International Journal of
Emerging Research in Management &Technology
ISSN: 2278-9359 (Volume-3, Issue-5)

Research Article | **May 2014**

Data Fragmentation in Multi-Cloud

Tanvi Sharma
CSE/A.I.T.M (MDU)
India

Dr.Deepti Sharma
HOD CSE/A.I.T.M (MDU)
India

Abstract—

Cloud computing is an innovation of existing technology which provides long-dreamed vision of Computing as utility. The emergence of this novel technology in IT business has decoyed most of organizations in both private and public sector. Cloud services can be availed without capital investment as they are commoditized. Cloud users get services in pay per use fashion and enjoy many benefits of cloud including low cost and accessibility from anywhere in the world. However, users have security concerns as they outsource their valuable business data to cloud and treat the cloud as "untrusted".Ensuring the security of cloud computing is a ,as the users often store sensitive information with cloud storage providers may be un trusted. Dealing with single cloud providers is predicted to become less popular with customers due to risks of service availability failures and possibility of malicious insiders in single cloud. A movement towards "Multicloud" has emerged recently. Moving towards multiple clouds can address security problems. In this paper we will study and evaluate what was the need to shift from single cloud to multi cloud and data fragmentation as a service to facilitate enormous data processing, and introduce some functioning enhancement on data distribution to improve the cloud system performance.

Keywords— Single Cloud, Multi cloud, Security, Frgmentation

I. INTRODUCTION

Cloud computing is a phrase used to describe a variety of concepts that involve a large number of computers connected through a real-time communication network such as the Internet. In science, cloud computing is a synonym for distributed computing over a network, and means the ability to run a program or application on many connected computers at the same time [20]

Cloud computing is a general term for anything that involves delivering hosted services over the Internet. These services are broadly divided into three categories: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS). The name cloud computing was inspired by the cloud symbol that's often used to represent the Internet in flowcharts and diagrams. A cloud service has three distinct characteristics that differentiate it from traditional hosting. It is sold on demand, typically by the minute or the hour; it is elastic ,a user can have as much or as little of a service as they want at any given time; and the service is fully managed by the provider (the consumer needs nothing but a personal computer and Internet access). Significant innovations in virtualization and distributed computing, as well as improved access to high-speed Internet and a weak economy, have accelerated interest in cloud computing [6]

A cloud can be private or public. A public cloud sells services to anyone on the Internet. (Currently, Amazon Web Services is the largest public cloud provider) A private cloud is a proprietary network or a data center that supplies hosted services to a limited number of people. When a service provider uses public cloud resources to create their private cloud, the result is called a virtual private cloud. Private or public, the goal of cloud computing is to provide easy, scalable access to computing resources and IT services [9].

II. SECURITY ISSUES IN SINGLE CLOUD

Cloud service providers can offers benefits to users, but security risks play a major role in the cloud computing environment.[9] Users who use online data sharing or network facilities are aware of the potential loss of privacy.[10]

According to a recent IDCI survey, the top challenge for 74% of IT execution on CIO's of cloud computing adoption is related to security matters.[11] Protecting private and important information from attackers or malicious insiders is of critical importance. Moving database to a large data centre involves many security challenges such as Virtualization vulnerability, accessibility vulnerability, privacy and control issues related to data accessed from a third party, integrity, confidentiality, and data loss or theft. In different cloud service models, the security responsibility between users and providers is different. The impact of security issues in the public cloud is greater than the impact in the private cloud. As the cloud services have been built over the internet, any issue that is related to internet security will also affect the cloud services. Resources in the cloud are accessed through the internet; consequently even if the cloud provider focuses on today's world cloud has evolved as a boon But truly said everything has its pros and Cons. Point comes what are the Cons of such a beneficial technology that has exploded market with great capabilities and seems ever promising. Cloud Computing has motivated industries, academia, businesses to adopt cloud computing to host high computationally intensive application down to light weight application and services.[2]

