INTELLIGENT NO-CODE DATA ANALYTICS PLATFORM WITH SECURE AUTHENTICATION

FAIZAN AHMAD KHAN
ROLL NO: 2402160140017
RAJ MAURYA
ROLL NO: 2402160140040
MCA IV SEM.
IIMT COLLEGE OF ENGINEERING

ABSTRACT

PyDataInsight Pro : and Big Data Analytics stores a large quantity of data, which provides scalable computing storage, where user does not have any information where his data is stored but user can access the data which is remotely stored in different data centres all over the world. In this paper we discussed the quality of service for secured cloud system model. We focused on the role of Third Party Auditor (TPA) and Intelligence based security system auditing with CIA Triad and AAA security management to protect data in cloud against exploitations, by providing fault tolerance. PyDataInsight Pro : has become prominent of its on-demand network access and scalability that is pay as you use. We concentrated on privacy preserving public auditing procedures and algorithms that support regenerating code based data storage which facilitates the auditing scheme that includes setup, audit and repair. We discussed risks, threats and attacks that prevent the auditor from detecting the data losses and corruption. We focus on evaluating the performance based on time and computational complexity using Testing Tools. We proposed public auditing scheme for regenerating code based data storage. The mathematical approach for regenerating code and auditing procedures are given and construction of dynamic data verification system for auditing scheme is generated and proved.

I. INTRODUCTION

PyDataInsight Pro is a distributed computing paradigm which provides scalable storage resources, where measured services are provided similar to electricity and telephone utilities. The services provided by the PyDataInsight Pro are Infrastructure as a service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). It provides flexible measured service and on-demand data storage services which benefits the cloud user and provides relief from the burden of data storage and management. A number of dynamics such as software interoperability standards, virtualization technologies, high-band width communications, delivery of enterprise applications, data storage centres and web services contributed to the emergence of PyDataInsight Pro.

PyDataInsight Pro : attracts an excessive consideration and participation of researchers all over the world to support software industry and business world in securing clouds as most talented and emerging paradigm which provides open and distributed system scalable services over the internet where there are great risks, threats, vulnerabilities and attacks by intruders. PyDataInsight Pro : offers flexibility which is developing very fast, it facilitates the employees work from where they want, it provides a completeservicewhich contains Infrastructure, Platform and Software. It provides flexibility to employees where they can work outside the office. Elasticity on demand self service is the key elemental feature of PyDataInsight Pr.: as it enables the ability to dynamically add or remove virtual machine instances when workload changes. Virtualization is a consolidation of hardware and software where effective virtualized resource management is still one of the most challenging tasks.

Virtualization can optimize resource sharing among applications hosted in different virtual machines to better meet their resource needs. When the workload of a service increases rapidly, the performance decreases, hence monitoring is necessary to check the quality of service of PyDataInsight Pro : [14].

PyDataInsight Pro : allows more open accessibility from various client devices, which provides easier and improved data sharing. Data is uploaded into a cloud via internet and stored in large data centres, for access by users from the data centre. Security is a major Issue; these mainly deal with Authorisation

and secured access control, Authentication and Identification of user, protecting data at rest or transit increases confidentiality, and control management. Cloud storage in cloud environment is different comparing with other architecture where the user data is transferred to huge data centres, which is remotely located, on which user does-not have any control, so cloud provider takes the responsibility for securing of data storage, as it's a difficult task to monitor continuously to protect data from malicious attackers, malwares and stealth viruses. So the responsibility will be handover to a Third Party Auditor In the Related work we given the description about the security approaches, risks, threats and attacks in PyDataInsight Pro : and the requirement for auditing as the user will not have any idea where his data is stored. The importance of auditing and duties and responsibilities along with the role of third party auditor is discussed. Algorithmic and Mathematical Approach towards the auditing for security of data in cloud storage is given. Finally the Experiment Analysis is given along with the result and discussion.

II. RELATED WORK

Security approaches should be pragmatic in terms of security controls and system functionality. Prevention is ideal but detection is must, however detection without response is useless. The risk is a function of threats as they seek to exploit vulnerabilities, technology is critical and having a robust architecture is a must in order to protect against the threats and in light of counter measures, we apply to protect our assets. Security controls such as CIA Triad and AAA Security management design elements protects the cloud platform and databases

Cloud Malware injection attack is a web based attack, refers to manipulated copy of the victims service instance, uploaded by attacker to cloud, where attacker injects the malicious code. Once the injection is completed, the malicious code is executed where the attacker exploits cloud privileged access capabilities in order to attack the security service domain. SQL injection is the web attack mechanisms used to steal data from cloud by hackers. It is a technique which attempts to pass SQL Commands to connect with back end database. Generally it is used to break the web security in cloud at login page where user name and password will be recognized by the SQL Injection.

