SECURED DISTRIBUTED DATA DEDUPLICATION WITH IMPROVED RELIABILITY

Darshana Bhosale, Shweta Shitole, Manas Kumar,Rahul Gangwani
SSBT's Co**ll**ege of Engineering and Technology,
Jalgaon,Maharashtra

Abstract : Information deduplication is also called as intelligent compression or one warehousing example. It is a unconscious process used to eliminate duplicate extra transcript of data by deliverance just single transcript of the data and replacing other copies. Lotion of it in swarm formula aims at minimizing the warehousing place with bandwidth usage during transportation of the file cabinet. Just single copy of duplicated filing cabinet is retained while rest other single file are been deleted. Existence of this duplicate files is decided from metadata. Many companies are frequently using this technique in backup man and recovery practical application, where it is used to free certain space in primary storage. Files are clustered into bins according to their size of it. Then further segmented, de-duplicated, compressed and then saved on storage space. Bin restricts identification number of segments and their size so that it is optimum for each file size. On request for a file, compressed segments of the file are sent along with the file to segment to senses of map Finally they are uncompressed and combined for creating complete file, hence it minimizes the overall bandwidth requisite.

Index Terms- **Deduplication, cloud, warehouse, compression, bin**

I. INTRODUCTION

Secure data distributed systems with improved dependability is a technique for eliminating duplicate copy of data, and has been widely used in cloud storehouse to reduce storage space and upload bandwidth. However, there is only one copy for each filing cabinet stored in cloud even if such a file is owned by a huge number of users. As a result, de-duplication system improves storage utilization while reducing reliability.

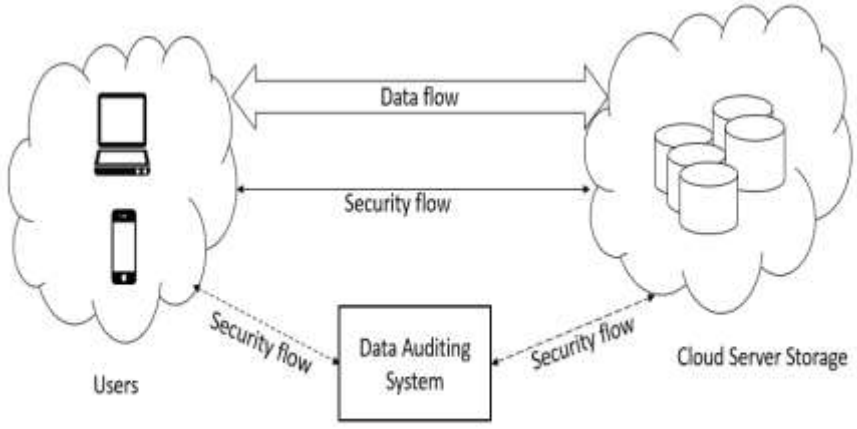

Figure: Cloud Architecture

1.1 Motivation

Information deduplication is ideal for highly redundant operations like backup, which requires repeatedly written matter and storing the same data chunk for multiple times. Data deduplication memory board has unique instance of data. The redundant data get eliminated and replaced with a pointer to the unique data copy. The benefit of data deduplication is obvious. Eliminating redundant data can significantly shrink storage requirements and berth storage senses of cost. Deduplication also improve the network bandwidth efficiency and save the processing power.

1.2 Problem Definition

The techniques that were previously used in duplication of data had wastage of storage space. In order to overcome this problem de-duplication technique is used, which checks duplicate copies of data, if found it then eliminates these duplicate copies of data to reduce storage space and upload bandwidth. Only one copy of data will be stored on cloud and that copy will be access by many users.

II. LITERATURE REVIEW

The literature survey for the project identifies the need of project. Also, any work done relating to project. Techniques used to guide the implementation project are also mentioned.

2.1 **A view of cloud computing**

AUTHORS: M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia

Cloud computation, the long -clench ambition of computing as a public utility company has the potential to transform a large part of the IT industriousness, making computer software even more attractive as a service and shaping the way IT hardware is designed and purchased. Moreover,fellowship with large muckle-oriented project can get solution as quickly as their programs can scale.