Cloud computing reduces IT Costs and increase capabilities and reach ability of delivered services. As per the Gartner's survey the cloud market was worth USD 68 billon dollar in 2010 and will reach USD 148 by this year. This revenue

proves that Cloud Are a promising platforms. As everything comes with certain Cons in Cloud computing also there is tradeoff between security and high computability still there are a lot of open issues that impact the cloud computing model credibility and pervasiveness.[1][3] Some of these Issues are Cloud security, Vendor Lock In, Multi-Tenancy, Data Integrity, Data Intrusion, and Service Availability

III. MOVING TOWARDS MULTICLOUD

The very first question that arises when once hears about Multi Cloud. What is multi cloud?
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.
The migration of cloud computing from single toward multi-clouds to ensure the security of user's data is extremely important. The term "multi-clouds" is similar to the terms "intercloud" or "cloud-of-clouds" that were introduced by Vukolic.[13] Moving from single cloud to multi-clouds is reasonable and important for many reasons. According to Cachin et al.[14] "Services of single cloud are still subject to outage". Vukolic assumes that the main purpose of moving to multi-clouds is to improve what was offered in single cloud by distributing the realibitiy, trust and the security among multiple cloud providers.

The basic underlying idea is to use multiple distinct clouds at the same time to mitigate the risks of malicious data manipulation, disclosure, and process tampering.[8] By integrating distinct clouds, the trust assumption can be lowered to an assumption of no collaborating cloud service providers. Further, this setting makes it much harder for an external attacker to retrieve or tamper hosted data applications of a specific cloud user.[5]

Table 1: Comparison between Single cloud / Multi cloud

	Data Intrusion	Data Integrity	Service Availability	Data Status	
				Safe	Lost
Single Cloud	If password hacked?	If data hacked?	If system down?	no	yes
Multi Cloud	If password hacked from one Cloud service provider?	If data hacked from one Cloud service provider?	If one cloud down?	yes	no

IV. NEED OF MULTI CLOUD

Many applications today such as web applications, mobile applications are not limited to any geographical boundaries. The customers of these applications may be at far flung places. In such cases if one wants to expand globally, then multi cloud strategies are a boon. While this enables us to reach distributed markets, new challenges related to latency, performance, pricing, and availability crop up. Every customer residing at far flung places need good performance not only for certain period of time, but a consistent good performance is required.

Ensuring Consistent performance and high are high availability are the two major challenges for global expansion. A single cloud service provider or a single delivery network cannot be trusted for this task. Delivering maximum performance globally round the clock is not possible even by most resilient cloud service provider. The only solution is adopting multi-clouds which would distribute the performance and availability threats across global public and private data centers. Global performance is not a representation of a single cloud service providers but a judicious assimilation of many. In order to drive revenue and other benefits that are very closely related with performance and availability, a multi-cloud strategy is essential. When options are available, performance and availability differences are in dollars. One cannot rely on one cloud service provider to fulfill the requirements of customers belonging to different geographic regions. So, what is required is an assimilated network of multiple clouds. Recent high-profile cloud outages are unmitigated manifestation of the need for multi-sourcing, although many companies are still relying on single-source providers. It's an avoidable threat.

A multi-cloud strategy permits one to manager traffic across data centers, clouds and delivery networks to manage costs and optimize price-to-performance ratios.

An unhappy thing about life is that things collapse. This is true for Cloud Computing also: no matter how much better uptime or availability or performance is offered by cloud providers, these services collapse eventually. Preparedness is the only thing we can do. Building redundancy in cloud based application is a part of preparedness, but in clouds this redundancy is limited to running several redundant copies on separate data centers of the same cloud provider. Having multiple data centers at different geographical locations by large cloud service providers is one approach to the solution of this problem. Another possible solution could be adoption of multi-cloud strategy. By using services of multiple cloud service providers, redundancy is achieved at a new abstraction level. In order to host our cloud servers, if we select data centers from different providers, we can effectively do away with the threats related with business continuity of the cloud service provider, threats concerning electricity suppliers, "data center" managerial issues, networking providers. This is possible as each service provider works independently. Other threats which are correlated with a single cloud service provider are also reduced. Cloud works on virtualization. If in case, any vulnerability is detected on our infrastructure provider, and a multi-cloud strategy is adopted, one can immediately switch on to the other provider without any impact to the operations.

