ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

# Research on Key Technologies of Data Mining For Cloud Computing On Data Storage Management Analysis Secret Sharing Algorithm.

**Mr.J.Muthukrishnan**
**MCA, M.Phil,**
**Department of Computer Science,**
**Swami Vivekananda art and Science College,**
**Villupuram.**

**Dr.G.T.Shrivakshan**
**MCA, M.Phil, Ph.D Head,**
**Research and Post Graduate,**
**Department of Computer Science,**
**Thiruvalluvar university, Vellore.**

***ABSTRACT:***
***In recent years data mining us used for extracting potentially useful information from raw data on integration of data mining techniques into normal day to day activities has become common places everyday people are confront with together data mining techniques help businesses to more efficient by reducing cost to retrieve meaningful information from virtually integrated data due to attractive features of cloud computing users are storing large amount of data on cloud storage, these data may be users personal of secret. Users remotely store their personal or secret information on cloud storage on enjoy best features of cloud application without burden on local hardware and software management .After storing data on cloud storage users can access their data by very thin liens. The three main service forms of cloud computingSAAS, PAAS, IAAS comparison. But there is a drawback of outsourcing data on cloud storage is data security risks because users has no control over outsourced data even users are unknown about location of their data. So,in this situation achieving data confidentiality on outsourced data is an open challenge for researches. Today some system are available by using that we can store and retrieve encrypted data but if we have large amount of data to store on cloud than it is difficult to retrieve them efficiently Another problem is to store and manage encryption key efficiently. To sole the data confidentiality and key management problem together we are using ranking function on inverted indexes and advanced secret sharing algorithm.***

***KEY WORDS:*** ***Cloud computing, cloud data storage, Data encryption symmetric, Data mining Key management.***

## 1 INTRODUCTION:

According to NIST [1] Cloud computing is a model for convenient, on-demand network access to a shared pool of congurable computing resources (e.g., net-works, servers, storage, applications, and services) that can be rapidly provisionedand released with minimal management eort or service provider interaction. In cloud computing data and applications are maintained with the use of central remote server and internet and allow consumers to use the applications with-out installating and also with the help of internet cloud computing allows customers to access their personal les which are stored in some other computer. Yahoo email, Gmail, or Hotmail etc are examples of cloud computing. The email management software and the server are fully managed and controlled by the CSPGoogle, Yahoo etc and are all on the cloud (internet).

### 1.1. DEPLOYMENT MODELS

Based on the deployment the cloud can be of following types. Different types are clouds are described below:

**1.1.1. PUBLIC CLOUD:** The cloud vendor hosts the computing infrastructure at its own premises and the customer dont have any control on where the computing infrastructure is hosted. The computing infrastructure is open for use by general public or deferent organizations share this computing infrastructure.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

**1.1.2. PRIVATECLOUD:** Computing infrastructure is not shared between the organizations. It is dedicated to a particular organization and is therefore more secure but expensive than public clouds. The two types of private clouds are described below. Externally hosted private clouds are hosted by some third party that specializes in cloud infrastructure and are also exclusively used by one organization. On-premise private clouds are hosted by the enterprise itself and are more expensive as compared to externally hosted private clouds.

**1.1.3. HYBRID CLOUD:** The cloud infrastructure in hybrid cloud is a synthesis of two or more unique cloud infrastructures that are bounded together by some standardized technology that empowers application and data portablility.

**1.1.4. COMMUNITY CLOUD:** The cloud infrastructure is solely utilized by consumers from organization that have related or imparted concerns. Community cloud may be overseen, managed and worked by a third party or one or a greater amount of organizations in the community or some consolidation of them.

### 1.2. SERVICE MODELS

Deferent service delivery models in Cloud Computing.

#### 1.2.1. SOFTWARE AS A SERVICE (SAAS):

In SaaS model consumer has the procience to utilize applications of supplier that are running on a cloud infrastructure. A program interface can be used to access the applications through various client devices. Operating systems, storage, servers, network or other underlying cloud infrastructure are not overseen by the consumer.

