Cloud Dedup Using Proxy Reencryption

N Bharath Kumar [1], C Jyothi Priya [2], P Mounica [3], E Janaki [4]
[1, 2, 3] B.E., Under Graduate, Department of Computer Science and Engineering, Misrimal Navajee Munoth Jain Engineering College (Affiliated to Anna University), Chennai, Tamil Nadu, India.

[4] B.E., Under Graduate, Department of Computer Science and Engineering, Sri Muthukumuran Institute of Technology (Affiliated to Anna University), Chennai, Tamil Nadu, India.

Abstract – Cloud computing suggest an innovative way of service provision by re-arranging a variety of assets over the Internet. The majority significant and accepted cloud service is information storage. In sort to protect the confidentiality of information controller, information are often stored in cloud in an encrypted form. However, encrypted data introduce new challenges for cloud data deduplication, which becomes crucial for big data storage and processing in cloud. Traditional deduplication schemes cannot work on encrypted data. Existing solutions of encrypted data deduplication suffer from security weakness. They cannot flexibly support data access control and revocation. Therefore, few of them can be readily deployed in practice. In this paper, we propose a scheme to deduplicate encrypted data stored in cloud based on ownership challenge and proxy re-encryption. It integrates cloud data deduplication with access control. We evaluate its performance based on extensive analysis and computer simulations. The results show the superior efficiency and effectiveness of the scheme for potential practical deployment, especially for big data deduplication in cloud storage.

Index Terms – Cloud, Deduplication, Proxy, Reencryption, Internet.

1. INTRODUCTION

1.1. DOMAIN INTRODUCTION

1.1.1 DATA MINING

Data mining is an interdisciplinary sub field of computer science. It is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics and database system. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization and online updating.

Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD. The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown, interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection), and dependencies (association rule mining). This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection data preparation, nor result interpretation and reporting is part of the data mining step, but do belong to the overall KDD process as additional steps.

1.1.2 Cloud

Cloud computing is arecent trending in IT that where computing and data storage is done in data centers rather than personalportable PC's. It refers to applications delivered as service over the internet as well as to the cloud infrastructure-namely the hardware and the system software in data centers that provide this service. The sharing of resources reduces the cost to individuals.

The best definition for cloud is given as "The large pool of easily accessible and virtualizedresources which can be dynamically reconfigured to adjust the variable load, allowing also for optimum for scale utilization". The most widely used definition of cloud is made by NIST as "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction".

2. LITERATURE SURVEY

2.1 DupLESS: Server aided encryption for deduplicated storage-(2014)

Author: N. Kaaniche and M. Laurent

Cloud storage service providers such as Dropbox, Mozy, and others perform deduplication to save space by only storing one copy of each file uploaded. Should clients conventionally encrypt their files, however, savings are lost.

Message-locked encryption (the most prominent manifestation of which is convergent encryption) resolves this tension.

However it is inherently subject to brute-force attacks that can recover files falling into a known set.

3. SYSTEM ANALYSIS

3.1 Existing System:

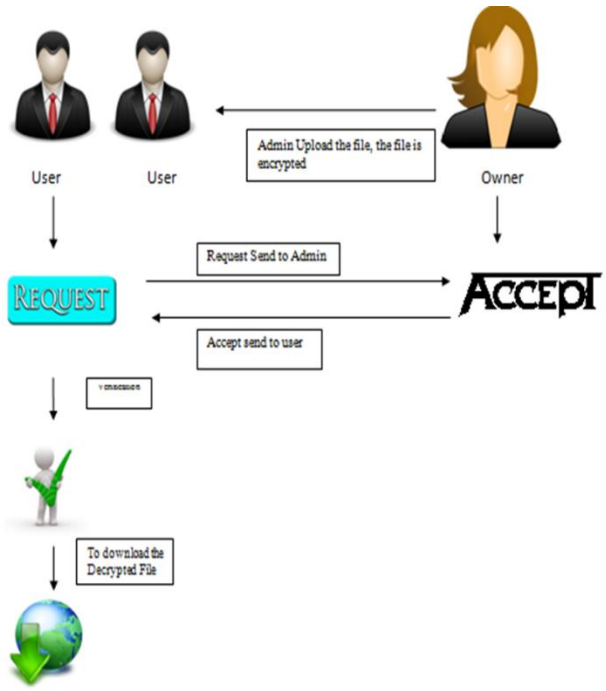
In existing scheme encrypted information deduplication undergo from security weakness it cannot flexibly support to right to use the information and revocation. it is inflexible to allow records owner to handle deduplication due to a amount of basis. First, data holders may not be forever online or obtainable for such a organization, which could cause storage delay. Second, deduplication could become too complicated in terms ofcommunications and computations to involve data holders into deduplication process

3.2 Proposed System:

We proposed a practical scheme to manage the encrypted big data in cloud with deduplication based on ownership challenge and PRE. Our scheme can flexibly support data update and sharing with deduplication even when the data holders are offline. Encrypted data can be securely accessed because only authorized data holders can obtain the symmetric keys used for data decryption. Extensive performance analysis and test showed that our scheme is secure and efficient under the described security model and very suitable for big data deduplication.

