Addressing Data Security and Performance Issues on the Cloud

Shikha Banga
Department of Information Technology
Maharaja Agrasen Institute of Technology
Delhi, India

Abhishek Kapoor
Department of Information Technology
Maharaja Agrasen Institute of Technology
Delhi, India

ABSTRACT

Cloud computing has, with its advent, brought what IT always dreamt of adding capabilities on the fly without any new investments, thereby also reducing the cost. Be it an individual or a big organization, everyone is using cloud computing either knowingly or unknowingly. If we talk about the services and advantages of cloud computing, the benefits are astonishing like pay-as-you-go service, flexibility, accessibility, scalability, etc. but what is it that is still bothering the cloud providers and its customers? Some famous and prominent attacks indicate the lack of security in the cloud which ultimately leads to mistrust of customers toward the cloud provider. Although, cloud computing provides security to some extent, but the rising demand and use of big data require a solution to the poor security in the cloud. Confidentiality and privacy are the main reasons that people still hesitate to use the services of cloud in-spite of such promising technology and positive outcomes. For this purpose, we are proposing a technique in which the centralized data is fragmented based on its sensitivity and thereby improving the performance also.

General Terms

Cloud Computing, Data Security, Fragmentation

Keywords

Cloud computing; cloud security; data privacy; fragmentation; database-as-a-service; distributed computing; PostgreSQL.

1. INTRODUCTION

Cloud services are widely used in the computer industry and their features and effectiveness clearly testifies that they will be a driving force for businesses in the coming generations. The definition of cloud computing- cloud as a metaphor for internet and cloud computing as internet based computing- has become a cliché. The definition goes a lot bigger and broader. Some experts have defined cloud computing as utility computing or even distributed computing, so there are still many aspects of cloud computing yet to be determined. The main features of cloud computing are flexibility, elasticity, scalability, no dependency on location, reliability and sustainability [1]. The main technology around which the concepts of cloud computing revolve is known as virtualization [2].

There are three types of clouds, namely

1. **Public clouds:** Public clouds are open to the people for use and deployment, like Google and Amazon. The services like application and storage are available to the people through the internet [3].

2. **Private clouds:** Private clouds are built for the specific organizations so as to provide them a more secure environment in which only the specific clients can operate [3].

3. **Hybrid clouds:** Hybrid clouds are a combination of public and private cloud, in which users get the functionalities and features of both the cloud. Here, organizations can create their own cloud and can give control to others as well [4]. Our proposed method will make use of the hybrid model of clouds.

The cloud provides us with a variety of services, which include,

1. Software-as-a-Service (SaaS): SaaS provides applications that can be used directly without installing anything on the system, for example, Google Apps and Microsoft Office 365.

2. Platform-as-a-Service (PaaS): PaaS provides cloud application platform for the developers, like Windows Azure and Google App Engine.

3. Infrastructure-as-a-Service (IaaS): IaaS provides cloud infrastructure in terms of hardware, storage, memory, processor, etc. Its examples are Windows Azure, Amazon EC2, Google Compute Engine, etc. [5].

The services stated above, SaaS, PaaS and IaaS have some security issues of their own as well [6].

4. Database-as-a-Service (DBaaS): DBaaS provides database related services to the customers so that they do not need the database experts or the DBA i.e Database Administrator. In this paper, we will be working mainly on the DBaaS.

Cloud services are mainly used as a means for storing data. The customers can store the data on the cloud provider's servers and thus, they are free from the installation of their own. A major advantage here is that they can pay only for the amount of storage they need, which also saves their investment in the infrastructure [7]. Here comes the daunting task of providing efficient storage and security of the data because of the presence of various restrictions like 1. Restriction on the domain 2. Law restrictions 3. Providing suitable security mechanisms 4. Locality restrictions 5. Security level restrictions, etc. The customers are faced with security issues such as data security, locality of data, integrity, confidentiality and authenticity [8]. Usually, to protect the data, various cryptographic techniques are used which leads to a problem of heavy computational overhead and cypher-key management concerns [9]. So, the main concern of our research is to improve the security of clouds consequently enhancing the performance and efficiency of it [10].

