure as a service.[4]



Business could not afford services that can be provided by cloud formation solutions. Business can also use these development solutions as a test measure before adopting a new technology company wide.



4.1 PAAS-Platform as an administration

Service is the application that occurs throughout an area in computing to give stages for the process of advancement and utilisation of custom applications. [4] The PAAS process incorporates application plan, application testing, organising, facilitating and other types of related advancement gadgetry. organisations achieve cost reserve funds by utilising PAAS through establishing usage of cloud-based stage over sorts of applications. [4] Other preferences of utilising PAAS contain bringing down the dangers by utilising pretested in a purposeful way improve the quality of programming security introducing down ability as a prior condition for new concept and framework improvement. PAAS gives Corporation org stage for creating custom usage expected to discover expansive amount about the unstructured information demand for an extraordinarily little effort and in a protected domain.

4.2 SAAS-Software as an administration

Programming as an administration provides the applications that are moved away and kept running on virtual servers in the cloud. Business is charged for necessary items for the purpose of transmitting capacity for the time and number of client requirements.[4] The principle is to choose standpoint of SAAS so that the arrangement enables organisation to move the dangers related to programming security while at the same time moving IT from being responsive to proactive. Advantages of utilising it are less demanding programming organisations, programming similarity over business, worldwide openness, less demanding joint effort.[4] Distinction among SAAS and PAAS is that SAAS would not give a tweak the arrangement though PAAS will enable organisation to build up an answer customised to organisations need.

4.3 IAAS-Infrastructure as an administration

In the IAAS, a customer business will pay on for every utilisation reason for utilisation of gear to bolster registering tasks including capacity, equipment, servers and systems administration gear.[4] Infrastructure as an administration hey is the distributed computed model accepting most considerations from market which desire 25% and were wanting to receive an administration supplier for IAAS.[4] Administrations accessible to organisations show incorporable debacle recuperation, stop filing as an administration, virtual work area framework which is giving pinnacle stack ability to variable forms. Advantages include expanded budgetary adaptability, business readiness, expanded security.

V. WHICH CLOUD FOR YOUR DATA?

The sort of cloud an organisation utility relies upon the organisation's needs and assets. General society cloud is analysed as the base anchored of the 3 types, with administrations and assets, ready to be gained over the internet by the regular system taken up by the supplier.[1] The

interchanges conventions received by the supplier are not required to be secure, the decision of utilising secure or non-secure conventions is up to the given organisation. The general population cloud is likewise the base exorbitant of the cloud types, with the cost reserve funds in the piece of data innovation association, administration, and support. The private cloud gives office to the administration or organisation representatives by an intranet. [1] On the off chance that versatile representatives can get to the private cloud, the entrance is commonly through secure correspondence conventions.

All administrations what is more, assets give is modified by the necessities of the business, and the business has the full authority over the office, assets and the administrations. Due to the monetary and HR expected to convey, oversee, and keep up the data innovation assets and administrations, the private cloud is the costliest sort of cloud, at the point when a business utilises a mixture cloud the business claims its centre data innovation assets and benefits and will give the assets and administrations in-house.[1] Non-basic administrations are redistributed and are kept on an open cloud. Normally, centre data innovation assets and administrations are mission-basic and are regularly classified. In this manner, assets and administration that should be secure are facilitated and kept up on the private cloud, with people in general cloud utilised for different administrations as a cost sparing measure.

