ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

Improved Efficient Data Classification Model for Ensuring Data Security in Cloud Environment

Meenal Jain, Ashok Verma

Research Scholar, Software System, Dept. of Computer Science and Engg, Gyan Ganga Intt. of Tech. and Science Jabalpur (M.P.), India

Professor, Computer Science & Engg, Dept. of Computer Science and Engg, Gyan Ganga Intt. of Tech. and Science Jabalpur (M.P.), India

ABSTRACT: Cloud computing assumes a stimulating job by golf stroke away the knowledge ANd it o.k. could also be masterminded by an outsider. the $64000 disadvantage within the cloud field is protection and security problems. one amongst the first problems is that the data security and protection of knowledge place away and ready at the cloud specialist co-op's frameworks. In spite of each one amongst these administrations given by cloud, it slacks within the real aspect of security. Distributed computing primarily assumes employment within the a part of powerful quality usage and administration utilization. freelance of the type of mists every specialist organizations focuses on the knowledge abode in cloud servers. Be that because it might, shoppers still have vital security and protection worries regarding their re-appropriated data as a results of conceivable unapproved access within the specialist organizations. this arrangements inscribe all information utilizing the distinctive key size while not mulling over the privacy dimension of data. during this analysis, we tend to propose a protected distributed computing model addicted to data order. The projected cloud model limits the overhead and getting ready time expected to safeguard information through utilizing numerous cryptography elements with variable key sizes to grant the correct secrecy level needed for the knowledge. during this exploration, a secure distributed computing model addicted to data grouping is projected. The increasing volume of individual and crucial information raises a lot of spotlights on golf stroke away the knowledge safely. data will incorporate basic shopper knowledge, vital reports, and alternative client's connected substance for arrangement of client's data. characterised data are disorganized by utilizing numerous AES cryptography calculations like AES 128, AES 192 and AES 256. finally disorganized data are place away distinctive style of cloud model as indicated by their information classification. The projected model is tried with numerous metric, informational collections and cryptography calculations. Re-enactment results incontestible the responsibility and proficiency of the projected structure.

KEYWORDS: Cloud Computing; Data Security, Secure Cloud Storage, Encryption, Cryptography, Multilevel Data Classification, AES

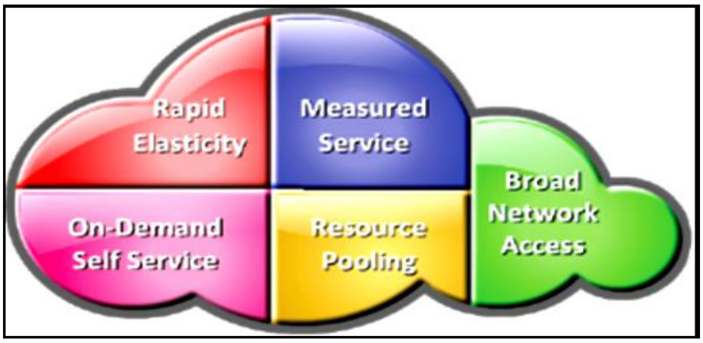
I. INTRODUCTION

Cloud computing is a standout amongst the most utilized innovation in this day and age. Distributed computing just depicts that it is a hard circle that given to the client by some client certifications, which the client can get to the information and store the information through web. In the cloud the client can build the limit of the capacity, so enormous measure of information can be put away. Cloud frameworks comprise of three unique layers Saas, Paas, Iaas which gives various types of administrations. These days, distributed computing is a developing territory that includes wide scope of new advances and applications that contacts pretty much every house occupants. They additionally encourage information sharing among clients and synchronization of different gadgets. Be that as it may, there is indispensable information that is prepared and put away in the cloud frameworks. Losing or uncovering this important information will have colossal terrible effect on the information proprietors being people or associations. Thus, there is an expanding request to secure information over the cloud frameworks. Clients dread from transferring private and secret documents to the online reinforcement because of worries that the specialist organization may utilize them improperly. Adding to that, there are worries about their information being hacked and bargained because of the spread of distributed storage fruitful assaults. Treating the low and high secret information by a similar route and at a similar security level will include pointless overhead and increment the handling time.

