

A Survey on Data Masking In Cloud

Santhosh Pawar

Dept of CS
KSWU, Vijayapura

Abstract- *Cloud computing is a technology, that provides cost effective , scalable computational resources and cloud service to perform the digital business activities, whenever we need we can increase or decrease the resource quantity allocated based on our requirement and use. Cloud provides reliable on demand service whenever needed and at low cost for development and research. when we access the cloud services for our needful work the security issues will arise during accessing of cloud through internet , we can store data in cloud which will have some security issues , cloud computing security is one of the important aspect in cloud , here we try to explore some of the security techniques by data masking.*

Keywords- Cloud resource, Cloud Computing Service Models, , Data Masking , Deployment Models, Data Security.

I. INTRODUCTION

Cloud Computing is a technology which resembles the working with client server technology , cloud computing uses the internet and central remote servers to provide the cloud services to its data and applications. We can access the applications installed in cloud server directly to our client machine.[1]It also offers online data storage, infrastructure and application. It is architecture for providing computing services via internet on demand and pay per use access to a pool of shared resources for the network storage, services and applications. It is totally an internet based technology in which client data is stored and maintained in data center of cloud provider like Google, Amazon, and Salesforce.com etc. The resources in cloud system are transparent for the application and the client do not know the place of resource. The client can access your application from anywhere. The amount of resources provided in the cloud system for the cloud system for the client is increased when their requirements are high and decreases when their requirements are less. The cloud computing can be seen as the important change of information industry and will make more impact on the development of information technology for the society.

Cloud Computing Advantages:

1. Low Cost

Cloud technology is paid incrementally, saving the money of

organizations.

2. More storage

Organization can store more data than on private computing system

3. Automated

No longer do IT personnel need to worry about keeping software up to date.

4. Flexible

Cloud computing offers much more flexibility than past computing methods.

II. SERVICE MODELS OF CLOUD COMPUTING

There are three types of services provided by cloud computing and these are called cloud computing models:

- Software as a Service (SaaS)
- Platform as a Service (PaaS)
- Infrastructure as a service (IaaS)

i) Software as a Service (SaaS): It is the top layer provider in which customer with ready to use applications running on the infrastructure provider. SaaS can be explained as a process by which Application Service Provider (ASP) provide different software application over the internet. SaaS applications are pattern for end users, deliver over the internet. It enables the customer to eliminate of installing and operating the application on his own computer and also get rid of the immense load of software maintenance. With SaaS a provider licenses an application to the customer as a service on demand through subscription.[2] Generally the customer is only able to modify parameters of the application that have been exposed by the provider. The customer should have knowledge protecting against administrative access by the provider. The customer should understand the data encryption methods which are applied to data. The customer needs to be aware of how secure data, as defined in their data classification, is to be handled in general and by configuration options[3]. Salesforce, Zoho, workday are instances of SaaS which are used for email, billing etc.

SaaS applications run on a SaaS provider's servers. The provider manages access to the application, including security, availability, and performance.[3] SaaS customers have no need of hardware or software to buy, install, maintain, or update. Its access to applications is easy. User just needs an Internet connection. This cloud computing delivers a single application through the browser to thousands of customers using architecture. On the customer side it means no investment in servers or software licensing and on the provider side just one app to maintain, costs are low as compared to conventional hosting. Office software is the best example of businesses in SaaS.

SaaS Benefits:

SaaS helps to managed software from a central location. The user can sign up and quickly start using ingenious business apps. Software delivered in a 'one to many' model. There is no setup costs with SaaS, as these are available with other applications.

ii) Platform as a Service(PaaS): It is a middle layer which provide platform oriented service. In this customer has responsibility for application deployment and to provide securing access to the application itself. PaaS is especially useful for situation where multiple developers working on a development project.. Here the consumer does not control the underlying cloud infrastructure including network, servers, operating systems, or storage, but it control over the deployed applications and possibly configuration settings for the application-hosting environment. Google App Engine, Load Storm are the instances of PaaS for executing web applications over internet.[4]PaaS is a combination of a development platform and a solution stack, delivered as a service on demand. It provides framework on which software developers can build new applications or extend existing ones without the cost and complexity of buying and managing the hardware and software. The consumer uses a hosting environment for their applications. Most cloud offerings, PaaS services are generally paid for on the basis of agreement with clients .

PaaS Benefits:

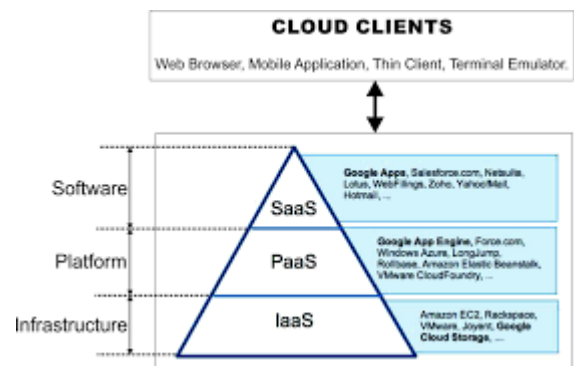
we can share the common platform or can have integrated access from different application developer.

