ISSN: 2321-7782 (Online)
Volume 3, Issue 2, February 2015
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Securing Data Management Based on Key Technologies in Cloud Computing

K. Padmini
CSE Department
Kallam Haranadha Reddy Engineering College
Guntur, Andhra Pradesh, India

Abstract: ***Cloud computing has elevated IT to newer limits by offering the market environment data storage and capacity with flexible scalable computing processing power to match elastic demand and supply, whilst reducing capital expenditure. However the opportunity cost of the successful implementation of Cloud computing is to effectively manage the security in the cloud applications. It provides people the way to share distributed resources and services that belong to different organization. Since cloud computing uses distributed resources in open environment, thus it is important to provide the security and trust to share the data for developing cloud computing applications. In this paper we assess how can cloud providers earn their customers' trust and provide the security, privacy and reliability, when a third party is processing sensitive data in a remote machine located in various countries? A concept of utility cloud has been represented to provide the various services to the users.***

Emerging technologies can help address the challenges of Security, Privacy and Trust in cloud computing.

Keywords: Cloud computing; Security; Public cloud, Private cloud, Hybrid Cloud, Security, Data protection, Data Management.

I. INTRODUCTION

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics**,** three service models, and four deployment models. With the development of multi-core processors, virtualization, distributed storage, broadband Internet and automatic management, a new type of computing mode named cloud computing is produced. It distributes computation task on the resource pool which consists of massive computers, so the application systems can obtain the computing power, the storage space and software service according to its demand. It can concentrate all the computing resources and manage them automatically by the software without intervene. This makes application offers not to annoy for tedious details and more absorbed in his business. It will be advantageous to innovation and reduce cost. It's the ultimate goal of cloud computing to provide calculation, services and applications as a public facility for the public, So that people can use the computer resources just like using water, electricity, gas and telephone. Currently, the understanding of cloud computing is developing and changing constantly, cloud computing still has no unanimous definition. This paper describes three main service forms of cloud computing: SAAS, PAAS, IAAS, compared the definition of cloud computing which is given by Google, Amazon, IBM and other companies, summarized the basic characteristics of cloud computing, and emphasized on the key technologies such as data storage, data management, virtualization and programming model.

II. Essential Characteristics

a) ***On-demand self-service****:* A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service's provider.

b) ***Broad network access****:* Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, laptops, and PDAs)

c) ***Resource pooling:*** The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or data centre). Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.

d) ***Rapid elasticity:*** Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out, and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.

e) ***Measured Service:*** Cloud systems automatically control and optimize resource use by leveraging a metering capability1 at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.

III. Types of Cloud

There are different types of clouds that you can subscribe to depending on your needs. As a home user or small business owner, you will most likely use public cloud services.

2.1 Deployment Models

a) ***Private cloud:*** The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.

b) Community cloud: The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or a third party and may exist on premise or off premise.

c) Public cloud: The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.

d) Hybrid cloud**:** The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

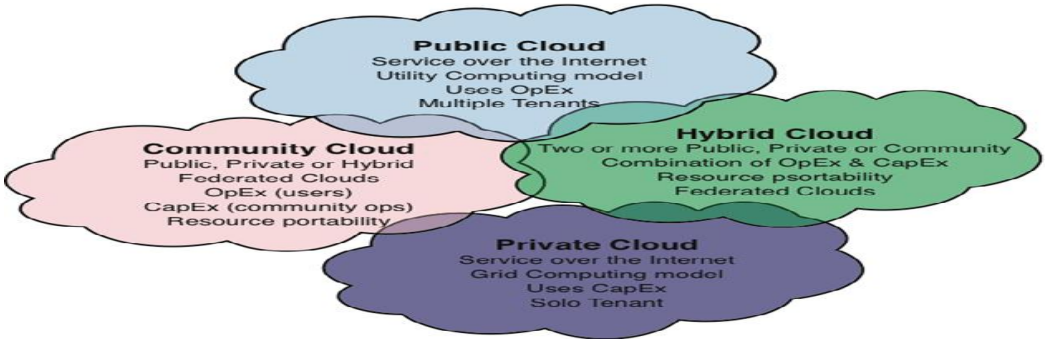

Figure 1: Types of Deplyoment models.

2.2 Service Models in Cloud Computing

Service delivery in Cloud Computing comprises three different service models, namely Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS). The three service models or layer are completed by an end user layer that encapsulates the end user perspective on cloud services. The model is shown in figure 2.

