NEW TECHNIQUES FOR LOAD BALANCING AND DEDUPLICATIONS

1.Salunkhe Prachi 2.Shinde Ashwini
3.Thorat Mayuri 4.Kolapkar Shrikumar

[1, 2,3,4] **Department of Computer Engineering,**
Rajgad Dyanpeeth Technical Campus,Pune University,
Dhangwadi Pune 412205,Maharashtra, India.

1.Salunkheprachi09@gmail.com 2.shindeashwini10295@gmail.com
3.mayurithorat095@gmail.com 4.shrikumar4rko@gmail.com

ABSTRACT :
Private cloud storage is mustly to a particular organization or company and data securitys are Less compared to the public cloud storage hence private cloud storage is built by experts that commodity machines within the organization and the important data is stored it in. When the utilization of such a private cloud storage space increases, there will be an increase in the storage space demand. It does to the expansion of the cloud storage with additional storage nodes. During such expansion, storage nodes in the cloud storage need to be balanced. In order to maintain load across the several storage nodes, this data migration consumes more network bandwidth. The key idea behind this project is to provides new technique for develop a dynamic load balancing algorithm based on deduplication to balance the node across the storage nodes during the expansion of private cloud storage.

***KEY WORDS** : Load balancing , Deduplication , Cloud storage , Space management.*

1. INTRODUCTION

Recent years there is increasing popularity of cloud computing, mobile computing and the Internet of things, which brings about the tremendous growth of data. To meet people's demands for low-cost and convenient storage services, cloud storage has already been a typical storage system. Cloud storage integrates various kinds of storage devices together through application cluster, web technology, and distributed file system to provide storage services, such as Amazon S3, Google Drive. According to the statistics of the IDC, the total amount of global data has reached 1.8ZB (1ZB = 109TB) in 2011; it is expected to hit 35ZB by 2020.The storage is increasing on cloud storage day by the day. IDC found that nearly 75% of data in the information systems can be reduced, which causes vast storage resources consumption, while data deduplication can be identified and eliminate duplicated data.

Nowadays, data deduplication is adopted in the backup and archive systems, which can achieve very high deduplication ratio. There are many different deduplication strategies depending on the range of deduplication, the position of deduplication (at the client or server side) the time of deduplication. User stores data on cloud and take advantage of on demand high quality applications. 'Pay per use model' is main feature of cloud. While using cloud, user is free from stress of storage management. Even there are many advantages of storing data on cloud but it does not guarantee about data integrity. Enabling public auditability for cloud storage is important. So that Third Party Auditor (TPA) can challenge cloud server to verify integrity of user's data. Use of TPA for auditing process is easier and affordable way. Privacy protection should be considered while external auditor that is TPA is used for auditing purpose.

When there is deduplication on server side no same file will be uploaded on the cloud, so there will be space management done by using this deduplication strategy. After that the data which client is going to upload on cloud that needs to distribute. This data distribution among clouds will balance the load of uploaded data and client can easily perform operations like upload, delete and download the file. So for this reason there is concept of load balancing.

Data deduplication is an effective data reduction techniques. Data deduplication stores only single copy of data. Data deduplication splitted the uploaded file into various chunks. Dedup server placed between client/user and cloud data. Data can be stored into private or public cloud. Public cloud has less security as compared to private cloud.

2. LITRATURE SURVEY:

In this paper [1], focuses on data deduplication strategies and terminologies with respect to the some of the storage systems. Data deduplication technique is new trend in market for data compression.

In this paper [2], work deals with the inherent tension between well established storage optimization methods and end to end encryption. Files transition from one mode to the other in a seamless way as soon as they become popular.

In this paper [3],solution is based on a cryptographic use of symmetric encryption used for enciphering the data and asymmetric encryption for meta data files, due to high sensibility of these information towards several intrusions.

In this paper [4], most of the main issues with cloud computing have addressed to degree that clouds have become interesting for full commercial exploitation. This however does not mean that all problems listed above have actually been solved, only that the according risks can be tolerated to a certain degree.

In this paper [5], Due to increase in the usage of cloud computing there is need for the efficient and effective resource allocation algorithm which can be used for proper usage of the resources and also check that the resource is not wastage. we propose a priority based resource allocation algorithm which can be used for proper allocation of resources and also the resources are allocated efficiently and effectively.

