Load Balancing with Disaster Recovery using Multi Cloud

Radha[1], Dr. Rekha Patil[2]

[1]Department of Computer Science and Engineering M.Tech (CSE), Poojya Doddappa Appa College of Engineering, Gulbarga, Karnataka, India

[2]Department of Computer Science and Engineering, Poojya Doddappa Appa College of Engineering, Gulbarga, Karnataka

Abstract: ***Several trends are opening up the era of Cloud Computing, which is an Internet-based development and use of computer technology. Moving data into the cloud offers great convenience to users since they don't have to care about the complexities of direct hardware management. The problem in single cloud is that, it takes more time to recover the lost data. In our proposed work, Load balancing with disaster recovery using multi cloud where we can reduce the time and space. In multiple cloud based disaster recovery service, resources of multiple cloud service providers can be utilized cooperatively. A simple and unified interface is exposed to the end users to adapt the cloud service providers and the internal process between the clouds is made invisible to users. Disaster recovery Cloud proposes a priority scheduling strategy to balance the disaster recovery objectives, such as, high data reliability, low backup cost and short recovery time which are transparent to the users. Further proposed work also focused on providing load balancing which avoids overloading of single resources. It also provided security by using Email option. Using this scheme recovering the data and providing security is improved compared with the existing system.***

Keywords: Cloud storage, backup, disaster recovery and replication

1. Introduction

Cloud computing is an attracting technology in the field of computer science. The cloud is changing our life by providing users with new types of services. Users get service from a cloud without paying attention to the details. More and more people pay attention to cloud computing. A short period of down time or small data loss could result in huge economic loss. Therefore important techniques is used to protect the important data[8]. Building and running a data center for disaster recovery is quite time-consuming and costly i.e,cost of data center includes buying infrastructure, maintenance, servers & human resource, but there is no difference in cost whether the service is in use or it is stand by, and it requires huge investment [9]. As we know Cloud computing is becoming more popular and large number of services are built and these utilizes various resources of cloud platform with pay as you use model[10].The cost of cloud resources which is used to perform data disaster recovery is also called as disaster recovery as a cloud service which requires less cost compared to the cost of building and maintaining recovery data in its own datacenter[11]. Rapid development in cloud computing is motivating more industries to use variety of cloud services. However , many security challenges and issues have been raised such as, risk management, trust & recovery mechanism which needs to be taken into account to provide business continuity and better user satisfaction. Disaster can be natural (like Tsunami and Earth quake), hardware/software failures or manmade (human error), which leads to expensive service disruption. Cloud based recovery solution is an increasing trend because its ability to tolerate disasters and achieve high data reliability and availability, so the popular distributed storage systems are used in cloud platform such as AmazonS3 [12], GoogleFileSystem [13] and HadoopFileSystem [14] by default each of these have adopted three replicas redundant data mechanism. In our proposed work the disaster recovery service utilizes multiple data centers from the different cloud service provider. In this method it it keeps a data backup to its own data center and also to the other cloud service providers data centers. By using data backup and recovery procedures, achieved better effect and minimizes the cost of service quality. Customers deal with cloud using common and simple interface. Also focuses on security and load balance which avoids the overloading of any single resources.

2. Organization

The paper is organized as: section 1 discusses the introduction, section 3 discusses related work, Section 4 discusses the proposed work. Section 5 discusses results & snapshots Finally, section 6 presents conclusion.

3. Related Work

This section mainly focuses on the literature review that has been carried out for this work. Cloud computing is Internet based technology where the users can subscribe high quality of services from data and software that resides solely in the remote servers. This provides many benefits for the users to create and store data in the remote servers thereby utilizing fewer resources in client system.

Kave Eshghi,Darrell.D.E and Mark.L[1]the author proposed Extreme Binning, a scalable de-duplication technique for non-traditional backup workloads that are madeup of individual files with no locality among consecutive files in a given window of time. Extreme Binning exploits file similarity instead of locality and makes only one disk access for chunk lookup per file ,which gives reasonable throughput. Each file is allocated using a stateless routing algorithm to only one node, allowing for maximum parallelization and each backup node is autonomous with no dependency across nodes. **Rajiv.R and RodrigoN.C**[2] the author has noticed the problem that do not support dynamically co-ordating load distribution among different cloud based data centers .To overcome this problem the author proposed a federated cloud computing environment that that facilitates scalable provisioning of application services, QoS target and various workload, resources and network condition. The overall goal creates a computing environment that supports dynamic expansion for handling variations in service. **Jih-sheng.C and Ruay shiung.C**[3] the author proposed a novel replica consistency decision model which is aimed to lower the access delay as well as saving network bandwidth by duplicating original data in distributed manner by using Naïve Bayesian classifier in order to improve the system performance in data grids.