Integration with web services and databases via common standards Reduces the complexity of integration with middleware as a service. Teams in various locations can work together Makes development possible for _non-experts.

iii) Infrastructure as a Service (IaaS): IaaS can be utilized by enterprise customers to create cost effective and easily scalable IT solutions where the complexities and expenses of managing the underlying hardware are outsourced to the cloud provider. The user can buy the infrastructure according to the requirements instead of buying the infrastructure that might not be used for months. IaaS operates on a Pay as you go|| model .For a startup or small business; one of the most difficult things to do is keep capital expenditures under control. In cloud you have the ability to scale as if you owned your own hardware and data center that the users pay for only what they are using. Virtualization enables IaaS providers to offer almost unlimited occurrence of servers to customers and make cost-effective use of the hosting hardware

IaaS Benefits :

1. Infrastructure scales on demand to support dynamic workloads. Generally include multiple users on a single piece of hardware. Flexible and innovative services are available on demand.No need to invest in your own hardware.Physical security of data centre locations.



Deployment Models:

Deployment models define the type of accesses to the cloud i.e. how the cloud is located? Cloud can have any four type of access: Public, private, Hybrid and community.

I. Public cloud: Public cloud which is based on standard cloud computing, services may be free or offered on a pay-per-use model.[4] The public cloud allows system and services to be easily accessible to general public. Public cloud may be less secure because it is open to everybody. Public clouds offers service, usually over an internet connection. A public cloud is lying on the internet and designed to be used by any user with an internet connection to provide a similar range of capabilities and services. Public cloud users are mostly residential clients and connect to the public through an internet service provider's network. Google, Amazon and Microsoft

are examples of public cloud who offer their services to the general public.

Public cloud providers manage the infrastructure and resources required by its users. Organization can utilize public clouds to make their operations significantly more efficient, for example, with the storage of non-sensitive content, online document collaboration and webmail. While one of the biggest obstacles facing public cloud computing is security, the cloud computing prototype provides opportunities for foundation in provisioning security services that hold the prospect of improving the overall security of some organizations. Organizations should require that any selected public cloud computing solution should be configured, deployed, and managed to meet their security and other requirements.

Benefits of the public cloud:

1. Public cloud provides ultimate scalability.
2. Cloud services like IaaS, PaaS, SaaS follow the public cloud, so it is more flexible.
3. It is Location independence means its services are available through internet whenever the client is required.
4. Public cloud is also cost effective because it brings together resources which are shared by all subsequent cloud services.

II. Private cloud: A private cloud provides more security than public clouds. It is set up within an organization's internal enterprise data centre. The scalable resources and virtual applications provided by the cloud vendor are merging together which are available for cloud users to share and use. The use of private cloud can be much more secure than that of the public cloud because of its specified internal exposure. The organization can access to operate on a specific Private cloud. Corporations are deciding that the private cloud proves less risky. The ability of Private cloud is to virtualize services maximizes hardware usage, ultimately reducing costs and complexity. Most important resources of any organization are its resources and its data. The major drawback of private cloud is its higher cost. When comparisons are made with public cloud the cost of purchasing equipment, software often results in higher costs to an organization in private cloud. However, under the private cloud model, the cloud is only accessible by a single organization providing that organization with control and privacy. A private cloud which is also called an Internal Cloud resides with in the company environment and its access is restricted usually to company employee and business partner.

Benefits of the Private cloud:

1. Flexibility and scale which meet client demands.
2. Resource sharing among a large number of users.
3. Payment according to use of the services.
4. Use of technologies and internet protocol to access cloud resources.

III Hybrid Cloud: A Hybrid Cloud is an integrated cloud services which use both private and public cloud to perform distinct functions within the same organization. It can also be defined as multiple cloud systems that are connected in a way which allows programs and data to be moved easily from one system to another [9]. It is a configuration of at least one private cloud and at least one public cloud.

This computing model combines the security benefits of a private cloud as well as public cloud. .

Hybrid Cloud provides more secure control of the data and applications which allows various parties to access information over the Internet.

A hybrid cloud offered in one of two ways: a vendor has a private cloud and forms a partnership with a public cloud provider, or a public cloud provider forms a partnership with a vendor that provides private cloud.[9] In hybrid cloud, an organization manages some resources in house and some out-house. Typically, the hybrid approach allows a business to take advantage of the scalability and cost-effectiveness that a public cloud computing environment offers without exposing data to third-party vendors.