4. SYSTEM DESIGN

4.1 ARCHITECTURE DESIGN



4.4 SYSTEM DESIGN

Design is multi-step process that focuses on data structure software architecture, procedural details, (algorithms etc.) and interface between modules.

4.4.1 INPUT DESIGN

Input design is the process of converting a user-oriented description of the inputs to a computer based business system into a program-oriented specification.

The objectives in the input design:

- To produce a cost-effective method of input.
- To achieve a highest possible level of accuracy.
- To ensure that input is acceptable to and understood by the user staff.

INPUT STAGES:

Several activities have to be carried out as a part of the overall input process. They include

- Data Recording – Collection of data at its source.
- Data Description – Transfer of data to an input form
- Data Conversion – Conversion of the input data to a computer acceptable medium.
- Data Verification – Checking the conversion
- Data Control – Checking the accuracy and controlling the flow of data to the computer.
- Data Transmission – Transmission or transferring the data to the computer.
- Data Validation – Checking the input data by program when it enters the computer system.
- Data Correction – Correction the errors that are found at any early stages.

4.4.2 OUTPUT DESIGN

Output design generally refers to the results and information that are generated by the system for many end-users; output is the main reason for developing the system and the basis on which they evaluate the usefulness of the application. In any system, the output design determines the input to be given to the application.

The output design is an ongoing activity almost from the beginning of the project, and follows the principles of form design. Effects and well define an output design improves the relationship of system and the user, thus facilitating decision-making. A major form of output is a hard copy from the printer, however soft copies re available.

The Types of output used in the system are: -

Internal outputs:

Whose destination is within the organization and is the user's main interface with the computer.

Interactive outputs: -

Which involves the user in communicating directly with the computer.

External outputs: –

Whose destination is outside the organization and which require special attention since they project the image of the organization.

5. SYSTEM REQUIREMENTS

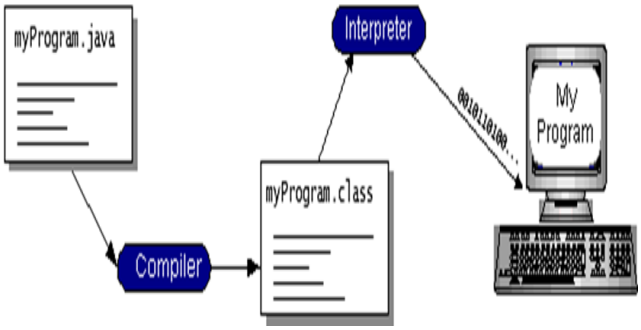
5.1 Software Requirements:

Front End	:	NETBEANS
Code Behind	:	JAVA
Back End	:	SQL SERVER 2008
Operating System	:	Windows 7.

5.2 Hardware Requirements:

Hard disk	:	40 GB
RAM	:	512mb
Processor	:	Pentium IV
Monitor	:	17" Color monitor

6. SYSTEM IMPLEMENTATION



7. CONCLUSION

7.1 CONCLUSION

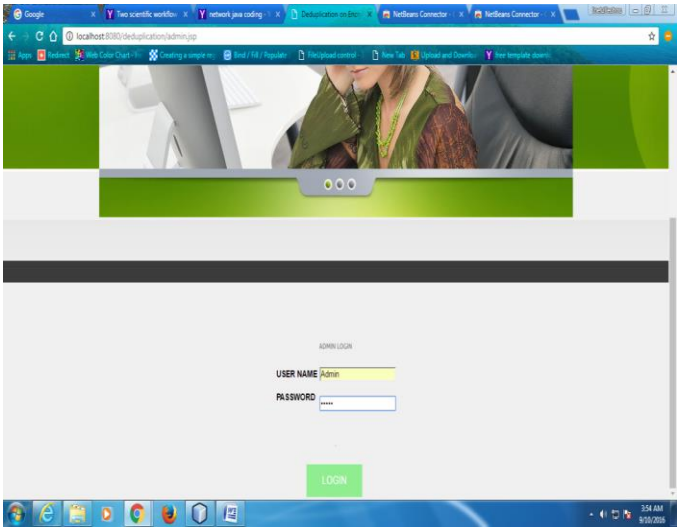
Encrypted information is able to be securely right to use since merely authoritative records controllers be able to achieve the symmetric keys used for information decryption. Extensive presentation investigation and analysis demonstrated that our system is protected and resourceful below the explained protection model and very suitable for big data deduplication.

