



JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A FRAMEWORK DESIGN ON DATA ACQUISITION AND COMPRESSION WITH LEAKAGE PREVENTION

Ms Sarika K Meshram

Dr. Sunil R Gupta

ME student, Department of Computer Science

Professor, Department of Computer Science

&Engineering,PRMIT&R

&Engineering,PRMIT&R

Abstract : Large information based securing and stockpiling framework (ASS) assumes a significant part in the plan of modern information stage. Numerous huge information structures have been incorporated pressure and serialization strategy. These strategies can't address the issues of modern creation data the executives for demanding tedious and mass capacity. In light of the current huge information structures, we propose an upgraded modern enormous information stage to decrease the information handling time while requiring less information extra room. In particular, this paper centers around assessing the effect of various pressure and serialization techniques on enormous information stage execution and attempts to pick ideal pressure and serialization strategy for the modern information stage. Many plans have been as of late progressed for putting away information on various mists. Appropriating information over various distributed storage suppliers (CSPs) naturally gives clients a specific level of data spillage control, for no single mark of assault can release all the data. Be that as it may, spontaneous dispersion of information pieces can prompt high data revelation even while utilizing numerous mists .Big measure of information dealing with is extremely significant to keep up with so performing dynamic pressure technique on this is essential

Keywords- Data, acquisition, serialization, compression, leakage.

I. INTRODUCTION

Information securing and stockpiling framework (ASS) is significant piece of the plan of modern information stage. modern information stage is the center part of modern information stockpiling, calculation and investigation for the administration of clever plant. Information obtaining is a course of social event, sifting and cleaning information before the information is placed in information some other stockpiling arrangement. The obtaining of large information is generally normally isolated in four of the Vs: volume, speed, assortment, and worth Most information procurement situations accept high-volume, high-speed, high-assortment, yet low-esteem information, making It essential to have versatile and time-effective assembling, separating, and cleaning calculations that guarantee that main the high-esteem pieces of the information are really handled by the information distribution center examination. Notwithstanding, for certain associations, most information is of possibly high worth .For such associations assume focal part after the information securing. The information stage remembers six layers for terms of information stream. These six layers are gadget layer, procurement layer, capacity layer, registering layer, administration layer and show layer, which compare thus to information securing, information capacity, information examination, administration bundle and front finish of modern information. This study centers around the procurement layer. we are examining information obtaining on modern information stage, enormous information stage incorporate three modules, which are information procurement module, information capacity module and information calculation module . The information procurement module gives an information source to information examination of the huge information stage and the information stockpiling module gives information source and extra room of the information calculation module. There are many types of modern information circulation procures a ton of information put away in social data sets in the business, and the constant prerequisite of these information isn't high. For this situation, at some point information spillage issue happens and consequently proposed framework manages information spillage discovery .

1.1 Data Acquisition:

Data acquisition is a process of gathering, filtering ,and cleaning data before the data is put in data any other storage solution. The acquisition of big data is most commonly governed by four of the Vs: volume, velocity, variety, and value. Most data acquisition scenarios assume high-volume, high-velocity, high-variety, but low-value data, making It important to have adaptable and time-efficient gathering, filtering, and cleaning algorithms that ensure that only the high-value fragments of the data are actually processed by the data-warehouse analysis. However, for some organizations, most data is of potentially high value as it

can be important to recruit new customers. For such organizations, data analysis, classification, and packaging on very high data volumes play the most central role after the data acquisition .

1.2 Data leakage :

The data leakage, detection and its impact on organizations. Data leakage may be defined as the illegal transfer of valuable/sensitive data by an entity to unauthorized entities. Data leakage detection is the process of finding the data leaker by using various techniques ranging from interrogation, watermark/fake data addition to other modern techniques. Data loss, which means a loss of data that occur on any device that stores data. It is a problem for anyone that uses a computer. Data loss happens when data may be physically or logically removed from the organization either intentionally or unintentionally. The data loss has become a biggest problem in organization today where the organizations are in responsibility to overcome this problem. Data Leakage is an incident when the confidentiality of information has been compromised. It refers to an unauthorized transmission of data from within an organization to an external destination. The data that is leaked out can either be private in nature and are deemed confidential whereas Data Loss is loss of data due to deletion, system crash etc. Totally both the term can be referred as data breach, has been one of the biggest fears that organization face today. Data loss / leakage prevention is a computer security term which will be used to find, watch, and protect data in use, data in motion, and data at rest. [9]The Data Leakage Prevention provides sensitive asset classification, sensitive asset audits, identity and access management audits, applying encryption to sensitive assets, applying enterprise digital rights management privileges to sensitive assets. Cloud Computing is a model for enabling ubiquitous, convenient, on demand network access to a shared pool of configurable computing resources that can be rapidly positioned with minimal management effort or service provider interaction.

II. LITERATURE SURVEY

Big data analysis of industry is considered as a necessary aspect for further improvement in order to improve the profit margin of industrial production and operation, and represents the next frontier of innovation, competition and productivity [1]Nowadays, industrial data platform is the core component of industrial data storage, computation and analysis for the management of intelligent plant.. According to industrial data acquisition and processing requirements, this paper designs an industrial big data platform. The data platform includes six layers in terms of data flow. These six layers are device layer, acquisition layer, storage layer, computing layer, service layer and display layer, which correspond in turn to data acquisition, data storage, data analysis, service package and front end of industrial data.[2] The acquisition of big data is most commonly governed by four of the Vs: volume, velocity, variety, and value. Most data acquisition scenarios assume high-volume, high-velocity, high-variety, but low-value data, making it important to have adaptable and time-efficient gathering, filtering, and cleaning algorithms that ensure that only the high-value fragments of the data are actually processed by the data-warehouse analysis.[3] We describe a network-based data-leak detection (DLD) technique, the main feature of which is that the detection does not reveal the content of the sensitive data. Instead, only a small amount of specialized digests are needed. Our technique referred to as the fuzzy fingerprint detection can be used to detect accidental data leaks due to human errors or application flaws. [4]This world completely depends on the data will be generated by the humans for their own use. The companies are totally depends on the people who are generating the data. So, the data is a hot currency to the IT world. Due to this the hackers or intruders or the inside employees are trying to gain the monetary benefits, they are trying to get access to the data with or without authentication. In this case, Security plays a major role to protect the data or sensitive information which organizes needs to protect. [5] .