Prakhar.Y,Dhirajsingh.T[4]the author proposed ALG-de-duplication, which is an application aware local-global source de-duplication scheme for cloud backup in the personal environment to improve the de-duplication efficiency and it is also designed to exploit file semantics to minimize computational overhead and maximize de-duplication effectiveness using application awareness. It combines both local de-duplication and global de-duplication to balance the effectiveness and latency of de-duplication. **Windsor.W.H, Alan jay.s,Honesty.C.Y**[5]the author proposed automatic locality-improving storage which is a introspective storage system that automatically reorganizes selected disk blocks based on the dynamic reference stream to increase effective storage performance. **Alysson.B, Miguel.C,Paulo.S**[6]the author proposed DepSKY i.e, dependable & secure storage in a cloud-of-clouds, a system that improves the availability, integrity & confidentiality of information stored in the cloud through encryption, encoding and replication of the data on diverse clouds that forms a clouds-of-clouds. **Bo Mao, Yinjin . Li, B. He, Q. Luo, and Y. Ke.** [7]the author has noticed the problem on the storage systems in the cloud environment. The data de-duplication technology has been demonstrated to be very effective in shortening the backup window and saving the network bandwidth and storage space in cloud backup, archiving and primary storage systems such as VM platforms. To address this problem, author proposed SAR, an SSD Assisted Restore scheme, that effectively exploits the high random-read performance and low power-consumption properties of SSDs and the unique data sharing characteristic of de-duplication-based storage system by storing in SSDs the unique data chunks with high reference count, small size and non-sequential characteristics. **M.Vrable,S.Savage**[15] the author focused on backup file system to the cloud storage by using the Least-common-denominator method on the cloud interface and also supports many different kinds of cloud services. But here the author uses only one cloud to maintain one backup and focused on the local file system mechanism not on the cloud platform. **T.Wood, H.A.Lagar**[16] the author overcomes the effects of speed of light delays by using the pipelined synchronous replication mechanism. This mechanism is designed to increase the throughput and reduce the response time while it provides zero data loss consistency. But it uses only one cloud to store the data replicas. **T.Nugyen, A.Cutway**[17] the author proposed differentiated replication strategy which can handle customers /users different requirements. The core of this work is mainly focused on, firstly the differentiated services for data center and designed simple but powerful APIs for high level users. Secondly proposed four different replication strategies on the server side and thus enabling differentiated services in terms of data availability. Third implemented the prototype for differentiated replication and evaluates the replication strategies in terms of availability using synthetic and failure traces. The strategy provides data reliability assurance for different service types and differentiated backup schemes. It also enhances the utilization of cloud resource**. D.Bermbach** [18] the author focused on meta storage system, a federated architecture that utilizes diverse cloud storage providers. Meta storage implements a replication scheme based on Amazon's Dynamo, but elevates concepts to network storage providers. Meta storage focused on the data consistency and communication latency of data replicas among multiple cloud providers. They established data access priority queue to reduce the communication delay between clouds and the smaller the file size, better the assurance of data consistency. Meta storage increases the over availability compared to individual providers. **C.Cachin,R.Haas**[19]the author analyzed the data integrity, data security, confidentiality and consistency. The goal here is to build a more dependable cloud services and system in the inter cloud storage. So they adopted fault tolerant and secure access control protocol to ensure data integrity and confidentiality in multi cloud platforms.