3. SYSTEM OVERVIEW:

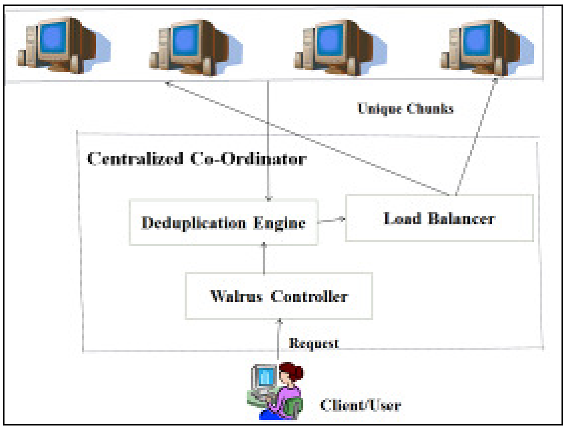


Fig.1. system overview

In this system,the system consist of client or user. Client or user send to request of file uploading, modification, access and deletion through dedup server. Firstly, client makes request for file uploading. This request accept by walrus controller. Centralized Co-ordinator is placed between uploaded files and client /user. All the metadata of files are stored in walrus controller. Walrus has its own database but is currently co-hosted by cloud controller. Walrus controller gives this data to deduplication engine. Deduplication engine encrypts the uploaded file using AES algorithm. Load balancer splitted uploaded file into various chunks using SHA algorithm. When user/client want to download or delete file,then deduplication engine decrypts file from various chunks and gives to client/user. Each cipher encrypt and decrypt data in 128 bit block size and the key size is the 128,192,256 bits. The sender and receiver knows private key. In this system, for encrypt and decrypt the data dedup server must use private key and also for upload,download,delete file user/client also mustly use private key

.

4. PROPOSED ARCHITECHURE:

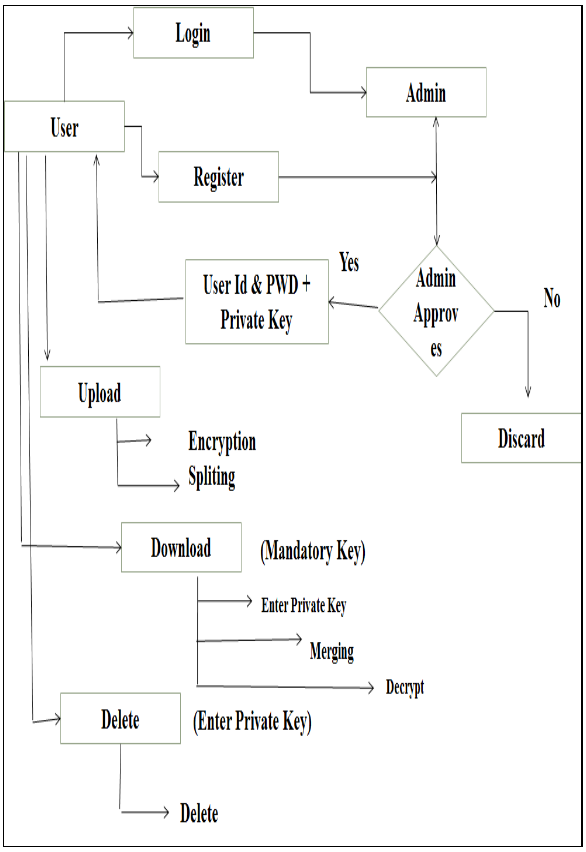


Fig 2. Proposed Architecture

This System has a functionality to ask information for the customer to the login and send the username, password and private key to the user with the help of the admin. Those have a login credentials as well as private key for the login who can easily perform upload, delete, and download operations. Using the Advanced Encryption standards (AES) and Secure Hash Code (SHA) algorithm the data security and load balancing will be manage. The Hash Code is used to create code according to the file data and stored into the database if the code is same then Duplicate file message will be arrive otherwise the code is unique then file split into three different chunk and stored it into three Different location.