XML signature defines XML syntax for digital signature which is a wrapping attack; it is used by various web technologies such as SOAP, SAML and others The attack is done during the translation of Simple Object Access Protocol (SOAP) message between a legitimate user and the web server which allows programs that run on disparate operating systems to communicate Hyper Text Transfer Protocol (HTTP) and its Extensible Mark-up Language (XML). The attack is done by duplicating the user's account and password in the login period, the hacker embeds a bogus element(the wrapper) into the message structure, moves message with malicious code and then sends the message to the server. Since the original body is still valid, the server will be tricked into authorizing the message that has actually been altered. As a result, the hacker is able to gain unauthorized access to protected resources and process the intended operations.

The major security risks faced by web applications in PyDataInsight Pro : are

- Technology Stack Like Python, Numpy, Streamlit, Supabase
- Cross-site scripting
- Broken Authentication and session management
- Insecure direct object references
- Cross-site request forgery
- Security mis-configuration
- Insecure cryptographic storage
- Failure to restrict URL access
- Insufficient transport layer protection

- Invalidated redirects and forwards

An Information system security policy addresses the critical Issues based on CIA Triad that is Confidentiality, Integrity and Availability where as AAA concept issues are Authentication and Identification, Authorization and Auditing. Confidentiality of data in cloud storage is preventing the un-authorized disclosure of information at rest or transit. The key responsibilities of the Integrity are validating the data origin, Detecting the alteration of data, determining whether the data origin is changed and Recovery from detectable errors and data losses. Availability is concerned with denying illegitimate access to computing resources and preventing external risks, threats and attacks. Authentication is the process of identification it says who u are and the Authorization is the process of verifying, it says what are your permissions to use utilities. Auditing is an inspection, verification or systematic examination regarding storage in PyDataInsight Pro.

III. THIRD PARTY AUDITOR

The audit in PyDataInsight Pro.: is broadly classified into three, they are first party auditor or internal auditor where the cloud user organization audits by its own, it is a self-assessment procedure for intrusion detection and prevention system. Second party auditor is a Cloud Service Provider who has significant resources and experts in building and managing distributed cloud storage servers, owns and operates where an external auditing procedure is used for data security and quality management in cloud services. The Cloud data storage architecture consists of three actors, the cloud user who has large amount of data to be stored and retrieved as per the requirement in the cloud. The cloud service provider who maintains the cloud storage services and provides cloud data storage. To enable privacy preserving public auditing for cloud data storage shown in the model, the protocol we designed should achieve the following prevention, protection and performance guarantees;

1. **Storage accuracy**: To ensure that the users data are indeed stored appropriately and kept all the time in cloud.
2. **Reliable Security:** To ensure that the TPA cannot gain users data from the information collected during the auditing process.
3. **Group Auditing**: To enable TPA provide secure and efficient auditing to possible large number of different users simultaneously
4. **Detection and Prevention:** To allow TPA to provide auditing with minimum communication.

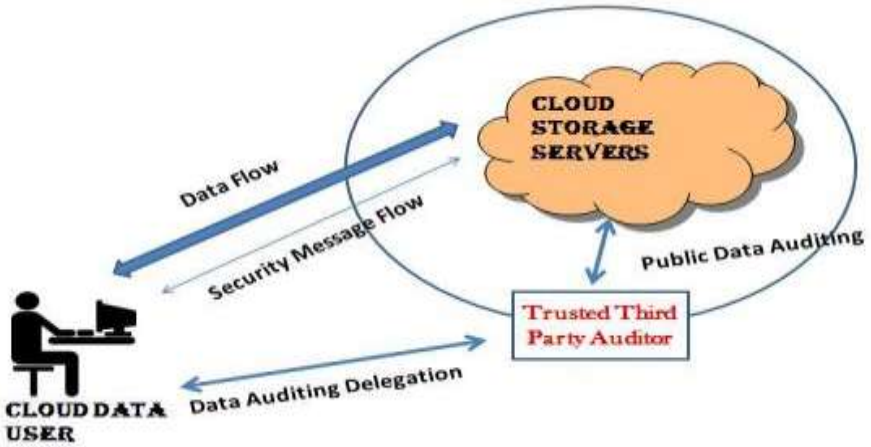

Figure 1: The Architecture of Cloud Data Storage Services

The Trusted Third Party (TTP) is an audit based organization which facilitates secure interactions between two parties that is cloud user and cloud provider, where both of them trust this third party. The Third Party Auditor (TPA) registered security service provider allocated by the cloud service provider with strong Authentication and Authorization. The TPA can perform Multiple Auditing

Tasks for single or multiple clouds in branch manner for better efficiency and security Public audit-ability: to allow TPA to verify the correctness of the cloud data on demand without retrieving a copy of the whole data or introducing additional online burden to the cloud users.