This paper examines the phenomenon of data leakage, detection and its impact on organizations. Data leakage may be defined as the illegal transfer of valuable/sensitive data by an entity to unauthorized entities. Data leakage detection is the process of finding the data leaker by using various techniques ranging from interrogation, watermark/fake data addition to other modern techniques.[6] Data loss, which means a loss of data that occur on any device that stores data. It is a problem for anyone that uses a computer. Data loss happens when data may be physically or logically removed from the organization either intentionally or unintentionally. The data loss has become a biggest problem in organization today where the organizations are in responsibility to overcome this problem. Data Leakage is an incident when the confidentiality of information has been compromised. In fact, the data deduplication technique, which is widely adopted by current cloud storage services like Dropbox, is one example of exploiting the similarities among different data chunks to save disk space and avoid data retransmission It identifies the same data chunks by their fingerprints which are generated by fingerprinting algorithms such as SHA-1, MD5. Any change to the data will produce a very different fingerprint with high probability. [7]

III. Analysis of Problem

In the current the different information pressure methods accessible however picking the right procedure of pressure is a lot of significant on the grounds that the proficiency will fluctuate from one calculation to another so it is important to get the variation information examination and pressure cost in single component is an excessive amount of significant. [8]It ought to likewise affect in variety of the information with the goal that the administration of information will get improve to considerably more degree. Delicate information in organizations and associations incorporate protected innovation (IP), monetary data, patient data, individual Visa information, and other data relying upon the business and the business. Information spillage represents a major issue for organizations as the quantity of episodes and the expense for those encountering them keep on expanding. [10]Information spillage is improved by the way that communicated information (both inbound and out-bound), including messages, texting, site structures, and document moves among others, are generally unregulated and unmonitored en route to their objections.

IV. Proposed work and Objectives

In this we are proposed a powerful execution module in which the framework will chip away at dynamic choice of the pressure strategy too assessment of examination apparatuses which make the client more valuable and developed with more angle so this framework will assists the client with performing effective investigation, serialization and pressure in view of inputted information work. Furthermore, the cloud streamlining finished with anticipation of spillage. In which the entire information is get put away in encoded structure. It will gain admittance to the proprietor as it were.

Objective:-

1. To apply encryption over the data in storage
2. Data sharing with encrypted format
3. Lossless data compression technique implementation
4. Apply data acquisition in cloud.

V. System Design

System implemented by using SHA 1 algorithm and Huffman coding algorithm.

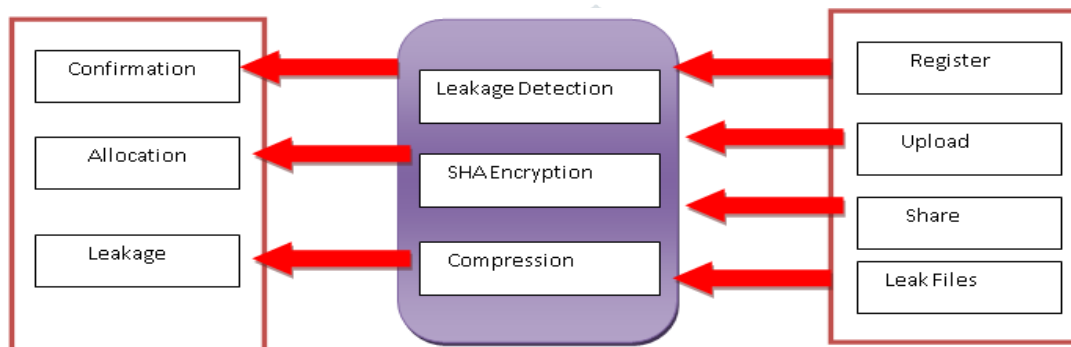
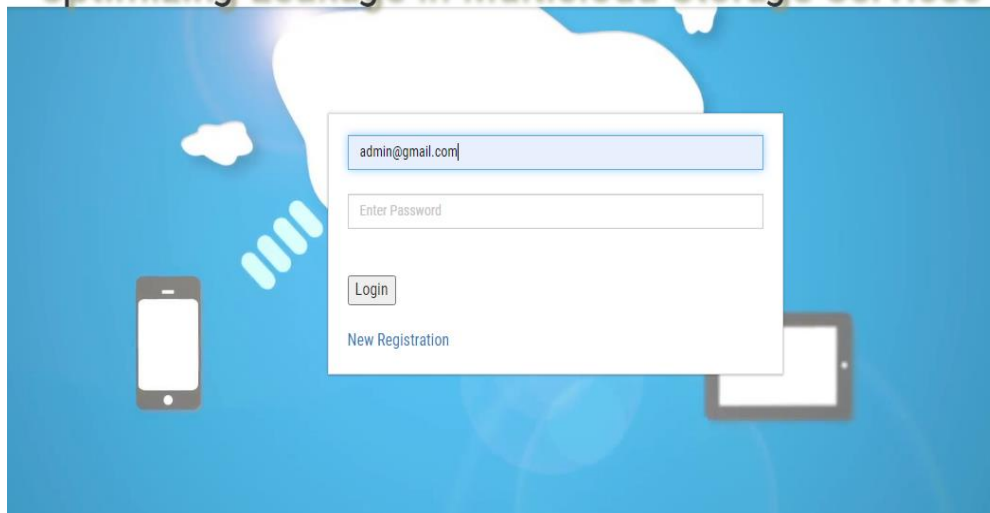