7.1.1 Future Enhancements:

Future work embraces optimizing our plan and achievement for convenient exploitation and learning certifiable estimation to guarantee that CSP perform as predictable in deduplication organization.

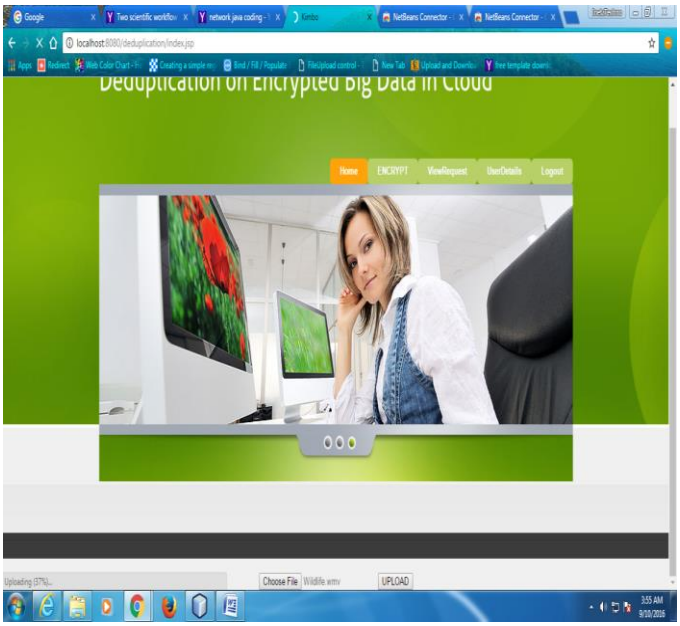
8. SAMPLES

Admin:



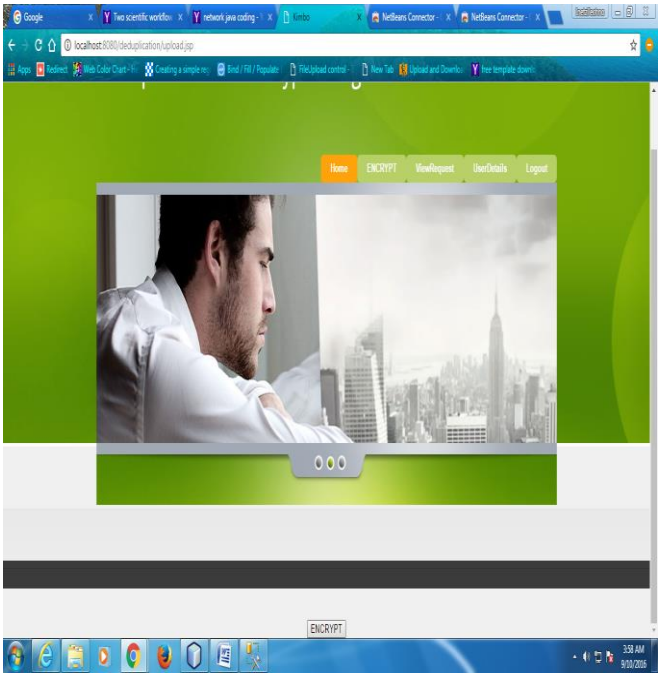
To enter into cloud, admin has to login the page using username and password.

Upload Video:



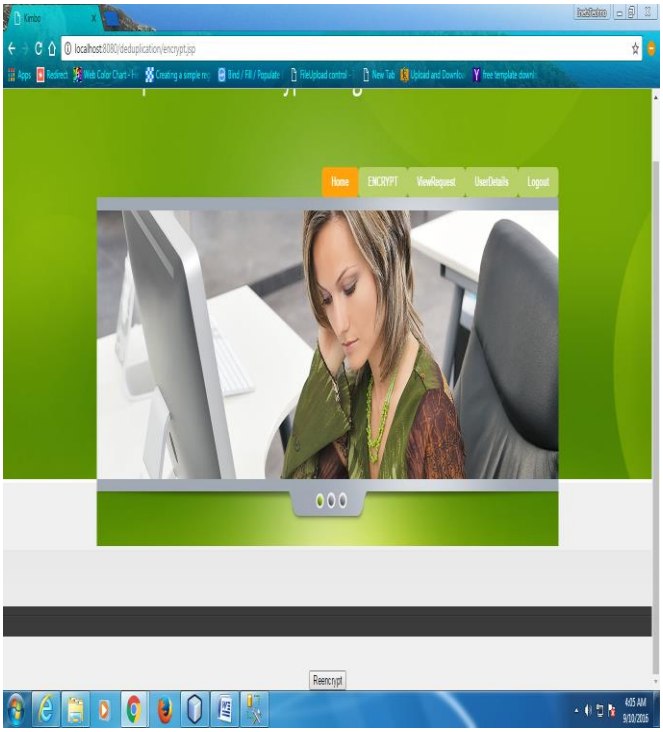
The admin uploads the file such as image, video, document, music etc….

Encrypt:



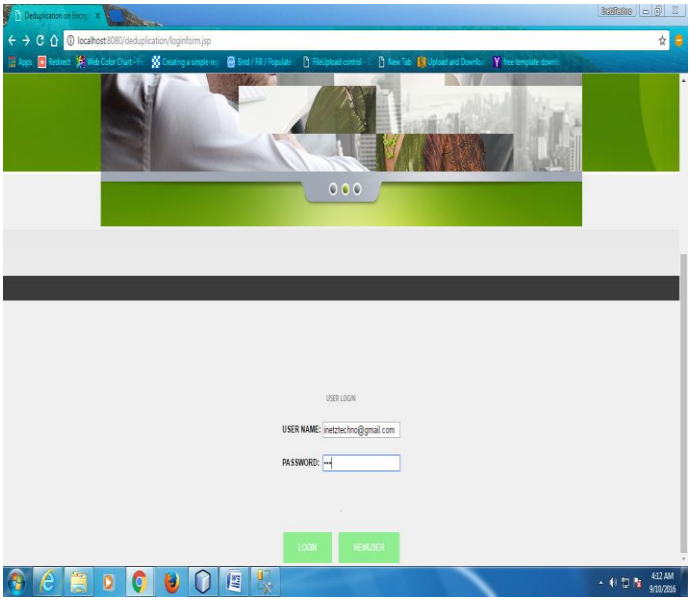
The uploaded file is encrypted.

Double Encryption:



The encrypted file is again encrypted using proxy re-encryption for security purpose.

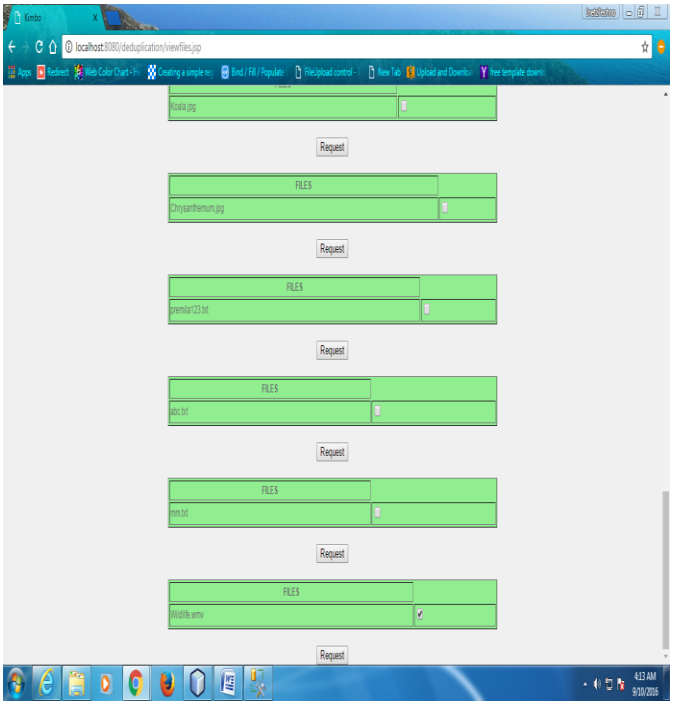
User Login:



The authorized user enters the cloud using their user name and password.