If the user try to Delete or Download the file without Private Key and its login credential it gets fails. If the Login credential gets match then all of the three chunks gets merged into a single file and Delete/Download Operations performed this makes the faster and more secure. Main motivation of the system is to remove a load on cloud base servers and avoiding data Duplications using the some methodologies and algorithm. This system is basically perform on Hash Code detection techniques which is used for avoiding multiple storage of the files on the Cloud Server.

For the load balancing techniques system split the file into three chunks and stored into the three different location and the access is only for the valid person's or authorized persons only who has login credentials with the valid user key which is given by the admin.

5. ADVANTAGES:

- Faster access and effective.
- Secure data via private key encryption.
- Avoid data loss while data cloud server failure.
- Maintain data redundancy.

6. ALGORITHMS TO BE USED :

6.1 AES - Advanced Encryption Standard
6.2 SHA - Secure Hash Algorithm

6.1 Advanced Encyption Standard:

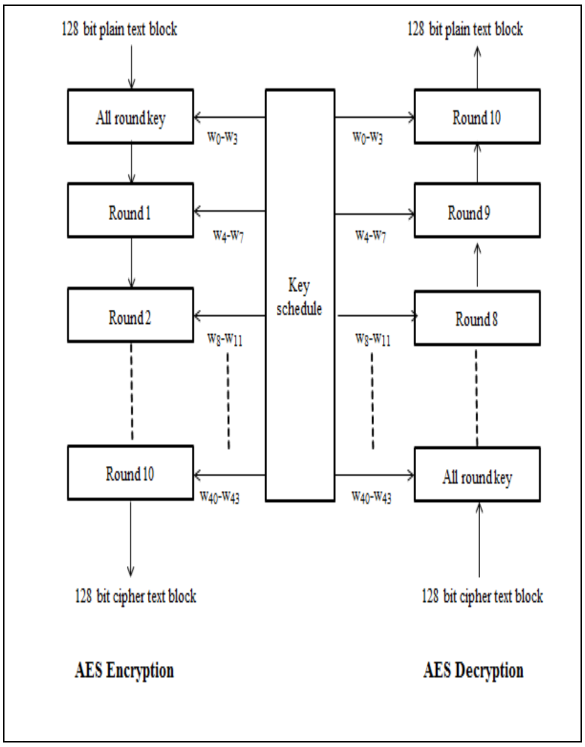


Fig 3. Advanced Encryption Standard

AES is a block cipher with a block size is 128 bits.AES algorithm is used to the encrypt and decrypt the data. AES comprises the three block ciphers. Each cipher encrypts and decrypts data in blocks of 128 bits using cryptographic keys. private key ciphers use the same key for encrypting and decrypting, so both the client/user and the dedup server must know and use the same private key. All key lengths are sufficient to protect classified information up to the Secret level with Top Secret information. There are 10 rounds for 128-bit keys, 12 rounds for 192-bit keys, and 14 rounds for the 256 bit keys a round consists of several processing steps that include substitution, transposition and mixing of the input plaintext and transform it into the final output of cipher text.

The **features** of AES are as follows :

- Symmetric key symmetric block cipher
- 128-bit data, 128/192/256-bit keys
- Stronger and faster than Triple-DES
- Provide full specification and design details
- Software implementable in C and Java.