4. Proposed Work

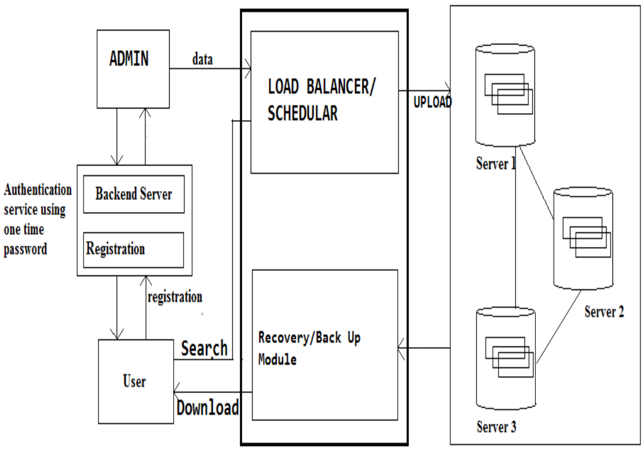

Figure 1: Block Diagram of Load balancing with disaster recovery using multi cloud Architecture

The system architecture of multi cloud based disaster recovery with load balancing is shown in Fig1. It consists of three entities as user/client, Administrator (ADMIN) and cloud service provider i.e, server1, server2, server3. The admin is the person who has the data files and uploads the file on the cloud for storage maintenance and computation. The client is an end user who tries to access the file which is uploaded by the admin. All of these three entities need to register and then admin will login and will give the file name which he wants to upload and choose the file and upload it on a cloud. The uploaded file is verified by the admin and also views the user details and each server capacity. Load balancing service can be used to automatically failover in the event that a virtual machine instance is down. The load balancer accepts the traffic through single global external IP address and then distributes it accordingly. Service interrupting events can happen at any time. When things go wrong, it's important to have a robust, targeted and well disaster recovery plan. The load balancer is a software program that is listening on the port where external clients connect to access services. The load balancer forwards request to one of the backend servers, which usually reply to the load balancer. This allows the load balancer to reply to the client without the client ever knowing about the internal separation of functions. It also prevents clients from contacting backend servers directly, which may have security benefits by hiding the structure of the internal network and preventing attacks on the kernels network stack or unrelated services running on other ports. In case if the file is corrupted in server1, as we know its replicates are also maintained in other two servers, the file is fetched from the other server and recovered with in short period of time to the user.

4.1 Data backup procedure

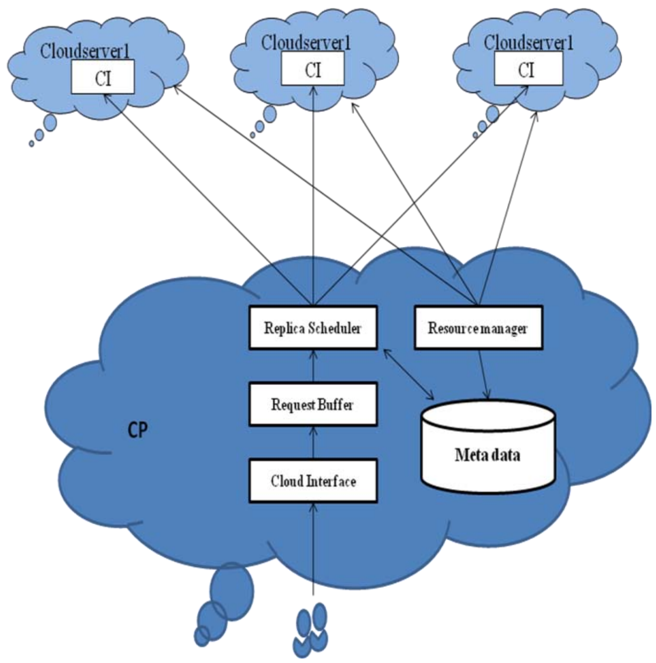

Figure 2: Data Backup Model

Cloud Backup delivers an intelligent data backup and recovery to protect organizations' business critical data in one simple unified solution. As organizations need to protect increasing data environments in decreasing backup windows. Cloud Backup allows customers to intelligently meet business objectives by backing up critical data and retaining it in a secure infrastructure. As IT environments increase in complexity, with growing physical and virtual infrastructures combining with an increasing number of applications that are critical to the functioning of organizations, IT managers are under pressure to enhance their data protection strategies. Cloud Backup helps achieve backup and recovery goals in a simple to deploy, enterprise class, intelligent managed service.