Benefits of the Hybrid Cloud :

The hybrid cloud provides security as the private cloud element of the hybrid cloud model provides the security where it is needed for sensitive operations and also satisfy customer requirements for data handling and data storage where it is applicable.

- i. Supplies support for cloud-bursting.
- ii. The hybrid cloud provides flexibility as the availability of both secure resource and scalable cost effective public resource can provide organizations with more opportunities to explore for different operations.

System Architecture:

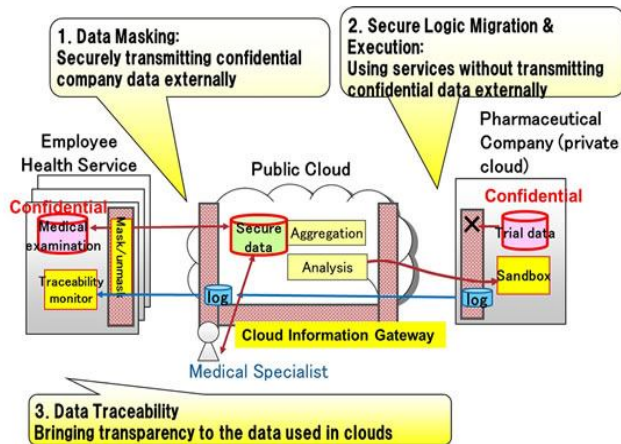
The architecture of cloud computing refers to refers to the components and subcomponents required for cloud computing. These components typically consist of a front end platform which consist of fat client,thin client, mobile device back end platforms which consist of servers, storage. These combined, components make up cloud computing architecture. Security is most important issue in cloud computing. Data masking is the process of hiding original data with random characters or data. The main purpose of data masking is to protect data that is classified as personal identifiable data or sensitive data. In data masking data may be altered in different methods including encryption, character stuffing and character of word substitution. The overall practice of Data Masking at an organizational level should be tightly coupled with the Test Management, underlying Methodology and should incorporate processes for the distribution of masked test data subsets

1. Working of proposed System:

This process shows how we encrypt the data so that the invader does not know what the actual data is about.

In this we use Data Masking and along with it the padding of data is applied

III. DATA SECURITY IN THE CLOUD



In the above diagram, the encryption process takes place at the socket layer of the sender side. The diagram shows us that the mask is applied to the data so that the original data is not being reflected to invader. After masking, the data which is masked is applied with the padding and hence the data is much more secure than it was before. It gives us the double encryption of the data and hence the data is much more securely transferred to the receiver.[6]

Receiver side:

At the receiver side, reverse the process of the sender happens. In this as shown in the diagram, the data which is encrypted doubly is being decrypted doubly. At first the Padding of the data is removed and after the removal of the padding we obtained the masked data. Now the masking of the data is being removed and hence the receiver obtains the original data.

Need of Data Masking:

- i)When copy sensitive data outside of production environment.
- ii)Moving the test data to cloud.
- iii)Sending data to vendors.
- iv) Leverage off-shore development/consultant.

Different types of Masking:

Static data masking: Static data masking is used by most organization when they create testing and in fact is the only possible masking method when using outsourced developers in a separate location or a separate company. In these cases it is necessary to duplicate the database. When doing so, it is critical to use a static data masking tools. These tools make sure that all sensitive data is masked before sending it out of the organization. Static data masking provides a basic level of data protection by creating an offline or testing database using a standard ETL procedure. [7]The static data base can be updated repeatedly, for example on a daily or weekly basis. This is not a security risk, but it often has conclusion for a variety of tests and development issues.

Dynamic data masking: Dynamic Data Masking (DDM) is a strategy for controlling or limiting unauthorized access to data, where data streams from a database or production environment are transformed or masked as they are requested. Dynamic data masking provides result for the cases where individuals are working close to the production environment, but should not have access to the original data. . For example, contractors and staffers may be trying to troubleshoot or update a production database.

[8]It is important that they do not have access to sensitive information such as individual health data, credit card numbers, etc. — with DDM, the information is tangled or otherwise altered, so that these technicians are working with harmless data as they manipulate a database.

Data Masking and the cloud: In recent years, organizations develop their new applications in the cloud. The cloud solution as of now allows organization to use IaaS, PaaS, SaaS. There are various models of creating test data and moving it to the cloud. Data masking becomes the part of these processes in SDLC as the development environments.

Data Masking Techniques

Substitution: Substitution technique is the most effective method of applying data masking and able to preserve the authentic look of the data records. This technique consists of randomly replacing the contents of a column of data with information that looks similar but is completely unrelated to the real details. For example, the surnames in a customer database could be sanitized by replacing the real last names with surnames drawn from a largish random list.[8] Substitution data can sometimes be very hard to find in large length - however any data masking software should contain datasets of commonly required items. For example, to sanitize surnames by substitution, a list of random last names must be available.[7] Then to sanitize telephone numbers, a list of phone numbers must be available. The substitution method need to be applied for many of the fields[11] in DB structure such as telephone numbers, zip codes, credit card numbers and other card type numbers like Social Security numbers .