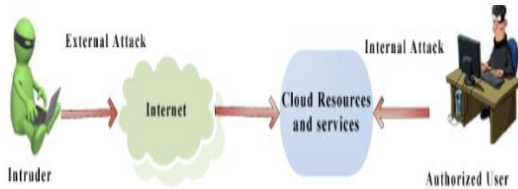
VI. DATA INTEGRITY IN CLOUD COMPUTING

Data security means that digital information is uncorrupted and is easily accessed or modified by only those who are authorised to do so. It involves 2 main steps which include maintaining the consistency accuracy and trustworthiness of the data over its entire life. Data must not be changed in the transit and various measures should be taken to ensure that the data is not accessed or changed by an unauthorised person.[2] Access control and version control are implemented for such reasons. Other methods include checksums and cryptographic checksum to verify integrity. When the data becomes corrupted backups must be available to restore the data do it correct state. Also, measures can be taken by controlling the physical environment of network terminals and servers because data accuracy and trustworthiness can be threatened by the environmental hazards such as heat, dust or electrical problems. Some devices must be in place to detect any changes in the data which are happening as a result of non-human activities or events such as electromagnetic pulse or server crash. Practices followed to prevent and protect the data integrity in physical environment includes keeping transmission media covered so that they cannot be tapped, protecting hardware and storage media from power surges, electrostatic discharges and magnetism. Data owner can perform verification at any time he or she wants. Hash value of data present at the cloud and present at the owner end is matched at the last.[2] When the data is matched if it matches the previous data then the data is secured and if it does

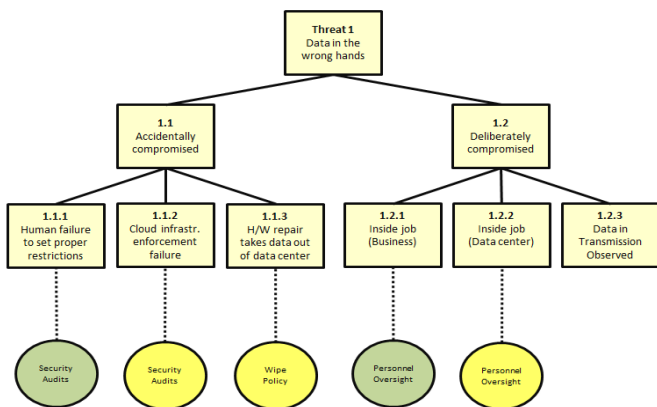
not match the previous data then the data has been modified and a change is there in the previous data. In this manner the owner can easily take a hold over his data and can verify it any time as and when wanted. In this way the data is kept secured and its integrity is maintained.

VII. THREAT MODEL FOR CLOUD

An abstract view of threat model for Cloud computing is shown in Figure. There are two types of security threats facing by cloud clients: External and Internal attacks.[2]



External network attacks in the cloud are increasing. Evil-minded user outside the Cloud performs DoS or DDoS attacks to affect the availability of Cloud services and resources. Port scanning, IP spoofing, DNS poisoning, phishing is also executed to gain access of Cloud resources. An Evil-minded user can capture and analyse the data in the packets sent over this network. IP spoofing occurs when evil-minded user impersonates a legitimate users IP address where they could access information that they would not have been able to access otherwise.[2] Availability is most important. Not having access to services can be a disaster for anyone especially in the case of being denied service. This can take place when exhaustion of the host servers causes requests from legitimate consumers to be denied.[2] This can cost a company big amount of money and time if the services they depend on to operate are not available. Internal attacker can easily get access to other user’s resources without getting revealed. An insider has higher privileges and knowledge than the external attacker. Therefore, it is easy for an insider to attack than external attackers.



VIII. VULNERABILITIES TO CLOUD COMPUTING

In Cloud, existing vulnerabilities, threats, and associated attacks boost several security concerns. Vulnerabilities in Cloud defined as the loopholes in the security architecture of Cloud, which can be used by an adversary via sophisticated techniques to access the network and other infrastructure resources. Now we discuss major Cloud specific vulnerabilities, which creates serious threats to Cloud computing.[2]

8.1 Session Riding and Hijacking

Session hijacking defines the use of a valid session key to take advantage of unauthorised access for the information or services located on a computer system.[2]It also defines theft of a cookie used to authenticate a user to a remote server and it is applicable to web application technologies weaknesses in the web application structure at their disposal that gives the chance to hackers in order to achieve a wide variety of Evil-minded activities.[2] While session riding defines the hackers sending commands to a web application on the place of the targeted user by just sending that user an email or tricking the user into visiting a specially crafted website. Session riding deletes user data, executes online transactions, sends spam to an intranet system via internet and changes system as well as network configurations.

