

# **An Efficient Data Replication and Disaster Recovery Technique for Cloud Data**

**Pradeep Fale<sup>1</sup>, Punam Wadhi<sup>2</sup>, Chetana Ghonge<sup>3</sup>, Smita Chauhan<sup>4</sup>**

<sup>1</sup>*Assistant Professor, Priyadarshini College of Engineering,  
Nagpur, India, 440019*

<sup>2,3,4</sup>*B.E. student, Dept. of Information Technology, Priyadarshini College of Engineering,  
Nagpur, India, 440019*

**Abstract** – Salesforce is one of the largest cloud computing technology, which is available on cloud, no need to install any software for it. Now a day's cloud computing is fastest growing technology. It is observed that businesses and individuals are moving their data to the cloud because fault-tolerant data storage is becoming more important. As large number of data nodes used in the cloud storage system, the probability of some data nodes failure increases. In our paper we use data replication strategies to increase the availability and reliability of data in cloud. For this we integrate the Salesforce environment with the Amazon Web Service (AWS) cloud where we stores the replica of the particular data for reliable access.

**Keywords-** Salesforce, Cloud computing, Data replication, AWS cloud.

## **INTRODUCTION**

Salesforce is a fastest growing and very demanded emerging technology, that provided services like Infrastructure for organization that service comes under IaaS (Infrastructure as a Service) as well as it provides PaaS (Platform as a Service) for building an Application. Also for online service we don't need to install particular software because this cloud provide a SaaS (Software as a Service). Salesforce make CRM easy to use for small business as well as large business. That approach makes Sales Cloud so much popular in the world. The CRM is not only one service they provide; they also provide platform for building fully native business to business (B2B) application as well as business to customer (B2C) application. Salesforce provides different cloud for setup and management of

business the clouds are: Sales Cloud, Service Cloud, Marketing Cloud, IoT Cloud, and Community Cloud & App Cloud.

Cloud computing has been a huge-scale parallel with shared computing system. In cloud computing, data generated in electronic form are large in amount. To maintain this data efficiently, there is a necessity of data recovery services.

Cloud computing has been envisioned as the next-generation architecture of IT enterprise. In contrast to traditional solutions, where the IT services are under proper physical, logical and personnel controls, cloud computing moves the application software and databases to the large data centers, where the management of the data and services may not be fully trustworthy.

This unique attribute, however, poses many new security challenges which have not been well understood. In this project, we focus on cloud data storage security, which has always been an important aspect of quality of service

However, the Cloud environment constitutes a heterogeneous and a highly dynamic environment. Failures on the data centers nodes are normal rather because of the large scale of physical resources and data.

As a result, the cloud environment requires an efficient adaptive data replication management in order to cope with the inherent characteristic of the Cloud environment. In this paper, we propose a data replication strategy which adaptively selects the data

files for replication in order to improve the overall reliability of the system and to meet the required quality of services. Further, the proposed strategy decides dynamically the number of replicas as well as the effective data nodes for replication. The popular data files are selected for replication based on employing a lightweight time-series technique, which analyzes the recent pattern of data files requests, and provides predictions for the future data requests.

To cater this, The objective of proposed algorithm is twofold; first it help the users to collect information from any remote location in the absence of network connectivity and second to recover the files in case of the file deletion or if the cloud gets destroyed due to any reason. The time related issues are also being solved by proposed Algorithm such that it will take minimum time for the recovery process.

As we know Salesforce cloud is more popular in world but it has some disadvantage like it store very limited data, hence to overcome this we use the AWS cloud by integrating it with the Salesforce environment. Amazon Web Service (AWS) cloud is provided by the Amazon. It store large amount of data. We can easily do the replication in Amazon Web Service cloud. Its cloud is located in different countries. We basically use the two different countries cloud for storing the replicas of our data.

**METHODOLOGY**

We propose an adaptive replication strategy in a cloud environment that adaptively copes with the following issues:

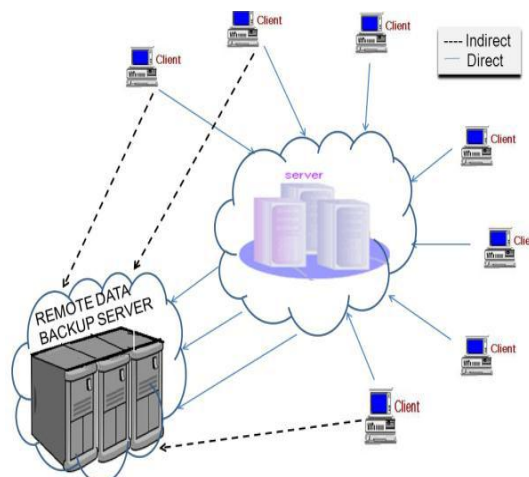
- What to replicate to improve the non-functional QoS. The select process is mainly depends on analyzing the history of the data requests using a lightweight time-series prediction algorithm. Using the predicted data request, we can identify what data files need replication to improve the system reliability.
- The number of replicas for each selected data.
- The position of the new replicas on the available data centers.
- The overhead of replication strategy on the Cloud infrastructure. This is the most important factor of the proposed adaptive replication strategy where the Cloud has a large number of data centers as well as a large-scale data.

Hence, the adaptive replication strategy should be lightweight strategy.