2.2 **Secure and constant cost public cloud storage auditing with deduplication**

AUTHORS: J. Yuan and S. Yu

Data unity and warehousing efficiency are two important requirements for swarm storage. Proof of Retrievability (POR) and Proof of Data Self-Possession(PDP) techniques assure information unity for cloud storage. Proof of Ownership (POW) improves storage efficiency by securely removing unnecessarily duplicated information on the storage server. However, trivial combination of the two techniques, in order to achieve both data integrity and storage efficiency, event in non-trivial duplication of metadata (i.e., authentication tags), which contradicts the objectives of POW. Recent epoch attempts to this job introduce tremendous computational and communication costs and have also been proven not secure.

III. SYSTEM ARCHITECTURE

1) File Uploading Protocol

This protocol allows clients to upload files via the auditor. It includes three phases:

Phase 1 (cloud client cloud server): client performs the duplicate check with the cloud server to confirm if such a file is stored in cloud storage or not before uploading a file.
Phase 2 (cloud client auditor): client uploads files to the auditor, and get acknowledgement from auditor.
Phase 3 (auditor cloud server): auditor helps to generate a set of tags for file and send these tags along with file during uploading it.

2) Integrity Auditing Protocol

Main aim of this protocol is integrity verification. Here, the cloud server plays the role of prover, while the auditor or client works as the verifier. This protocol includes two phases:

Phase 1 (cloud client/auditor cloud server): verifier generates a set of challenges and sends them to the prover.
Phase 2 (cloud server cloud client/auditor): based on the stored files and file tags, prover tries to prove that it exactly owns the target file by sending the proof back to verifier.

3) Proof of Ownership Protocol

This protocol used for verifying that the client exactly owns a claimed file. This protocol is typically triggered along with file uploading protocol to prevent the leakage of side channel information. On the contrast to integrity auditing protocol, in Proof of Ownership the cloud server works as verifier, while the client plays the role of prover. This protocol also includes two phases:

Phase 1 (cloud server client): cloud server generates a set of challenges and sends them to the client.
Phase 2 (client cloud server): the client responds with the proof for file ownership, and cloud server finally verifies the validity of proof.

Figure: System Model

IV. SYSTEM ALGORITHM

Ramp's Secret Sharing Scheme

There are two algorithmic rule in a secret sharing scheme, that are Share and Recover. The secret is divided and shared by using Share. With enough shares, the secret can be extracted and recovered with the algorithm of Recover. In de-duplication System implementation , The system will use the Ramp secret sharing scheme (RSSS) to secretly split a secret into shards.
Specially, the (n, k, r) RSSS (where n,k,r = 0) generates n shares from a secret so that (i) the secret can be recovered from any k or more shares, and (ii) no information about the secret can be deduced from any r or less shares. Two algorithms, Share and Recover, are defined in the (n, k, r)-RSSS.
Share divides a secret S into (k -r) pieces of equal size, generates r random pieces of the same size, and encodes the k pieces using a non-systematic k-of-n erasure code into n shares of the same size.
_ Recover takes any k out of n shares as inputs and then outputs the original secret S.

V. CONCLUSION

The major goal of this WWW application is to help the users to store their information on the cloud with confidentiality and security. Delaware-duplication of data is the main focus in the entire web application. Providing storage of data on a large plate with multipl e data file share-out. Auditing helps the user to check the wholeness of the data. In this paper, we have proposed a new auditing mechanism for information stored on cloud which provide Delware installation like efficient data de-duplication while maintaining shared data integrity. In this mechanism we can be able to achieve properties like correctness and scalability while improving the data de-duplication.

VI. RESULT

Here are some of the images of the system that we have designed. In this, in order to access, first new user have to get registered then that registered user can upload and download his/her file. Besides this, we can get details of particular user profile and even details of downloaded files.