In fig2, the CP represents the data recovery service provider. Where all the clients are the users of the service provider, which can be individual or other cloud service providers. All they have appropriate accounts and of CP .Cloud server1, cloud server2 and cloud server3 are the other cloud service providers which provide common cloud resources to CP. The interface of each cloud server send/receives a data from/to users. Request buffer of CP holds the data backup request which is arrived and waiting to schedule. Replica scheduler reads the requests from the request buffer and makes 3-replicas for each cloud servers and then sent to the CP. Resource manager, monitors the changes of resources of all cloud service providers. Meta data is a database containing the information of the replica's location and resource usages.

The flow of the data backup procedure is shown in Fig2, is as follows:

- The user send data backup requests to the cloud service provider which needs to be stored in CP through its interface.
- Cloud provider receives the user requests and maintains them in Request Buffer and then checks customers' account and refuses illegal requests.
- Replica Scheduler it's a scheduling strategy once per unit time or when there are certain numbers of requests in Request Buffer.
- Replica Scheduler reads all requests, data size and store duration parameters from Request Buffer, makes three replicas for each data, and determines storage location of each replica. Then it generates an overall data backup scheme.
- According to the scheme, Replica Scheduler sends all replicas to destinations respectively, which may be CP's own data centers or other CPs' data centers.
- When finished with the transmission of data replicas , clod service provider saves all data replicas location to Metadata database. Then deletes all fulfilled requests from Request Buffer, and records those requests information for charging user later.
- Service provider then sends acknowledgement to users to return corresponding data, which can be used to restore certain data from CP.

4.2 Data Recovery Procedure

The fig3, shows the data recovery model, in which cloud server1, cloud server2 and cloud server3 are the three cloud service providers, which contains the replicas of the data recovery request. Recovery manager receives and checks the recovery requests and then selects the suitable cloud disaster recovery service provider which contains at least one replica of that request. Finally recovery proxy is installed at the user side, which is responsible for restoring the data cloud service provider.

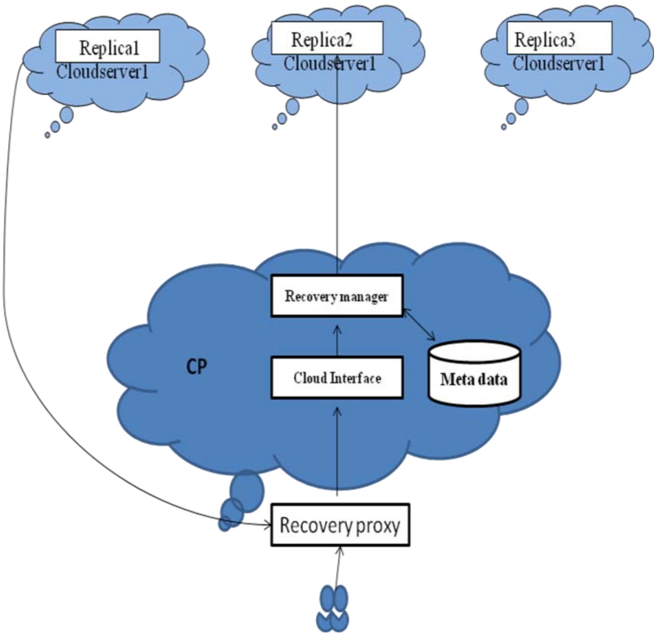

Figure 3: Data recovery model

The flow of the data recovery procedure is shown in Fig.3, also described as follows:

- A user sends a data recovery request to service provider through its Cloud Interface, with the corresponding data handle got from CP before.
- CP receives the request, checks the user's account and privilege, and refuses the request if it's illegal.
- Recovery Manager of CP looks up the data in Metadata database to find backup locations of that data.
- CP compares the 3 replicas' locations of that data, and chooses the fastest location, i.e., the one with the highest recovery bandwidth. Supposing the location is cloud server1.
- Recovery Manager informs cloud server1 to set up a temporary access authorization, which can only be used to read that replica, and can only be used by the Recovery Proxy of that user. Then Cloud server1 establishes the authorization, notifies to CP.
- CP sends the data location and authorization information to the Recovery Proxy of that user.
- The Recovery Proxy pulls data from cloud server1,and notifies CP.
- CP notifies cloud server1 to destroy that temporary access authorization and records the recovery request's information for charging the user later.
- Cloud server1 records network traffic consumption of the recovery request, which is used to charge CP later. In particular, if cloud server1 is CP itself, then no temporary access authorization needs to be set up by CP, the Recovery Proxy will use the user's account of CP directly.