6.2 Secure Hash Algorithm:

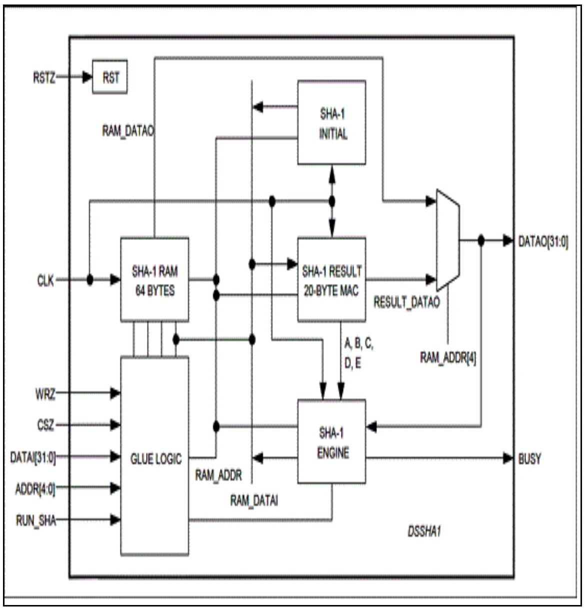
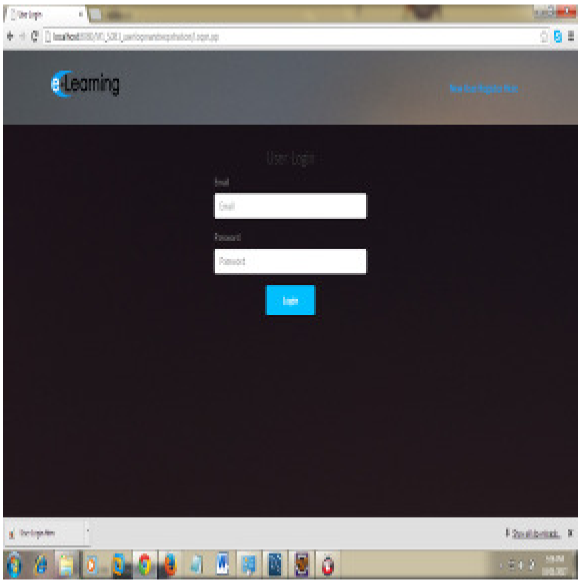


Fig 4. Secure Hash Algorithm

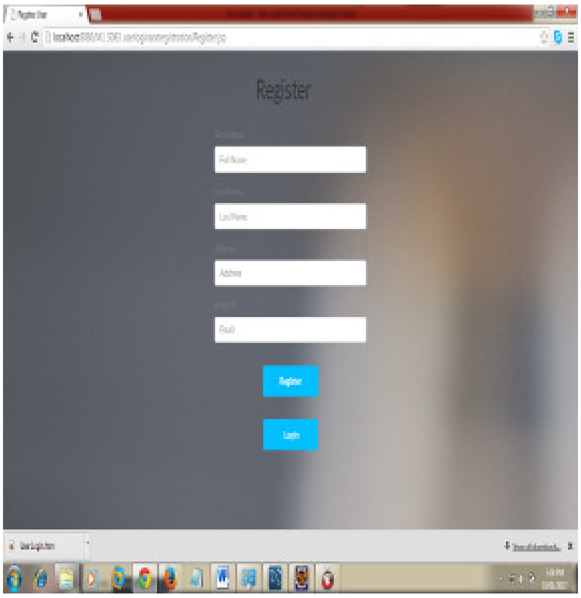
SHA algorithm generates a hash code on the basis of file content. Cryptographic hash functions are mathematical operations run on digital data by comparing the computed hash that is the output from execution of the algorithm to a known and expected hash value a person can determine the data integrity. For example, computing the hash of a downloaded file and comparing result to previously published hash result can show whether the download has been modified or tampered with. The key aspect of cryptographic hash functions is their collision resistance: nobody should be able to find two different input values that result in the same hash output.

7. RESULTS:

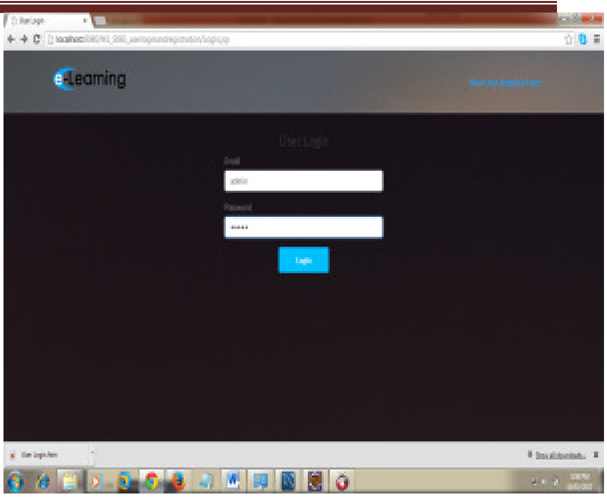
This is our login page. Firstly user have to register to the system.



