Saveetha Institute of Medical and Technical Sciences

Saveetha Scholl of Engineering, Thandalam, Chennai – 602 105.



Capstone Project

Ensuring Security and Compliance of a Large-Scale Big Data Infrastructure in a Healthcare Organization

Done by:-

Name: V. Tamizhselvan

Register. No: 192224029

Guide: Dr. Balamanigandan

Ensuring Security and Compliance of a Large-Scale Big Data Infrastructure in a Healthcare Organization

1. **Aim: -**

To design and implement a comprehensive security solution to protect sensitive healthcare data within a large-scale big data infrastructure, ensuring data privacy, preventing unauthorized access, and achieving compliance with HIPAA regulations.

1. **Scope: -**

The scope includes securing the Hadoop ecosystem (HDFS, YARN, etc.), Spark, and Kafka components of the infrastructure. The solution will cover authentication and authorization, data encryption, auditing and logging, security monitoring, and staff training.

1. **Problem Statement: -**

A healthcare organization’s big data infrastructure is vulnerable to unauthorized access and data breaches. The infrastructure must comply with HIPAA regulations, requiring robust security measures to protect sensitive patient data, ensure data confidentiality, and maintain comprehensive logging and monitoring capabilities.

1. **Risk Assessment: -**

**Identification of Potential Security Vulnerabilities: -**

1. Weak authentication mechanisms.

2. Inadequate access controls.

3. Unencrypted sensitive data.

4. Lack of comprehensive audit logs.

5. Insufficient monitoring and alerting systems.

**Identification of Compliance Gaps: -**

1. Non-compliance with HIPAA data protection and privacy requirements.

2. Lack of regular compliance audits and reporting mechanisms.

3. Insufficient documentation of security policies and procedures.

**Authentication and Authorization: -**

Implementation of Kerberos Authentication for Hadoop Services: -

1. Configure Kerberos for Hadoop Components: -

* Secure inter-service communication within HDFS, YARN, and MapReduce.
* Set up a Key Distribution Centre (KDC) and update Hadoop configuration files (core-site.xml, hdfs-site.xml, yarn-site.xml) to enable Kerberos authentication.

Implementation of LDAP Integration for User Authentication and Authorization: -

1. Integrate LDAP with Hadoop: -

* Centralize user identity management and authentication.
* Configure Hadoop, Spark, and Kafka to authenticate users against the LDAP directory.

**Configure Role-Based Access Control (RBAC) Policies: -**

1. Define and Enforce RBAC Policies: -

* Apply RBAC across Hadoop, Spark, and Kafka.
* Assign roles and permissions based on job functions and data access requirements.

**Data Encryption: -**

**Enable Encryption at Rest Using Hadoop's Native Encryption** **Features: -**

1. Hadoop Transparent Data Encryption (TDE): -

* Encrypt HDFS data at rest.
* Configure encryption zones within HDFS and manage encryption keys using the Hadoop KMS (Key Management Server).

**Configuration of SSL/TLS Encryption: -**

1. Secure Data in Transit: -

* Configure SSL/TLS for communication across Hadoop, Spark, and Kafka components.
* Generate and deploy SSL certificates to secure communication channels.

**Auditing and Logging: -**

**Configure Audit Logging for Hadoop, Spark, and Kafka Components: -**

1. Enable Detailed Logging: -

* Capture user activities and data access logs for HDFS, YARN, Spark, and Kafka.

**Centralize Log Management Using Tools Like Apache NiFi: -**

1. \*Implement Apache NiFi: -

* Aggregate logs from Hadoop, Spark, and Kafka.
* Process and store logs centrally.

**Review Audit Logs & Conduct Compliance Audits: -**

1. Regular Log Review: -

* Identify suspicious activities and potential security incidents.

2. Periodic Compliance Audits: -

* Ensure adherence to HIPAA and other regulatory requirements.

**Security Monitoring and Alerting: -**

**Integrate Security Monitoring Tools: -**

1. Real-Time Analysis: -

* Use tools like Apache Metron or Splunk to analyse log data in real-time.
* Set up alerts and notifications for potential security incidents.

**Implement a Security Information and Event Management (SIEM) System: -**

1. Correlate and Respond: -

* Correlate security events and respond promptly to incidents.

**Documentation and Training: -**

**Develop Comprehensive Documentation: -**

1. Security Policies and Procedures: -

* Detail configurations and security measures implemented.

**Create Training Materials and Conduct Regular Training Sessions: -**

1. Educate Staff: -

* Train staff on security best practices and compliance requirements.

2. Roles and Responsibilities: -

* Ensure all employees understand their roles in maintaining security.

**Conclusion: -**

Implementing a comprehensive security solution for a healthcare organization's big data infrastructure requires a multi-faceted approach. By conducting a thorough risk assessment, implementing strong authentication and authorization mechanisms, encrypting data at rest and in transit, configuring robust auditing and logging practices, and integrating security monitoring and alerting systems, the organization can significantly enhance its data protection capabilities. Additionally, continuous staff training and documentation will ensure that security measures are effectively maintained and compliance with HIPAA regulations is achieved.