1.1 Characteristics of Cloud Computing

NIST further specifies that cloud computing exhibits the following five characteristics in its operation

- On-Demand Self Service
- Broad Network Access
- Measured Service
- Rapid Elasticity
- Resource Pooling



Figre 1.1 Characteristics of Cloud Computing

i. On demand Self Service

On interest self-administration implies that clients can get to the functionalities and administrations without reliance. Customer can give abilities of processing, for example, server time; arrange capacity, programmed Information Technology (IT) assets that require no human collaborations. The client can likewise change or trim the functionalities and arrangement its offices dependent on the client's needs.

ii. Broad Network Access

This implies assets can be gotten to in stages like cell phones, workstation, and work area from any far (remote) area over the system. Organizations incline toward private cloud administration since they will be applicable about their data spills in external territory.

iii. Measured Service

The significant need of distributed computing is the deliberate administration. This is reference to administrations where the cloud supplier advises the arrangement of administrations for some sorts of reason including charging, successful utilization of assets, or arranging expectation all through the framework.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

iv. Rapid Elasticity

Quick flexibility is one of the fundamental key qualities of distributed computing, where the capacities can be flexibly orchestrated and discharged naturally dependent on the internal and outward on relating request. At the end of the day, the client can increment or decline the product necessities as per his need which will be cost effective.

v. Resource Pooling

The assets of cloud are pooled and exhibited to more than one client suing the Multi-Tenant Model. This Multi-Tenant model caused the client to have a similar corporal equipment to more people so the client to can utilize the product with less budgetary expenses. Numerous mirror duplicates are put away and reflecting procedure is done to diminish the loss of information [2].

II. DATA SECURITY ISSUES OR CHALLENGES

In big business figuring, information is put away inside their association and it is completely under the control of the venture [5]. In distributed computing, the information is put away outside the client's place (in the CSP's side). Thus, distributed computing must utilize extra safety efforts separated from the customary security checks to guarantee that information is sheltered and no information breaks because of security vulnerabilities.

2.1. Data Security Basics

There are six phases in the existence cycle of information: Create Store, Use, Share, Archive and Destroy. When the information is made, it can move openly between any stages. Information ought to be verified in every one of an amazing phase's cycle from its creation to its annihilation. One of the disregarded issues is information after-erase [1] and this is additionally called as information remanence.

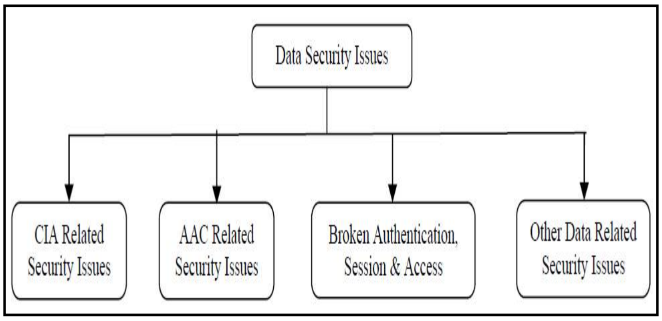

Data Security Issues
CIA Related Security Issues
AAC Related Security Issues
Broken Authentication, Session & Access
Other Data Related Security Issues

Fig. 1.2 Types of Data Security Issues

2.2. Security Challenges in the CIA Triad

Privacy, Integrity and Availability (CIA) misfortunes can have a major effect in the matter of the distributed computing in light of the fact that the information is the canter part for any business. Information honesty is the confirmation given to the computerized data is uncorrupted and just be gotten to by those approved clients. In this way, uprightness includes keeping up the precision, consistency and dependability of information over its whole life cycle [6].

2.3 Security Challenges in the Authentication and Access Control (AAC)

Validation and Access Control (AAC) is the procedure of check and affirmation on client's personality to interface, to access and utilize the cloud assets. In big business processing, the accreditations are put away in the server as Active Directory (AD) or Lightweight Directory Access Protocol (LDAP). AAC security challenges:

- Apply single-sign-on approach any place conceivable.

• Multi-factor verification can be utilized which empowers both character and access the board and it is utilized Amazon Web Services (AWS).
• Biometric verification can possibly be the most secure type of single-sign-on validation.
• RSA cryptosystem can be embraced which can acknowledge diverse validation models like two-factor verification, information based confirmation and versatile validation. The above strategies are extremely successful for information insurance in distributed computing.
• Intrusion Detection System (IDS), firewalls just as isolation of commitments can be executed on the distinctive system and cloud layers to empower legitimate access control in distributed computing for better information assurance.
• There are some outsider personalities the executive’s arrangements in the market. Utilize some outsider arrangements like Microsoft Azure Active Directory, Okta character the board, McAfee cloud personality chief, and so forth. As of late, Identity-Management-as-a-Service (IDaaS) arrangements are getting increasingly mainstream in the corporate framework.
• Cloud applications should utilize open principles where relevant, for example, Security Assertion Markup Language (SAML), a XML-based OASIS (Organization for the Advancement of Structured Information Standards) open standard for trading validation and approval information between security spaces and Open Authorization (OAuth), an open standard for approval, enabling clients to share their private assets utilizing tokens as opposed to qualifications.