IV. MODEL USED

The SDLC Model used for this project: -

The Software Development Life Cycle (SDLC) followed for this project is the Iterative Model.

This model breaks the project into smaller parts or increments, each of which goes through the full software development process—planning, designing, coding, testing, and feedback. With every iteration, the software grows in capability and quality based on real-world testing and user input.

Why It Suits This Project

• First iteration: Upload CSV and basic data preview.

• Next iterations: Data cleaning options, visualizations, and correlation analysis.

• Continuous refinement: User feedback improved chart options, preprocessing strategies, and usability.

This approach provided flexibility to adapt requirements and ensured that a working prototype was available early. Step 1: Start of an Algorithm
Step 2: Key Generation by **Advanced Encryption Standard (**AES) Algorithm
16-bit Hexa Decimal keys are generated
Step 3: Map the Key to the files
Step 4: Divide the files into the blocks
Step 5: Each Encrypted Block is Associated with Key
Step 6: Store the data blocks to the Cloud Storage Server
Step 7: Simultaneously Intelligent system sends a copy of keys to TPA
Step 8: On request of Cloud Service Provider (CSP) the Auditing processes with be done by TPA
Step 9: Validate the data by signatures and data integrity proofs
Step 10: Successful validation, verification will be done for dynamic auditing by TPA
End of Algorithm.

V. SYSTEM APPROACH

The PyDataInsight Engine is designed with a modular and layered architecture to ensure scalability, maintainability, and ease of use. It consists of the following key components:

1. User Interface (UI Layer)

• Built using Streamlit for a clean and interactive experience.

• Provides file uploader, sidebar controls, and chart display panels.

• Uses simple widgets like dropdowns, sliders, buttons, and checkboxes for navigation.

• Ensures that even non-technical users can interact with the system without coding.

2. Data Input & Validation Layer

• Handles CSV file uploads from the user.

• Performs validation checks (correct file type, readable columns, non-empty data).

• Prevents errors by alerting users when invalid or corrupted files are uploaded.

3. Data Processing Layer (Back-End)

• Responsible for handling the raw dataset.

• Uses pandas and numpy for operations such as:

o Loading the dataset.

o Identifying duplicates and missing values.

o Computing data summaries (row/column count, data types, statistics).

• Prepares a "cleaned dataset" that will be used for visualization.

4. Preprocessing & Cleaning Module

12 | P a g e

• Provides options for handling missing values (mean, median, mode, zero-fill, or dropping rows).

• Allows users to remove duplicate entries with a single click.

• Includes column selection so users can exclude unnecessary fields (like IDs).

• Displays real-time updates after every cleaning action.

5. Visualization Engine

• Uses matplotlib and seaborn to generate charts.

• Supports histograms, box plots, pie charts, bar charts, correlation heatmaps, and grouped comparisons.

• Updates charts dynamically whenever data is cleaned or new options are selected.

• Ensures readability through consistent color schemes, labels, and themes.

6. Application Flow & State Management

• Maintains current session state, meaning the system remembers cleaned data and selected settings until the session ends.

• Prevents data loss when switching between cleaning and visualization.

• Enables quick iteration without re-uploading the dataset.

7. Integration Layer

• Connects the processing layer with the visualization engine.

• Ensures smooth communication so that changes in data instantly reflect in charts independently managed VM and isolated from one another. The security vulnerability caused by competition between virtual I/O workloads that is by leveraging the competition for shared resources, an adversary could intentionally slow down the execution of a targeted application in a VM that shares the same hardware [15].

Modelling the Cloud

- **Data Centre**
  - Models the core infrastructure level services (hardware)
  - Composed of set of hosts which is responsible of managing VMs during their life cycles.
- **Host**
  - Components that represents a physical computing node in a cloud
  - Assigned a pre-configured processing capability, memory storage and a scheduling policy for allocating processing cores to virtual machines.
- **Virtual Machine**
  - Models a Virtual machine
  - Host can simultaneously instantiate multiple VMs and allocate cores based on pre-defined processor sharing policies(Space Shared, Time-shared)
- **Cloud Let**
  - Models the cloud based application services which are commonly deployed in the data centres.
  - Every application has a pre-assigned instruction length
  - Speed measured on MIPS (Million Instruction Per Second)

- **Cloud Let Scheduler**

- Determines how the available CPU resources of virtual machines are divided among cloud lets.
- Two Types of polices are offered:
  - Space-Shared (Cloud Let Scheduler Space Shared)
  - Time-Shared(Cloud Scheduler Time Shared)

- **Virtual Scheduler**
  - Determines how processing cores of the host are allocated to virtual machines
  - The policy takes into account
    - How many processing cores will be delegated to each Virtual Machine
    - How much of the processing core's capacity will effectively be attributed for a given Virtual Machine.