From the above description, the data is restored from directly, not retransmitted by CP, which reduces recovery time when disaster occurs.

4.3 DFD of the System Working

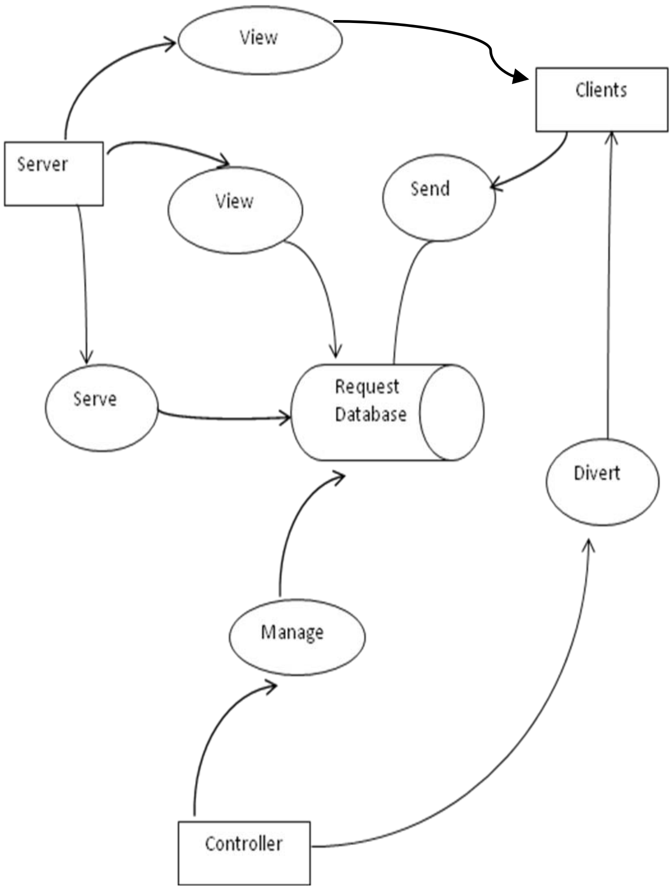

Figure 4: Data Flow Diagram of the Working System

The fig 4 shows the overall flow diagram of the working system. Server views the clients request and also the request database. Client sends the request to the database and server will serve the request database for fulfilling the client request. Controller will manage the request database.

5. Results and Snapshots

Screen shots of An approach to multi-cloud based disaster recovery with load balancing.

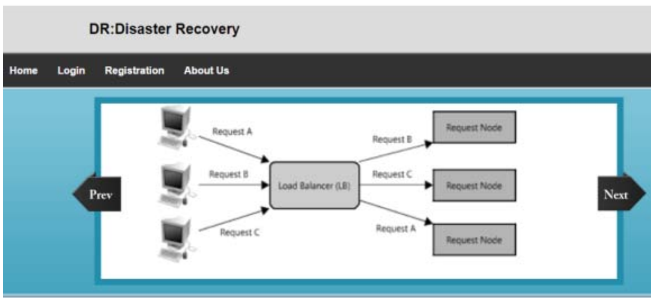

Figure 1: Home page

Fig **1** shows home page of Multi-Cloud Disaster Recovery with Load balancing.

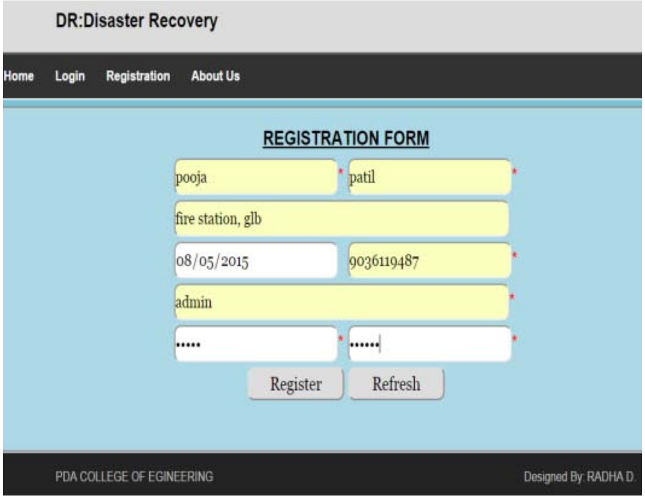

Figure 2: Registration page (admin and user)

Fig **2** shows registration page of admin and user. How the owner and user registers into cloud by entering name, Age, Email, Password, mobile number and gender.

