

SURVEY ON CONFIDENTIALITY IN CLOUD

M.K.S Karthikeya¹, Arif Mohammed Abdul²

¹Student, Department of Computer Science, GITAM University, Hyderabad

²Assistant Professor, Department of Computer Science, GITAM University, Hyderabad

Abstract

Organizations store their sensitive information in cloud along with their trust. A different organization selects the database models according to usage and by majorly considering the aspect like confidentiality and recovery of the data in database. The main theme of this paper is that to show trends in the cloud along with demerits which helps to increase the level of confidentiality and security in cloud databases.

Keywords: Data models, virtualization, confidentiality, survey, security

1. INTRODUCTION

Databases are repositories of information with links that help to make the data searchable. The database models run independently with the help of virtual machine image and maintained by the cloud service provider. The databases that are available in the cloud are SQL Databases, NOSQL Databases. The use of databases includes availability, performance, flexibility and mainly price. Performance is observed and it will only affect the information but not completely on the database. When an input is given by the user then it is considered and then the time taken between the input and output is considered as a performance factor. Availability is the process of getting the data from the system when it is required by the user. Flexibility is the term used when the user shifts from one system to other system the data is to be modified without any loss of data in it. Price is considered as a factor only when the amount required by the small network is equal to amount required by the large one and is considered as less expensive. In order to run the database on the cloud we have two ways namely virtual machine and Database as a service.

Virtual machine Image: In this it helps the enterprises to make more efficient use of hardware resources. It also facilitates great degree of abstraction by using the servers which exist as a single file and the data can be transferred to servers easily from one piece to other by duplicating them and it helps to create a more scalable infrastructure. When virtualization was not present a single application is run and a single OS image is observed which are running a multiple application on same machine and often creates a conflicts and costly infrastructure.

When virtualization came into picture hardware independence operating system and system applications are observed. In this it can manage OS and application as a unit by encapsulating them into virtual machines. Virtualization has transformed individual server into a pool of computing resources which result in decrease of costs and increase of efficiency.

Database as a service: It is also called as DBaaS. DBaaS enabled the optimized commodity platform to reduce software and hardware databases to build database applications. This database software is self-serve and elastic which empowers to implement, configure and maintain individual database from a single database. The main theme of DBaaS is that it is capable of lowering cost and significantly to reduce hardware requirements and thus by sharing hardware resources with multiple databases. In DBaaS automated backup and recovery module is present which helps to update the data periodically. DBaaS is a perfect for applications that require quick provisioning and dropping of data in the databases. DBaaS works well for dynamic applications that require custom data and provides cost efficient platform.

2. DATA MODEL

Data model is nothing but data modeling which is representation of the data in the cloud for small and medium sized organizations. Data models are mainly classified into two types namely SQL database and No SQL database.

SQL database: SQL stands for Structured Query Language which helps for managing the data in the Relational Database Management System (RDBMS). In this cloud based protection is given to the databases with high availability and distributed resources. SQL database can run either as a virtual machine image or as a service. The table shown below gives the example of SQL data model in the Virtual machine and Database as a Service

Data model	Virtual Machine Deployment	Database as a Service
SQL Data Model	Postgre SQL	clustrix

NOSQL Database: Databases like Mongo DB are another type of databases which run on the cloud and they are built to serve read and write operations. In the NOSQL database the database administrator have relied on buying bigger servers as the data load increases rather than distributing the

data across multiple hosts as load increases. NOSQL databases are generally designed to have the properties like automatic, repair and less time management to recover. Minor changes can be made to the RDBMS to change the data model to non relational. The table shown below is an example of NOSQL data model

Data Model	Virtual Machine	Database as a service
NOSQL Data Model	Mongo DB	Database.com

3. LITERATURE SURVEY

Loss age of data and data leakage is the major problem in the cloud computing. Even though the security issues are categorized in the SPI model (SaaS, PaaS, IaaS) identifying the main vulnerabilities is most important. In the SQL data model the performance is better but we can just display what we get from the database and translate the code into the value. SQL databases are also called as relational databases and database consists of tables which have predefined schema and are vertically scalable. SQL database uses Structured Query Language for defining and manipulating the data. The best examples of SQL databases are My SQL, Oracle, Sqlite and Postgres. SQL database posses ACID properties (Atomicity, Consistency, Isolation and Durability). SQL databases can be open source and closed source based upon the usage of vendors. MS SQL is user friendly database which has stability, reliability and scalability factors. The strengths that are supported by the SQL DB is that disaster recovery mechanism which includes RAID mirroring and clustering along with RAID partition. A backup is also provided which is helpful to recover the data from the database.