1.1 Fragmentation

Fragmentation is a technique of partitioning relations into two or more sub-relations such that by combining the relations we get the original relation without any loss of information [11]. When these fragments are allocated at different sites in the hybrid cloud, the irrelevant data access is reduced to a much smaller amount, thus reducing the number of disk accesses. So, we can get a much faster and efficient system. Fragmentation is of three types- Horizontal, Vertical and Hybrid fragmentation.

1. Horizontal Fragmentation: If the relation is divided along the rows, i.e. horizontally such that all attributes are present in the subsets after fragmentation (like tearing a page horizontally), we call it horizontal fragmentation. If in a relation, a subset of rows is present at site1 and another subset is present at site2, original relation can be obtained by taking the union of the subsets from both the sites.

2. Vertical Fragmentation: When a relation is fragmented along the columns (like tearing a page vertically), it is called vertical fragmentation. The subsets of vertical fragmentation contain sets of columns with at least one attribute common to all the sets (primary key). The original relation can be obtained by taking the natural join of the subsets.

3. Hybrid Fragmentation: The combination of horizontal and vertical fragmentation is termed as hybrid fragmentation. Here, a relation is divided into blocks depending on the requirements. The fragments are allocated to a specific block and this is the most complex type of fragmentation. It consists of horizontal fragmentation followed by vertical fragmentation or vertical fragmentation followed by horizontal fragmentation. Mixed fragmentation provides more security to the data as compared to horizontal and vertical fragmentation.

Illustrating the concept of fragmentation with the help of an example given below-

Let’s say we have a customer relation with the following attributes, C_ID, NAME, AREA, PAYMENT_TYPE and SEX. Let’s say we have a customer relation with the following attributes, C-ID, NAME, AREA, PAYMENT_TYPE and SEX.

Table 1: Customer Relation

C-ID	NAME	AREA	PAYMENT_TYPE	SEX
001	SHIKHA	NORTH	CREDIT CARD	FEMALE
002	ABHI	EAST	CASH	MALE
003	MEGHA	SOUTH	CASH	FEMALE

Horizontal Fragmentation

1. **Fragment 1**

Table 2: Horizontal Fragment 1

C_ID	NAME	AREA	PAYMENT_TYPE	SEX
001	**SHIKHA**	**NORTH**	**CREDIT CARD**	**FEMALE**
002	**ABHI**	**EAST**	**CASH**	**MALE**

2. **Fragment 2**

Table 3: Horizontal Fragment 2

C_ID	NAME	AREA	PAYMENT_TYPE	SEX
003	MEGHA	SOUTH	CASH	FEMALE

Vertical Fragmentation

1. **Fragment 1**

Table 4: Vertical Fragment 1

C_ID	NAME	AREA	SEX
001	SHIKHA	NORTH	FEMALE
002	ABHI	EAST	MALE
003	MEGHA	SOUTH	FEMALE

2. **Fragment 2**

Table 5: Vertical Fragment 2

C_ID	PAYMENT_TYPE
001	CREDIT CARD
002	CASH
003	CASH

Hybrid Fragmentation

Table 6: Hybrid Fragment

<table>
<tr><th>C_ID</th><th>NAME</th><th>AREA</th><th>PAYMENT_TYPE</th><th>SEX</th></tr>
<tr><td colspan="4">Fragment1</td><td>Fragment2</td></tr>
<tr><td colspan="2">Fragment3</td><td colspan="3">Fragment4</td></tr>
</table>

2. ISSUES IN CLOUD

Information ranging from general to sensitive and from small companies to big organizations is being stored in the cloud. Considering its benefits, people are more likely to use its services in the future. Before completely relying on it, it is very important to know about the issues related to cloud computing [12]. Some of them are stated as follows-

1. Security Issues: There are many security issues related to cloud computing like data loss, data breaches,

Phishing and botnet that poses serious threats to the data stored in the cloud. Many users of cloud share the same network address which makes the data vulnerable to many attacks [13]. Moreover, cloud computing utilizes virtualization in which user's data is scattered in many locations rather than staying at a single physical location. This increases the threat of data leakage. Under security, features like confidentiality, availability, integrity, control, audit, etc. are to be provided by the cloud vendors [14].

2.Long-term availability: The cloud providers are also companies or organizations like Google, Amazon, et al. These cloud providers should ensure the long-term existence of our data and available on demand in the case of company collaboration, disintegration or other inside issues.