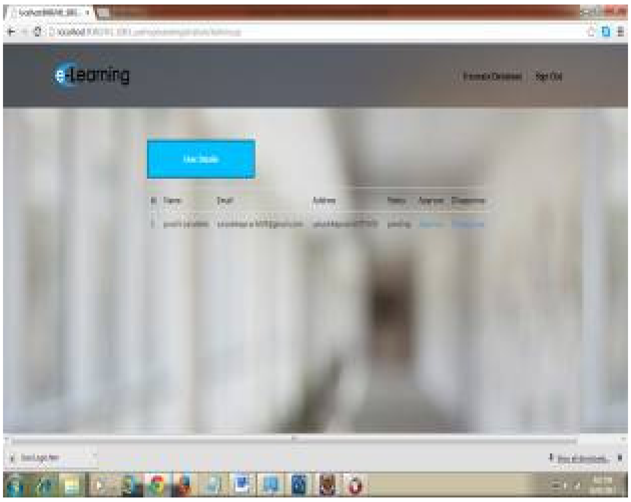
This is our registration window. User must fill the information like First name,Last name,user address and email address.



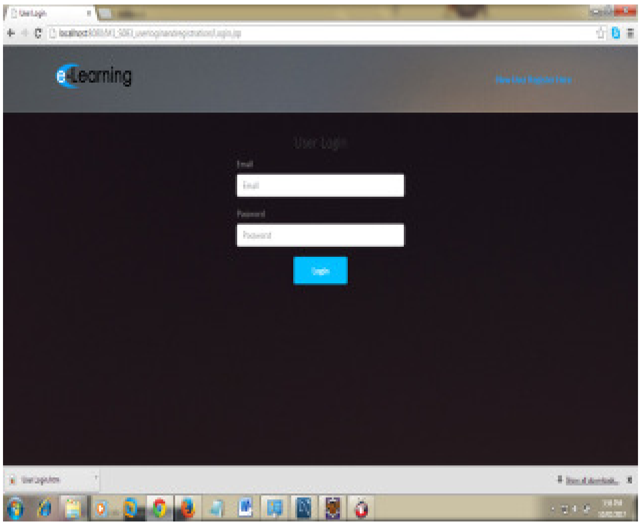
In this window, admin done login.



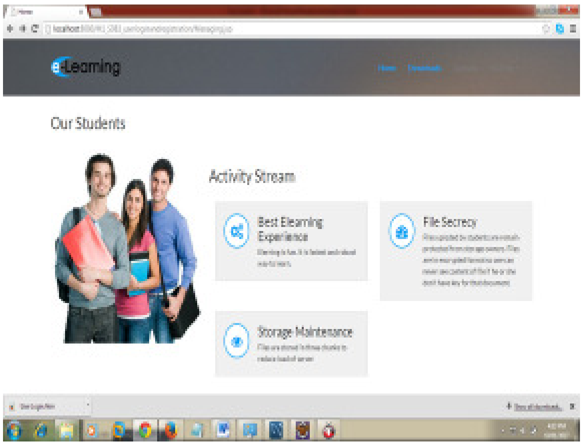
In this window, shows user details to the admin. Admin can approve or disapprove of user. If admin approves the user then it validate for this system otherwise it is invalidate for system.



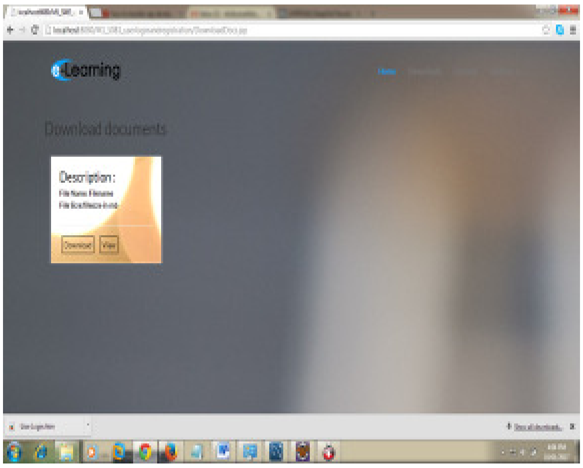
In this window, validate user can login to the system.