**1.2.2. PLATFORM AS A SERVICE (PAAS):**Consumer is given the capability to deploy onto the cloud infrastructure created by the consumer or applications acquired that are created using services, libraries, programming languages and tools supported by the provider.The underlying cloud infrastructure including storage, operating systems, servers, network controlled by the consumer

**1.2.3. INFRASTRUCTURE AS A SERVICE (IAAS):**Consumer has the capability to provision computing resources like networks, storage, processing etc where the consumer can deploy and run arbitrary software, which can include applications and operating systems. The underlying cloud base is not overseen or controlled by the customer however the consumer has control over deployed applications, storage and operating systems.

## 3. KEY TECHNOLOGIES OF CLOUD COMPUTING

Cloud computing systems use many technologies, of which the programmingmodel, datamanagement, datastorage, virtualization are the key technologies.

### 3.1 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 resourcesand virtual resources, improving the flexibility of the system, reducing the cost, improving the service andreducing 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

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

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 .

### 3.2MASSDISTRIBUTED 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.shows the system structure of GFS. A GFS clusters includes master server and many chunkservers, and it can be visited by several clients. The file is spilt into fixed size blocks. When creating ablock, the server distributes an unchanged and globally unique 64 handles to identify it. The block servertreats 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

## 4. SECURITY ISSUES IN CLOUDCOMPUTING

### 4.1. Security of Cloud Implementation Models

Basically, the deployment of a cloud is managed in house (Private Cloud) or over a third-party location (Public Cloud). While, for various reasons, it is deployed as an integrated private-public cloud (Hybrid Cloud) [1,8].A “Community Cloud” is a fourth type of cloud implementation models, where the infrastructure spreads over several organizations and is accessed by a specific community [8]. The different cloud implementation models are shown in Figure 2.In private cloud configuration an organization may have

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

control over its infrastructure or delegate that to a thirdparty, being physically on-site or off-site [1,8]. Securing the in-house cloud infrastructure is controllable and requires no need for extra trust mechanisms. While having a third-party service provider running the private cloud is prone to several doubts [8]. Users adopt a private cloud implementation to increase the security level. That decreases the isolation level between the services and the infrastructure. For instance, managing the security of the provided service in conjunction with the existing firewalls and protection services. Furthermore, operating over a secure virtual private network is an option to isolate the private cloud hosted by a third-party. Despite the benefits of a private cloud, several issues need attention as unbalanced resources utilization. An idle infrastructure is a wasted resource for example [17].Public cloud implementation is a model in which a service provider, third-party, offers public services on pay-per-use manner. Some of the benefits of this model are the economies of scale, ability to have short-term usage and greater resources utilization [1]. Secure use of the shared public cloud is more challenging compared to private clouds. For that, public cloud suits more incidental or less vulnerable applications [17].

In [8] the authors stated that trust is an important issue for public clouds, hence the management is governed by a third-party. A trusted third party auditor (TPA) is proposed in [18] to solve the trust issues in public clouds. A TPA is expected to analyze the public cloud services and provide an adequate report. Public cloud service providers are up posed to prove the credibility of their systems, guarantee service availability, ensure a high level of data protect and handle security breach attempts efficiently [1,8,18]. An organization reluctant to fully trust the public community cloud, due to security issues, may think of having a hybrid cloud implementation. Without compromising the security of essential data, they have the possibility of keeping only sensitive data on a relatively small private cloud [1,8,17]. The best of less-sensitive


Figure 2. Cloud implementation model

### 4.2. SECURITY OF SERVICE DELIVERY MODELS

Cloud service providers mainly offer three delivery models that are the SaaS, PaaS, and IaaS, lternatively called provision and distribution models. Figure 3 illustrates the delivery models and their basic components. Other delivery models exist, for example in [19] the authors mentioned the Human as a Service (HuaaS), and the Support Services. Here the main security requirements for the three delivery models are exhibited. IaaS layer provides the primary infrastructure of the cloud as a service to the customers.

Infrastructure is the main hardware components and their management software that includes servers, network, storage, file system and operating systems [1,17]. Customers using IaaS havea limited control over the actual infrastructure, as their usage is based on pay-per-use only [3]. Securing the IaaS layer is divided into two main areas, the virtual environment and the physical environment [8,12]. Several security requirements need to be present at the virtual level, which includes controlling the access, data encryption, secure communication channels, and virtual protection [8].