V. DATA FRAGMENTATION IN MULTI CLOUD

When we Consider DaaS it become one's prime responsibility to manage this data accurately.
When Cloud is considered, it just give you infrastructure where you can keep data safe, Cloud will not ensure Non redundant data, data integrity on data. It depends on the organization or the User what policies or constraints they set on Data before uploading it on to cloud. Data can be managed on Multiple Cloud taking into account the security issues discussed with single cloud service Providers.

Many cloud computing providers have their data centres spread worldwide to maintain data availability which is typically achieved by replication processes. Amazon's cloud simple storage service [1] replicates data across different geographical regions so that data and applications can continue even in the face of failures of their location. This is likely to be help in running applications on data warehouses, but not transactional data management systems [2]. Yahoo [3] and Amazon [4] both implement data replication through PNUTS and SimpleDB cloud data services over distributed network sites. They designed to run analytical applications on data warehouses, but not for transactional data applications. Similarly, Google [5] implements a replicated database, but does not offer a complete relational Application Programming Interface and weakens the data atomicity. The cloud API is written as series of XML-based messages, and executed on the cloud servers to utilize remote web-based applications and reduce the number of calls between the client and the distributed servers [6].

Microsoft SQL Server [7] cloud data service is implemented over distributed network sites. However, as it doesn't apply commit protocols, the distributed system presents lack of data consistency.

The design of a distributed database introduces three new issues:
• How to partition the database into fragments.
• Which fragments to replicate?
• Where to locate those fragments and replicas.

DATA FRAGMENTATION

Data fragmentation allows you to break a single object into two or more segments or fragments. The object might be a user's database, a system database, or a table. Each fragment can be stored at any site over a computer network. Information about data fragmentation is stored in the distributed data catalog (DDC), from which it is accessed by the TP to process user requests.

There are three types of data fragmentation strategies:

• **Horizontal fragmentation** refers to the division of a relation into subset (fragments) of tuple (rows). Each fragment is stored at a different node, and each fragment has unique rows. However, the unique rows all have the same attributes (columns). In short, each fragment represents the equivalent of a SELECT statement, with the WHERE clause on a single attribute.
• **Vertical fragmentation** refers to the division of a relation into attribute (column) subsets. Each subset (fragment) is stored at a different node, And each fragment has unique columns—with the exception of the key Column, which is common to all fragments.
• **Mixed fragmentation** refers to a combination of horizontal and vertical Strategies. In other words, a table may be divided into several horizontal Subsets (rows), each one having a subset of the attributes (columns).

DATA REPLICATION

Data replication refers to the storage of data copies at multiple sites served by a computer network. Fragment copies can be stored at several sites to serve specific information requirements.

Three replication scenarios exist:

• A **fully replicated** database stores multiple copies of each database fragment at multiple sites. In this case, all database fragments are replicated. A fully replicated database can be impractical due to the overhead it imposes on the system.

• A **partially replicated** database stores multiple copies of some database fragments at multiple sites. Most DDBMSs are able to handle the partially replicated database well.

• An **UN replicated** database stores each database fragment at a single site. Therefore, there are no database fragments.

Several factors influence the decision to use data replication:

DATA ALLOCATION

Data allocation describes the process of deciding where to locate data. Data allocation Strategies are as follows:

• With **centralized data allocation**, the entire database is stored at one site.

• With **partitioned data allocation**, the database is divided into two or more disjointed parts (fragments) and stored at two or more sites.