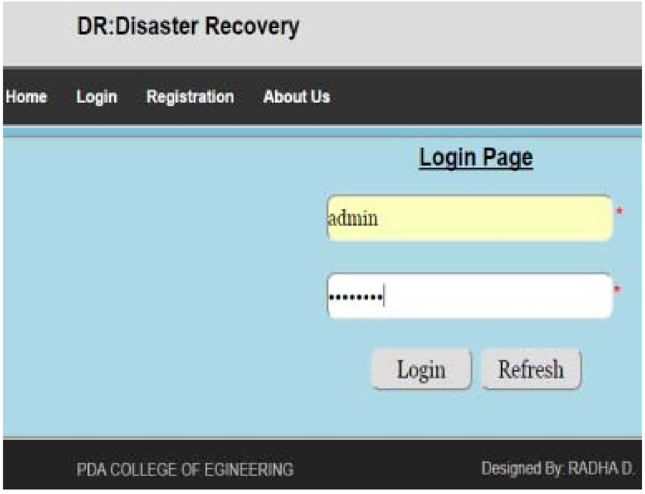

Figure 3: Login page(Admin login)

Fig 3 shows login page of disaster recovery with load balance in cloud computing. In which Administrator (Admin) will login.

Figure 4: File Upload

Fig 4 shows how the admin uploads the file into the cloud. In this phase admin will login by using email id and password then will give file name and then chose one of the file from the folder and upload it on the cloud.

Figure 5: View Files

Fig **5** shows View Files. In which the Admin verifies whether the file has been uploaded or not.

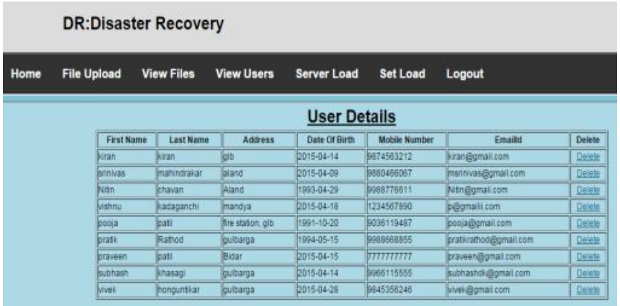

Figure 6: View Users

Fig **6** shows View Users. In which the Admin can view the registered user details.

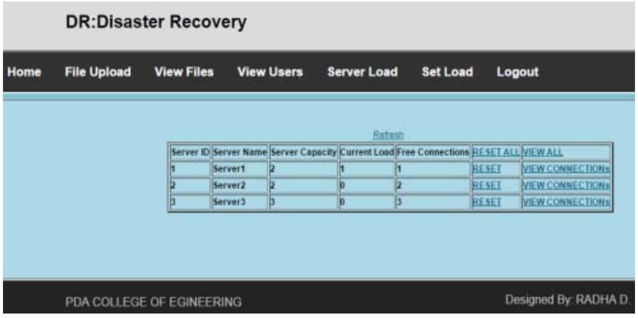

Figure 7: Server Load

Fig **7** shows the sever load, i.e, the admin can view each of the servers capacity, current load and number of free connections. Admin can also RESET all the server's load or can be RESET for individual server's.

Figure 8: Set Load

Fig **8** shows Set load. In which the admin can modify the server capacity, by inserting appropriate server name with the value.

Figure 9: Client Registration

Fig**9** shows client registration form. By providing the user first name, last name, address date of birth, mobile number, e-mail address and password.

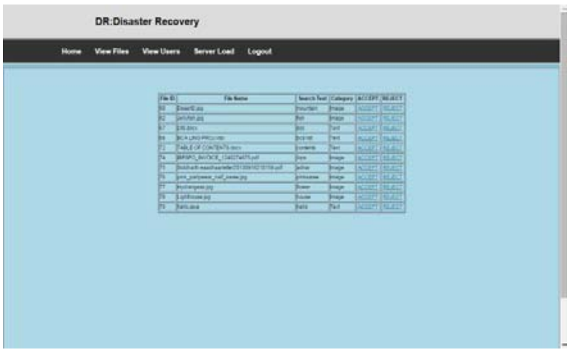

Figure 10: Server Login

Fig **10** shows the server login. In which the server can view filename, file category. Server can also accept or reject the file.