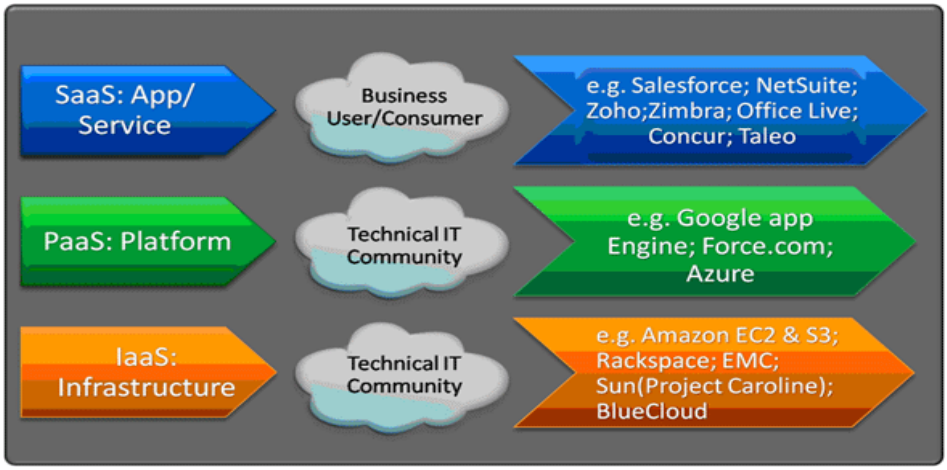

Figure 2: Service models and end user layer

If a cloud user accesses services on the infrastructure layer, for instance, she can run her own applications on the resources of a cloud infrastructure and remain responsible for the support, maintenance, and security of these applications herself. If she accesses a service on the application layer, these tasks are normally taken care of by the cloud service provider.

a) SaaS

Software-as-a-Service provides complete applications to a cloud's end user. It is mainly accessed through a web portal and service oriented architectures based on web service technologies. The services on the application layer can be seen as an extension of the ASP (application service provider) model, in which an application is run, maintained, and supported by a service vendor. The main differences between the services on the application layer and the classic ASP model are the encapsulation of the application as a service, the dynamic procurement, and billing by units of consumption (pay as you go). However, both models pursue the goal of focusing on core competencies by outsourcing applications.

b) PaaS

PaaS comprises the environment for developing and provisioning cloud applications. The principal users of this layer are developers seeking to develop and run a cloud application for a particular platform. They are supported by the platform operators with an open or proprietary language, a set of essential basic services to facilitate communication, monitoring, or service billing, and various other components, for instance to facilitate start up or ensure an application's scalability and/or elasticity . Distributing the application to the underlying infrastructure is normally the responsibility of the cloud platform operator. The services offered on a cloud platform tend to represent a compromise between complexity and flexibility that allows applications to be implemented quickly and loaded in the cloud without much configuration. Restrictions regarding the programming languages supported, the programming model, the ability to access resources, and persistency are possible downsides.

c) IaaS

The services on the infrastructure layer are used to access essential IT resources that are combined under the heading Infrastructure-as-a-Service (IaaS). These essential IT resources include services linked to computing resources, data storage resources, and the communications.

Physical resources are abstracted by virtualization, which means they can then be shared by several operating systems and end user environments on the virtual resources – ideally, without any mutual interference. These virtualized resources usually comprise CPU and RAM, data storage resources (elastic block store and databases), and network resources.

IV. CLOUD COMPUTING SECURITY ISSUE

Security is considered one of the most critical aspects in everyday computing, and it is no different for cloud computing due to the sensitivity and importance of data stored in the cloud .Cloud computing infrastructures use new technologies and services, most which haven't been fully evaluated with respect to security. The main problems cloud computing faces are preserving confidentiality and integrity of data in aiding data security.

a) Privileged user access - information transmitted from the client through the Internet poses a certain degree of risk, because of issues of data ownership; enterprises should spend time getting to know their providers and their regulations as much as possible before assigning some trivial applications first to test the water. Regulatory compliance - clients are accountable for the security of their solution, as they can choose between providers that allow to be audited by 3rd party organizations that check levels of security and providers.

b) Data location - depending on contracts, some clients might never know what country or what jurisdiction their data is located

c) Data segregation - encrypted information from multiple companies may be stored on the same hard disk, so a mechanism to separate data should be deployed by the provider.

d) Recovery - every provider should have a disaster recovery protocol to protect user data

e) Investigative support - if a client suspects faulty activity from the provider, it may not have many legal ways pursue an investigation

f*) Long-term viability* - refers to the ability to retract a contract and all data if the current provider is bought out by another firm.

Given that not all of the above need to be improved depending on the application at hand, it is still paramount that consensus is reached on the issues regarding standardization.

