Volume 4, Issue 2, February 2014 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

A Novel Approach for Multi Cloud Storage

Elpula Vijaya Kumar*, V. Rama Krishna, P. Raja Sekhar Reddy

CSE Dept, AGI, India

Abstract— Cloud data storage redefines the issues targeted on customer's out-sourced data (data that is not stored/retrieved from the costumers own servers). Cloud Computing is gaining the attention of today's users and businesses, Cloud Computing provides suitable on-demand network access to a shared pool of computing resources. Cloud Computing allows to store and maintain data on remote servers that are managed by Cloud Service Providers. This data can then be accessed throughout the globe. But as more and more information of individuals and companies is placed in the cloud, security concerns are beginning to grow about how safe an environment is. In order to overcome the security issues multi cloud is a best alternate where we can store the data in more number of CSP's .In this Approach an individual CSP cannot make utilise of the data what he had, until unless it combined or joins with other set of data which is available with other CSP's we propose a traditional technique followed to distribute data to multi-cloud storage model in cloud computing which holds an economical distribution of data among the available SPs in the market, to provide customers with data availability as well as reliability . Data fragmentation plays an important role in data distribution

Keywords— Cloud computing, storage, Cloud service provider, Security, Fragmentation

I. INTRODUCTION

Multi-Cloud strategy is the concomitant use of two or more cloud services to minimize the risk of widespread data loss or downtime due to a localized component failure in a cloud computing environment. such a failure can occur in hardware, software, or infrastructure. a multi-cloud strategy can also improve overall enterprise performance by avoiding "vendor lock-in" and using different infrastructures to meet the needs of diverse partners and customers. reasons for an adverse cloud event can vary from a single cable connector failure to an EMP (electromagnetic pulse), or from a natural disaster to an act of cyber warfare. Even the failure of a single hard disk/drive unit can result in a large-scale network outage if the malfunction takes place at a critical point in the system such as a host computer. As customer bases and device types grow increasingly diverse (yet at the same time increasingly specialized), organizations face a complex array of challenges in their quest to satisfy the demands of all end users. In particular, the speed with which a given website loads has a huge impact on customer satisfaction. recent research has revealed that the average user expects a webpage to load just as fast on a mobile device as it would on their home computer (two seconds or less), because faster Page loading results in more frequent and longer visits to a given website, page loading time can indirectly affect rankings in search engine. A Multi-cloud strategy can help an organization to minimize page loading times for all types of content. A multi cloud approach can offer not only the hardware, software and infrastructure redundancy necessary to optimize fault tolerance but it can also steer traffic from different customer bases or partners through the fastest possible parts of the network. Some Clouds are better suited than others for a particular task. For example a certain cloud might handle large numbers of requests per unit time requiring small data transfers on the average ,but a different cloud might perform better for smaller numbers of requests per unit time involving large data transfers on the average. Some organizations use a public cloud to make resources available to consumers over the internet and a private cloud to provide hosted services to a limited number of people behind a firewall. A third type of cloud called a hybrid cloud, may also be used to manage miscellaneous internal and external services.

II. CLOUD COMPUTING ARCHITECTURE

This section describes the architectural, business and various operation models of Cloud Computing.

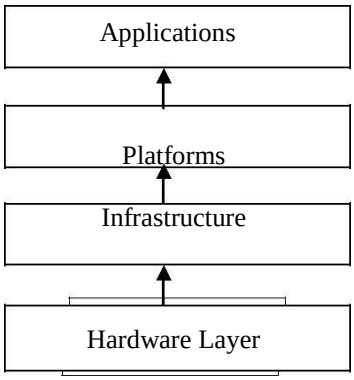
A Layered Model of Cloud Computing:The Cloud Computing architecture can be divided into 4 layers [8]: the hardware/data center layer, the infrastructure layer, the platform layer and the application layer, as shown in Figure 1. We describe each of them in detail:

The Hardware Layer: This layer is implemented in the data centres and used for managing the physical resources of the Cloud, including physical servers, routers, switches, power and cooling systems. A data center has thousands of servers that are interconnected through switches, routers or other fabrics and are organized in racks. Hardware layer deals with issues like configuring hardware, fault tolerance, traffic management, power and cooling resource management [8].