3.User Control: The storage of data is done by the cloud providers only; the role of users is restricted here. The users should be given control over the data so that they are aware of the way in which the data is stored and protected by the vendors.

4.Reliability: Although, the organizations is providing their best services to their cloud customers, still at times shutdowns or slowdowns occur with clouds thereby restricting the services.

5.Legal Issues: The ease of use and benefits of cloud like pay-as-you-go, scalability, access anywhere, et al makes it look much simpler and easier, but behind the scenes, there are many legal issues to resolve and under these legal issues, organizations need to know the data location, especially from a legal standpoint [15].

Like these, there are many voids that the organizations are trying to fill so as to build a good customer-relationship, win their trust and come to the leading positions in the competitive cloud market. These issues are important and must be addressed if a good quality cloud service is expected to be provided to the customers by the cloud-vendors.

3. PROPOSED METHOD

This paper presents a safe and efficient method of data storage in the cloud environment using fragmentation. Our method uses minimum encryption so as to reduce mathematical computations. We have used horizontal fragmentation in our technique to improve data security in clouds. The data to be stored in the cloud is partitioned among various sites depending on the locality of precedence of the attributes in a relation [11]. The level of security is provided according to the nature of the data which is of two types-

1.General

The type of data which poses little or no threat to the organization's reputation, resources or individuals, are known as general data. General data require the lowest level of protection. It can be conveniently stored in the public clouds.

No fragmentation is required in this case as any complicated procedure will only exceed the present costs of the system while providing futile results.

2.Sensitive

The type of data on which any unauthorized disclosure or access may have serious adverse effect on the organization's reputation, resources, services, or individuals are called as the sensitive data. It includes data protected under certain regulations, or due to proprietary, ethical, or privacy considerations. Sensitive data require the highest level of protection [16]. This data needs to be secured using techniques like encryption, fragmentation and normalization.

3.1 Methodologies for storing data:

3.1.1 Without Fragmentation

1. **Public cloud:** In public clouds, many clients share the same infrastructure, for example, well known public cloud storage platform like Dropbox and Google Drive. Public cloud services are much use for the people who do not need a high level of security. The resources offered in the public clouds can be stretched as per the usage and requirement. This feature of unlimited scalability makes them affordable and cost effective for the clients. This is a benefit for the small business owners as they can utilize the functionality of cloud in an affordable manner [17].

2. **Private Cloud:** Private cloud provides security at a higher level as compared to the public cloud, so it can be used in businesses where security is a priority. Another advantage in using private cloud is that they can be customized according to the needs of the user, so it also offers some control over the server. Thus, it becomes difficult to access the data stored in private clouds from remote locations. Therefore, the initial costs involved in setting up the private clouds can make it difficult for the small business owners to adopt it [18].

3. **Hybrid cloud:** It allows us to protect the sensitive data while maintaining a lower cost and flexibility of the cloud. The resource allocation to the cloud is much lower as compared to the alterations that would have been required for the infrastructure changes.

The major drawback of hybrid cloud implementation is the dependency on IT infrastructure.It is also difficult to maintain compliance between the public and private cloud providers. Also, there is a need to ensure that the transferred data are protected from fraud or malpractices [19].

3.1.2 With Fragmentation

1. Public Cloud: If data is fragmented in the public clouds, it will be placed at distinct places, but still open for everyone because clients are sharing the same infrastructure. The security won't increase too much in this process and it will also make public clouds quite expensive. It's like providing semi-security of the data, but can be done if the data is of moderate quality, i.e. it does not require high protection because using public clouds will also save costs.

2. Private Cloud: Private Clouds store the data that is sensitive, i.e. it needs extra protection. Fragmenting the data on private clouds can enhance the privacy and confidentiality of data. This will add up to the original cost but will improve the security. Data availability is a problem when storing the data on private clouds and it's also cost ineffective to store all the data in private clouds.

3. Hybrid Cloud: Hybrid Clouds can store the data separately in public clouds and private clouds depending upon the level of security required by the data. This process can easily reduce the costs as the general data can be stored in public clouds and sensitive data can be stored and fragmented in private clouds. Data availability will not be affected much due to the arrangement of data in public and private clouds.