• With **replicated data allocation**, copies of one or more database fragments are stored at several sites.

In distributed relational database systems, the transactions on the applications are usually subsets of relations (fragments), so using these fragments and distributing them over the network sites increases the system throughput by means of parallel execution. Therefore, an efficient cloud API fragmentation web service is presented to access and manage data relationships, and enhance both the speed and simplicity of the distributed database functionality. This web service is used to retrieve raw data from the cloud data centers by external programs like Java applications. Moreover, it helps to reduce the cost of accessing data over distributed network sites and increases the distributed system performance through data allocation processes.

VI. DATA FRAGMENTATION ARCHITECTURE (DFA)

The requested data in DFA are identified by means of transactions triggered as queries, in the above diagram queries were run on the Cloud or say here in our example a Database or a knowledge base, which determine the specific information that should be extracted from the cloud database servers. The transactions are executed and result in redundant data records as two or more different queries may require the same data records this will lead to redundant data. The redundant data are eliminated and the remaining data records are then partitioned (fragmented) so as to generate the minimum number of fragments which are neither replicated nor intersected, hence the fragments obtained are disjoint.

The architecture of DFA is recognized by the domain knowledge and three main processes; eliminating data redundancy, defining transactions, and fragmenting data records. Figure 1 describes the Data Fragmentation Architecture (DFA) service that will be used for generating data fragments, supporting the use of knowledge extraction, and helping to achieve the effective use of small fragments.

The domain knowledge in DFA describes and categorizes the essential and representative elements of the distributed database systems, specifically, for the databases fragmentation. The purpose of the DFA domain knowledge is to ensure that all data elements are available and consistent for database fragmentation process. In addition, it is used to prepare data elements that are valid from one transaction to another, from one application to another, and from one database to another in distributed database systems.

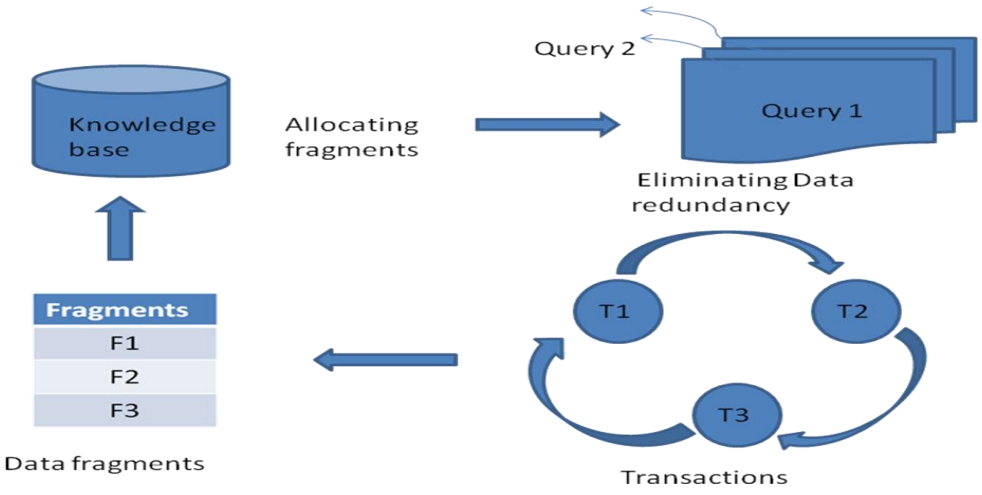

Fig 2.Proposed Data Fragmentation Architecture

The fragmentation process starts looking for any two data records over the same relation having intersection records between them. From any two intersected data sets, three disjoint fragments will be generated; the intersection fragment which represents the common records in both sets, the fragment that represents the records in the first set but not in the second intersected set, and the fragment that represents the records in the second set but not in the first intersected set. Then, the intersected sets are deleted from the data sets list. This process is continued until no more intersections between the data sets still exist. The subsequent fragmentation algorithm describes the processes of generating disjoint fragments from the intersected data records for each relation in DDBS.