V. PROTECTING DATA IN CLOUD

a) Trust

Trust between the Service provider and the customer is one of the main issues cloud computing faces today. There is no way for the customer to be sure whether the management of the Service is trustworthy, and whether there is any risk of insider attacks. This is a major issue and has received strong attention by companies. The only legal document between the customer and service provider is the Service Level Agreement (SLA). This document contains all the agreements between the customer and the service provider; it contains what the service provider is doing and is willing to do . However, there is currently no clear format for the SLA, and as such, there may be services not documented in the SLA that the customer may be unaware that it will need these services at some later time.

b) Confidentiality

Confidentiality is preventing the improper disclosure of information. Preserving confidentiality is one of the major issues faced by cloud systems since the information is stored at a remote location that the Service Provider has full access to. Therefore, there has been some method of preserving the confidentiality of data stored in the cloud. The main method used to preserve data confidentiality is data encryption; however encryption brings about its own issues, some of which are discussed later.

b) Authenticity (Integrity and Completeness)

The authenticity of a subject or object is defined as its genuineness and credibility; these can be verified on the basis of its unique identity and characteristic features. Information is authentic if it can be reliably assigned to the sender, and if it can be proved that this information has not been changed since it was created and distributed. A secure technique for identifying the communication partners and mechanisms for ensuring authenticity are essential here. These mechanisms must be capable of confirming or disproving the authenticity of the protected information. None of the system participants can create or distribute messages and data on behalf of another subject.

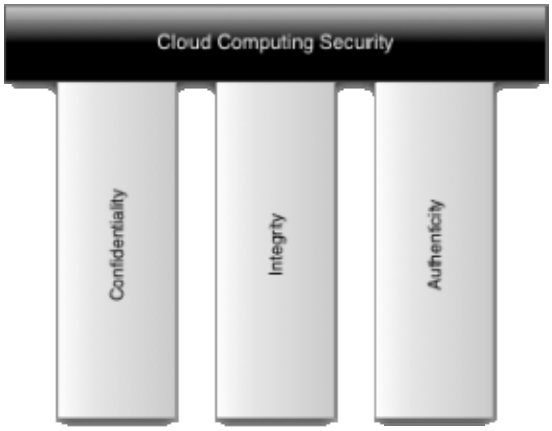

Figure 3: Data Security Protection.

VI. KEY TECHNOLOGY OF CLOUD COMPUTING

Cloud computing systems use many technologies, of which the programming model, data management, data storage, virtualization are the key technologies.

Virtualization

Virtualization is a method of deploying computing resources. It separates the different levels of the application system including hardware, software, data, networking, storage and so on, breaks the division among the data center, servers, storage, networking, data and the physical devices, realize dynamic architecture, and achieves the goals of managing centralized and use dynamically the physical resources and virtual resources, improving the flexibility of the system, reducing the cost, improving the service and reducing the risk of management. In the cloud computing environment, all virtualization solutions are system integration solutions including servers, storage systems, network devices, software and service. They include multiple layers of virtualization technologies such as hardware virtualization, network infrastructure virtualization, application virtualization and desktop virtualization, and combine several layers flexibly to realize the different models of virtualization solutions according to the application environment. In the whole cloud computing virtualization strategy, We can make use of various mechanisms which is provided by virtualization technique, quickly imitate different environment and experiment without important hardware and physical resources, and achieve the purpose of building operate system and application, raising the safety and realizing management environment, for later in a more simplified and effective way to put them into the production environment. Thus provide greater flexibility and quickly identify potential conflicts. In the meantime, We can make use of server virtualization technique to integrate a large number of scattered and underutilized physical servers to less independent

and aggregate physical servers, even make a large network virtual machine to replace thousands of server and make it run under the high utilization in long time, thus bitterly manage IT cost, maximize energy efficiency and advance using rate of resource. We can also make use of storage virtualization technique to support the varied disk storage system in network environment, through integrating the storage capacity to a storage resources pool, help IT system to simplify storage foundation structure, manage the life cycle of information system and maintain business continuity. We also make use of application and desk virtualization technique to provide application infrastructure virtualization function, lower the cost of establish, management and run the application, and achieve the purposes of improving flexibility and agility, ensuring business process integrity, raising application function and bitterly manage running status of the application. In addition, virtualized system management and supervision service can help us detect, monitor and manage all the virtual and physical resources including system and software through a common access point, and provide complete cross-enterprise service management, decrease the amount of managing tool which is used to support various type servers .