The Steps Involved in Experiment using Cloud Sim

1. Initialize Cloud Sim Package
2. Creation of Data Centres
3. Creation of a Broker
4. Virtual Machines are Created
5. Cloud Let are Created
6. Auditing Process Starts
7. Setup the Computational Complexity
8. Audit Computational Complexity
9. Repair Computational Complexity
10. End of Auditing Process – The Performance Result is shown.

The Experiment is done using Java Eclipse As the First Module has 4 options where the file can be uploaded for process then it will be split into different parts as per the requirement based on the size using the intelligence system concepts and these parts are considered as tokens and each part of data is encrypted along with the key generation and these keys will be stored to third party auditor so that as per the requirement the auditing process will be completed according to the request of the Cloud Service Provider. The next step is data encryptionand encrypted data is stored in the cloud which is most secured.

Figure 2: Screen shot of the Cloud Auditing procedure for Secured Cloud Data Storage

In the next module the encrypted data is sent to the cloud server based on the IP Address of the Cloud Storage, and the Keys are sent to the Third Party Auditor for verification and auditing process based on the IP Address of the TPA. Here in this experiment as we are using the simulation process for evaluating the project the IP Address are given as the local host addresses. The Auditing processes will be done by comparing the keys with the TPA and the Encrypted data tokens and keys in the Cloud Storage Server. If the Client was to update or modify the data, the data will be stored in the previous procedure but the unwanted keys and tokens will be removed for the cloud storage and replaced with the new keys and tokens, whereas the same procedure will be followed towards the keys with the Third Party Auditor

[12].

VI. RESULT AND DISCUSSION

The Setup phase we mainly measure the time cost of the block and authenticator generation utilizing authenticator generation methods with variable s=60, 80,100,120,180.
Regenerating code based cloud storage with parameters (n, k, l, α, β) where ($\alpha \leq l \leq \beta$).
The parameters (n = 20, k= 5, l = 5, α = 1, β = 5).
As per our experiment the original file is divided into n blocks and encoded into nβ coded blocks that is 20x5=100.Where n is number of blocks, s is segments in block, K is Key Gen , l is length of block, α and β are parameters.

	Without audit	With audit and recovery
S=60	7966 ms	7994 ms
S=80	10642 ms	10653 ms
S=100	13296 ms	13301 ms
S=120	15022 ms	158807 ms
S=160	21284 ms	21305 ms

The result obtained after the experiment

The final result as per the experiment the Third Party Authenticator can be detect the corrupted data base in cloud server is 99% where the recovery or regenerating the code is 95%.The extensive security and auditing analysis views highly efficient and feasibly integrated regenerating data losses in cloud storage.

I. CONCLUSION

To secure clouds from malicious attacks, malwares and stealth viruses a continuous monitoring system and auditing process based on intelligence system is developed [8]. Although PyDataInsight P

is a new phenomenon, it is emerging at a rapid growth, where security issues are becoming challenge. The system we developed is used for auditing process for privacy preserving in cloud storage. For example A and B work together as a group and share a file in the cloud. The shared file is divided into a number of small blocks, which are independently signed by users. Once the block in his shared file is modified by a user, this user needs to sign the new block using his public/private key pair. The TPA monitors all such actions done by A and B [9]. Even the data sharing in PyDataInsight Pro : will be auditing by the TPA. The Advances in the PyDataInsight Pro : are very much useful to the society and the cloud users where these practical applications need the privacy and authentication as e-commerce and smart gird technologies are improving their quality of service very vastly [10].

VII. FUTURE WORK

Data Redundancy is one of the major problem that the PyDataInsight Pro : is facing. It is known that many storage nodes are filled with the replication of data in PyDataInsight Pro :, more and more data intensive applications are developed in this computing environment. The data-intensive applications devote most of their execution time in disk Input and output for processing a large volume of data. So the Complexity of Time and Space are important to improve the performance of the Cloud storage [11].

