Data Mining and Cloud Computing Security

Dr.Padmavalli Mesa[1], Prof. K.Sreenivasa Rao[2]

[1] Department of Computer Science, S.K.University, Anantpur, India

[2] Rayalaseema University, Dept of OR&SQC, Kurnool, India

ABSTRACT

Data mining has been considered as an essential component in business domain. The objective is to gain the understanding of the project objectives and business requirements, and then converting this knowledge into a data mining problem definition and a preliminary plan to achieve the objectives. But the current needs of IT make the Cloud computing comes into existence. Cloud computing provides means to improve or add abilities on- demand without making an investment in setting up infrastructures, training new employees. Reasoning processing involves any subscription-based or pay-per-use support that, immediately over the Online, expands ITs current abilities. Recently cloud computing has been facing lots of security issues regarding privacy of data. To protect it from unethical managers and attackers, researches are being conducted on application of a variety of cryptography systems such as searchable encryption and proxy re-encryption to Cloud storage system. However, existing searchable encryption technology is inconvenient in the cloud storage environment in which the user uploads data in person, and those data are shared with others, whenever it is necessary to do, and those with whom data are shared change frequently.In this paper we discuss several technologies to make the data stored in cloud safe. These technologies can be divided into two parts: Storage protect and Access protect.

Keywords: Data mining, Cloud computing, privacy, databases and security, Encyption

1. INTRODUCTION

Data mining prospered in 1990's but has a long history behind its evolution. History of data mining can be traced in classical statistics, artificial intelligence and machine learning. Statistics includes standard distribution, standard derivation, standard variance, cluster analysis, discriminate analysis etc. All these play role in analyzing data and data relationships. Artificial intelligence applies to processing human thought to statistical problems. Certain commercial products like RDMS have used query optimization modules which is a concept of Artificial Intelligence. Another concept of machine learning allows program to study data and then make decisions on the basis of data studied. Programmers uses statistics for fundamental concepts adding advanced AI heuristics, algorithms for the above given purpose. Thus data mining is nothing but application of machine learning techniques to business applications. Techniques like AI, statistics, machine learning are used to find previously hidden trends or practices or data. It is nothing but data mining. Association Rule which finds association between data and various objects by finding dependence amongst data is often used method in data mining. One thing has to be discussed here that business must hide complexity of data mining from end-user to make a successful application.

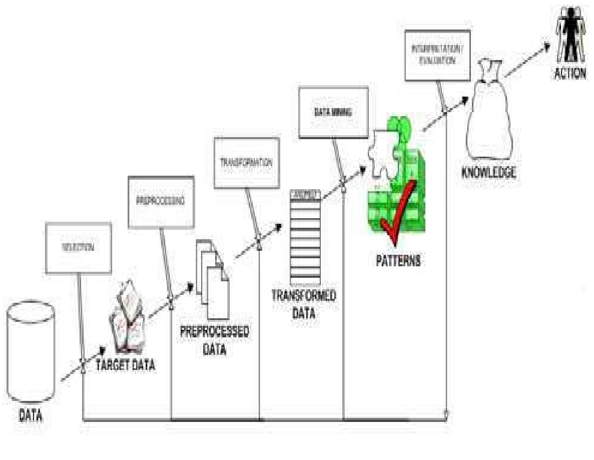


Figure 3.2: Steps in Data Mining Process

Data mining is a break through invention also known as knowledge Discovery is database.It is used to search significant patterns of data from large volumes of data. Noted areas in Data mining are frequent pattern mining, Association Rule mining etc. Cloud in cloud computing stands for network or Internet or in other words which is present at remote location. It provides services available on WAN, LAN or VPN. For e.g. web conferencing, email etc.

2. WHAT IS CLOUD COMPUTING

The Internet is turning into an increasingly vital tool in our everyday life as its users are becoming more numerous. It is not amazing that business is progressively led over the Internet. Maybe a standout amongst the most revolutionary concepts of recent years is Cloud Computing. The cloud computing is a general term for anything that includes conveying facilitated administrations over the Internet. The distributed computing administrations are considered Infrastructure-as-a-Service (Iaas), Platform-as-a-Service (Paas) and Software-as-a-Service (Saas).

Various associations are picking as an elective to building their specific IT base to have databases or programming so the association may have entry to its data and customizing over the Internet. The use of Cloud Computing is getting conspicuousness due to its transportability, enormous approachability and insignificant expense. It joins matrix registering, utility processing, virtualization, and bunching and so on. Distributed computing blankets a share of the thoughts of conveyed, framework and utility registering. Cloud is by and large a virtualization of assets that administers and oversees itself. The cloud computing is fundamentally entering the assets and administrations required to perform capacities with alterably evolving requirements. The administration engineer appeals access from the cloud instead of a particular endpoint or named asset.