The Infrastructure Layer: This layer is also known as the virtualization layer, the infrastructure layer creates a pool of storage and computing resources by partitioning the physical resources using virtualization technologies such as Xen, KVM and VMware. Issues like dynamic resource mangement are dealt through virtualization technologies [8].

The Platform Layer: Above infrastructure layer, the platform layer exits and it consists of operating systems and application frameworks. The main aim of platform layer is to reduce the burden of deploying applications directly into VM containers. For example, Google App Engine operates at the platform layer to provide API support for implementing storage, database and business logic of typical web applications [8].

The Application Layer: At the highest level of the hierarchy, their exit an application layer Different from traditional applications, Cloud applications can leverage the automatic-scaling feature to achieve better performance, availability and lower operating cost. Compared to traditional service hosting environments such as dedicated server farms, the architecture of Cloud Computing is more modular. Each layer is loosely coupled with the layers above and below, allowing each layer to evolve separately. [8].



III. CONTROL OF RESOURCES IN A MULTI CLOUD

It is sometimes affirmed that when compared to traditional on premises computing, Cloud Computing requires consumers to give up two important capabilities to their providers: [7,9]

Control: it is the responsibility of the provider to decide, who and what is allowed to access consumer data and programs, and who has the ability to perform actions like erasing data or disconnecting a network and also that what type of actions have been taken that would not challenge the consumer's intent.

Visibility: the ability to monitor, with high confidence, the status of a consumer's data and programs and how consumer data and programs are being accessed by others. The extent to which consumers may need to surrender control or visibility depends on a number of factors including physical ownership and the ability to configure protective access boundary mechanisms around a consumer's computing resources

Data Privacy

Privacy [9] addresses the confidentiality of data for specific entities, such as consumers or others whose information is processed in a system. Privacy carries legal and liability concerns, and should be viewed not only as a technical challenge but also as a legal and ethical concern. Protecting privacy in any computing system is a technical challenge; in a Cloud, setting this challenge is complicated by the distributed nature of clouds and the possible lack of consumer awareness over where data is stored and who has or can have access.

System Integrity

Clouds require protection against intentional subversion of the functionality of a cloud. Within a Cloud there are stakeholders: consumers, providers, and a variety of administrators. The ability to partition access rights to each of these groups, while keeping malicious attacks at bay, is a key attribute of maintaining Cloud integrity. In a Cloud setting, any lack of visibility into a cloud's mechanisms makes it more difficult for consumers to check the integrity of cloud-hosted applications.[9]

Security Concerns

Cloud Computing presents specific challenges to privacy and security. When using cloud-based services, one is entrusting the third party for their data storage and security. Cloud-sourcing involves the use of many services, and many Cloud based services provide services to each other, and thus cloud-based products may have to share your information with third parties if they are involved in processing or transferring of your information. Each cloud-based service has its own terms and conditions, or service level agreement, that the user agrees to before signing to a Cloud, and these services agreements are often updated. There must be proper awareness about privacy and security issues around Cloud Computing. People need to be aware of terms and conditions of the Cloud service providers as well as keep up with updates. The information stored by Cloud services is subject to the legal, regulatory and policy environments of the country .As more and more information is stored in the Cloud these issues become significant, and Cloud Computing will continue to offer challenges to national policy and regulation as well as to internet governance, on how best to resolve privacy and security issues.