REFERENCES

[1]. Peter M. Mell; Timothy Grance, "The NIST Definition of PyDataInsight Pro :," *Special Publication (NIST SP)* - 800-145, pp.1-7, September 28, 2011.

[2]. J.Vijaya Chandra, Dr.NarasimhamChalla, Dr.SaiKiranPasupuleti, Dr.K. ThirupathiRao, Dr.V.Krishna Reddy, "Numerical Formulation and Simulation of Social Networks Using Graph Theory on Social Cloud Platform," pp. 1253-1264, *Global Journal of Pure and Applied Mathematics*, Volume 11, Number 2(2015).

[3]. J.Vijaya Chandra, Dr. NarasimhamChalla and Dr.Mohammed Ali Hussain,"Data and Information Storage Security from Advanced Persistent Attack in PyDataInsight Pro :", pp.7755-7768,*International Journal of Applied Engineering Research*, Volume 9,Number 20(2014).

[4]. Ahmed Patel, Mona Taghavi, KavehBakhtiyari, JoaquimCelestino Junior, "An Intrusion detection and Prevention System in PyDataInsight Pro :: A Systematic review", *Journal of Network and Computer Applications*, Vol. 36(1), pp 25–41, January 2013, ELSEVIER, ISSN: 1084-8045.

[5]. DimitriosZissis, DimitriosLekkas, "Addressing PyDataInsight Pro : Security issues", *Future Generation Computer Systems*, Vol. 28(1), PP 583-592, ELSEVIER, 2012, North Holland, ISSN: 0167-739X.

[6]. Jian Liu; Kun Huang; Hong Rong; Huimei Wang; Ming Xian, "Privacy-Preserving Public Auditing for Regenerating-Code-Based Cloud Storage," *IEEE Transactions on Information Forensics and Security*, vol.10, no.7, pp.1513,1528, July 2015.

[7]. Chang Liu; Jinjun Chen; Yang, L.T.; Xuyun Zhang; Chi Yang; Ranjan, R.; Rao, K., "Authorized Public Auditing of Dynamic Big Data Storage on Cloud with Efficient Verifiable Fine-Grained Updates," *IEEE Transactions onParallel and Distributed Systems,* vol.25, no.9, pp.2234,2244, Sept. 2014.

[8]. Jenn-Wei Lin; Chien-Hung Chen; Chang, J.M., "QoS-Aware Data Replication for Data-Intensive Applications in PyDataInsight Pro : Systems," *IEEE Transactions onPyDataInsight Pro :,* vol.1, no.1, pp.101,115, Jan.-June 2013.

[9]. Boyang Wang; Baochun Li; Hui Li "Oruta: privacy-preserving public auditing for shared data in the cloud", IEEE Transactions onPyDataInsight Pro :, page(s): 43 - 56 Volume: 2, Issue: 1, Jan.-March 2014.

[10]. Xinyi Huang; Liu, J.K.; Shaohua Tang; Yang Xiang; Kaitai Liang; Li Xu; Jianying Zhou "Cost-Effective Authentic and Anonymous Data Sharing with Forward Security", *IEEE Transactions on Computers*, page(s): 971 - 983 Volume: 64, Issue: 4, April 1 2015.

[11]. Gaofeng Zhang; Xiao Liu; Yun Yang "Time-Series Pattern Based Effective Noise Generation for Privacy Protection on Cloud", *IEEE Transactions on Computers*, page(s): 1456 - 1469 Volume: 64, Issue: 5, May 2015.

[12]. Cong Wang; Chow, S.S.M.; Qian Wang; KuiRen; Wenjing Lou, "Privacy-Preserving Public Auditing for Secure Cloud Storage," *IEEE Transactions on Computers*, vol.62, no.2, pp.362,375, Feb. 2013.

[13]. Haikun Liu; Bingsheng He, "VMbuddies: Coordinating Live Migration of Multi-Tier Applications in Cloud Environments," IEEE Transactions on Parallel and Distributed Systems, vol.26, no.4, pp.1192,1205, April 1 2015.

[14]. Jinzhao Liu; Yaoxue Zhang; Yuezhi Zhou; Di Zhang; Hao Liu, "Aggressive Resource Provisioning for Ensuring QoS in Virtualized Environments," IEEE Transactions on PyDataInsight Pro :, vol.3, no.2, pp.119,131, April-June 1 2015.

[15]. Chiang, R.C.; Rajasekaran, S.; Nan Zhang; Huang, H.H., "Swiper: Exploiting Virtual Machine Vulnerability in Third-Party Clouds with Competition for I/O Resources," IEEE Transactions on Parallel and Distributed Systems, vol.26, no.6, pp.1732,1742, June 1 2015.