Interrelation between data mining and cloud computing have their own pros and cons: Data mining can be used as a tool to provide better services by a cloud provider. But on the other hand outside attackers can use data mining techniques to unauthorizely access private data by interacting it. Interaction of data can involve two factors; one is suitable amount of data and other is appropriate mining algorithms. There are number of mining algorithms which are helpful to interact to private data and hence threat to data privacy. For example association rule mining algorithms can be used to locate relationship between huge numbers of business transaction records..

A. Pros

1) Reduced Cost: In cloud technology, we pay for just what we utilize, which prompts sparing associations cash in the short run. Spared sum should be utilized for other paramount assets.
2) Increased Storage: Associations can store more information on cloud than on private machine frameworks.
3) Highly Automated: Cloud engineering is profoundly computerized as IT faculty are not required to stay up with the latest.
4) More Mobility: Workers can access data wherever they are as opposed to being stick to work area.

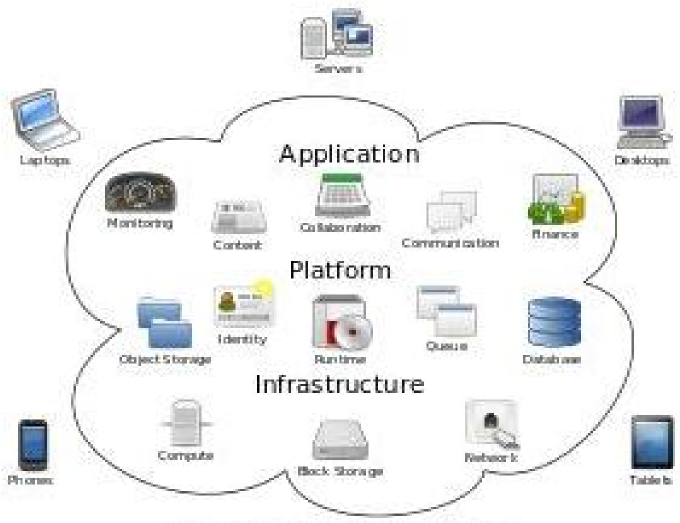


Fig 2: Cloud Computing Logical Diagram

3. CLOUD SERVICES

There are three types of cloud services in which they are as follow:-

- Infrastructure as a Service
- Platform as a Service
- Software as a Service.

A. *IaaS*

Convey PC foundation as an utility administration, regularly in a nature. I t is also known as utility computing. Provide enormous scalability.

B. PaaS

Approach to lease fittings, working frameworks, space and system limit over the web to create provisions Sits on a top of the Iaas construction modelling and joins with improvement and middleware proficiencies and database, informing and queuing capacities.

C. SaaS

This is the place clients basically make utilization of a client interface to gain entrance to programming that others have created and offered as an administration over the web. It is built on underlying IaaS and PaaS Layer.

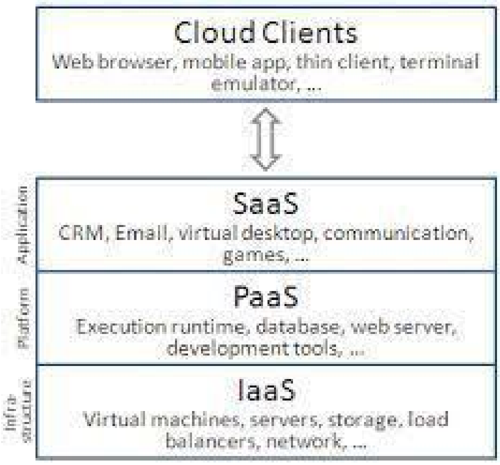

Cloud Clients
Web browser, mobile app, thin client, terminal emulator, ...
Application
SaaS
CRM, Email, virtual desktop, communication, games, ...
Platform
PaaS
Execution runtime, database, web server, development tools, ...
Infra-structure
IaaS
Virtual machines, servers, storage, load balancers, network, ...

Fig 3: Layers of Cloud Computing

Cloud Computing architecture comprises of many cloud components, which are loosely coupled. We can broadly divide the cloud architecture into two parts:

- Front End
- Back End

Each of the ends is connected through a network, usually Internet. The following diagram shows the graphical view of cloud computing architecture:

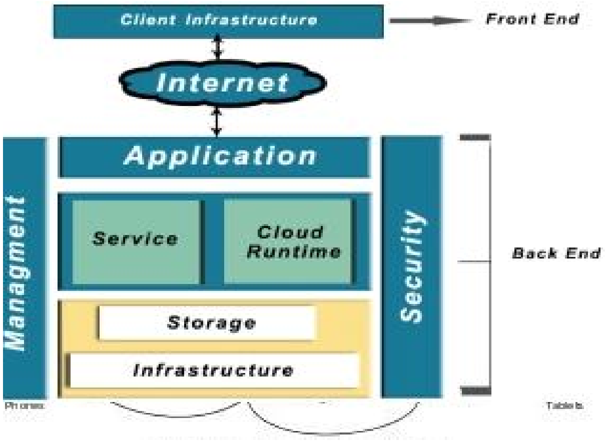


Fig 4: Cloud Computing Architecture

4. CLOUD COMPUTING DEPLOYMENT MODELS

Cloud computing architects give following basic service models:

• Public cloud
• Private cloud
• Hybrid cloud
• Community Cloud

Conveying distributed computing can vary relying upon necessities, and the accompanying four arrangement models have been recognized, each with particular attributes that help the requirements of the administrations and clients of the mists specifically ways. There exist four separate sorts of mists on the groundwork of who claims and utilization them:

1) **Public Clouds:** A public cloud encompasses the traditional concept of cloud computing, having the opportunity to use computing resources from anywhere in the world. Public clouds are frequently hosted away from customer site, and they provide flexible infrastructure to cut down customer risk and cost.
2) **Private Clouds:** Private clouds are assembled for utilization of one customer solely, furnishing the most extreme control over information, security, and nature of administration. Here, the organization claims the foundation and has control over how requisitions are circulated on it. Private mists could be manufactured and deliver the goods by an organization's IT association or by a cloud supplier. In this model, an organization can introduce, arrange, and work the base to help a private cloud inside an organization's undertaking data centre.
3) **Hybrid Clouds:** Hybrid clouds join the aspects of both open and private cloud models. They can help to give on-interest, remotely provisioned scale. The capacity to incorporate a private cloud with the assets of an open cloud might be utilized to administer administration levels. A half breed cloud likewise could be utilized to handle arranged workload spikes. Half and half mists present the intricacy of figuring out how to disperse provisions crosswise over both an open and private cloud. The cloud foundation comprises of various billows of any sort, yet the mists have the capability through their interfaces to permit information or provisions to be moved starting with one cloud then onto the next.
4) **Community Clouds:** In Community Cloud the cloud base is imparted by numerous associations that have imparted contemplations. It is ought to be overseen by the associations or a third gathering and might as well exist on-premises or off-premises.

5. CLOUD COMPUTING SECURITY

Cloud makes it possible to store and access your data from anywhere anytime without worrying about maintenance of hardware software and storage space. All these services are provided to user at low cost. User has to pay according to storage space he is using. Due to this flexibility everyone is transferring his data on cloud. Security becomes big issue when any one stores its important information to a platform which is not directly controlled by the user and which is far away. While sending of data and during storage data is under threat because any unauthorised user can access it, modify it, so there is need to secure data. A data is secure, if it fulfils three conditions (i) Confidentiality (ii) Integrity (iii) Availability. Confidentiality means the data is understandable to the receiver only for all others it would be waste; it helps in preventing the unauthorised disclosure of sensitive information. Integrity means data received by receiver should be in the same form, the sender sends it; integrity helps in preventing modification from unauthorised user. Availability refers to assurance that user has access to information anytime and to any network. In the cloud confidentiality is obtained by cryptography.

Since data stored in cloud can be accessed from anywhere, we must have a mechanism to isolate data and protect it from client's direct access.Below are the key mechanisms for protecting data.

- Access Control
- Auditing
- Authentication
- Authorization

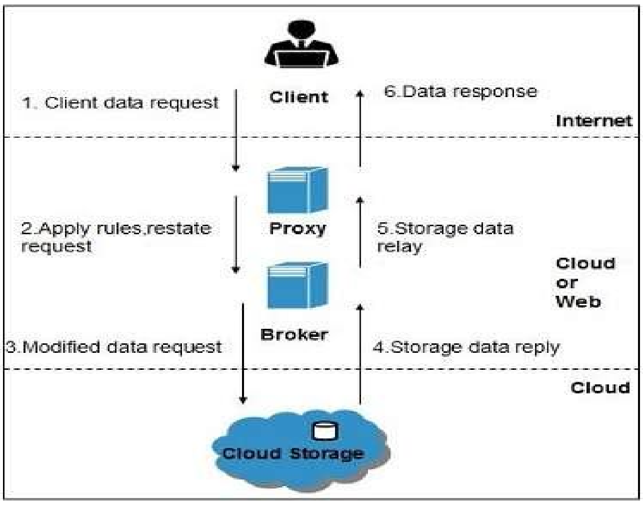
All of the service models should incorporate security mechanism operating in all above-mentioned areas.

Data in cloud can be effectively secured by encrypting it. Direct access of client can be restricted by using proxy and brokerage services.