Fig. 5.1 Proposed System Architecture

According to the proposed framework seen there is two primary modules Admin and User. Our first module is User. Client can perform different activities our framework .first and foremost the client register itself by utilizing new client enlistment process. After that client can perform transfer information document in network, client can send the information or message to another client that client can be inside organization . Presently the Most significant second module is Admin. here administrator play out the principle job of our project. In our framework administrator have authorities to play out all activity in our framework. Administrator can make the hubs that is networks. Admin perform principle task our framework. Whenever the clients are enrolled itself administrator check all the data and affirmed the client registration. Well after affirmation of client administrator allocated cloud to that client .the cloud which are giving the stockpiling of client for putting away their information documents and data. In our framework administrator make different cloud for putting away data which are apportioned to client for putting away the data and offer the data to other client. By sharing data some time data leaking. Here we are figured out the spillage. Here we are utilizing two calculations initially is SHA1 Algorithm and second is Huffman coding calculations. For observing spillage in framework we will figure out the size of information and data. The size of information which are showing contrasts the size then we tracked down the spillage.

VI. Implementation

Optimizing Leakage in Multicloud Storage Services



Screenshot 4.3.1 Admin Login page

Screenshot 4.3.1 shows the admin login area where admin will login using their own credentials such as username and password. Here the admin can perform various tasks like confirmed user registration, create cloud, allot the cloud to the user. And the main aim of our admin is to find out leakage.

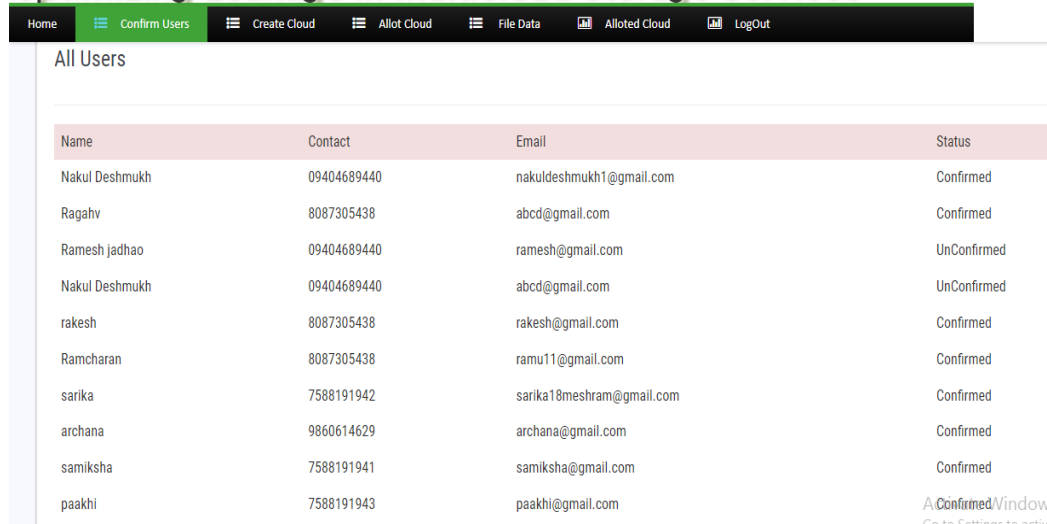
Optimizing Leakage in Multicloud Storage Services

New Registration

Screenshot 4.3.2 New Registration Window

Screenshot 4.3.2 shows the Registration form for new user. If the user is new then he/she first register their name, contact and Email. Then after registration user can login by entering their username and Password.

Optimizing Leakage in Multicloud Storage Services



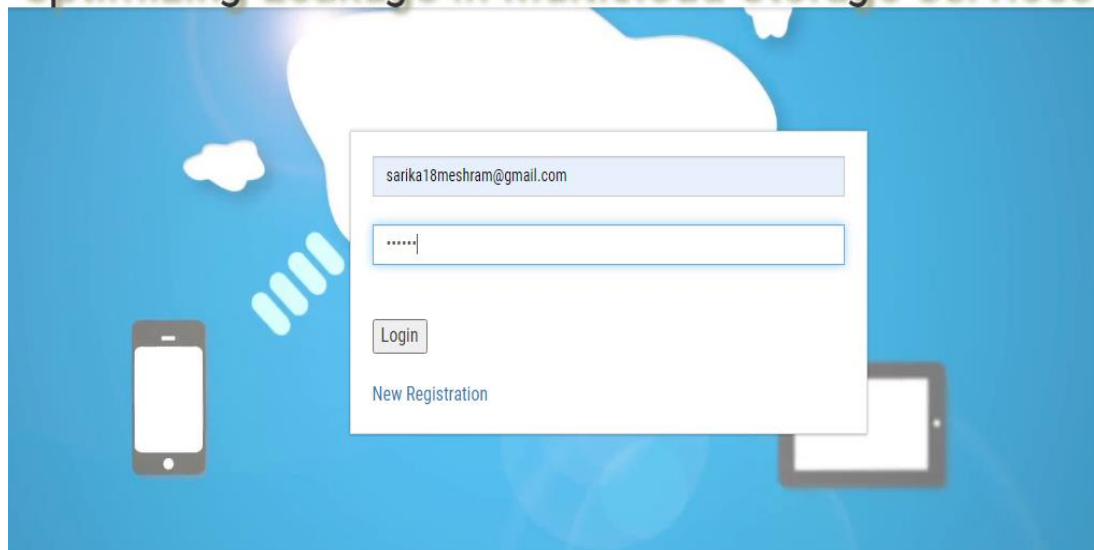
The screenshot shows an admin dashboard with a navigation bar at the top containing: Home, Confirm Users, Create Cloud, Allot Cloud, File Data, Alloted Cloud, and LogOut. Below the navigation bar is a section titled "All Users" containing a table with the following data:

Name	Contact	Email	Status
Nakul Deshmukh	09404689440	nakuldeshmukh1@gmail.com	Confirmed
Ragahv	8087305438	abcd@gmail.com	Confirmed
Ramesh jadhao	09404689440	ramesh@gmail.com	UnConfirmed
Nakul Deshmukh	09404689440	abcd@gmail.com	UnConfirmed
rakesh	8087305438	rakesh@gmail.com	Confirmed
Ramcharan	8087305438	ramu11@gmail.com	Confirmed
sarika	7588191942	sarika18meshram@gmail.com	Confirmed
archana	9860614629	archana@gmail.com	Confirmed
samiksha	7588191941	samiksha@gmail.com	Confirmed
paakhi	7588191943	paakhi@gmail.com	Confirmed

Screenshot 4.3.3 Admin Dashboard

Screenshot 4.3.3 shows the admin dashboard after login, admin can see the list of users. Also can perform some operation on user and nodes such as confirm user, allot cloud, create cloud, file data .After performing all operations admin will logout .

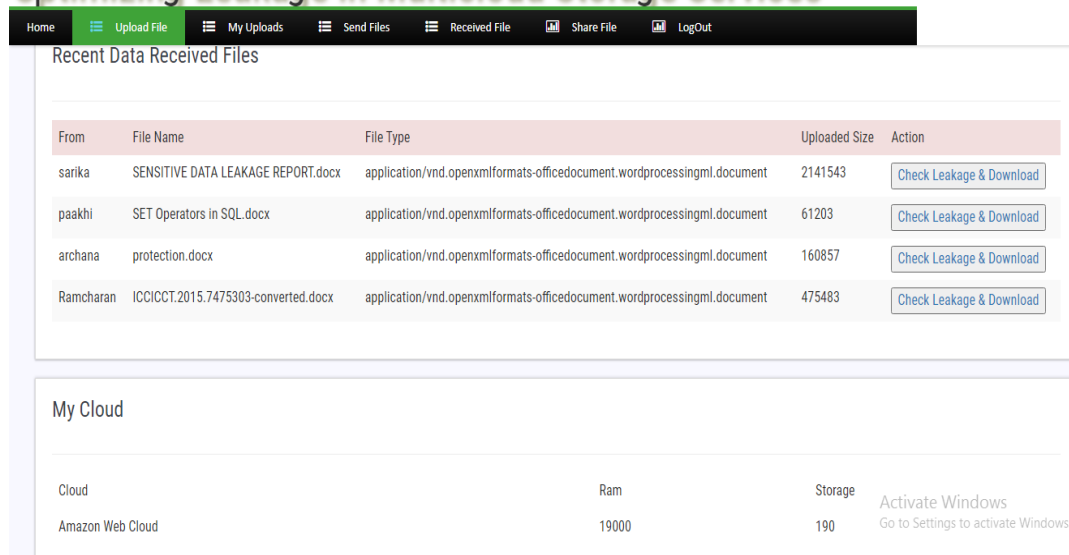
Optimizing Leakage in Multicloud Storage Services



Screenshot 4.3.4 User Login Page

Screenshot 4.3.4 shows the user login area where user can login by using their credentials such as username and password. And then perform their task after login.

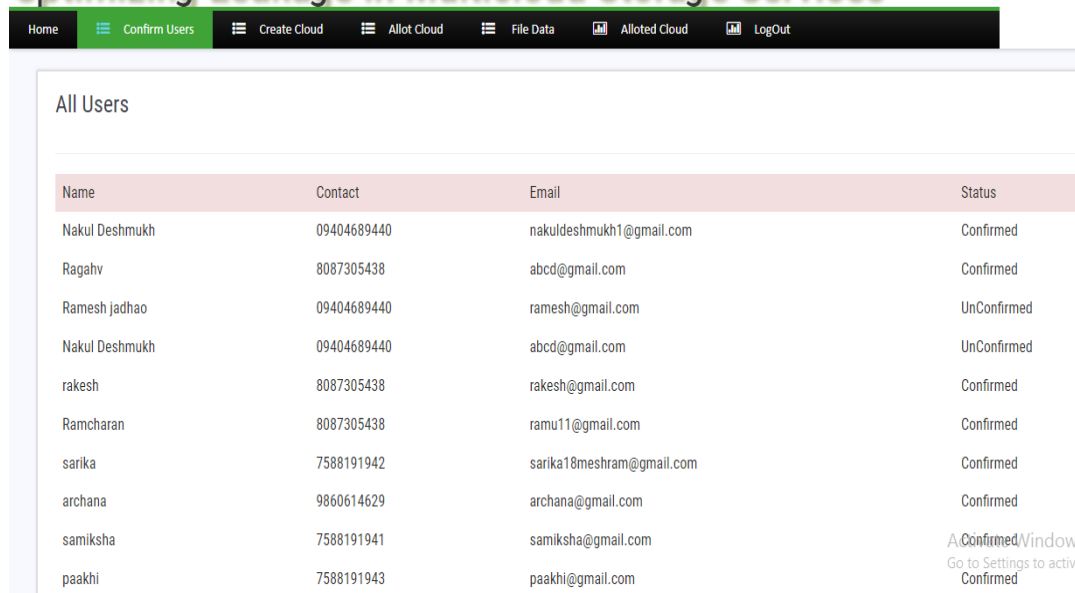
Optimizing Leakage in Multicloud Storage Services



Screenshot 4.3.5 User Dashboard

Screenshot 4.3.5 shows the user dashboard after login User. User can recently received data and allotted cloud information. And perform multiple operations such as, send file ,upload file ,share file etc.