The technique used for fragmentation makes use of the nature of data before applying fragmentation on it. If the data is highly critical, we need to provide fragmentation and encryption on the data before storing it on the clouds. If the data is of moderate nature, i.e. not too confidential, we need to apply only fragmentation and then store it in the proper sites in the clouds. If the data is general, there is no need to apply any kind of fragmentation or encryption because there is no threat to the leakage of data.

3.2 Proposed Algorithm

The algorithm takes in the relations to be fragmented and then partition it according to the attributes with respect to which fragmentation of the relation is required (All this has been implemented in PostgreSQL).

The algorithm is given below:

Given: A relation R, the set of predicates P.

Output: The set of fragments of R, FR = {R1, R2,…, RN} that follows the fragmentation rules.

Step 1.) Construct the relation R with the required attributes.

Step 2.) Construct the child tables for the relation R according to the attribute with respect to which fragmentation is required.

Step 3.) Form primary keys for the child tables.

Step 4.) Form indexes for the child tables.

Step 5.) Create triggers for the child tables.

Step 6.) Allocate values to the relation R.

Step 7.) Allocate the fragments R(R1, R2, R3,….) to different sites based on the priority of the attributes at the particular site.

To illustrate the above algorithm, we have taken the following example. We have taken two relations without_partition and with_partition. In this example, we want to fragment the data according to the months in a year.

Without_partition is the relation in which fragmentation is not performed whereas with_partition is the table on which fragmentation is performed. Both relations are required so as to check the performance difference between the two.

1. The with_partition table is created with the following attributes-

Create table with _partition(name bytea not null, wtime timestamp without time zone not null);

2. Next, we will inherit tables from with_partition-

Create table with_partition _y2000m01 (CHECK (wtime >= DATE '2000-01-01' AND wtime < DATE '2000-01-31')) inherits (with_partition);

Create table with_partition_y2000m02 (CHECK (wtime >= DATE '2000-02-01' AND wtime < DATE '2000-02-29')) inherits (with_partition);

………………… are created for the whole year.

3. Now, we make primary keys for the child tables

Alter table without_ part_y2000m01 add constraint without_part_y2000m01k primary key (wtime, name);

Alter table without_ part_y2000m02 add constraint without_part_y2000m02k primary key (wtime, name);

…………… are created for the whole year.

4. Next, we have to create indexes for the child tables

Create index with_part_y2000m01_id on with_part_y2000m01 (wtime);

Create index with_part_y2000m02_id on with_part_y2000m02 (wtime);

…………are created for the whole year.

5. The final step will be to create triggers so as to direct data in child tables (i.e to activate partitions)

If (new.wtime>=2000-01-01 and new.wtime<2000-01-31) then insert into with_partition_y2000m01 values (new.*);

elseIf (new.wtime>=2000-02-01 and new.wtime<2000-02-28) then insert into with_partition_y2000m02 values (new.*);

………. So on for the whole year.

6. The with_partition table is now loaded with data.

7. The without_partition table can be made using the same attributes and inserting the same number of rows. Also, in without_partition, we do not have to perform any inheritance from the mother table. We do not need to create any child tables as created in with_partition table. Only a single table is created and data is inserted into it [20]. Now, to observe the performance difference between the two databases, we can execute similar queries and notice the runtime difference between the two (All the simulations are performed on PostGreSQL) [6].