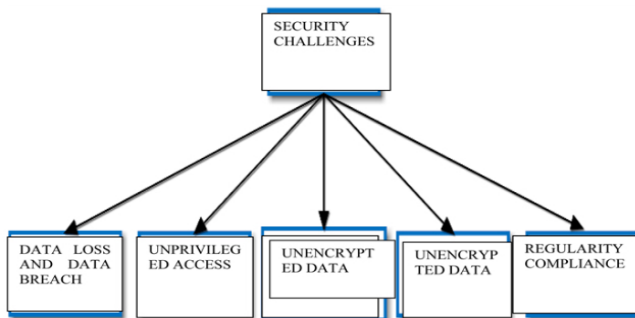
8.2 Insecure Cryptography

Attackers’ can solve any cryptographic mechanism or algorithm by hacking them. It’s common to find major flaws in cryptographic algorithm implementations, which can turn strong encryption into weak encryption or sometimes no encryption at all.[2] For example, in cloud virtualisation providers uses virtualisation software to partition servers into images that are provided to the users as on-demand services. Although utilisation of those VMs into cloud providers' data centres provides more flexible and efficient setup than traditional servers but they don't have enough access to generate random numbers needed to properly encrypt data. This is one of the fundamental problems of cryptography.

IX. SECURITY

9.1 Secure data storage and Secure data backup

Secure data storage collectively refers to the manual and automated computing processes and technology used to ensure stored data security and integrity.[3] It can physical protection of the hardware in which the data is stored, as well as security software. Secure data storage applies to the data which is stored in computer hard disks, and portable devices like – USB drives, online/cloud, network-based storage area network (SAN).



Secure data storage is achieved in the following ways-

- Physical storage device and infrastructure security.
- Data encryption
- Enforcement and implementation of layered storage security architecture.
- Protection against viruses and other data corruption threats.

This is essential for the organizations which deal with sensitive data, both in order to avoid data theft, and as well as to ensure uninterrupted operations.

Data backup refers to saving additional copies of data in separate physical locations from data files in storage.[3] Data is uploaded on cloud as it is necessary to keep this data secure for future usage so that the data will not be damaged or lost.

9.2 Secure data sharing-

The data sharing process in which data owners share their data with other users. For sharing user access list should be generated by the owner so the list is previously generated which contains the name of the authorised users, for secure communication private key is sent through email to the users so that the permit is hidden.[3] The access permission is in two forms read only and read and write

- Read only: the user with this permission can only read the data and they cannot make any modifications to it.
- read and write: the users with this permission can only read and write or any other on the file.

X. CONCLUSION

Big data merged all the data that is difficult to store, process and work on. Cloud computing has been accepted by many companies to handle the huge amount of big data easily. The velocity and the demand of the data is also increasing in the new trend markets. The mentioning of cloud computing with the term big data has played remarkable in storing and analysing the data. Cloud modify the upcoming IT industries techniques of handling the data. It helps in handling big amount of data in a quick and easy go. In our research paper, the key features of cloud computing targeting big data has been shown. The

classification of big data that is volume, value, velocity and variety has also been explained in the paper. The profit of cloud on big data and how the industries, academics are being enhanced by using new virtual technologies like cloud are also shown in this paper. We have also discussed the characteristics of a cloud security that contains threats/attacks and vulnerabilities. Explanation of data integrity and Security is also there in our research paper.

XI. REFERENCES

- [1] https://www.researchgate.net/publication/256888844_Big_data_using_cloud_computing
- [2] https://www.researchgate.net/publication/276196135_Secure_Cloud_Architecture
- [3] https://www.researchgate.net/publication/304289783_Data_security_and_integrity_in_cloud_computing_based_on_RSA_partial_homomorphic_and_MD5_cryptography
- [4] https://www.researchgate.net/publication/319913597_Cloud_Computing_and_Big_Data_is_there_a_Relation_between_the_Two_A_Study
- [5] <https://www.ibm.com/big-data/us/en/> Bello-Orcas G, Jung JJ, Camacho D. Social big data: Recent achievements and new challenges. Information Fusion. 2016 Mar 31; 28:45-59.
- [6] <https://www.sciencedirect.com/science/article/abs/pii/S0306437914001288>
- [7] https://www.researchgate.net/publication/280026494_A_Survey_of_Big_Data_Cloud_Computing_Security
- [8] https://acme.able.cs.cmu.edu/pubs/uploads/pdf/IoTBD_2016_10.pdf
- [9] https://www.researchgate.net/publication/326073288_A_Review_Paper_on_Cloud_Computing
- [10] http://apps.aima.in/ejournal_new/articlesPDF/Uday-Salunkhe.pdf