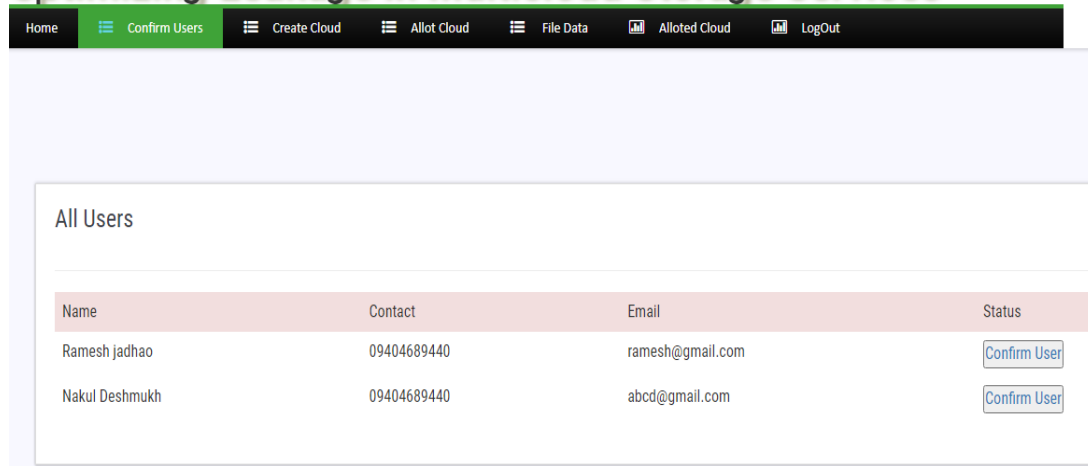
Optimizing Leakage in Multicloud Storage Services



Screenshot 4.3.6 view window

Screenshot 4.3.6 shows the view node window where admin can see all the user registration after adding new user.

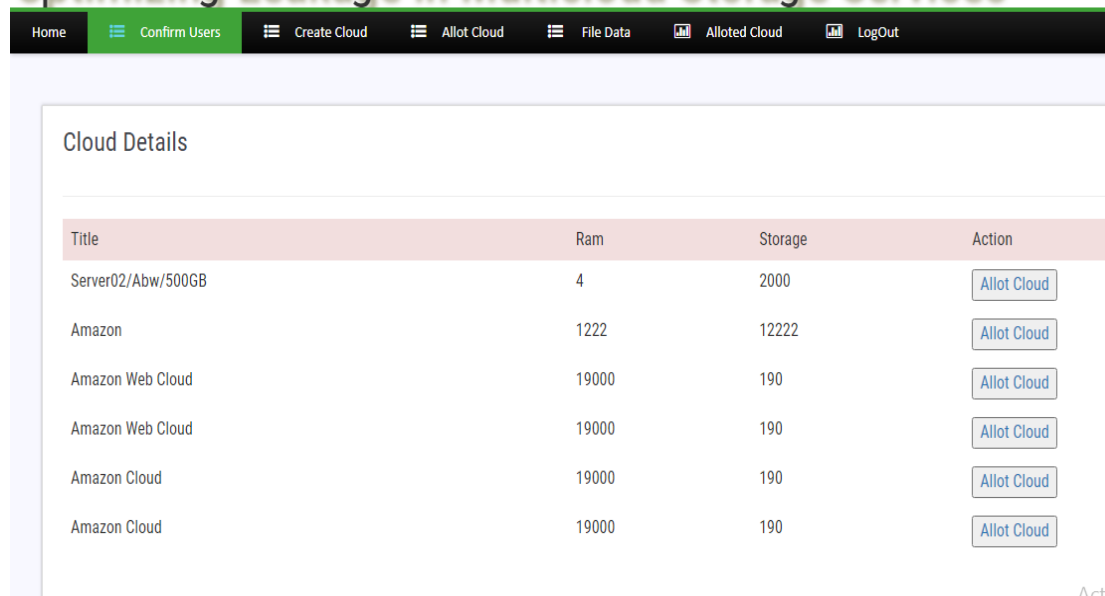
Optimizing Leakage in Multicloud Storage Services



Screenshot4.3.7 Confirm User Window

Screenshot 4.3.7 shows the confirm user window where admin can confirm user.

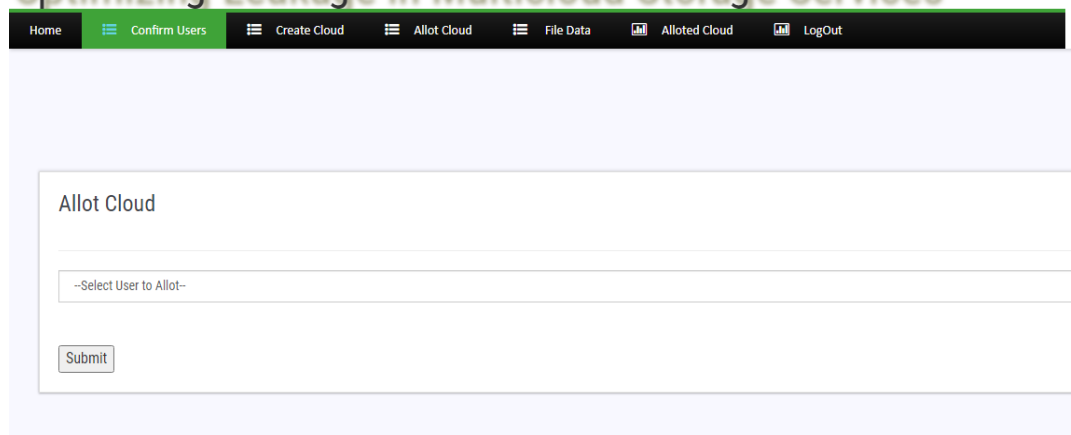
Optimizing Leakage in Multicloud Storage Services



Screenshot 4.3. 8 Allot Cloud Window

Screenshot 8 shows the allot cloud window where admin can allot cloud by pressing the option allot cloud.