2.4. Other Data Related Security Issues

Other minor information related security issues can happen through Data area, Multi-occupancy and Backup in distributed computing. Coming up next are a portion of the strategies to defeat other information related security issues:
• CSC should know the sensible and the physical area of the information if not in any event which state, nation and server farm because of all the potential administrative, authoritative and other jurisdictional issues.
• Establish area and jurisdictional strategies to administer the information area.
• Adopt clever information isolation strategies to isolate the information from various clients.
• Use solid encryption procedures for the reinforcement information to dodge information spillage.

III. DATA CLASSIFICATION

Information grouping is a productive strategy to ensure the information as indicated by its significance and affectability. In this area, we present four order levels (high, medium, low, ordinary) as per the diverse degree of the potential side impact on authoritative activities, institutional resources, or people.
□ Confidentiality: A shortage of privacy is the **unauthorized leak** of data.
□ Integrity: A scarcity of integrity is the **unauthorized alteration or damage** of data
□ Availability: A scarcity of availability is the **commotion of access** to or use of data.

3.1 High sensitivity

On the off chance that the missing of CIA of information is foreseen to have a serious or damaging symptom, the grouping of information is high affectability. Genuine or dangerous outcomes show, for example: (I) The serious reduction or loss of mission capacities results in the association losing its capacity to satisfy its essential capacity. (ii) Serious harm to design resources. (iii) Significant budgetary misfortune. or on the other hand (iv) r Serious or disastrous harm to people. Instances of high affectability information incorporates yet not constrained:
1) Core business data: Core business data is the unmistakably significant data for association, the loss of which is significant on the mission, money related, notoriety, etc. For instance:
a) The mid-long haul improvement projects and exceptional plans
b) Safety insurance conspire
c) Key data for significant framework
d) Core organize topology

2) Personal delicate data: Personal touchy data is firmly identified with individual rights or interests, the exposure or maltreatment of which may imperil property security and life, unfriendly effect on close to home notoriety, prompting prejudicial treatment and other delicate data. For instance:
a) Identification card number, international ID number, and so on.
b) Credit/check card data and financial balance data
c) Protected Health Information
d) Electronic marks, biometric data, secret word data, and private encryption keys e) Criminal foundation.

3) Industry-explicit touchy data: From viewpoint of national security, some Industry-explicit delicate data, for example, oil, gas, coal and power plant related data ought to be ensured carefully. For instance:
a) Oil and gas creation data
b) Petrochemical industry significant creation materials import plan and outside trade sum c) Distribution of Urban Power Network Pipeline
d) Power transmission and change gear unwavering quality parameter.

3.2 Medium sensitivity
At the point when the loss of information of CIA is foreseen to have a genuine symptom, the characterization of information is medium affectability. An extreme symptom demonstrates that, for example:
(i) The noteworthy reduction of mission capacities to satisfy its essential capacity and the adequacy of the capacities is decreased
(ii) Substantial harm to institutional resources;
(iii) Substantial budgetary misfortune; or
(iv) Huge harm to people that includes neither under mortality nor lethal wounds. Instances of medium affectability information incorporate however not constrained:

1) Important business data:
Significant business data is the data the loss of which will result in genuine antagonistic effect on the mission, monetary, notoriety, etc. For instance:
a) Important pointer of advertising activities
b) Operation observing centre markers and operational information
c) Computer programming source code
d) Important business understandings or contracts and the applicable data

2) Personal typical data:
Individual data b(Personal data is data that is recorded electronically or something else, which can be utilized alone or in mix with other data to recognize natives, including however not constrained to residents' names, birth dates, character report numbers, individual biometric data, phone number, etc.) is partitioned into individual touchy data and individual typical data. Therefore, if individual data isn't touchy, it might be typical data. For the most part, individual typical data can't distinguish an individual alone, yet in blend with other data together.
a) Email address or substance
b) Phone number, place of residence
c) Information used to approve personality, for example, name, date of birth, mother's name, and so on.
d) Other data not named as delicate data (organize client account, race, ethnicity, conjugal status).