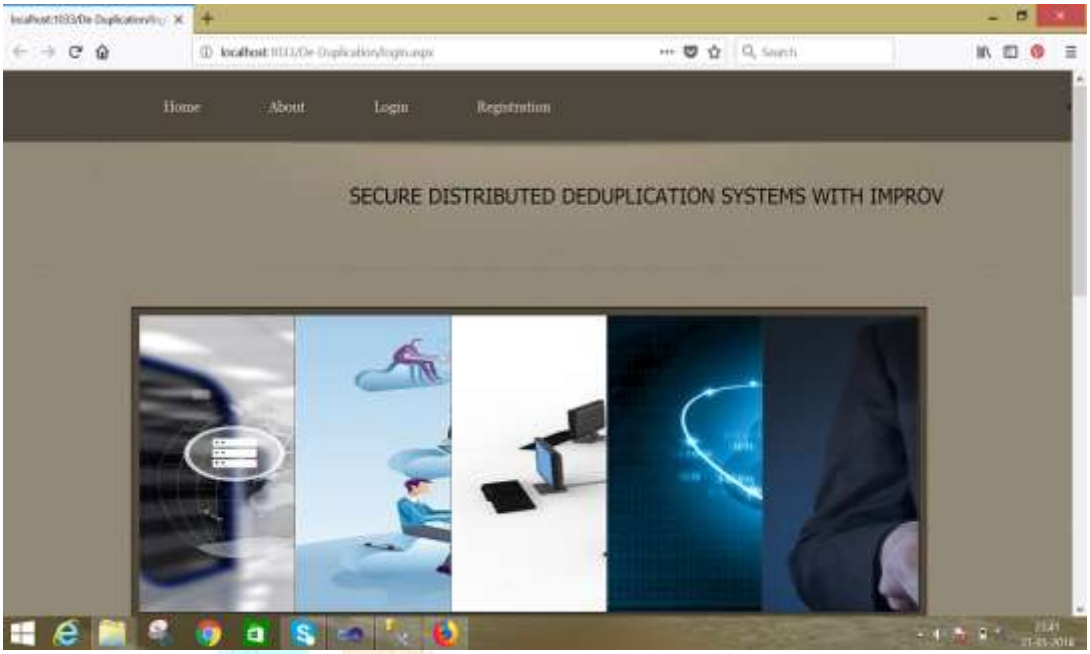

Figure above shows the image of Homescreen

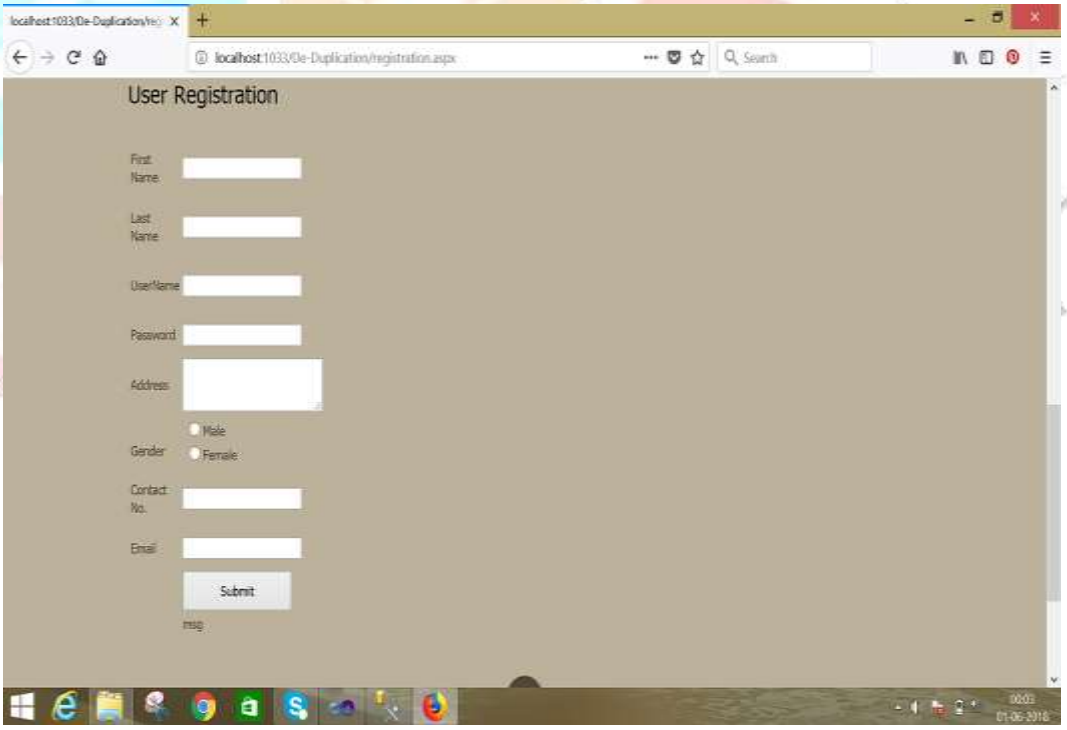

Figure above shows image for New User Registration

Figure above shows User Login