Mass Distributed Storage

In order to ensure high credibility and economy, cloud computing adopts distributed storage to save data, using redundancy storage to ensure the reliability of stored data and using high credible software to make up the incredibility of the hardware, therefore providing the cheap and credible mass distributed storage and computing system. The data storage system of cloud computing are Google File System (GFS) and Hadoop Distributed File System (HDFS) which is developed Hadoop team.

1) GFS:

GFS is a distensible distributed file system. It is used in large and distributed applications which need to access mass data. The designing ideology of GFS is different from the traditional file system, which is designed for dealing large-scale data and the application property of Google. It runs on the cheap and common hardware, but it can provide fault tolerance function. It can provide high-performance service to a great deal of users.

Figure 4 shows the system structure of GFS. A GFS clusters includes master server and many chunk servers, and it can be visited by several clients. The file is spilt into fixed size blocks. When creating a block, the server distributes an unchanged and globally unique 64 handles to identify it. The block server treats the blocks as Linux files and saves them at the local hard disk, read or write block data according to specified handles and byte range. To ensure the credibility, each block will be copied to several block severs, and default saving three copy. The master server manages all metadata of the file system, including namespace, access control information, the reflecting information from file to blocks and position of the blocks. The code of GFS Client is embedded in each program, realizes Google file system API, which helps the application communicate with master server and block server, and then read or write data. Exchange between client and server is only the operations of metadata, all the data communication directly contact with block server, which consumedly raised the efficiency of system, and keeping master server from overloading.

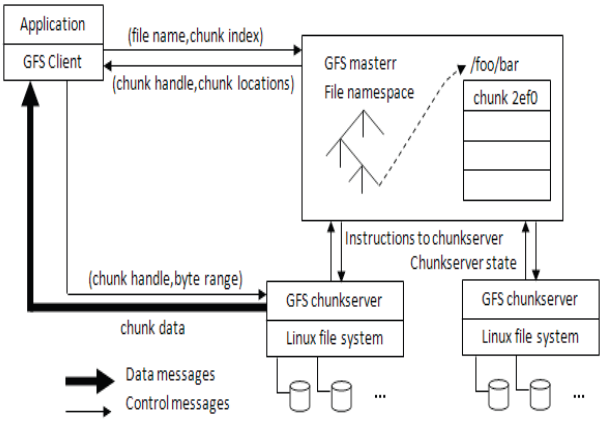

Figure 4: Goole File System Architecture.

Parallel programming model

To enable users efficiently to use cloud computing resources and more easily enjoy services that cloud computing brings about; cloud computing programming model must make task scheduling and parallel execution transparent to users and programmers. Cloud computing adopts MapReduce programming model, which decomposes the task into multiple subtasks, and through two steps (Map and Reduce) to realize scheduling and allocation in the large-scale node. MapReduce is a parallel programming system developed by Google. It puts parallelism and fault tolerance, data distribution, and load balance in a database, and all the operations of data are summarized in two steps: Map and Reduce. When the Programmer submitted his parallel processing procedures to MapReduce, he just need to definite two functions: Map and Reduce. According to the size of input data and configure information, MapReduce can automatically initialize them to several same Map tasks and Reduce tasks, and then process them using different data blocks by calling Map function and Reduce function. MapReduce system mainly consists of three modules: client, master and worker. The client is responsible for submitting parallel processing assignments composed by the users to master node; master node will automatically decompose user's task into Map missions and Reduce missions, and delivered to worker nodes; worker nodes request to the master node for the work tasks, at the same time, the distributed file system consisting of many worker nodes will be used for storing input and output data of MapReduce.

MapReduce is mainly used in mass data processing. One of the features of the task scheduling strategy is scheduling priority the task the node which the data belong. This kind of scheduling scheme which is based on data position enables Map tasks to read and process data locally when the worker node which request task saves the data needing to process, thus reduces the network overhead and improve the performance of the system.

Data management

Cloud computing needs to process and analyze mass and distributed data, therefore, data management technology must be able to efficiently manage large data sets. There are two kinds of data management technology in cloud computing system: BigTable of Google and HBase developed by Hadoop team. BigTable is based on GFS, Scheduler, Lock Service and MapReduce. Each Table is a multidimensional sparse map. Row, column, Tablet and timestamp are the basic elements of BigTable. Tablet is a collection of rows. Data items in BigTable are ordered according to the sequence of keyword in the dictionary, with each row dynamically delivered to Tablets. Each node manages about 100 Tablets. Timestamp is a 64-bit integer, representing the different versions of data. Column family is an aggregation of several columns, whose granularity decides access authority. BigTable needs three components to execute: one database which is linked to each client, a master server, and several Tablet servers. The master server is responsible for arranging Tablets to Tablet servers, load balance and garbage collection, etc. Tablet servers are responsible for managing a group of Tablets, processing the requests of read or write, etc. To ensure the high scalability of data structure, BigTable adopts three-level hierarchical way to store location information, as shown in figure 5.