3.3 Low sensitivity
At the point when the loss of information of CIA is foreseen to have a constrained reaction, the characterization of information is low affectability. A constrained reaction demonstrates that, for example:
(I) A decrease in the capacity of doing undertakings to a degree and interim that the association can satisfy its premier capacities, just as perceptibly diminish the viability of the capacities

(ii) Negligible harm to institutional resources;
(iii) Trivial money related misfortune; or
(iv) Minor harm to people.

1) Normal business data:
Typical business data is the data the loss of which will result in restricted unfavorable effect on the mission, monetary, notoriety, etc. For instance:
a) Company address book
b) Final records and review report PC of ventures
c) Important business understandings or contracts and the significant data
d) Employee execution survey data

2) Personal administration data:
Individual administration data is the information and substance data that the association gathers in the administration procedure with individual security qualities. For instance:
a) Payment recodes, for example, gas, power, cell phone bill, and so on.
b) Device data, for example, Meter, power gathering terminal
c) Service content data, for example, business charges, client value program, client value methodology, and so forth.

3.4 Normal sensitivity

In the event that information is relied upon to be open, the arrangement of information is ordinary affectability. The greater part of the data from the web is typical delicate data.

IV. PROPOSED METHODOLOGY

Data Classification is the way toward characterizing different information levels and choosing a dimension of affectability to it. It is a basic action at different stages as it is being made, adjusted, put away, or transmitted. In a distributed computing condition information resource is essential relying upon the business and the administration conveyance models. To give the controlled access and approval, ordering information dependent on security level criteria getting to be zone of enthusiasm by numerous associations utilizing or giving cloud administrations. Here we have contemplated a lot of groupings given in the writing and recognized a lot of parameters dependent on the security necessities for cloud information. We have broke down certain informational indexes that can be utilized to give the security dependent on their utilization and access control as for distributed computing condition. Information grouping is the way toward recognizing information components concerning its incentive in the business. Esteem is distinguished dependent on their use and access control confinements.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

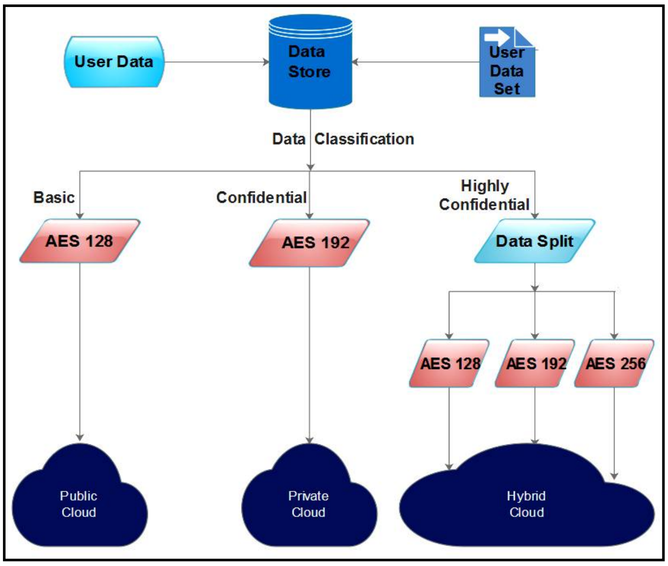


Figure 4.1 Proposed Methodology

Figure demonstrates the three sorts of attributes on which information must be characterized and in like manner security contemplations can be connected. Information grouping approach can be considered and its adequacy can be dictated by re-enacting on an example informational collection. Re-enacting the order of information dependent on close to home informational collection is broke down. Individual information components like name, addresses and so forth are taken as the example. We have utilized the emotional criteria to order them and in like manner security arrangements for the capacity and interchanges can be joined.

4.1 Execution Steps of Proposed Methodology

Step 1 – User data is stored in local data storage.
Step 2 – A user data set which is a set of user's information (like user's name, date of birth, address, Aadhar number, Driving License Number, Area PIN number, Favorite sports, Favorite Color, School name, University Name etc..) is stored in local level for data classification scheme.
Step 3 – Data Classification will take place for separation of user data according to data set which was provided by user at the time of registration.
Step 4 – Data will be classified as Basic Data, Confidential Data and Highly Confidential Data by using Data Set.
Step 5 – Now data will be encrypted by using Advanced Encryption Standard (AES) Encryption technique:-
I. Basic type of data will be encrypted by AES-128.
II. Confidential type of data will be encrypted by AES-192.
III. Highly Confidential type of data splits in 3 parts, First part will be encrypted by AES-128, Second part will be encrypted by AES-192, and Third part will be encrypted by AES-256.
Step 6 – After encryption, Basic type of data will be stored in Public Cloud, Confidential type of data will be stored in Private Cloud, and Highly Confidential type of data will be stored in Public Cloud.