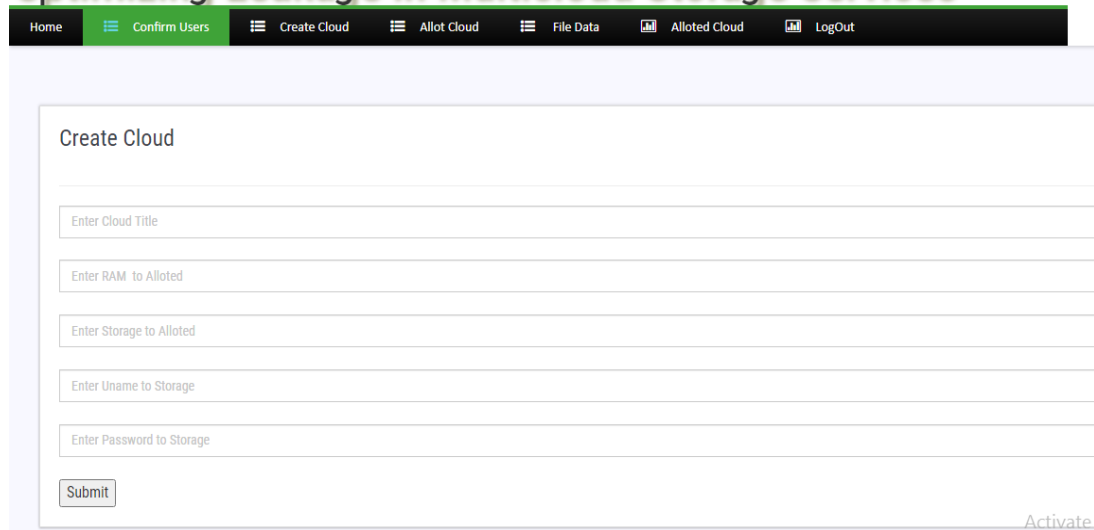
Optimizing Leakage in Multicloud Storage Services



Screenshot 4.3.9 Cloud Allotted Process Window

Screenshot 4.3.9 shows the window cloud allot process where admin can allot the cloud to selected user.

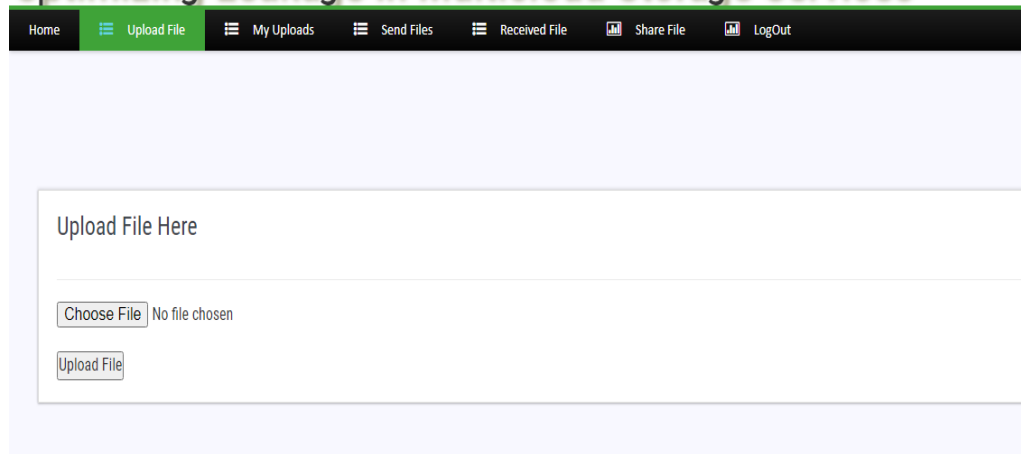
Optimizing Leakage in Multicloud Storage Services



Screenshot 4.3.10 create new cloud window

Screenshot 4.3.10 shows window create cloud window where admin can create new cloud.

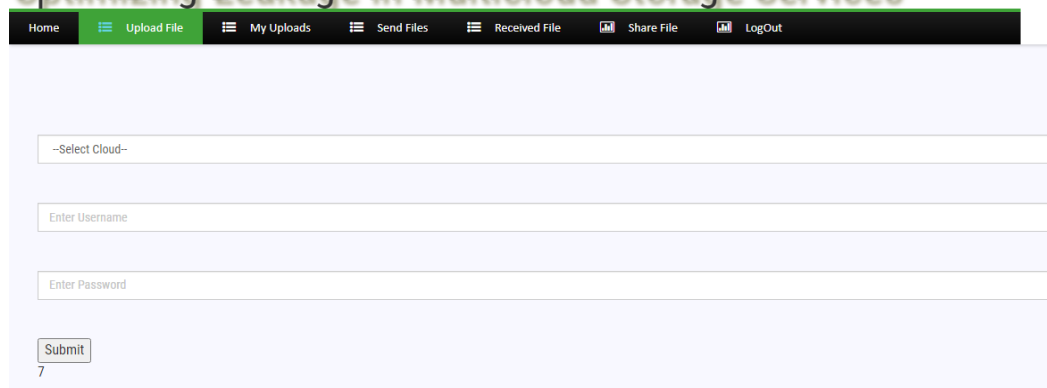
Optimizing Leakage in Multicloud Storage Services



Screenshot4.3. 11 Upload File window

Screenshot 4.3.11 shows the Upload file window where user can the upload files.