The proposed adaptive replication strategy is originally motivated by the fact that the recently most accessed data files will be accessed again in the near future according to the collected prediction statistics of the files access pattern. A replication factor is calculated based on a data block and the availability of each existing replica passes a predetermined threshold, the replication operation will be triggered. A new replica will be created on a new node which achieves a better new replication factor. The number of new replicas will be determined adaptively based on enhancing the availability of each file heuristically. However, we employ a lightweight time-series algorithm for predicting the future requests of data files. The replication decision is primarily based on the provided predictions. The heuristic proposed for the dynamic replication strategy is computationally cheap, and can handle large scale resources and data in a reasonable time.

**DESIGN**

Remote Data Backup server is a server which stores the main cloud’s entire data as a whole and located at remote place (far away from cloud). And if the central repository lost its data, then it uses the information from the remote repository. The purpose is to help clients to collect information from remote repository either if network connectivity is not available or the main cloud is unable to provide the data to the clients. As shown in Fig 1, if clients found that data is not available on central repository, then clients are allowed to access the files from remote repository (i.e. indirectly).



The Remote backup services should cover the following issues:

- 1) Privacy and ownership.
- 2) Relocation of servers to the cloud.
- 3) Data security.
- 4) Reliability.
- 5) Cost effectiveness.
- 6) Appropriate Timing.

1) Privacy and ownership

Different clients access the cloud with their different login or after any authentication process. They are freely allowed to upload their private and essential data on the cloud. Hence, the privacy and ownership of data should be maintained; Owner of the data should only be able to access his private data and perform read, write or any other operation. Remote Server must maintain this Privacy and ownership.

2) Relocation of server

For data recovery there must be relocation of server to the cloud. The Relocation of server means to transfer main server's data to another server; however the new of location is unknown to the client. The clients get the data in same way as before without any intimation of relocation of main server, such that it provides the location transparency of relocated server to the clients and other third party while data is been shifted to remote server.

3) Data security

The client's data is stored at central repository with complete protection. Such a security should be followed in its remote repository as well. In remote repository, the data should be fully protected such that no access and harm can be made to the remote cloud's data either intentionally or unintentionally by third party or any other client.

4) Reliability

The remote cloud must possess the reliability characteristics. Because in cloud computing the main cloud stores the complete data and each client is

dependent on the main cloud for each and every little amount of data; therefore the cloud and remote backup cloud must play a trustworthy role. That means, both the server must be able to provide the data to the client immediately whenever they required either from main cloud or remote server.

5) Cost effectiveness

The cost for implementation of remote server and its recovery & back-up technique also play an important role while creating the structure for main cloud and its correspondent remote cloud. The cost for establishing the remote setup and for implementing its technique must be minimum such that small business can afford such system and large business can spend minimum cost as possible.

6) Appropriate Timing

The process of data recovery takes some time for retrieval of data from remote repository as this remote repository is far away from the main cloud and its clients. Therefore, the time taken for such a retrieval must be minimum as possible such that the client can get the data as soon as possible without concerning the fact that remote repository is how far away from the client.

## CONCLUSION

This paper proposed efficient data replication for handling the situation of data lost with the help of Salesforce environment. In first section of paper we explained about what is the Salesforce. After that we describe the concept of cloud computing and how its store the data. We mainly focus upon the data replication concept for the disaster recovery technique with the help of AWS cloud. According to the need, performance, cost we store the replicas on different data center. Different issues like privacy and ownership, reliability, security, cost effectiveness, appropriate timing, etc. are overcome by applying proper strategy. In this way we successfully give the strategy that give desire solution for the fault-tolerant data.

## ACKNOWLEDGMENT

The authors thank all reviewers for improving this paper. We thank our college Priyadarshini College of Engineering, Nagpur who provided insight and expertise that greatly assisted the research .

#### REFERENCES

- [1] *M.Amutha1, R.Sugumar, "A Survey on Dynamic Data Replication System in Cloud Computing", International Journal of Innovative Research in Science, Engineering and Technology*
- [2] *Rakesh Kumar1, Yougeshwary Sharma2, Sonu Agarwal3, Pragya4, Bhanu Bhushan Parashar5, "Extremely effective CRM Solution Using Salesforce", JETIR (ISSN-2349-5162)*
- [3] *Krutarth Soni, Brijesh Vala, "Roadmap to Salesforce Security Governance & Salesforce Access Management"*
- [4] *Rajkumar Buyyaa, Chee Shin Yeo, Srikumar Venugopala, James Broberga, Ivona Brandicc, "Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility", Future Generation Computer Systems, Vol. 25, No. 6, pp. 599–616, 2009.*
- [5] *Vijaya -Kumar-C and Dr. G.A. Ramachandhra, "Optimization of Large Data in Cloud computing using Replication Methods", International Journal of Computer Science and Information Technologies, Vol. 5, No. 3, pp. 3034-3038, 2014.*
- [6] *Tehmina Amjad, Muhammad Sher and Ali Daud, "A survey of dynamic replication strategies for improving data availability in data grids", Future Generation Computer Systems, Vol. 28, pp. 337–349, 2012.*
- [7] *<http://www.salesforce.com/in/what-is-salesforce/>*
- [8] *Salesforce CRM Security Audit Guide, White Paper, Salesforce*
- [9] *Rakesh Kumar, Neha Gupta, Shilpi Charu, Kanishk Jain, Sunil Kumar Jangir" Open Source Solution for Cloud Computing Platform Using OpenStack"; IJCSMC, Vol. 3, Issue. 5, May 2014, pg.89 – 98; ISSN 2320–088X*
- [10] *Salesforce Tutorial: <http://salesforcetutorial.com/>*