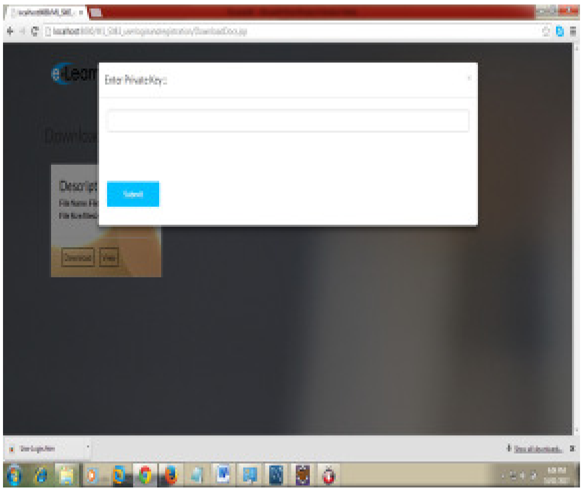


In this window, shows the users page. It includes options like home,upload,download,signout.

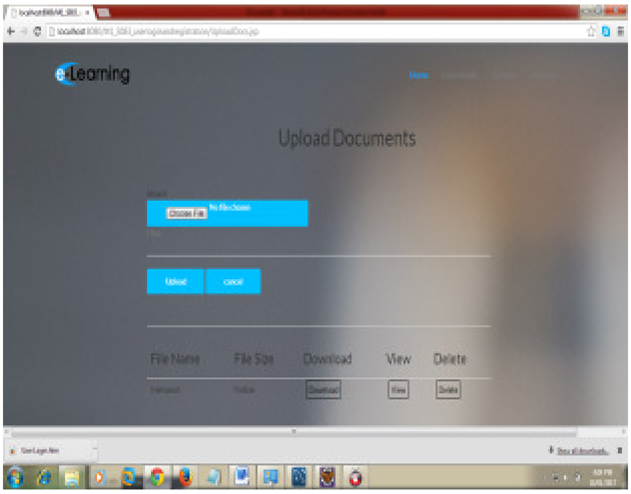


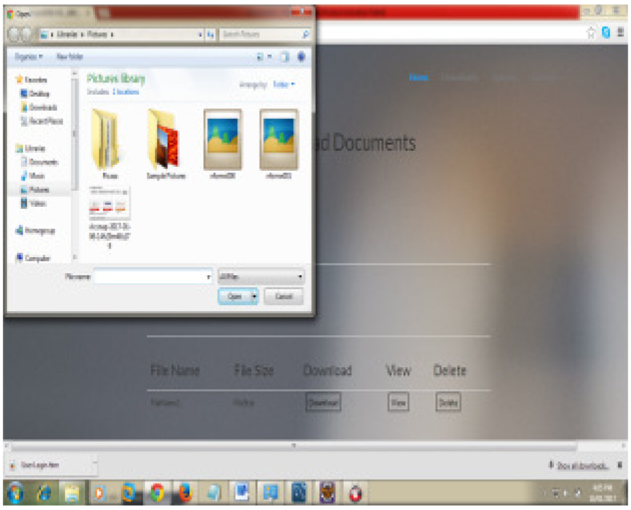
These windows includes the download document option. During downloading documents everytime it must have a private key i.e., that is provided by this system.



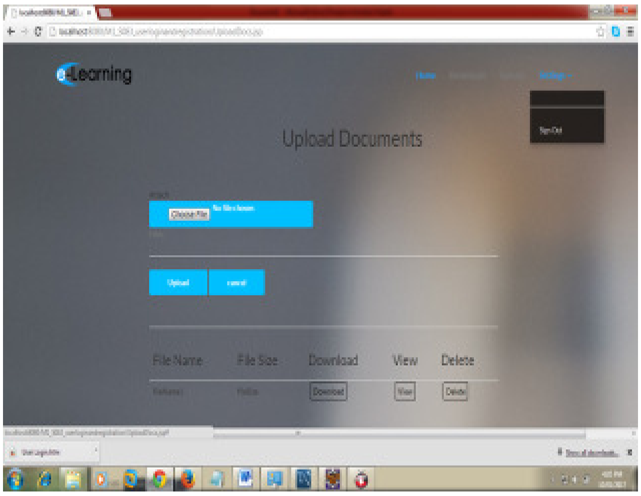


These window shows the uploading document option. It also required private key everytime for uploading document.