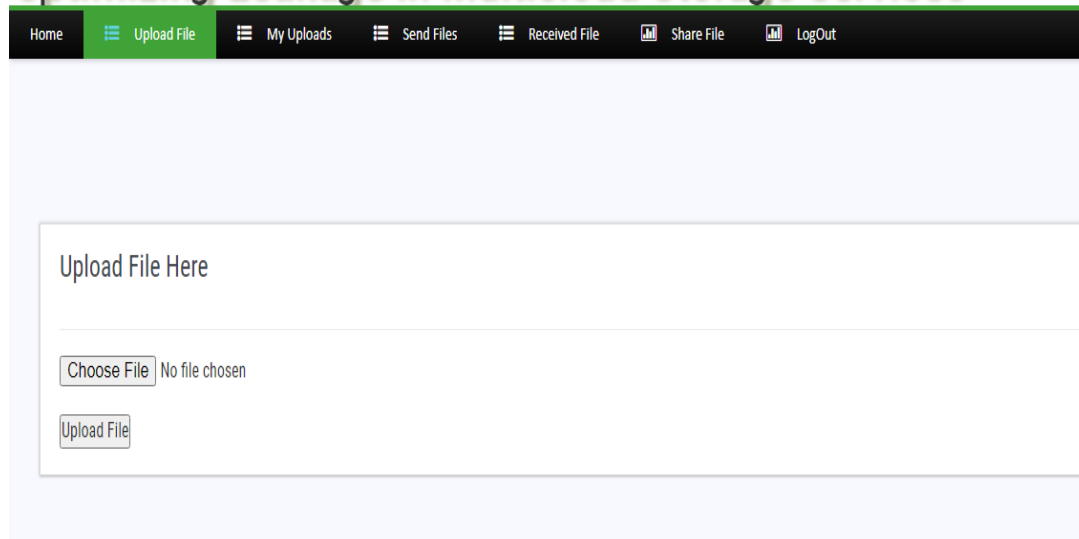
Optimizing Leakage in Multicloud Storage Services



Screenshot 4.3.12 Select Cloud Window

Screenshot 4.3.12 shows the select cloud window where user can save the files.which cloud proteted by using cloud user name and password.

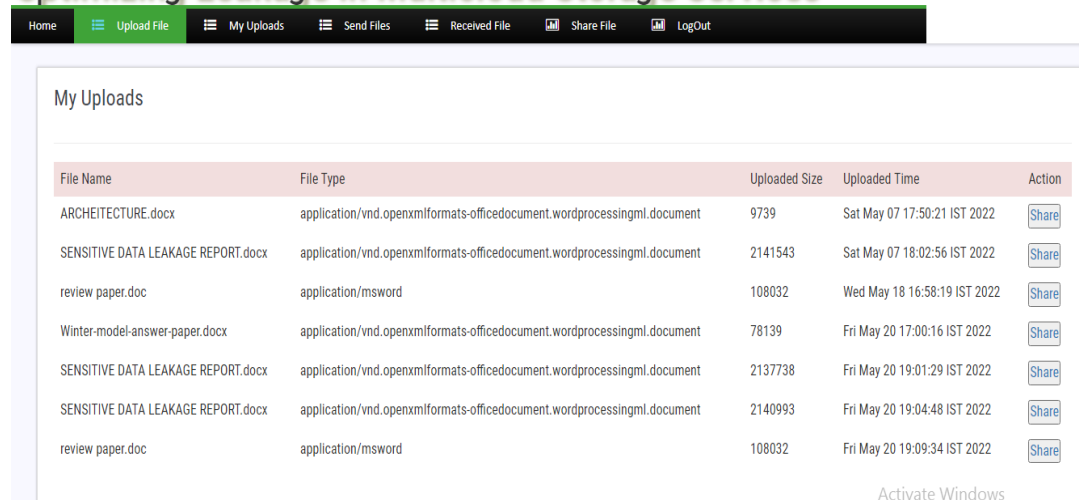
Optimizing Leakage in Multicloud Storage Services



Screenshot 4.3.13 list Upload File window

Screenshot 4.3. 11 shows the list Upload file window where user can see the uploaded files.

Optimizing Leakage in Multicloud Storage Services



Screenshot 4.3.14 Share File window

Screenshot 4.3.14 shows the Share file window where user can share data to any other User.

Data Compression & Data Leakage Detection System

Screenshot 4.3.15 File share Process Window

Screenshot 4.3.15 shows the File share process window where user can share to any other selected user.

Optimizing Leakage in Multicloud Storage Services

Screenshot 4.3.16 received file Window

Screenshot 4.3.16 shows the Received file window where user can check received file.

Optimizing Leakage in Multicloud Storage Services

Screenshot 4.3.17 send file Window

Screenshot 4.3.17 shows the send file window where user can check send file list.

Optimizing Leakage in Multicloud Storage Services

The screenshot shows a web application interface for checking file leakage. It features a navigation bar with options: Home, Upload File, My Uploads, Send Files, Received File, Share File, and LogOut. The main content area is titled 'Check For Leakage' and contains two tables. The first table, 'Check For Leakage', lists a file named 'SENSITIVE DATA LEAKAGE REPORT.docx' with an uploaded size of 2141543 and an uploaded time of Sat May 07 18:02:56 IST 2022. The second table, 'Actual File Size', compares the uploaded file with the received file, showing a difference of -550.0 and a green checkmark in the 'is Leak' column, indicating a leakage was detected.