6. Conclusion

In our proposed system that is load balancing with disaster recovery using multi clouds, resources of multiple cloud service providers can be utilized simultaneously by the data disaster recovery provider, and customers/users need to deal with that service provider itself using simple interface without concerning the internal process between heterogeneous clouds. Load balancing distributes the workloads among the resources and maximizes the throughput. Thus it also provides security by using the E-mail option. Thus DR-cloud with load balance ensures high data reliability and short recovery time. In this DR-cloud with load balance, achieved their optimization objectives effectively. In future, further investigation of other multi-objective optimization algorithms to explore better effects.

References

[1] KaveEshghi,Darrell.D.Elong and MarkLillibridge, Extreme Binning : Scalable, Parallel Deduplication for Chunk-based File Backup.

[2] Rajiv Ranjan , Rodrigo N. Calheiros and Rajkumar Buyya" InterCloud: Utility-Oriented Federation of Cloud Computing Environments for Scaling of Application Services"

[3] Cloud Computing and Distributed Systems (CLOUDS) Laboratory Department of Computer Science. The University of Melbourne, Australia.

[4] Jih-Sheng Chang and Ruay-Shiung Chang "An Innovative Replica Consistency Model in Data Grids" *Network Innovation Technology Laboratory.*

[5] Prakhar Yadav Dhirajsing Thakur and Shital Bhosale "A Study on Application-Aware Local-Global Source Deduplication for Cloud Backup Services of Personal Storage".

[6] *Windsor W. H, Alan Jay Smith* and *Honesty C. Youngy* "The Automatic Improvement of Locality in Storage Systems" *Storage Systems Department , Almaden Research Center IBM Research.*

[7] Alysson Bessani Miguel Correia Bruno Quaresma Fernando Andr´e Paulo Sousa "DEPSKY: Dependable and Secure Storage in a Cloud-of-Clouds".

[8] Bo Mao, Hong Jiang, Suzhen Wu, Yinjin Fu, Lei Tian" SAR: SSD Assisted Restore Optimization for Deduplication-Based Storage Systems in the Cloud" Dept. of Comput. Sci. & Eng., Univ. of Nebraska-Lincoln, Lincoln, NE, USA.

[9] C.Warrick and S.John, A disaster recovery solution selection methodology, IBM Corporation , February 2004.

[10] A.Greenberg , J.Hamilton, D.A Maltz, and P.Patel, The cost of cloud:Research problems in datacenter networks, ACM SIGCOMM computer communication.

[11] A.Fox ,R.Griffith, A.Joseph, R.Katz,Above the clouds: A Berkeley view of cloud computing.

[12] T.Wood, E. Cecchet ,K.K.Ramakrishnan, P.Shenony, J.Van der Merwe, Disaster recovery as a cloud service: Economics benefits & deployment challenges.

[13] Amazon corporation, Amazon simple storage service(Amazon S3),http://aws.amazon.com/s3,2008.

[14] S.Ghemawat , H.Gobioff and S.T.Leung , The Google file system,ACM SIGOPS operating system Rieview.

[15] D.Borthakur , The hadoop distributed file system:Architecture and design.

[16] M.Vrable,s.SavageandG.M.Voelkar,Culumus:Filesystem backup to the cloud, ACM Transactions on storage.

[17] T.Wood,H.A.LagarCavilla,K.K.Ramakrishnan,P.Shenoy ,and.J.Vander Merwe,Pipecloud: Using causality to overcome speed-of-light delays in cloud based disaster recovery, in proceedings of the 2nd ACM Symposium on cloud computing,2011.
[18] T. Nguyen, A.Cutway , and W.Shi , Differentiated replication strategy in data centers , in proc. The IFIP International Conference on Network and Parallel Computing,Guangzhou,China,2010.
[19] D.Bermbach, M.Klems , S.Tai and M.Menzel,Metastorage : A federated cloud storage system to manage consistency –latency tradeoffs, in IEEE International Conference on Cloud Computing,2011.
[20] C.Cachin, R.Haas, and M.Vukolic, Dependable storage in the Intercloud,IBM Research, vol3783.