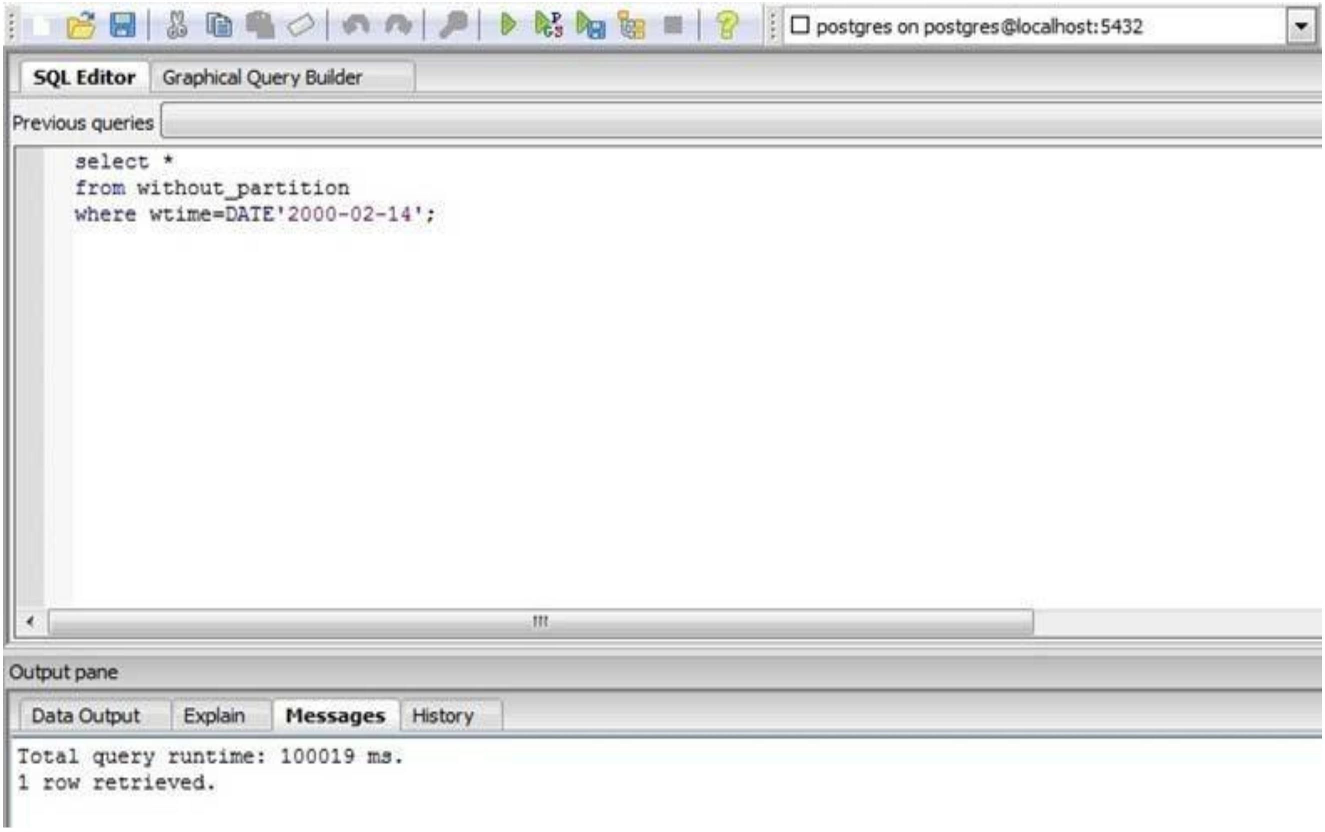
8. A simple SQL query such as

Select * from without_partition where wtime=DATE '2014-01-14';

And Select * from with_partition where wtime=DATE '2014-01-14';

It is sufficient to observe the performance difference.