Shuffling: Shuffling is similar to substitution except that the substitution data is derived from the column itself. In simple terms the data is randomly shuffled with the column. Shuffling is effective for small amounts of data. Another consideration is the algorithm used to shuffle the data. If the shuffling method can be determined, then the data can be easily —unshuffled|. [8] For example, if the shuffle algorithm simply ran down the table swapping the column data in between every group of two rows it would not take much work from an interested party to revert things to their unshuffled state. Shuffling is rarely effective when used on small amounts of data.[10] For example, if there are only 5 rows in a table it probably will not be too difficult to figure out which of the shuffled data really belongs to which row.[10] On the other hand, if a column of numeric data is shuffled, the sum and average of the column still work out to the same amount. It is sometimes useful.

Encryption: Encryption is one of the most complex methods to solve the data masking problem. The Encryption technique algorithmically mix-up the data. This usually does not leave the data looking realistic and can sometimes make the data larger. Encryption also destroys the formatting and look and feel of the data. Encrypted data rarely looks meaningful; in fact, it usually looks like binary data. This sometimes leads to character set issues when manipulating encrypted varchar fields. Certain types of encryption impose constraints on the data format as well. [7] This means that the fields must be extended with a suitable padding character which must then be stripped off at decryption time.

IV. CONCLUSION

Cloud computing is a technology and many organization needs the combination of dynamic and static database masking. We have the cloud service models, deployment models and security in cloud by using data masking techniques. Storage of data on the cloud refines the way we manage the storage of data and access the data from the cloud. In this paper it is also mentioned about the various cryptographic methods which help us to encrypting the data at sender side and then transferring it to the receiver side. This paper also explores the need of data masking in present information. Data masking will enable us to accomplish the following: (a) Increase protection against data theft. (b) Enforces 'need to access'. (c) Provides realistic data for testing, development and data sharing. (d) Provides a heightened sense of security to clients, employee and supplier.

REFERENCES

- [1] [1]Research on Cloud Data Storage Technology and Its Architecture Implementation Kun Liua, Long-jiang Donga
- [2] An Advanced Survey on Cloud Computing and State-of-the-art Research Issues Mohiuddin Ahmed1, Abu Sina Md. Raju Chowdhury2, Mustaq Ahmed3, Md. Mahmudul Hasan Rafee4 Department of Computer Science and Information Technology, Islamic University of Technology Board Bazar, Gazipur-1704, Bangladesh
- [3] Cloud Computing Security Issues And Challenges Pranita P. Khairnar, Electronics Department, Amrutvahini College of Engineering, Sangamner Prof. V.S. Ubale Asst. Prof. Electronics Department, Amrutvahini College of Engineering, Sangamner
- [4] Various Issues & Challenges of Load Balancing Over Cloud: A Survey Palak Shrivastava1 Sudheer Kumar Arya2, Dr. Priyanka Tripathi3 Research Scholar National Institute Of Technical Teachers Training & Research Bhopal, India shrivastava, Associate Professor National Institute Of Technical Teachers Training & Research Bhopal, India.
- [5] International Journal of Advanced Research in Computer Science and Software Engineering Identity – Based Secure and Flexible Data Sharing in Cloud Based Smart Grid R. Ruhina Student, (SE), Department of Computer Science and Engineering, JNTUA College of Engineering, Anantapuramu, Andhra Pradesh, India
- [6] DATA SECURITY IN THE CLOUD :Mr. Jiten Prithiani MCA Final year Student ,V.E.S. Institute of Technology, Mumbai, India jiten.prithiani@ves.ac.in, Mrs. Dhanamma

Jagli, Department of MCA, V.E.S.Institute of Technology
,Mumbai, India

- [7] Design of Data Masking Architecture and Analysis of Data Masking Techniques for Testing Ravikumar G K, Dr. B. Justus Rabi,Dr MGR University, Chennai, Tamil Nadu, INDIA
- [8] Data Masking: What You Need to Know What You Really Need To Know Before You Begin A Net 2000 Ltd.White Paper.
- [9] Swot Analysis of Mobile Cloud ComputingA.M.S.Zunaita Sulthana*, L.Clara Mary*, A.SangeethaDepartmentof computer science, MIET Institution, Trichy
- [10] A Survey on Recent Trends, Process and Development in Data Masking for Testing Ravikumar G K1 ,Manjunath T N2, Ravindra S Hegadi3,Umesh I M4
- [11] Data Sanitization TechniquesA Net 2000 Ltd. White Paper