The first level is chubby file which contains the position information of root Tablet. There is only one root Tablet, which contains all the location information of MetaData tablets; a MetaData contains the position information of many user tables.

When reading data, the client firstly gets the location of root Tablet from chubby file, and it gets the position information of MetaData tablet from root Tablet, then it gets the location of user table which contains the location information of the object data, and then it read the location information of the object data from user table, according to the information, read data from a special location in the server.

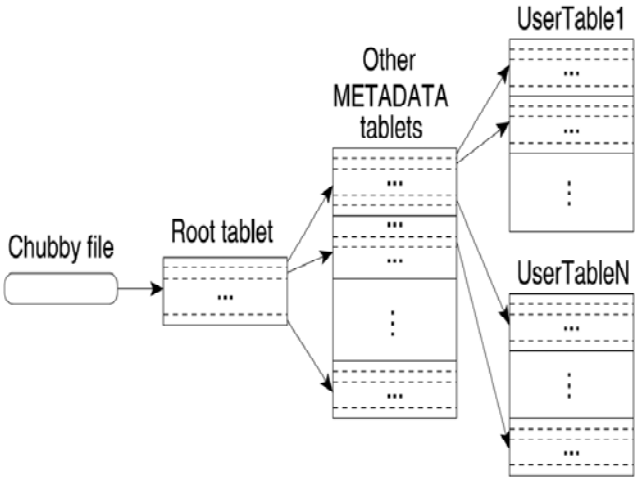

Figure 5.The Structure of Storing Tablet Location Information

VII. CONCLUSION

Although Cloud computing can be seen as a new phenomenon which is set to revolutionise the way we use the Internet, there is much to be cautious about. There are many new technologies emerging at a rapid rate, each with technological advancements and with the potential of making human's lives easier. However one must be very careful to understand the limitations and security risks posed in utilizing these technologies. Cloud computing is no exception.

Cloud computing is a new kind of commercial computing model developed on the basis of grid computing, public computing and SaaS. It can distribute computing tasks to the resources pool consisting of massive computers, enabling different application systems to acquire computing power, storage space and various software services according to needs. The ultimate goal of cloud computing is to provide calculation, services and applications as a public facility for the public, So that people can use the computer resources just like using water, electricity, gas and telephone. Therefore, the enterprises can save many costs purchasing hardware and software. This paper introduces the definition of could computing and its main service patterns, summarizes the characteristics, and focused on the key technologies such as the data storage, data management and programming model.

References

1. Al-Fares M et al (2008) A scalable, commodity data center network architecture. In: Proc SIGCOMM
2. SANJAY GHEMAWAT; HOWARD GOBIOFF; PSHUN-TAK LEUNG. The Google file system. Proceedings of the nineteenth ACM symposium on Operating systems principles. Oct. 2003
3. Jinesh Varia. Cloud architectures- Amazon web services [EB/OL]. ACM Monthly Tech Talk, http://acmbangalore.org/events/monthly-talk/may-2008--cloud-architectures---amazon-web-services.html, May, 2008
4. Amazon Elastic Computing Cloud, aws.amazon.com/ec2
5. Google Docs [URL]. http://docs.google.com/, access on Sep. 2008
6. Aymerich, F.M. Fenu, G. Surcis, S. An approach to a Cloud Computing network. Applications of Digital Information and Web Technologies, 2008
7. Zhang Weimin, Tang Jianfeng. Cloud Computing. Science Press, 2009
8. Dlamini M T, Eloff M M and Eloff J H P, 'Internet of People, Things and Services – The Convergence of Security, Trust and Privacy', 2009.
9. Balachandra R K, Ramakrishna P V, Dr. Rakshit A, 'Cloud Security Issues', 2009 IEEE International Conference on Services Computing. [10] Ahmed, M., Xiang, Y., and Ali, S "Above the Trust and Security in Cloud Computing: A Notion Towards Innovation", in 2010 IEEE/IFIP 8th International Conference on Embedded and Ubiquitous Computing.

AUTHOR(S) PROFILE

Mrs .K. Padmini, received the M.Tech in CSE, Software Engineering as specialization in 2009 with Distinction and B.Tech, Information Technology in 2006 from Aurora Engg College Affiliated to JNTU, Hyderabad. Working as Assistant Professor Since 6 years in Computer Department in Engineering Colleges. Her Research interest includes DataMining, Cloud Computing, Network Security and Big Data.