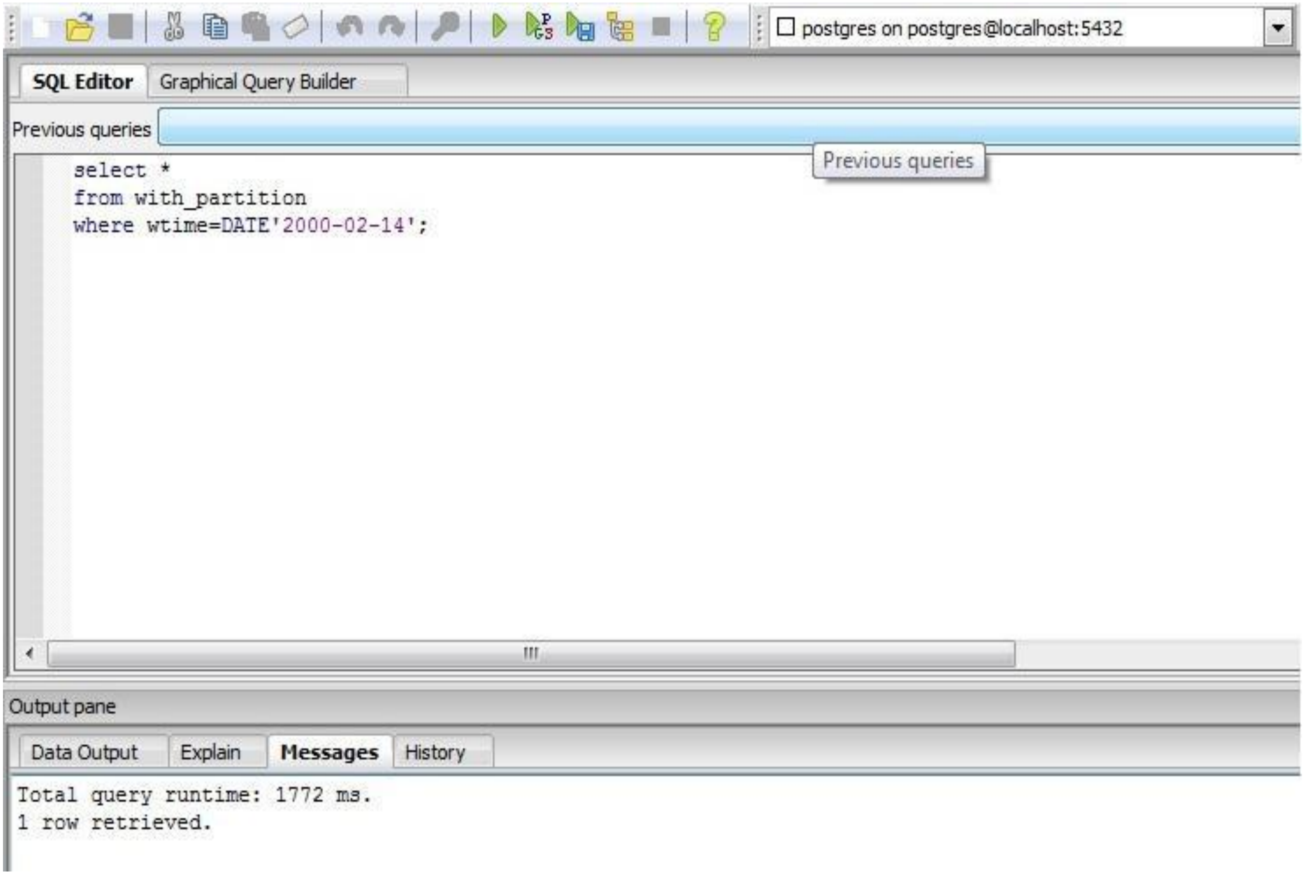
Result: The execution runtime of the first query is around **56.4 times** that of the second query. This performance difference is so large that it can even make the execution of complex queries unfeasible in the case of highly populated databases. The first query that involves the non-partitioned table takes **100019 msec** whereas the second query that makes use of the partitioned table takes only **1772 msec** as depicted in the screenshots below. In this case, we have populated both the tables with 2 million random records to simulate the runtime difference between the queries. With such a large amount of data, the use of non-fragmented tables becomes highly inefficient. So, the proposed technique reduces the query runtime and increases the system performance, especially when big data is involved.



```
select *
from without_partition
where wtime=DATE'2000-02-14';
```

```
Total query runtime: 100019 ms.
1 row retrieved.
```

Figure 1: Screenshot of PostgreSQL running a simple query on without_partion relation.



```
select *
from with_partition
where wtime=DATE'2000-02-14';
```

```
Total query runtime: 1772 ms.
1 row retrieved.
```

Figure 2: Similar query run on with_partion relation

3.3 Advantages of Horizontal Fragmentation

1. It allows parallel processing on the data.

2. The Splitting of a relation facilitates its location at a place where it is required the most.

3. Increase in query performance as evident from above.

4. Increase in granularity.

5. Increase in security and confidentiality of the data.

For simplicity purpose, we have taken only two relations as an example here and shown fragmentation and its effects on only one relation. The data are also added in a random manner into the relations. When working practically with the clouds and databases, we will take the actual data into the database; fragment the required tables according to the level of privacy required and the nature of data present. These fragments are stored at appropriate locations in the cloud where they are required the most, so that accessing time and cost, both are minimized.

The above described fragmentation model is depicted with the help of a table and diagrams given below. The table shows some information related to the data like the name of the data, its location and its sensitivity level. The diagrams describe its hybrid architecture with the help of illustrating data in public and private clouds and finally the fragments are allocated to the requisite location (sites).