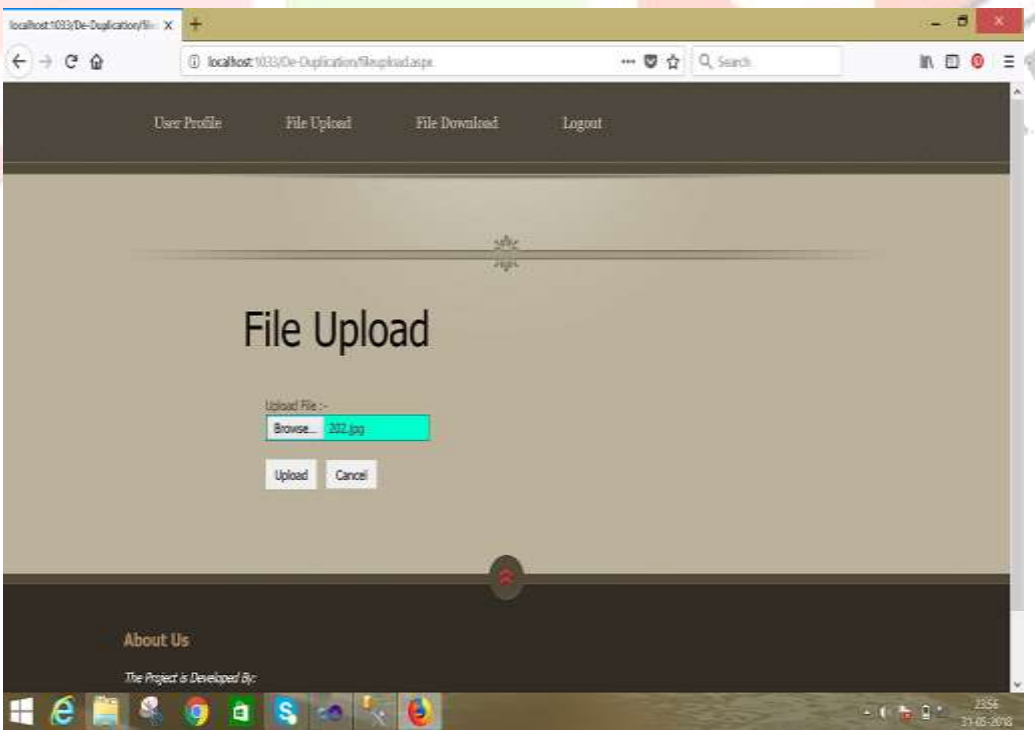

Figure above shows Uploaded File

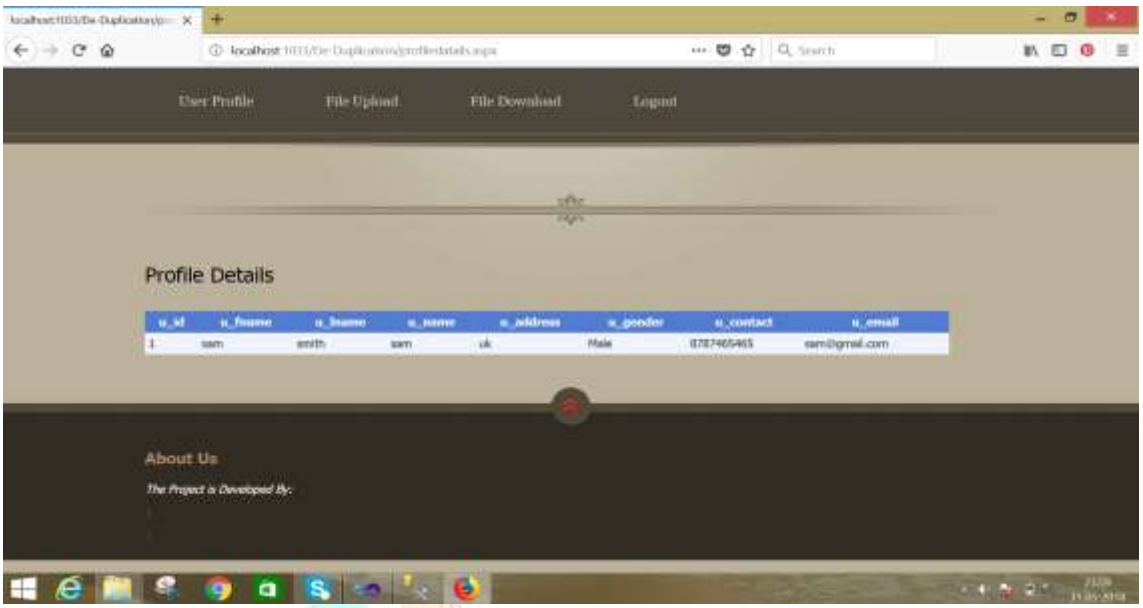

Figure above shows User Profile Details

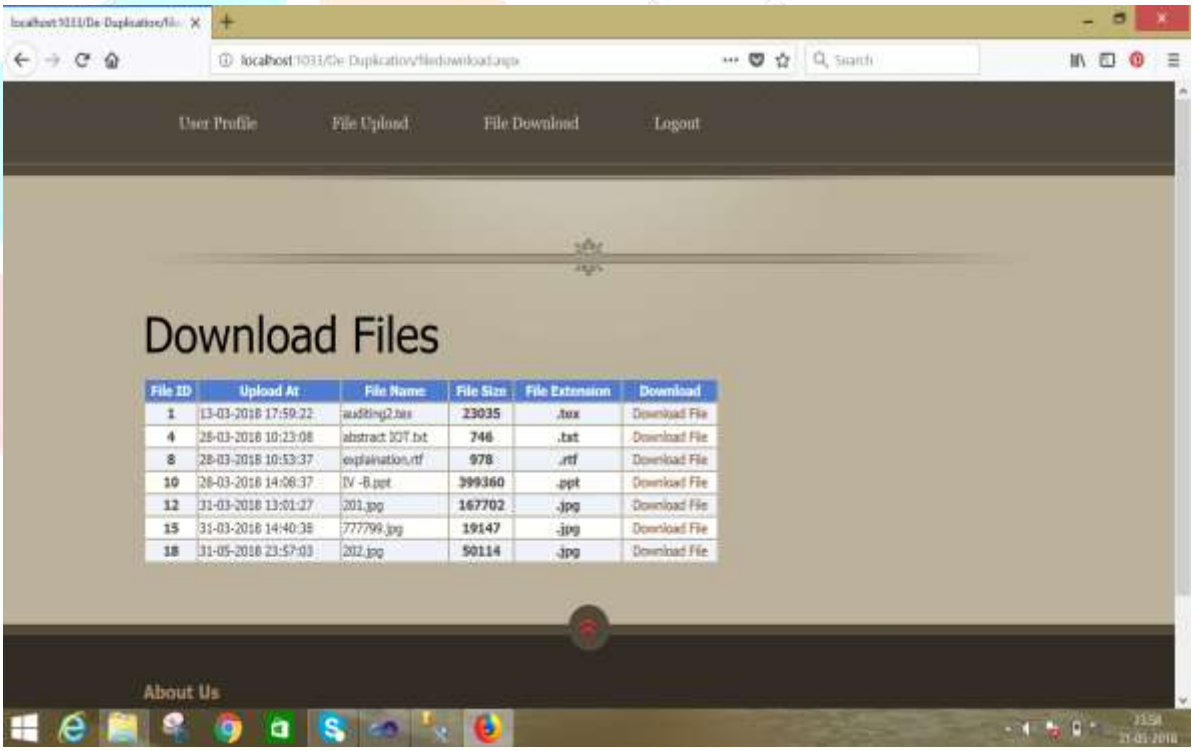

Figure above shows Downloaded Files

VII. REFERENCES

1. M Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In USENIX Security Symposium, 2013.
2. P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.
3. R Sravan Kumar,A.Saxena Data Integrity and Proofs in Cloud Storage