1. **Brokered Cloud Storage Access** is an approach for isolating storage in the cloud. In this approach, two services are created:
   - A broker with full access to storage but no access to client.
   - A proxy with no access to storage but access to both client and broker.

Working of Brokered Cloud Storage Access System : When the client issues request to access data:

- The client data request goes to the external service interface of proxy.
- The proxy forwards the request to the broker.
- The broker requests the data from cloud storage system.
- The cloud storage system returns the data to the broker.
- The broker returns the data to proxy.
- Finally the proxy sends the data to the client.



2. Encryption:

Encryption helps to protect data from being compromised. It protects data that is being transferred as well as data stored in the cloud. Although encryption helps to protect data from any unauthorized access, it does not prevent data loss..Cloud security is built around encryption methodologies. These are of three kinds: hashing encryption, symmetric cryptography, and asymmetric cryptography. Each method has several advantages and disadvantages and is used by Cloud service providers to ensure that user data is not tampered with or compromised in any fashion.

2.1. Hashing: This method uses a unique, fixed length signature to encrypt a data set. The hash is created using a hash function or an algorithm and each hash is compared with other hash sets to verify uniqueness of the data set. Since a small change in the data will result in the generation of a new hash, the data owner will be alerted to any security breaches that may have occurred.

Unlike other kinds of encryptions, the hash encryption is irreversible. This means that there are no decryption or de-hashing keys that can be used to reverse the process of hashing. This makes hashing secure. Hackers who have accessed the data will not be able to discover the contents of the data set even if the hash system is understood. A few hashing algorithms that are commonly in use are detailed in Message Digest 5 (MD5).

2.2. Symmetric Encryption: A *Symmetric* encryption uses the same key for both encryption and decryption. The key is known as the “private key” and must be kept secure by the user if the data set is to remain secure. This key may be ‘user defined’ or ‘system generated’. The encryption operation maybe performed on a “stream of data” (encryption of one byte at a time) or a “block of data” (encryption of one block at a time). Commonly used symmetric algorithms are DES, AES, and Blowfish.

2.3. Asymmetric encryption: An *Asymmetric* encryption uses two different keys for encryption and decryption of a data set. The encrypting key is known as the “public key” and the decrypting key is known as the “private key”. The public key is freely available and the private key is available only to the person authorized to decrypt the message. This use of two keys is said to be the weakness of the system. Examples of asymmetric algorithms are RSA and Diffle-Hellman.
Homomorphic algorithms are making their appearances with the growing popularity of the Cloud. A homomorphic algorithm is an encryption algorithm that allows the user to perform mathematical operations on the data set without decrypting the data.

6. CONCLUSION

Research on classification of data in cloud has already been extensively done; so now it is important to use the result of these researches and analyse the security requirements which are important for keeping data secure. Relying on cloud computing millions of users store their data on a cloud which possess lot many cloud storage risks like unauthorized access, data loss etc. Privacy of data is a major concern in people who use public cloud services, so an approach is proposed to keep data safe and secure also keeping sure only authorized personnel can access data. It is proposed to implement cloud security aspects for data mining by implementing cloud system. After implementing cloud infrastructure for data mining for cloud system, security measure for data mining in cloud will be evaluated. Threats will be fixed in data mining to Personal/private data in cloud systems. . Cloud computing has its advantages in a sense that end user need not invest on infrastructure. But it has to be kept in mind that it is fragile to data mining techniques used by attackers who gain unauthorized access, risking data privacy. Hence there is need to evaluate security measures and protect client.

REFERENCES

[1] Parikshit Prasad, Badrinath Ojha, Rajeev Ranjan shahi, Ratan Lal, 3 Dimensional Security in Cloud Computing, IEEE, 2011
[2] Uma Somani, Kanika Lakhani, Manish Mundra, Implementing Digital Signature with RSA Encryption to Enhance Data Security of Cloud in Cloud Computing, IEEE, 2010
[3] Fadi Aloul, Syed Zahidi, Wassim El-Hajj, Two Factor Authentication Using Mobile Phones, 2008.
[4] Balachandran Reddy, Cloud computing security issues and challenges, 2009
[5] Special Publications 800-145 “National Institute of Standard and Technology (NIST)”
[6] http://en.wikipedia.org/wiki/Cloud_computing
[7] http://cloudcomputing.sys-con.com/node/1744132
[8] Zhidong Shen, Qiang Tong, “The Security of Cloud Computing System enabled by Trusted Computing Technology”, 2010
[9] http://www.cloudcomputingchina.cn/Article/ /200909/306.html
[10] Pekka Riikonen, “RSA Algorithm”, 2002
[11] Torry Harris, “Cloud Computing an Overview”