V. IMPLEMENTATION AND RESULTS

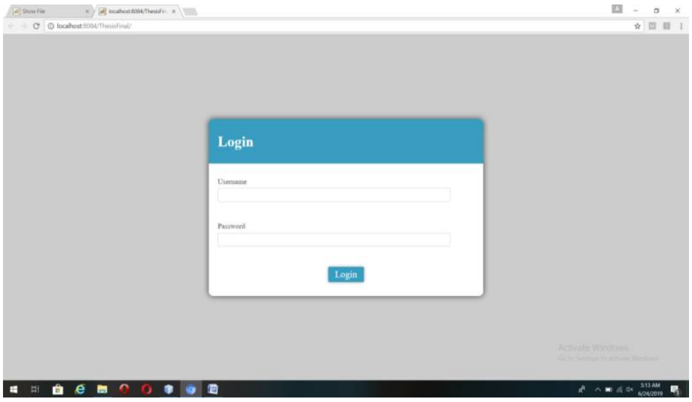


Figure 5.1 User Login Form

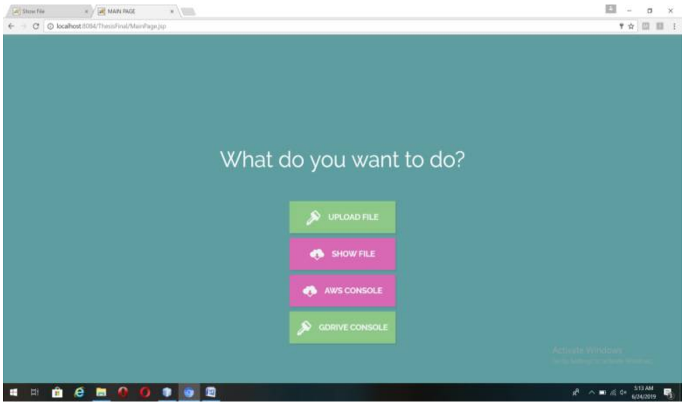
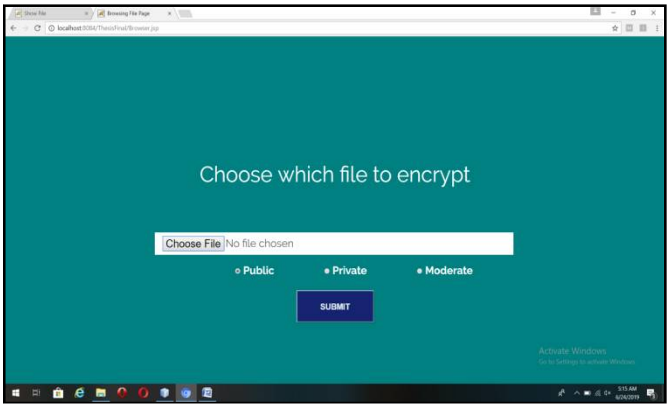

What do you want to do?