VII. CONCLUSIONS

The use of multiple cloud providers for gaining security and privacy benefits is nontrivial. In this paper, we proposed a different 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. Data fragmentation is one of the primary techniques used in partitioning and developing cloud computing services for distributed database systems. This research discussed the efficiency, usefulness, and the performance improvement achieved by the API fragmentation service in a cloud computing system. The experimental results emphasized the ability of this fragmentation method to minimize the data processed and transferred between the distributed database system network sites, reduce the storage size by eliminating data redundancy, and present significant performance improvements that increase distributed database network system throughput. In our Future work we will implement the DFA to obtain Disjoint, Non redundant and non Intersected fragments.

ACKNOWLEDGMENT

I would like to place on record my deep sense of gratitude to **Dr. Deepti Sharma** Head of department of Computer Science and Engineering, A.I.T.M, Palwal, India for her generous guidance, help and useful suggestions. I express my sincere gratitude to **Mr.P.S Bishnoi** Principal of A.C.T.M Palwal India, for his stimulating guidance, continuous encouragement and supervision throughout the course of present work.

I also wish to extend my thanks to **Mr. Mahesh Singh** senior asst professor A.I.T.M for attending my seminars and for their insightful comments and constructive suggestions to improve the quality of this research work.I am extremely thankful to **Dr. R.S.Chaudhary** Director A.E.I Palwal, for providing Me infrastructural facilities to work in, without which this work would not have been Possible.

Tanvi Sharma

REFERENCES

[1] Cloud Computing Architectures Based Multi Tenant IDS Elmahdi Khalil , Saad Enniari and Mo tapha ZbakhAuthors.
[2] Cloud security issues Balachandra reddy kandukari,Ramakrishna Paturi V and Dr. Atanu Rakshit
[3] Security architecture for cloud computing Platform Sanjaya Dahal
[4] Amazon Simple DB. http://aws.amazon.com/simpledb/ [Accessed 19th February, 2011].
[5] F. Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D. A. Wallach, M. Burrows, T. Chandra, A. Fikes, and R. E. Gruber. Bigtable: a distributed storage system for structured data. Proceedings of OSDI, 2006.
[6] A.Velte, T.Velte & R.Elsenpeter. Cloud Computing: A Practical Approach. McGraw-Hill. 2010.
[7] Microsoft SQL Server for Cloud Servers. http://www.rackspace.com/cloud/blog/2010/12/01/announcing-sql-server-licenses-for-cloud-servers. [Accessed 7th March, 2011]. *FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.
[8] Multi-Cloud Architecture to Reduce Security Risks in Cloud Computing Vinod Kumar Paidi* , P.Varaprasada Rao
[9] http://blogs.idc.com/ie/wpcontent/uploads/2009/12/idc_cloud_challenges_2009.jpg
[10] http://www.microsoft.com/presspass/press/2010/jan10/1-20BrookingsPR.mspx
[11] D. Talbot. Security in the Ether. Technology Review, pages 36–42
[12] Hassan Takabi and James B.D.Joshi,Security and Privacy Challenges in Cloud Computing Environments, University of Pittsburgh, Gail-Joon Ahn,Arizona State University.
[13] http://www.cepis.org/index.jsp?p=641&n=825&a=4758#sthash.7AZ6fRvt.dpuf
[14] http://en.wikipedia.org/wiki/Cloud_computing]
[15] A Brief History of Cloud Computing by JamesSteddum in Cloud, Technology
[16] Cloud Computing - Concepts, Architecture and Challenges By Yashpalsinh Jadeja and Kirit Modi
[17] Cloud Computing Security: From Single to Multi-Clouds By Mohammed A. AlZain, Eric Pardede, Ben Soh and James A. Thom
[18] Security Framework of Cloud Data Storage Based on Multi Agent System