Affter all that user wants to signout then sidnout option is there.



8. CONCLUSION:

This system proposes the architecture of deduplication system for cloud storage environment and give the process of avoiding deduplication in each stage. In Client, system employ the file-level and chunk-level deduplication to avoid duplication. The algorithm also supports mutual inclusion and exclusion. Load sharing algorithm which is having policy to partitions the system into various domains and also having concept of cache manager and information dissemination for the various cloudlets.

9. REFERENCE:

[1] P. Xie,"Survey on deduplication techniques for storage systems," Comput. Sci., vol. 41, no. 1, pp. 22-30, Jan. 2014.
[2] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, "A secure data deduplication scheme for cloud storage," in Proc. 18th Int. Conf. Financial Cryptography and Data Security, Christ Church, 2014, pp. 99-118.
[3] N. Kaaniche, and M. Laurent, "A secure Client side deduplication scheme in cloud storage environments," in Proc. 2014 6th Int. Conf. New Technol., Mobility and Security (NTMS), Dubai, 2014, pp. 1-7.
[4] R. Shen, "Research on the mechanism of avoiding storing data duplicate in cloud storage," M.S. thesis, Dept. Comput. Tech., Yunnan Univ., Kunming, Yunnan, 2013.
[5] C. S. Pawar, and R. B. Wagh, Priority Based Dynamic Resource Allocation in Cloud Computing with Modified Waiting Queue, 2013 International Conference on Intelligent Systems and Signal Processing (ISSP), 2013
[6] Prasenjit Kumar Patra,Harshpreet Singh, GurupreetSingh,"Fault Tolerance Techniques and Comparative Implementation in Cloud Computing",International Journal of Computer Applications (0975 – 8887) Volume 64– No.14, February 2013.
[7] Abhijit A. Rajguru, S.S. Apte,"A Comparative Performance Analysis of Load balancing Algorithms in Distributed System using Qualitative Parameters",International journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-1, Issue-3, August 2012.
[8] Y. Tan, H. Jiang, D. Feng, L. Tian and Z. Yan, "CABdedupe: A causality-based deduplication performance booster for cloud backup services," in Proc. 2011 IEEE Int. Parallel & Distributed Process. Symp. (IPDPS), Anchorage, 2011, pp. 1266-1277.
[9] Y. Fu, H. Jiang, N. Xiao, L. Tian, and F. Liu, "AA-Dedupe: An application-aware source deduplication approach for cloud backup services in the personal computing environment," in Proc. 2011 IEEE Int. Conf. Cluster Comput. (CLUSTER), Austin, 2011, pp. 112-120.
[10] R. Hu, Y. Li, and Y. Zhang, Adaptive Resource Management in PaaS Platform Using Feedback Control LRU Algorithm, International Conference on Cloud and Service Computing,2011.
[11] J. Wu, L. Ping, X. Ge, Y. Wang, and J. Fu, "Cloud storage as the infrastructure of cloud computing," in Proc. 2010 Int. Conf. Intell. Comput. Cognitive Inform. (ICICCI), Kuala Lumpur,2010,pp.380-383.
[12] J. Gantz and D. Reinsel, "The digital universe decade-Are you ready," IDC White Paper,http://www.emc.com/collateral/analyst-reports/idc-digitaluniverse-are-you-ready.pdf,2010.
[13] Nidhi Jain Kansal, InderveerChana,"Cloud Load Balancing Techniques : A Step Towards Green Computing",IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 1, January 2012 ISSN (Online): 1694-0814.
[14] AnjuBala1, InderveerChana2, "Fault Tolerance-Challenges, Techniques and Implementation in Cloud Computing",IJCSIInternational Journal Of Computer Scienceissues, vol. 9, issue 1, no 1, January 2012, ISSN (online): 1694-0814.
[15] N.Chandrakala,Dr.P.Sivaprakasam, "Analysis of Fault Tolerance Approaches in Dynamic Cloud Computing", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 2, February 2013.