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

On the other hand in terms of physical components, it is required to ensure the hardware reliability, and preventing physical intrusion [8,12]. PaaS is the application deployment level, where developers are supposed to develop their applications and implement them. Though, some authors [1,8] consider PaaS and IaaS to be at the same layer rather than two. A platform usually enables utilizing development platform, databases and middlewares [17].

Meanwhile, platform providers currently enable a limited number of specific development languages and API's. For example the limited development languages on Google AppEngine, Facebook Platform, Microsoft Azure and ZohoCreator [19]. The security requirements for PaaS are almost the same as those for the IaaS. Since both share the virtual environment characteristics.

The differences in the security measures, if any, are related to the components' level or the role of the service user, a developer or system administrator for instance [8]. SaaS is usually accessed over the internet by the end users (tenants) as employees, managers, clients and auditors [1,8,17]. It is possible that SaaS may run over an IaaS or PaaS operated by a different provider [17]. This service delivery level encompasses various on-demand applications ranging from automation and productivity to enterprise integration [3]. Being at the higher level of service delivery it requires a relatively different level of security. The main concerns include preserving the privacy, preventing impersonation, availability of services, secure communication and data protection [8,17]. Another important dimension of security to look at is the SaaS provider being a customer to IaaS or PaaS providers [1]. IaaS, PaaS and SaaS service delivery models together are considered the foundations of cloud computing. The complexity of the cloud services as an interrelated system is high, for that managing the security is complex as well. Depending on the service provided by the cloud, the security measurements may vary in application but they still an issue. Another security dimension to consider is the interface channels between the delivery models.


Figure 3. Cloud computing delivery models

## 5 SECRET SHARING SCHEMES

Deferent secret sharing schemes are discussed below:
Simple secret sharing

### 5.1. ADDITIVE SECRET SHARING

In additive secret sharing for a given secret SF,n 1 random integers R =r1, r2..... rn⍰1 are selected by the dealer uniformly from F. X then computes Sn = S ⍰ n⍰1 i=0 rimodF X sends the share si = ri to each player Pj 1 j n ⍰ 1 ,and the share sn is sent to Pm The secret S reconstruction is trivial and it can be generated simply by adding all the shares together.

S = ni=0

SimodF In this scheme contribution of shares from all the participants is required in order to reconstruct the secret. The original share cannot be recovered in case if one or more of the articipants are missing, A scheme in which contribution of all the participants is required is known as a perfect secret sharing scheme.

### 5.2. THRESHOLD SECRET SHARING SCHEME

In 1979 Shamir [7] and Blakley [16] introduced the concept in order to make the message more secure. In this scheme, the message MG is divided into n pieces

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

MG1;MG2;MG3; : : : ;MGn, with or witout ransformation of the message, in such a way that, for a specifed k, 14

## 6 SECURITY ISSUES AND RISKS IN CLOUD COMPUTING

Gartner in 2014 recognized seven security issues [4] that need to be tended to before organizations switch completely to the cloud computing model.

### 6.1 SECURITY ISSUES AND RISKS IN CLOUD COMPUTING

**6.1.1. DATA LOCATION:** While storing data in cloud some clients might not know where their data is actually located.

**6.1.2. REGULATORY COMPLIANCE:** Customers can choose between providers that permit to be examined by third party organizations that check levels of security provided by cloud service providers.

**6.1.3. DATA SEGREGATION:**

Since the data in encrypted form from deferent organizations may be stored in the same place, so a system is required that separates data from deferent organizations and it should be provided by the cloud service provider.

**6.1.4. LONG-TERM VIABILITY:** It alludes to the capability to withdraw an agreement and all information if the current supplier is bought out by another rm.

**6.1.5. INVESTIGATIVE SUPPORT:** In case a customer suspects defective movement from the supplier, he might not have numerous legitimate ways seek after an investigation.

**6.1.6. RECOVERY:** Each supplier ought to have a disaster recovery convention to ensure client data is protected in case of a disaster also.

**6.1.7.PRIVILEGED USER ACCESS:**Data transmitted from the customer through the Internet represents a certain level of risk, in view of issues of information possession; ventures ought to invest time getting to know their suppliers and their regulations however much as could be expected before allotting some trivial applications.