Security concerns have been raised due to the new computing model introduced by Cloud Computing, which is characterized by off-premises computing, lost control of IT infrastructure, service oriented computing, and virtualization, and so on. Security concerns from users can be briefly summarized as follows:

- System failure and Data availability: When keeping data at remote systems owned by others, data owners may suffer from system failures of the service provider.
- Data error: Client data should be error free on the cloud. The data is stored on the Cloud which is remote to the client. If the correct storage strategy is not used data might not be stored correctly on the storage server of the Cloud.
- Data Migratibility: Users that adopt Cloud Computing may want their data to be migrated to other Clouds but this procedure is subject to the risk that their data cannot be migrated to other clouds.
- Long-term availability: Consumer must be sure that data must be available every time even if the Cloud Computing provider will get acquired and swallowed up by a larger company.
- Data location: Client has no idea of the whereabouts of data or exactly where the data is hosted. In fact, one might not even know what country it will be stored in.
- Data segregation: Client has to be sure that encryption must be available at all stages, and that these encryption schemes were designed and tested by experienced professionals.
- Data Recovery: Clouds provides the data recovery to some extent in case of a disaster. Any offering that does not replicate the data and application infrastructure across multiple sites is vulnerable to a total failure.

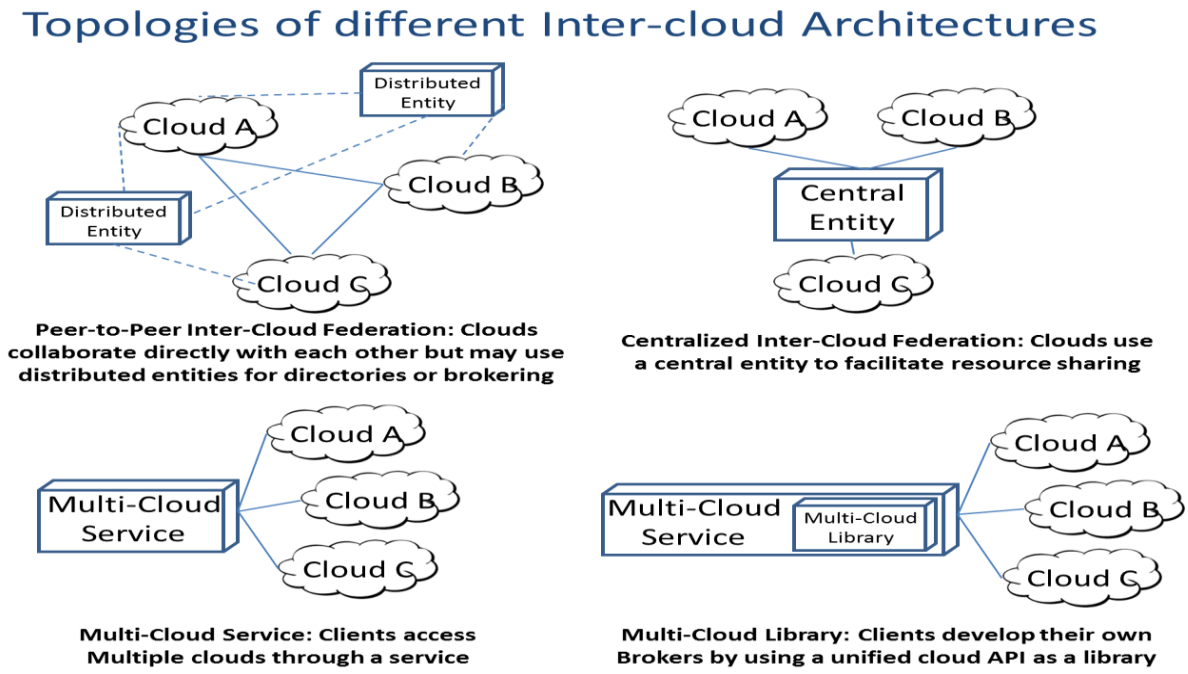
Security concerns for Multi Cloud Computing, can be briefly summarized as follows:

- System failure and Data availability: When keeping data at remote systems owned by others, data owners may suffer from system failures of the service provider, but data is resided in multiple clouds losing all copies have a very little chance.
- Data error: Client data should be error free on the cloud. The data is stored on the Cloud which is remote to the client. If the correct storage strategy is not used data might not be stored correctly on the storage server of the Cloud.
- Data Migratibility: Users that adopt Cloud Computing may want their data to be migrated to other Clouds but this procedure is subject to the risk that their data cannot be migrated to other clouds, as we are adopting the multi cloud storage we can have the flexibility to migrate data to different clouds with out violating SLA's
- Long-term availability: Consumer must be sure that data must be available every time even if the Cloud Computing provider will get acquired and swallowed up by a larger company.
- Data location: Client has no idea of the whereabouts of data or exactly where the data is hosted. In fact, one might not even know what country it will be stored in.
- Data segregation: Client has to be sure that encryption must be available at all stages, and that these encryption schemes were designed and tested by experienced professionals, and data becomes useful only when it joins with other cloud data.
- Data Recovery: Clouds provides the data recovery to some extent in case of a disaster. Any offering that does not replicate the data and application infrastructure across multiple sites is vulnerable to a total failure. This is most advantage when we maintain data in multi cloud as we can easily get back the data after failure ,because data is available in multiple clouds.

Distribution Model

Customers' stored data at cloud service providers is vulner-able to various threats. Previous studies in a cloud service provider can be a victim to Denial of service attacks or its variants[3],[4].The idea for distributing the data among two storage clouds such that, an adversary cannot retrieve the contents of the data without having access to both the storage clouds. Relaying entirely upon a couple of service providers for the storage Such an attack scenario is entirely passive, because the cloud user cannot detect that his information has been collectively retrieved from the service providers without his consent.. Let us assume that two cloud service providers are available for customer who want to store his own data securely. seeks a distribution of customer's data pieces among the available SP s in such a way that, at

least q number of SP s must take part in data retrieval, while minimizing the total cost of storing the data on SP s as well as maximizing the quality of service and availability of data provided by the SP s.



IV. DATA DISTRIBUTION

Data preservation and data integrity are two of the most critical security issues related to user data[2],[10]. In conventional paradigm, the organizations had the physical possession of their data, and thus have an ease of implementing better data availability policies. But in case of cloud computing, the data is stored on an autonomous business party, that provides data storage as a subscription service. The users have to trust the *cloud service provider* (*SP*) with security of their data. In the author discussed the criticality of the privacy issues in cloud computing, and pointed out that obtaining an information from a third party is much more easier than from the creator himself. One more bigger concern that arises in such schemes of cloud storage services, is that, there is no full-proof way to be certain that the service provider doe not retains the user data, even after the user opts out of the subscription. With enormous amount of time, such data can be decrypted and meaningful information can be retrieved and user privacy can easily be breached. In order to stop the SP to observe the data ,data can be fragmented and distributedtoseveralSP's

Why Fragment?

- Applications work with views rather than entire relations.
- Efficiency
- Data is stored close to where it is most frequently used.
- Data that is not needed by local applications is not stored
- Parallelism

Types of Fragmentation

- Four types of fragmentation:
- Horizontal,
- Vertical,
- Mixed,
- Derived.

Horizontal Fragmentation

- Each fragment consists of a subset of the tuples of a relation R.

- Defined using Selection operation of relational algebra:
  - $\sigma_p(R)$
- Example:
  - Relation: Sells(pub, address,price,type)
  - Fragments:
    - » SellsBitter= $\sigma_{type = \text{“bitter”}}$(Sells)
    - » SellsLager= $\sigma_{type = \text{“lager”}}$(Sells)

This strategy is determined by looking at predicates used by transactions.

- Involves finding set of *minimal* (*complete* and *relevant*) predicates.
- Set of predicates is *complete,* if and only if, any two tuples in same fragment are referenced with same probability by any application.
- Predicate is *relevant* if there is at least one application that accesses fragments differently.

Vertical Fragmentation

- Each fragment consists of a subset of attributes of a relation R.

Defined using projection operation of relational algebra

- $^{\Pi}a_1,\ldots a_n^{(R)}$

- Determined by establishing *affinity* of one attribute to another.
- Example:
  - Relation:
    Bars(name,address,licence,employees,owner

)
  - Fragments:
    - » $^{\Pi}$name,address,licence $^{(Bars)}$
    - » $^{\Pi}$name,address,employees,owner$^{(Bars)}$

Mixed Fragmentation

- We can also mix horizontal and vertical fragmentation.
- We obtain a fragment that consist of an horizontal fragment that is vertically fragmented, or a vertical fragment that is horizontally fragmented.
- Defined using Selection and Projection operations of relational algebra.