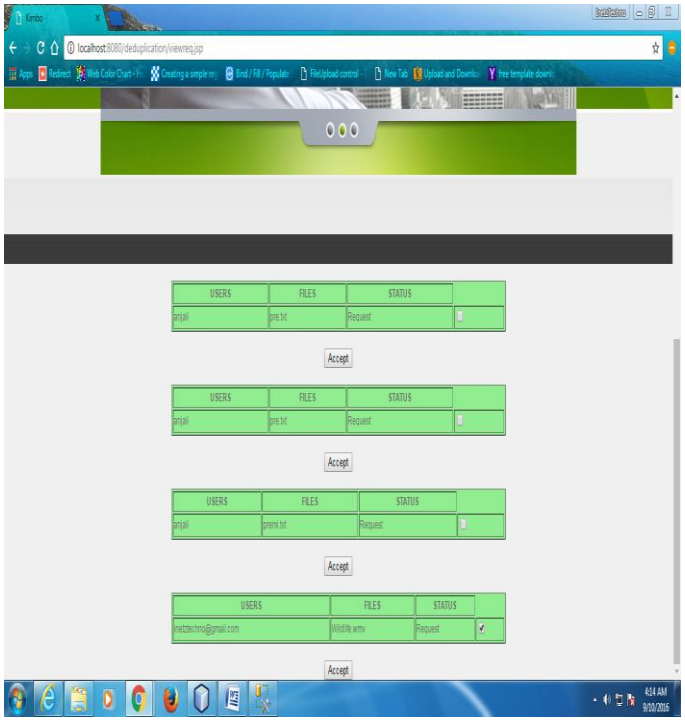
If any hacker or 3rd party tries to hack the user account using different password then an alert message is send to the respective user through mail .

File Request:



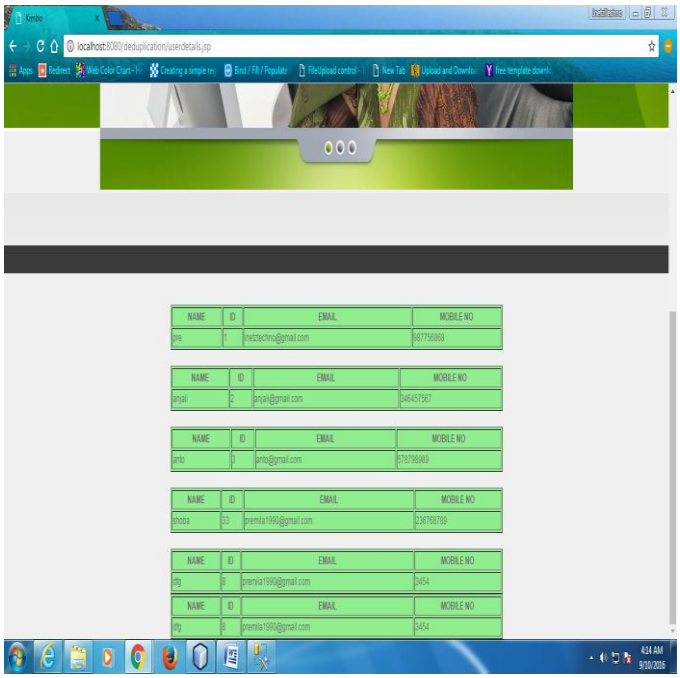
The user can view the files that are uploaded by the admin and send requests.

Admin Accept:



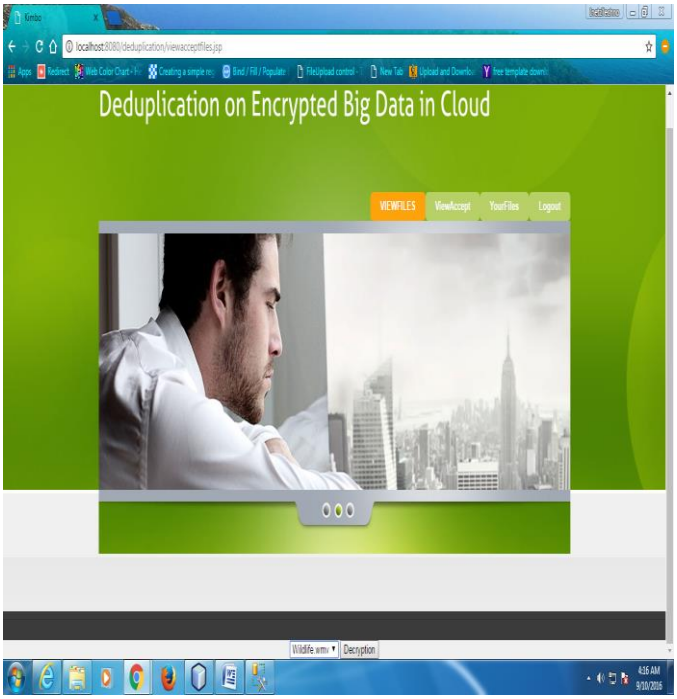
Admin verifies the user is authorized or not. After the verification, if the user is authorized then accepts the request otherwise deny the request.