File Name	File Type	Uploaded Size	Uploaded Time
SENSITIVE DATA LEAKAGE REPORT.docx	application/vnd.openxmlformats-officedocument.wordprocessingml.document	2141543	Sat May 07 18:02:56 IST 2022

File Name	File Type	Uploaded Size	Uploaded Time	Received Size	Diff	is Leak
SENSITIVE DATA LEAKAGE REPORT.docx	application/vnd.openxmlformats-officedocument.wordprocessingml.document	2141543	Sat May 07 18:02:56 IST 2022	2140993.0	-550.0	✓

Screenshot 4.3.18 Check leakage Window

Screenshot 4.3.18 shows the Check Leakage Window where user can check leakage. After receiving the file from user. Receiver check the file. Compare the size of file. Here receiver found any difference between sizes then receiver recognize leakage and found the leakage.

Optimizing Leakage in Multicloud Storage Services

The screenshot shows the same web application interface as Screenshot 4.3.18. The 'Check For Leakage' table lists a file named 'SET Operators in SQL.docx' with an uploaded size of 61203 and an uploaded time of Fri May 20 17:19:53 IST 2022. The 'Actual File Size' table compares the uploaded file with the received file, showing a difference of 0.0 and a red 'X' in the 'is Leak' column, indicating no leakage was detected.

File Name	File Type	Uploaded Size	Uploaded Time
SET Operators in SQL.docx	application/vnd.openxmlformats-officedocument.wordprocessingml.document	61203	Fri May 20 17:19:53 IST 2022

File Name	File Type	Uploaded Size	Uploaded Time	Received Size	Diff	is Leak
SET Operators in SQL.docx	application/vnd.openxmlformats-officedocument.wordprocessingml.document	61203	Fri May 20 17:19:53 IST 2022	61203.0	0.0	✗

Screenshot 4.3.19 Check leakage Window

Screenshot 4.3.19 shows the Check Leakage Window where user can check leakage. After receiving the file from user. Receiver check the file. Compare the size of file. Here receiver found any difference between sizes then receiver recognize leakage and found the leakage.

VII. CONCLUSION

This is to be infer that the carried out framework is more productive when contrasted with the given paper wherein the pressure strategy and anticipation method is get executed. In this the information spillage is add highlight remembered for proposed so the executed strategy is more successful. Appropriating information on numerous mists gives clients a specific level of data However, spontaneous dispersion of information piece examine lead to avoidable data spillage. We show that circulating information lumps in a cooperative manner can release client's information as high as 80% of the all out data with the proposed instrument will be useful in part reach out to build the information security in the cloud and work on the procedure of information spillage. In this the spillage information can be exercise all the more really when contrasted with existing framework and spillage control in that no single cloud supplier is protection to the whole client's information.

VIII. FUTURE SCOPE

In future we can use the cloud storage for storing our data and no need to worry about security issues. As the world grows technically day by day we require more security for the data and in future we can deploy this system for many organization and maintain privacy of all confidential documents so we can reducing the data leakage problems. Fast data communication will get implemented and Data acquisition get will extends the level to more extends The system which are integrated with high level of Security.

IX ACKNOWLEDGMENT

I owe my deep gratitude to my guide **Dr. S. R. Gupta (Assistant professor)**, who took keen interest on my dissertation work and guided us all along, till the completion of this work by providing all the necessary information for developing a good system.

REFERENCES

- [1] Daoqu Geng, Chengyun Zhang, Chengjing Xia, Xue Xia, Qilin Liu, Xinshuai Fu, "Big Data Based Improved Data Acquisition and Storage System for Designing Industrial Data Platform" Received March 13, 2019, accepted March 22, 2019, date of publication April 3, 2019, date of current version April 15, 2019.
- [2] Klaus Lyko, Marcus Nitzschke, and Axel-Cyrille Ngonga Ngomo(2016) "Big Data Acquisition" Shivakumara T1*, Rajshekhar M Patil2, Muneshwara M S3 " Review Paper on Dynamic Mechanisms of Data Leakage Detection and Prevention" Vol.-7, Issue-2, Feb 2019
- [3] Hao Zhuang, Member, IEEE, Rameez Rahman, Pan Hui, Member, IEEE, and Karl Aberer, Member, IEEE "Optimizing Information Leakage in Multicloud Storage Services" IEEE TRANSACTIONS ON CLOUD COMPUTING, VOL. 14, NO. 8, JANUARY 2016.
- [4] Bijayalaxmi Purohit, Pawan Prakash Singh "Data leakage analysis on cloud computing" Vol. 3, Issue 3, May-Jun 2013
- [5] R.Senthil kumar1, S.Kaviya priya2, M.Manimegalai3, P.Logesh4 "Privacy Preservation With Data Leakage Detection & Monitoring" Vol 6 Issue 3 March -2018
- [6] K. Manoj Kumar+ G. Shubhang+ G.Rajesh Chandra "DATA LEAKAGE DETECTION SYSTEM FOR CLOUD-BASED STORAGE SYSTEMS" Vol. No. 8, Issue No. V, November 2014.
- [7] Prof. Dipti Mathpal, Prof. Mittal Darji, Prof. Sarangi Mehta "A Research Paper on Lossless Data Compression Techniques" IJIRST –International Journal for Innovative Research in Science & Technology, Volume 4, Issue 1, June 2017.
- [8] H. Chen, Y. Hu, P. Lee, and Y. Tang, "Nccloud: A network-coding-based storage system in a cloud-of-clouds," 2013.
- [9] B. Hauer, "Data and Information Leakage Prevention Within the Scope of Information Security", december 16, 2015.
- [10] Nina Pearl Doe 1, Sumaila Alfa 2, V. Suganya, "Efficient method to prevent cloud" Volume 16, Issue 3, Ver. III (May-Jun. 2014)