Figure 5.2 Options for File Operations


Choose which file to encrypt

Choose File No file chosen

○ Public ● Private ● Moderate

Figure 5.3 Upload File for Public Cloud

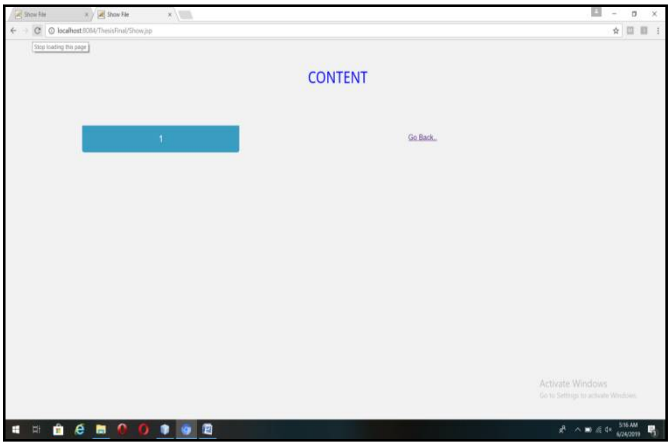


Figure 5.4 Show uploaded file in public cloud environment

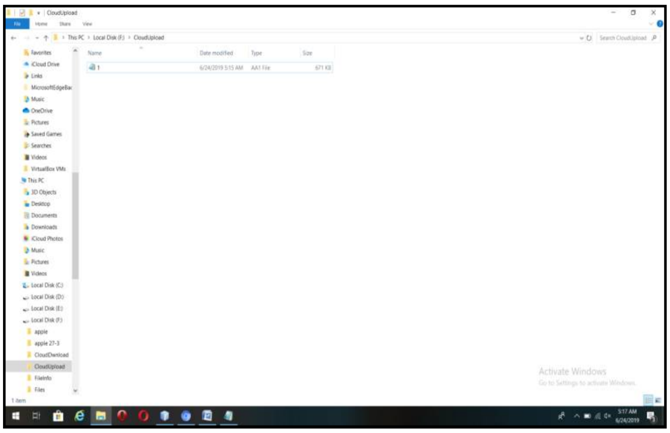


Figure 5.5 Encrypted File

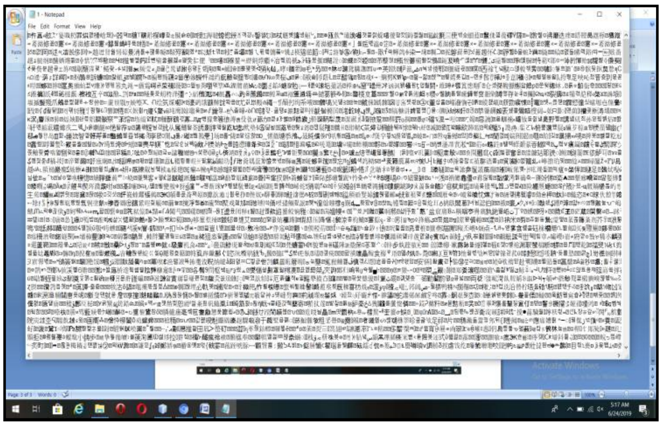


Figure 5.6 Encrypted Data

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

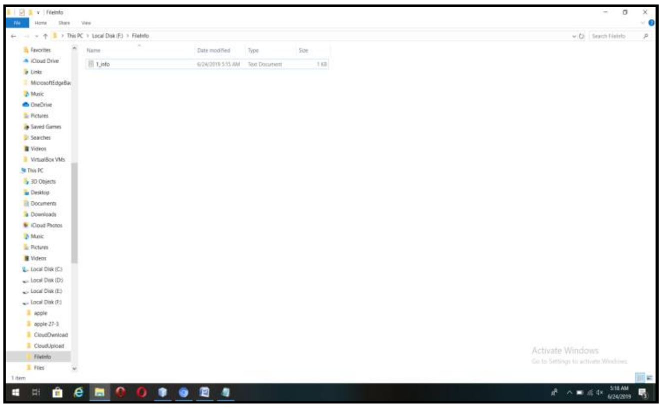
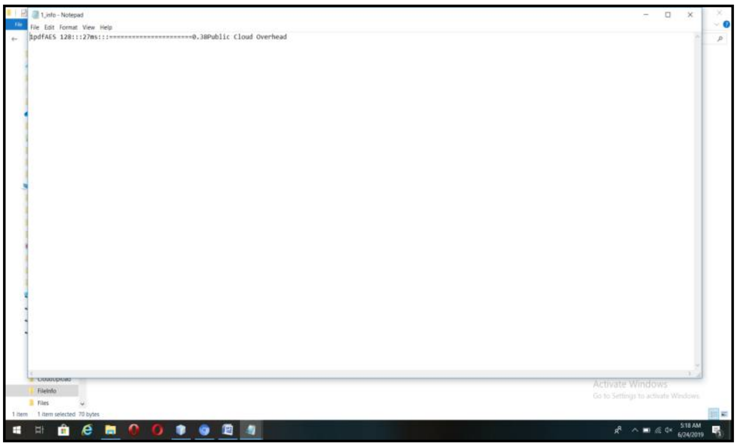
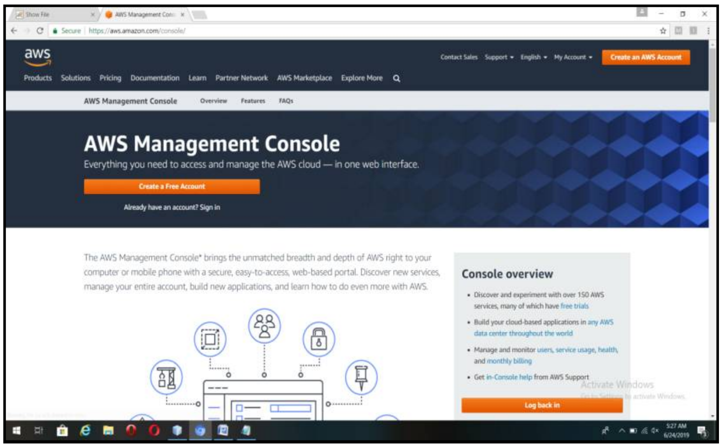