User Details:



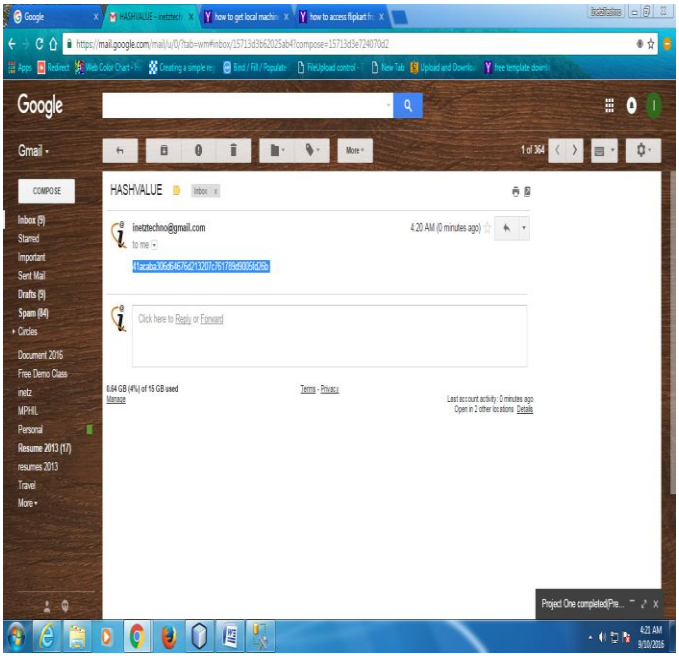
The user details are verified in this.

Decryption:



To decrypt the file, user needs to enter the key send to the mail id.

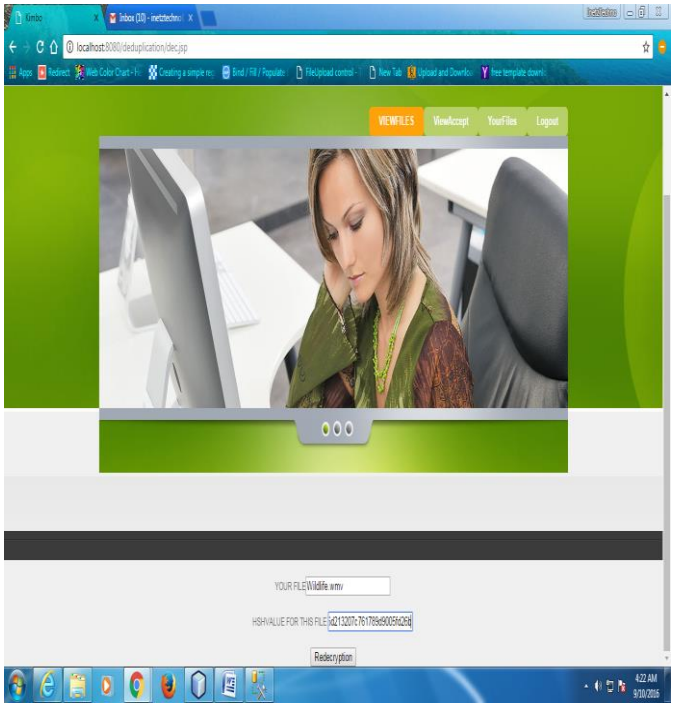
Verify Hash value of file:

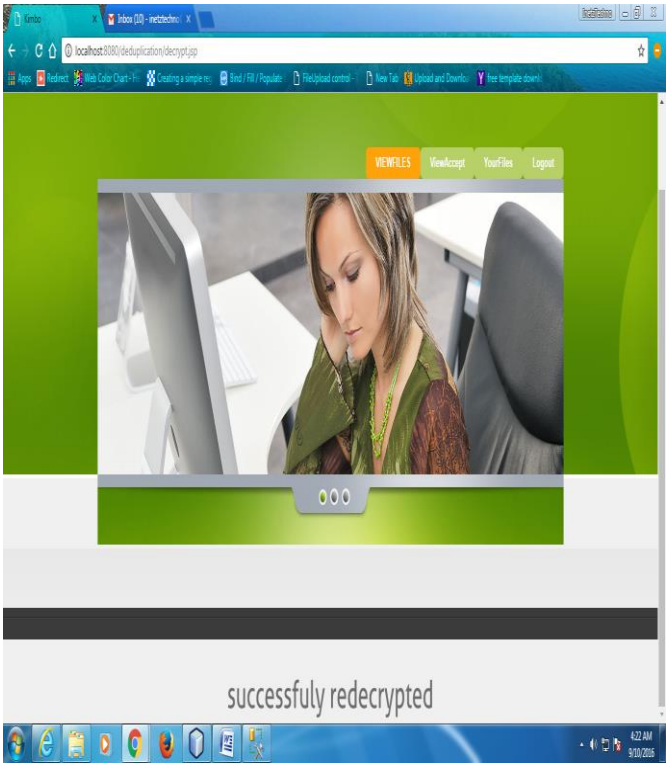


The hash value is sent by the admin to the authorized user through mail.