Table 7: Data Sensitivity Level

S.No	Data	Location	Sensitivity Level
1	D1	Site 2	1
2	D2	Site 3&1	2
3	D3	Site 1	3
4	D4	Site 4&3	4

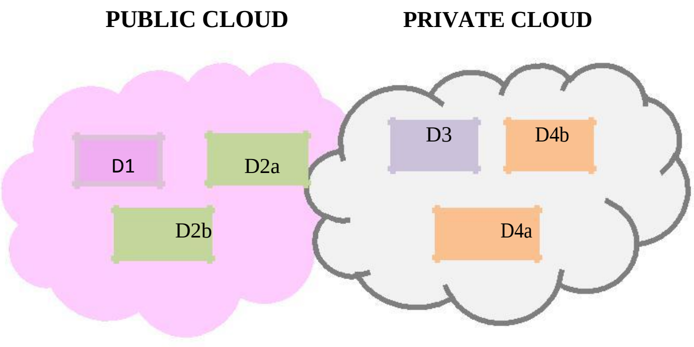


Figure 3: Hybrid Architecture

In the above figure, the data that requires the lowest level of security like D1 is kept in the public cloud without performing any fragmentation (because it is the general data) and the data D2 that requires moderate security is kept in the public cloud, but fragmentation is provided to it so as to give it the requisite amount of protection (D2a and D2b represents the fragmented parts of the data D2) . Finally, the data that requires a high level of security like D3 and D4 are stored in the private cloud. D4a and D4b represent the fragmented parts of the data D4.

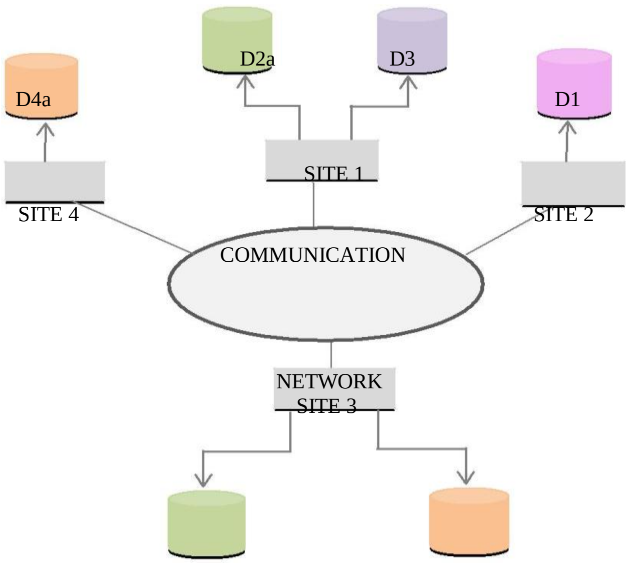


Figure 4: Distributed Model

As we know, the data is fragmented depending upon its security level. Now, referring to the table 7, the fragments or the complete data are allocated to the site where maximum access of the data is required. D1and D3 are allocated to site 2 and site 1 respectively. The fragmented parts D2a and D2b are allocated to Site 1 to Site 3 respectively. Finally, the highest confidential fragmented data, D4a is allocated to site 4 and D4b to Site 3.

4. CONCLUSION

In this paper, we have proven the benefits of fragmentation in cloud computing. First of all, The nature of data is important to judge before performing any technique. The data is segregated according to its quality and demand of the customer. After deciding the category of data, the technique is decided based on the amount of security it needs so as to avoid any wastage in costs and resources. Less vulnerable data should be kept on the public clouds because of the less threat to its leakage. But, if the data is confidential, proper planning and technique should be performed so as to ensure security. It could be fragmented and even encrypted (if, required) depending upon the desired amount of privacy. After performing fragmentation based on the technique discussed above, the partitioned data are placed at different locations or sites in the cloud environment. The data is located on the basis of its usage and priority at a particular site. The data is stored in those locations where it is frequently used to gain maximum advantage through fragmentation. This technique provides us many benefits, such as the efficiency in speed because of the low run time cost, parallel processing, et al. Although, the use of such procedures may inculcate additional costs in the system, but the benefits obtained after following this technique will improve the security of the system. This technique described above, if implemented efficiently, will reduce the cost drastically, because restoring lost or breached data results in high levels of wastage, both in cost and resources - if implemented at a later stage. Our method also improves the query performance, which helps in saving query processing time and building a good relationship with the customer.