### RISKS IN CLOUD COMPUTING

The six specie areas of cloud computing where substantial security attention is required is are as follows

1. Security of data in transit.

2. Security of data at rest.

## 7. CONCLUSION

Cloud computing has been a surpassing shift so far in terms of utilizing the current technologies. The trend of having cloud services as part of an organization seems to be gaining more importance. Especially in this era the cycle of introducing more technological innovations is getting shorter. For many reasons, including the reduction of capital expenditures, organizations need to consider utilizing cloud services as an essential part of their foundations. Nevertheless, various challenges are prohibiting the attainment of vast deployment and acceptance levels. The main drawback of the existing cloud service implementations is their inability to provide an accredited high security level.

Moreover, security assurance needs to cover the transmission channels which might include a third-party. To have better utilization of cloud services many issues need to be enhanced in a way ensuring high level of security, confidentiality, authenticity, integration, agility, scalability and trust. Possibly an automated SLA, third trusted party, or a novel innovation would be an interesting study area to cover the security issues related to cloud computing.

## REFERENCES

[1] Mell, Peter, and Timothy Grance. "The NIST denition of cloud computing (draft)." NIST special publication 800.145 (2011): 7.

[2] Jaydeep. "Security and Security and Privacy PrivacyPrivacy Issues in C loudComputing." http://arxiv.org/

[3]"CloudComputingachitecture".http://communication.howstuworks.com/cloudcomputingl.htm.

[4] Brodkin, Jon. "Gartner: Seven cloud-computing security risks." Infoworld(2008): 1-3.

[5] Calheiros, Rodrigo N., et al. "Cloudsim: A novel framework for modelingand simulation of cloud computing infrastructures and services."arXivpreprint arXiv:0903.2525 (2009).

[6] Ogbu, Richard Chukwu, and IfeanyiUgbagaNkole. "Cloud Computing: A review."

[7] Shamir, Adi. "How to share a secret." Communications of the ACM 22.11(19 79): 612-613.

[8] Liu, C.L. Introduction to Combinatorial Mathematics. McGraw- Hill, NewYork, 1968.

[9] Damgrd, Ivan, et al. "Secure key management in the cloud." Cryptography and Coding. Springer Berlin Heidelberg, 2013. 270-289.

[10] Mazieres, David, et al. "Separating key management from le system secu-rity." ACM SIGOPS Operating Systems Review 33.5 (1999): 124-139.35

[11] Zhang JianXun, GuZhiMin.Surey of research progress on cloud omputing.Application Research of Computers,2010,27(2).429-433.

[12] FENG DengGuo, ZHANG Min,ZHANGYan,XUZhen.Study on Cloud Computing Security.Journal of Software,2011,22(1).

13] Jamsa, K., Cloud Computing: SaaS, aaS, IaaS, Virtualization, Business Models, obile, Security and More, Jones &Bartlettearning, 2012,ISBN-10: 1449647391, ISBN-13: 978-1449647391.

[14] Mahowald RP, Sullivan CG. Worldwide SaaS and Cloud Software 2012–2016 Forecast and 2011 Vendor Shares. International Data Corporation, 2012.

[15] Peter Mell, and Timothy Grance, "The NIST Definition of Cloud Computing", The National Institute of Standards and Technology, USA, 2011, Link: http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.

[16] IT Strategists, "Top Cloud Computing Companies and Key Features", Link: http://www.itstrategists.com/Top-Cloud-Computing-Companies.aspx.

[17] Merriam-Webster Dictionary, "Definition of data mining", Link: http://www.merriam-webster.com/dictionary/data%20mining.

[18] ORACLE, "Oracle Data Mining Mining Techniques and Algorithms", Link: http://www.oracle.com/technetwork/database/options/advanced-analytics/odm/odm-techniques-algorithms-097163.html.

[19] BhagyashreeAmbulkar and VaishaliBorkar, "Data Mining in Cloud Computing", MPGI National Multi Conference 2012 (MPGINMC-2012), 7-8 April 2012, Link: http://research.ijcaonline.org/ncrtc/number6/mpginmc1047.pdf.

[20] Amazon. Amazon elastic compute cloud (Amazon EC2). 2009. http://aws.amazon.com/ec2/

[21] SANJAY GHEMAWAT; HOWARD GOBIOFF; PSHUN-TAK LEUNG. The Google file system. Proceedings of the nineteenth ACM symposium on Operating systems principles. Oct. 2003.