Figure 5.7 AES 128bit Encrypted file's information file



```
1pdfAES 128:::27ms:::======================0.38Public Cloud Overhead
```

Figure 5.8 Showing total executions overhead in public cloud



AWS Management Console

Everything you need to access and manage the AWS cloud — in one web interface.

The AWS Management Console* brings the unmatched breadth and depth of AWS right to your computer or mobile phone with a secure, easy-to-access, web-based portal. Discover new services, manage your entire account, build new applications, and learn how to do even more with AWS.

Console overview

- Discover and experiment with over 150 AWS services, many of which have free trials
- Build your cloud-based applications in any AWS data center throughout the world
- Manage and monitor users, service usage, health, and monthly billing
- Get in-Console help from AWS Support

Figure 5.9 AWS cloud environment for Private cloud and Hybrid Cloud

DOI: 10.15680/IJIRCCE.2019. 0704050

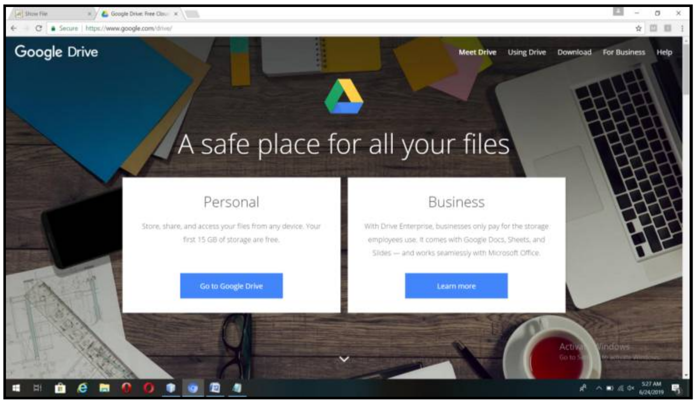
ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019



A safe place for all your files

Personal

Store, share, and access your files from any device. Your first 15 GB of storage are free.

Business

With Drive Enterprise, businesses only pay for the storage employees use. It comes with Google Docs, Sheets, and Slides — and works seamlessly with Microsoft Office.

Figure 5.21 Google Drive cloud environment for Public cloud and Hybrid Cloud

5.1 COMPARISON

Comparison between Existing system and proposed system.

5.1.1 Comparison based on Implementation

CRITERIA	EXISTING SYSTEM	PROPOSED SYSTEM
Data Classification – Decision Tree	Business Values (BV)	Data Set
Encryption Methods	**SSE** (Decryption is Not Available)	**PKI** (Decryption is Available)
Implementation Platforms	**MS .Net** (Licensed / Chargeable)	**Java** (Free / Open Source)
Server Scripting Language	**C#** (Licensed Plateform Dependency)	**JSP, Servlet** (Free / Open Source)
Database	**MS SQL** (Licensed Plateform Dependency)	**MySQL** (Free / Open Source)
Server Authentication Mechanism	**Token Based** (Less Secure)	**SSL** (Highly Secure)

Table 5.1 Comparison based on Implementation

5.2.2 Comparison based on Results and Output

Cloud Type	**Existing System (Overhead)**	**Proposed System (Overhead)**
Private Cloud	1	0.9
Public Cloud	0.7	0.6
Hybrid Cloud	0.4	0.38

Table 5.2 Comparison based on Results and Output

DOI: 10.15680/IJIRCCE.2019. 0704050

VI. CONCLUSION

The fundamental point of research is to group the information proficiently and secure that information relying on the security components of the information. This will isolate information into Basic, touchy and exceedingly delicate by improved arrangement system dependent on informational index grouping with information security. In this exploration, AES encryption strategies is utilized with AES-128, AES-192 and AES-256 security standard which gives better information security on arranged information contrasted and existing usage. Open source systems are utilized for usage which is savvy moreover. Distinctive cloud condition like Private Public and Hybrid cloud are executed and utilized for information stockpiling in this exploration. Complete framework is tried with various running conditions and informational indexe. All outcomes and testing demonstrates that the proposed framework is increasingly proficient and secure.