This key is used to decrypt the message.

Re-Decryption:

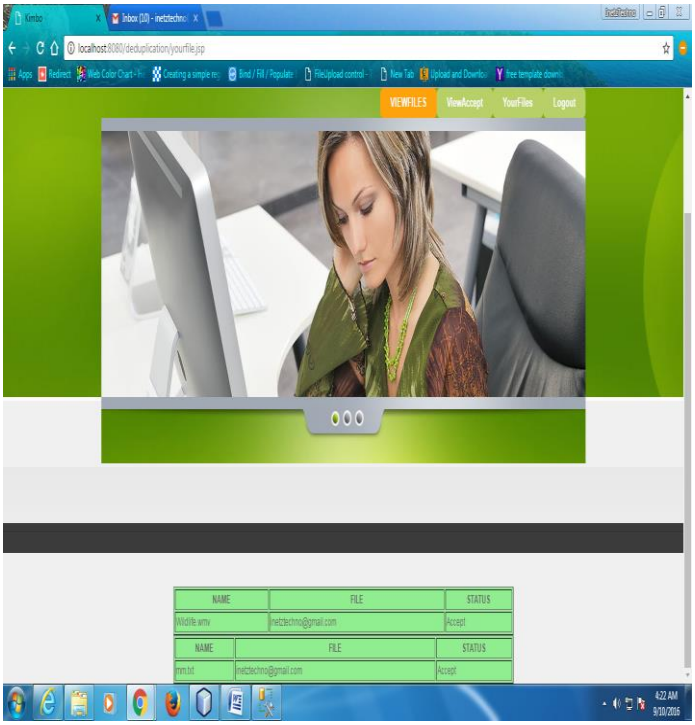




To redecrypt the file, another hash value is to be entered to get the original file.

This is also sent to the mail id to the user by admin.

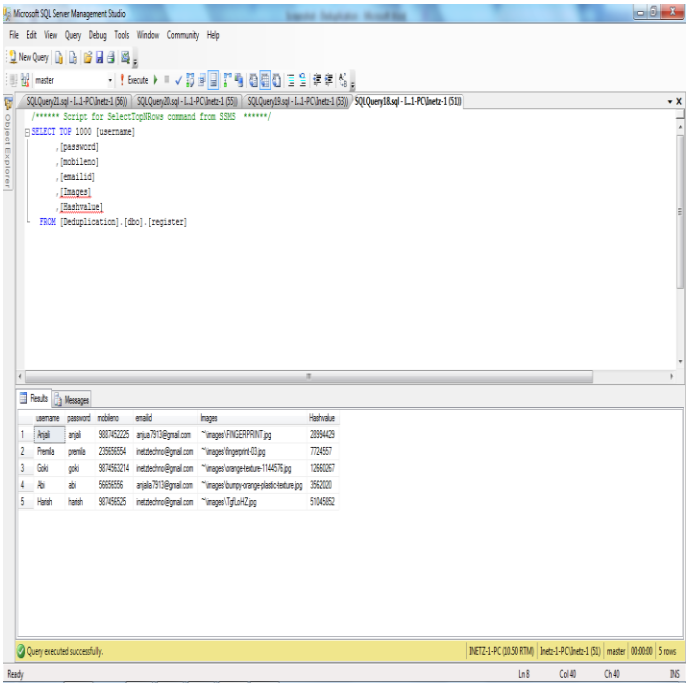
User Files:



The user can view the files that are existing in their account.