$^{\sigma}p^{(\Pi}a_1,\ldots a_n)\ ^{\Pi}a_1,\ldots a_n{}^{(\sigma}p)$

Derived Horizontal Fragmentation

- A horizontal fragment that is based on horizontal fragmentation of a parent relation.[10]
- Ensures that fragments that are frequently joined together are at same site.
- Defined using *Semijoin* operation of relational algebra:

  $R_i = R >_F S_i, \qquad 1 \leq i \leq w$

- If relation contains more than one foreign key, need to select one as parent.
- Choice can be based on fragmentation used most frequently or fragmentation with better join characteristics.

How can we define fragments correctly In defining fragments we have to be very careful.

Our model distributes the data pieces among more than one *service providers*, in such a way that no one of the *SP* s can retrieve any meaningful information from the pieces of data stored on its servers, without getting some more pieces of data from other service providers. Therefore, the conventional single service provider based techniques does not seem too much promising. Distributing the data over multiple clouds or networks in such a way that if an adversary is able to intrude in one network, still he cannot retrieve any meaningful data, because its complementary pieces are stored in the other network. Our approach is similar to this approach, because both aim to remove the centralized distribution of cloud data. Although, in their approach, if the adversary causes a service outage even in one of the data networks, the user data cannot be retrieved at all. This is why in our model; we propose to use a redundant distribution scheme in which at least a threshold number of pieces of the data are required out of the entire distribution range, for successful retrieval. Meaningful information from the data pieces allocated at their servers. Also, in addition, we provide the user with better assurance of availability of data, by maintaining redundancy in data distribution. In this case, if a service provider suffers service outage [1] or goes bankrupt, the user still can access his data by retrieving it from other service providers. From the business point of view, since *cloud data storage* is a subscription service, the higher the data redundancy, the higher will be the cost to be paid by the use.

V. CONCLUSIONS

In this we proposed a various data fragmentation schemes for multi cloud storage in cloud computing , which seeks to provide each customer with reliability, availability and better cloud data storage decisions

ACKNOWLEDGMENT

We are very much thankful for all the faculty and friends who helped in carrying out this paper with out which it could not happen.

REFERENCES

[1] Amazon.com, "Amazon s3 availablity event: July 20, 2008", Online at http://status.aws.amazon.com/s3-20080720.html, 2008.

[2] P. S. Browne, "Data privacy and integrity: an overview", *In Proceeding of SIGFIDET '71 Proceedings of the ACM SIGFIDET (now SIGMOD)*, 1971.

[3] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, ―A break in the clouds: towards a cloud definition,‖ SIGCOMM Computer Communication Review, vol. 39, pp. 50–55, 2008.

[4] M. Armbrust, A. Fox, R. Griffith, A.D.Joseph, R.Katz, A.Konwinski, G.Lee, D. Patterson, A. Rabkin, I. Stoica, and M.Zaharia, ―A view of cloud computing‖, Communications of the ACM, vol. 53, no. 4, pp. 50–58, 2010.

[5] Suraj Pandey, ―Scheduling and Management of Data Intensive Application Workflows in Grid and Cloud Computing Environments‖, Ph.D. dissertation, The University of Melbourne,Australia, 2010.

[6] S. H. Shin, K. Kobara, "Towards secure cloud storage", *Demo forCloudCom2010*, Dec 2010.

[7] P. S. Browne, "Data privacy and integrity: an overview", *In Proceeding of SIGFIDET '71 Proceedings of the ACM SIGFIDET (now SIGMOD)*, 1971.

[8] A. Cavoukian, "Privacy in clouds", *Identity in the Information Society*, Dec 2008. R. Gellman, "Privacy in the clouds: Risks to privacy and confidentiality from cloud Computing"

[9] W. Itani, A. Kayssi, A. Chehab, "Privacy as a Service: Privacy-Aware Data Storage and Processing in Cloud Computing

[10] Stefanoceri-Giuseppe-pelagati"Distributed Databases-Principles and systems"Tata MC Graw Hill 2008