In our future work, focus on the experimental evaluation of the proposed architecture and the test assessment of the proposed engineering and the appraisal of the forecast precision at scale. Additionally investigate the utilization of dispersed AI calculations for this application and evaluate them grouping exactness with progressively secure encryption calculations for better information security.

REFERENCES

[1]. Uma Somani, Kanika Lakhani, Manish Mundra, “Implementing Digital Signature with RSA Encryption Algorithm to Enhance the Data Security of Cloud in Cloud Computing”, 2010 1st International Conference on Parallel, Distributed and Grid Computing (PDGC - 2010), ©2010 IEEE.

[2]. Mr. Prashant Rewagad, Ms.Yogita Pawar, “Use of Digital Signature with Diffie Hellman Key Exchange and AES Encryption Algorithm to Enhance Data Security in Cloud Computing”, 2013 International Conference on Communication Systems and Network Technologies, © 2013 IEEE.

[3]. Kumar Pal Singh, Dr. Vinay Rishiwal, “Classification of Data to Enhance Data Security in Cloud Computing”, © 2018 by IEEE.

[4]. Vanya Diwan, Shubhra Malhotra,Rachna Jain, “ Cloud Security Solutions: Comparison among Various Cryptographic Algorithms”, © 2014, IJARCSS.

[5]. Gitanjali, Dr. Kamlesh, “Securing Big Data Over Cloud Using Classification and Encryption Techniques”, IJRECE VOL. 6 ISSUE 2 APR-JUNE 2018.

[6]. Rizwana Shaikh, Dr. M. Sasikumar, “Data Classification for achieving Security in cloud computing”, Procedia Computer Science 45 (2015) 493 – 498, ScienceDirect.

[7]. Lo’ai Tawalbeh , Nour S. Darwazeh, Raad S. Al-Qassas2 and Fahd AlDosari, “A Secure Cloud Computing Model based on Data Classification”, Procedia Computer Science 52 (2015) 1153 – 1158, ScienceDirect.

[8]. Amiza Amir, Bala Srinivasan and Asad I Khan,” A Communication-Efficient Distributed Algorithm for Large-scale Classification within P2P Networks”, SoICT 2015, December 03-04, 2015, Hue City, Viet Nam c 2015 ACM. ISBN 978-1-4503-3843-1/15/12.

[9]. Tina Francis, Dr. Muthiya Madiajagan and Dr. Vijay Kumar, “Privacy Issues and Techniques in E-Health Systems”, SIGMIS CPR ’15, June 4–6, 2015, Newport Beach, California, USA.

[10]. Spyridoula Lakka, Christos Michalakelis, Teta Stamati and Dimosthenis Anagnostopoulos, “A framework for the classification of Could Computing Business Models”, PCI 2015, October 01-03, 2015, Athens, Greece © 2015 ACM. ISBN 978-1-4503-3551-5/15/10.

[11]. Fara Yahya, Robert J Walters, Gary B Wills, “Protecting Data in Personal Cloud Storage with Security Classifications”, Science and Information Conference 2015.

[12] Lei Ding, Malek Ben Salem, “A Novel Architecture for Automatic Document Classification for
Effective Security in Edge Computing Environments”, 2018 Third ACM/IEEE Symposium on Edge Computing.

[13] Miraj Hossain, Md. Rafiqul Islam, “A Model for Ensuring Data Security to Distributed Financial System in Cloud Storage”, A Model for Ensuring Data Security to Distributed Financial System in Cloud Storage.

[14] Rasmeet Kour, Suparti Koul, Manpreet kour, “A Classification Based Approach For Data Confidentiality in Cloud Environment”, 2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS).

[15] Nasarul Islam.K.V, Mohamed Riyas.K.V, “Analysis of Various Encryptions Algorithms in Cloud Computing”, International Journal of Computer Science and Mobile Computing, Vol.6 Issue.7, July- 2017.