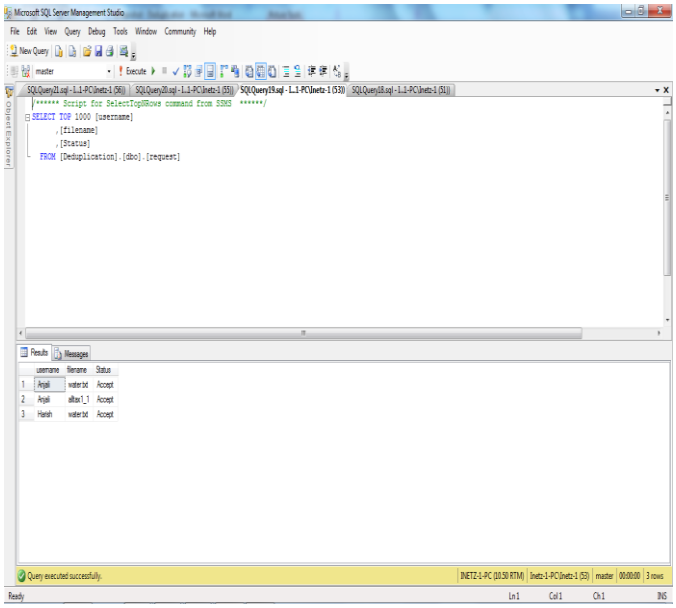
Database Tables

Register users

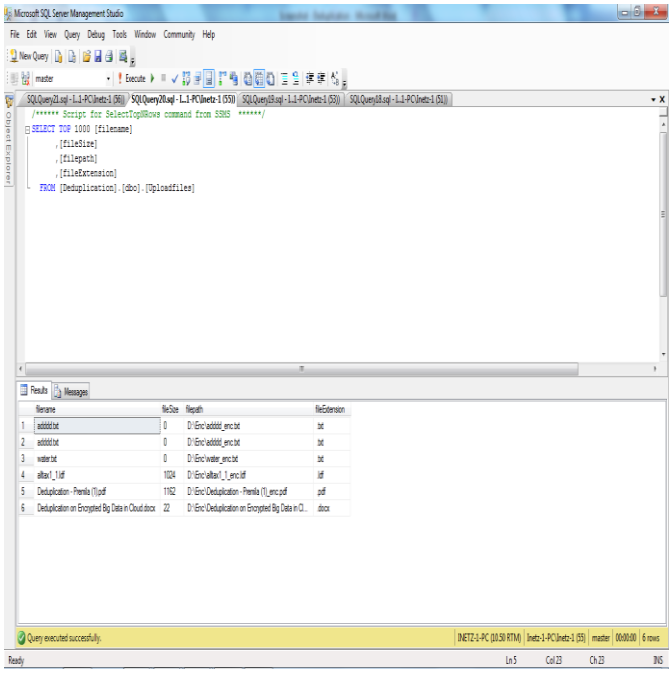


The registered users in the database.

Request:



Upload files:



The uploaded files are stored in the database

REFERENCES

[1] C. Yang, J. Ren, and J. F. Ma, "Provable ownership of file in deduplication cloud storage," in Proc. IEEE Global Commun. Conf.,2013, pp. 695–700, doi:10.1109/GLOCOM.2013.6831153.

[2] T. Y. Wu, J. S. Pan, and C. F. Lin, "Improving accessing efficiency of cloud storage using de-duplication and feedback schemes," IEEE Syst. J., vol. 8, no. 1, pp. 208–218, Mar. 2014, doi:10.1109/ JSYST.2013.2256715.

[3] C. Fan, S. Y. Huang, and W. C. Hsu, "Hybrid data deduplication in cloud environment," in Proc. Int. Conf. Inf. Secur. Intell. Control, 2012, pp. 174–177, doi:10.1109/ISIC.2012.6449734.

[4] J. W. Yuan and S. C. Yu, "Secure and constant cost public cloud storage auditing with deduplication," in Proc. IEEE Int. Conf. Communic.Netw. Secur., 2013, pp. 145–153, doi:10.1109/CNS.2013.6682702.

[5] N. Kaaniche and M. Laurent, "A secure client side deduplication scheme in cloud storage environments," in Proc. 6th Int. Conf.New Technol. Mobility Secur., 2014, pp. 1–7, doi:10.1109/ NTMS.2014.6814002.

[6] Z. Yan, W. X. Ding, and H. Q. Zhu, "A scheme to manage encrypted data storage with deduplication in cloud," in Proc. ICA3PP2015, Zhangjiajie, China, Nov. 2015, pp. 547–561.

[7] Z. Yan, X. Y. Li, and R. Kantola, "Controlling cloud data access based on reputation," Mobile Netw. Appl., vol. 20, no. 6, 2015,pp. 828–839, doi:10.1007/s11036-015-0591-6.

[8] T. T. Wu, W. C. Dou, C. H. Hu, and J. J. Chen, "Service mining for trusted service composition in cross-cloud environment," IEEE Systems Syst. J., vol. PP, no. 99, pp. 1–12, 2014, doi:10.1109/ JSYST.2014.2361841.