The described algorithm gives the desired amount of safety to the data stored in the clouds. It makes use of fragmentation and Database-as-a-Service. The performance improvement, one of the key factors in grading any algorithm, is good and will increase as more data is added to the system. The results are accomplished and implemented in PostgreSQL. This paper is based on the increasing expectations of the consumers for the security of their data while using the cloud platform.

This work can be extended to implement a hybrid type of fragmentation technique that includes the usage of both horizontal and vertical fragmentation. Furthermore, other security techniques like encryption, normalization, etc. can be used in conjunction with fragmentation, so as to provide a better level of security on the cloud. Various measures should be taken in this field and new algorithms can be developed, so as to address data security and performance issues on the cloud.

5. ACKNOWLEDGMENTS

We would like to thank the IT Department of our institution, Maharaja Agrasen Institute of Technology, which provided us with all the resources and help during our research. Also, we are grateful to the anonymous reviewers for their valuable comments and helpful suggestions.

6. REFERENCES

[1] Dimitrios Zissis and Dimitrios Lekkas, Addressing cloud computing security issues, Elsevier-Future Generation Computer Systems, 2012.

[2] Greg Boss, Padma Malladi, Dennis Quan, Linda Legregni and Harold Hall, Cloud Computing, IBM, 2007.

[3] Introduction to Cloud Computing Architecture, White paper, First edition, Sun Microsystems, 2009.

[4] Marios D. Dikaiakos and George Pallis, Cloud Computing-Distributed Internet Computing for IT and Scientific Research; IEEE Internet Computing, 2009.

[5] Olivier Brian, Thomas Brunschwiler, Heinz Dill, et al., Cloud Computing; Swiss Academy of Engineering Sciences, 2012.

[6] Mohamed Al Morsy, John Grundy and Ingo Müller, An Analysis of The Cloud Computing Security Problem; Swinburne University of Technology, 2010.

[7] Kevin Curran, Sean Carlin and Mervyn Adams, Security Issues in Cloud Computing, Elixir International Journal, 2011.

[8] S.Subashini and V. Kavitha, A survey of security issues in service delivery models of cloud computing, Elsevier-Journal of Network and Computer Applications, 2011.

[9] Aleksandar Hudic, Shareeful Islam, Peter Kieseberg and Edgar R. Weippl, Data Confidentiality using Fragmentation in Cloud Computing, International Journal of Communication Networks and Distributed Systems, Vol. 1 ,2012.

[10] Robert Taylor, Query Optimization for Distributed Database Systems, University of Oxford, 2010, Unpublished.

[11] P. R. Bhuyar, A. D. Gawande and A. B. Deshmukh, Horizontal Fragmentation Technique in Distributed Database, International Journal of Scientific and Research Publications, vol. 2, 2012.

[12] Santosh Kumar and R. H. Goudar, Cloud Computing – Research Issues, Challenges, Architecture, Platforms and Applications, International Journal of Future Computer and Communication, Vol. 1, 2012.

[13] Kevin Curran, Sean Carlin and Mervyn Adams, Security Issues in Cloud Computing, IGI Global, 2012.

[14] Amir Mohamed Talib, Rodziah Atan, Rusli Abdullah, Masrah Azrifah and Azmi Murad, Towards a Comprehensive Security Framework of Cloud Data Storage Based on Multi-Agent System Architecture, Journal of Information Security, 2012.

[15] searchcloudsecurity.techtarget.com/, Last accessed: -February 2014.

[16] Oscar Diez and Andrés Silva, Using Cloud Computing in Public Organizations, IEEE Technology and Society Magazine, 2013.

[17] http://bigideasblog.infusionsoft.com/public-and-private-cloud/

[18] exploreb2b.com/articles/pros-and-cons-of-hybrid-cloud

[19] www.idganswers.com/question/3551/what-are-the-main-advantages-disadvantages-of-a-hybrid-cloud-model-vs-the-public-cloud

[20] www.MKyong.com, last accessed March 2014.