NOSQL database model is also called as non-relational or distributed database where the databases are the collection of key-value pair, documents, graphs and wide column stores. In No SQL database uses UnQL (Unstructured Query Language) which varies from database to database. The best examples of NO SQL databases are Mongo DB, Big Table, Redis and Couch DB. No SQL database doesn't have standard interface to perform complex queries and it fits for hierarchical data as it has key-value pair way and database follows Brewers CAP theorem (Consistency, Availability and Partition tolerance). No SQL database are classified on the basis of way of storing data as graph databases, key value databases and column store databases. Mongo DB is one of the most popular document based NOSQL database as it stores data like documents with dynamic schema. The benefits of Mongo DB are it uses the simple queries which give good performance as all the related data are in the single document. In this the data is horizontally scalable and so we can reduce the work load by increasing the number of servers. Dynamic schema is observed which gives flexibility to evolve without modifying the existing data.

Security Overview

In the data models like SQL the performance is a major factor which leads to get the data from the database very slow and the structure of the data along with the meta is observed which leads to leakage of data from database. The portability of data is decreased which leads to vendor lock-in. In SQL databases encoding the values in the database is a bad form and localization is difficult. Where as in No-SQL even though the data is consistent the relation between the tables are not observed (mongo DB) and there are no stored procedures which leads to high complexity of getting data in the event of disaster recovery.

Even though we have benefits with the Virtual Machines the challenges with it are security, potential downtime. Virtual Machine enables server to perform tasks and run more applications it can also consume much more time. Security and access are to be monitored to avoid the loss or theft of the data in the runtime environment. The main challenge in the virtual machine is security that is if you are moving application from physical server to public cloud there are number of security risks since cloud servers will be hosted on the virtual machines in a shared infrastructure. Virtual machines are easy to harden attack than an operating system. The other challenge in the virtual machine is potential downtime which is unlikely that if a virtual resource fails at once then it can be more difficult to reboot virtual machines in the event of catastrophic failure. In the virtual machine when user gets the data it can be stored in the disk file. In middle if an attacker comes and compromises then integrity is lost. Here sensitive data is left on disks and they can be accessed to storage hosts which lead to compromise of integrity and data confidentiality of stored images.

In Database as a Service it has potential security and privacy issues which lead to potential loss of critical data or bankrupt of the cloud database service provider. DBaaS model also include a lack of control over network performance issues such as unacceptable latency and application failures. Some DBaaS doesn't support capabilities of the typical RDBMS such as data compression which leads to loss age of data. In DBaaS data residing in an unencrypted form may be vulnerable to bugs, errors and attacks from external entities confronting data confidentiality issues. In DBaaS super users are present who usually have privilege to access all resources for maintenance purposes. If the privilege is misused then it leads to loss of data. The other scenario in the DBaaS is that if the attacker is present outside and performs malicious actions like fraud, phishing, scamming will also lead to illegal transactions to launch Denial of Service (DoS). Confidentiality in network breaches and makes it prone to certain threats like data modification and eavesdropping. Recording the data and tracing refers to data provenance which helps to find out origin of data in-between the databases which leads to leakage of sensitive information.

4. CONCLUSION

Even though we have multiple numbers of databases and data models in the cloud, every database has data leakage at any particular point. As this confidentiality of data must be major part in the database where we must concentrate. Hence a new database must be proposed where the database must be strong enough to provide security to the data in the database. The database must be able provide the confidentiality at all points like during the data transfer from client to server, security at cloud even if the DBA compromises, the security of the data must be maintained.

FUTURE SCOPE

Levels of confidentiality is less in the cloud database which leads to leakage of sensitive information and security measures are to be considered by implementing proxy between organizations and data models in order to provide access to authenticated users. Usage of data models like SQL and NOSQL should be backed d up with a cryptographic schema so that the data will be encrypted and decrypted at the user interface and data is processed over an encrypted format by introducing a concept called onion layers which provides a high level of security and privacy to the data.

REFERENCES

- [1] Cloud Computing: Types, Architecture, Applications, Concerns, Virtualization, Role of IT Governance in Cloud
- [2] Role of virtualization in cloud computing Kamyab Khajehi
- [3] Client and Data Confidentiality in Cloud Computing Using Fragmentation Method by Thota Reshma Kishore, D.Akhila Devi, S.Prathyusha, D.Bhagyasri, Bhuma Naresh
- [4] Relational Cloud: A Database-as-a-Service for the Cloud
- [5] <http://wiki.hsr.ch/Datenbanken/files/MongoDB.pdf>
- [6] <http://en.wikipedia.org/wiki/MySQL>
- [7] <http://www.sqlcoffee.com/>
- [8] Supporting security and consistency for cloud database by Luca Ferretti, Michele Colajanni, and Mirco Marchetti
- [9] SQL and NoSQL Databases Vatika Sharma1 , Meenu Dave
- [10] <http://www.couchbase.com/press-releases/unql-query-language>
- [11] A Query Language and Optimization Techniques for Unstructured Data by Peter Buneman, Susan Davidson, Gerd Hillebrand, Dan Suciu
- [12] Protecting Confidentiality with Encrypted Query Processing by Raluca Ada Popa,