JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A REVIEW ON DATA LEAKAGE SECURITYOVER A HIGHLY SECURED INDUSTRIAL INFRASTRUCTURE

Ms Sarika K Meshram
ME student,
Department of Computer Science & Engineering , PRMIT&R

Dr. Sunil R Gupta
Professor,
Department of Computer Science & Engineering, PRMIT&R ,

Abstract : For designing an industrial data platform big data-based acquisition and storage system plays a vital role for proper utilization. Various frameworks of big data are using compressed and serialization method which meets the needs of any industrial production management for time consuming and large storage. Considering the importance of big data security and processing time we try to propose an easier big platform to consume less time while accessing data storage space. The main intention of the proposed system is to evaluate multiple compression and serialization methods for big data performance and tries for optimal compression in industrial platform. Multiple schemes was recently introduced for cloud storage by using distributed cloud storage providers with some degree of controlling information leakage. Although non avoidable data chunk distribution may lead to high security information disclosure even with high security cloud platform

Keywords: **Data, Acquisition, Compression, Serialization, Leakage.**

1.Introduction

Sharing of data in cloud and storing may leak it to unidentified user or source which creates havoc in the cloud storage. Leakage of data is a serious problem which happens with physically or logically removing it from the organization either intentionally or by any cause. Data loss became one of the biggest threats in the market and it is much necessary to resolve it as quickly as possible. Comprising the important and most vulnerable data is not accepted by any organization. And to make it most secured the proposed system design a system to avoid the data leakage issue in that organization and trying to reduce the maximum issue occurred due to it. Analysis of data in industry is considered as an important aspect for future improvement in order to gain maximum profit margin in industrial production and operations, and creates the next solution for innovation, in the competitive market for better productivity. Industrial data storage is one of the core components for any computational industry for intelligent analysis. Furthermore the ever increasing number of intelligent devices in various intelligent plant data is utmost important. The devices such as data gathering storage, cloud storage, IoT and mail linking are one of biggest sources of data and to secured it in proper manner so that the data will not leak is become a challenge to focus on. This system helps to processed large quantity of data where large infrastructure was available. Infrastructure from industry faces large problems such as defining different an efficient module for settings of various applications and maintaining specific models for analysis in industry. Many different methods have been developed for storage of data on multiple cloud storage and work on the principle of security concerned. During cloud storage many distributor of cloud service provider allows the users to provide the security with control of data leakage and hence single leakage will not leak the complete information.

2. Related Work

According to industrial data acquisition and processing requirements, this paper designs an industrial [9] 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. Nowadays, [1] industrial data platform is the core component of industrial data storage, computation and analysis for the management of intelligent plant. With the increasing number of intelligent equipment's used in intelligent plant, however, intelligent plant can acquire a large quantity of data of Radio Frequency

Identification (RFID) and intelligent equipment's, thus providing rich data sets for manufacturing industry. At present, industrial data lack effective interoperability standards or low-cost acquisition schemes. This leads to high costs of data collection and integration of equipment's and systems. The data acquisition module provides a data source for data analysis of the big data platform, and the data storage module provides data source and storage space of the data computation module. Before big data appear, database has become an important processing platform because of the data processing convenience. But when database is faced with non-relational or large-scale data, there is a difficulty dealing with them. Big data not only enhance the related computing services technologies but also change the traditional mode of many industries. Big data help people acquire knowledge from the massive, complex data, and become another focus after integrated circuit and Internet information technology. It provides data calculation, machine learning, graph analysis, data query, which is the core component of data analysis. Unlike traditional [3] data mining analysis, which consists of lean datasets (that is, datasets with few features), manufacturing has fat datasets. many platforms for industrial big data are developed based on Map Reduce computing framework. [6]With the increasingly rapid uptake of devices such as laptops, cell phones and tablets, users require an ubiquitous and massive network storage to handle their ever-growing digital lives.[2] To meet these demands, many cloud-based storage and file sharing services such as Drop box, Google Drive and Amazon S3, have gained popularity due to the easy-to-use interface and low storage cost. The information in users' data can be leaked e.g., by means of malicious insiders, back doors, bribe and coercion. One possible solution to reduce the risk of information leakage is to employ multi cloud storage systems in which no single point of attack can leak all the information [4]. For each update of local file, only chunks with changed hashes will be uploaded to the cloud.

Data leakage may be defined as the illegal transfer of valuable/sensitive data by an entity to unauthorized entities.[10] 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. The effects of data leakage could range from loss of valuable data, privacy, cyber- theft, to threat to economy and national security. In this paper a new approach using various techniques such watermarking, fake data addition, guilty party etc. to build a fail-safe data leakage detection especially suited to cloud-based data storage systems is proposed for enhanced security. [11]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. 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. When the data is in motion or transit the existing tools such as IDS, honey pot, firewalls, encryption, digital certificates, anti-virus tools, compression, authentication and monitoring tools helpful[7] to watch the data. The author , mainly focused on detecting the leakage of sensitive data over the mobile devices with the help of Labyrinth a run-time privacy enforcement system. Labyrinth supports both Android and IOS. the data duplication technique, which is widely adopted by current cloud storage services like Drop box, is one example of exploiting the similarities among different data chunks to save disk space and avoid data re transmission [10] This article focuses on data leakage prevention (DLP) and information leakage prevention (ILP) within the scope of information security (IS) which exceeds information technology (IT) security. [12] it is very important to implement DLP controls and information security controls to manage data loss risks. The terms DLP and ILP are not accurately specified in any official standard or regulation [5]. Cloud computing is transforming information technology. As information and processes migrate to the cloud, it is transforming not only where computing is done, but, fundamentally, how it is done.[8] As increasingly more corporate and academic worlds invest in this technology, it will also drastically change it professionals' working environment. Cloud computing, as a new technology, has also created new security challenges such as data breaches, data loss, account hijacking and denial of service.

3.Methods/Techniques/Methodology

This paper proposed the basic idea for big data industrial platform and helps to design the architecture with six layers of data flow pattern. All six layers are inclusive of 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 respectively. Our study mainly focuses on acquisition, storage and computing layer. Sqoop and Flume are the technology used for data acquisition module for industrial data distribution. The Sqoop relates with relational databases within the industry where real time requirement of such type is not of highest priority. Whereas Flume relates with data associated with real time devices in accordance with intelligent equipment's. Safety monitoring of dynamic data was handled by flume. Data mainly depends on the data that was generated by the humans for their individual use and at the same time company completely rely on the people who generate this type of data. So the data is one of the most important aspects of any firm or any industry which make it so vulnerable in all related parameters. And hence some hackers or intruders or insider make use of those data without authentication. So to avoid this issue a security structure need to design which plays a vital role to avoid data lose and leakage from the highly security infrastructure. This system will help any organization to provide the security measures for the sensitive data. It's actually a security terms to find, evaluate and protect vulnerable information. The asset classification, audits, management of data security and some privileges to assets can be provided with the proposed system. Transforming information in the cloud computing used to process and migrate in a cloud environment. Cloud computing, as a new technology, has also created new security instances which helps to avoid the issues such as data interruption, loss of data and hacking. Some advancement was done to improve the security measures in the proposed system and helps to provide more and more technology benefits.

4.Flowchart

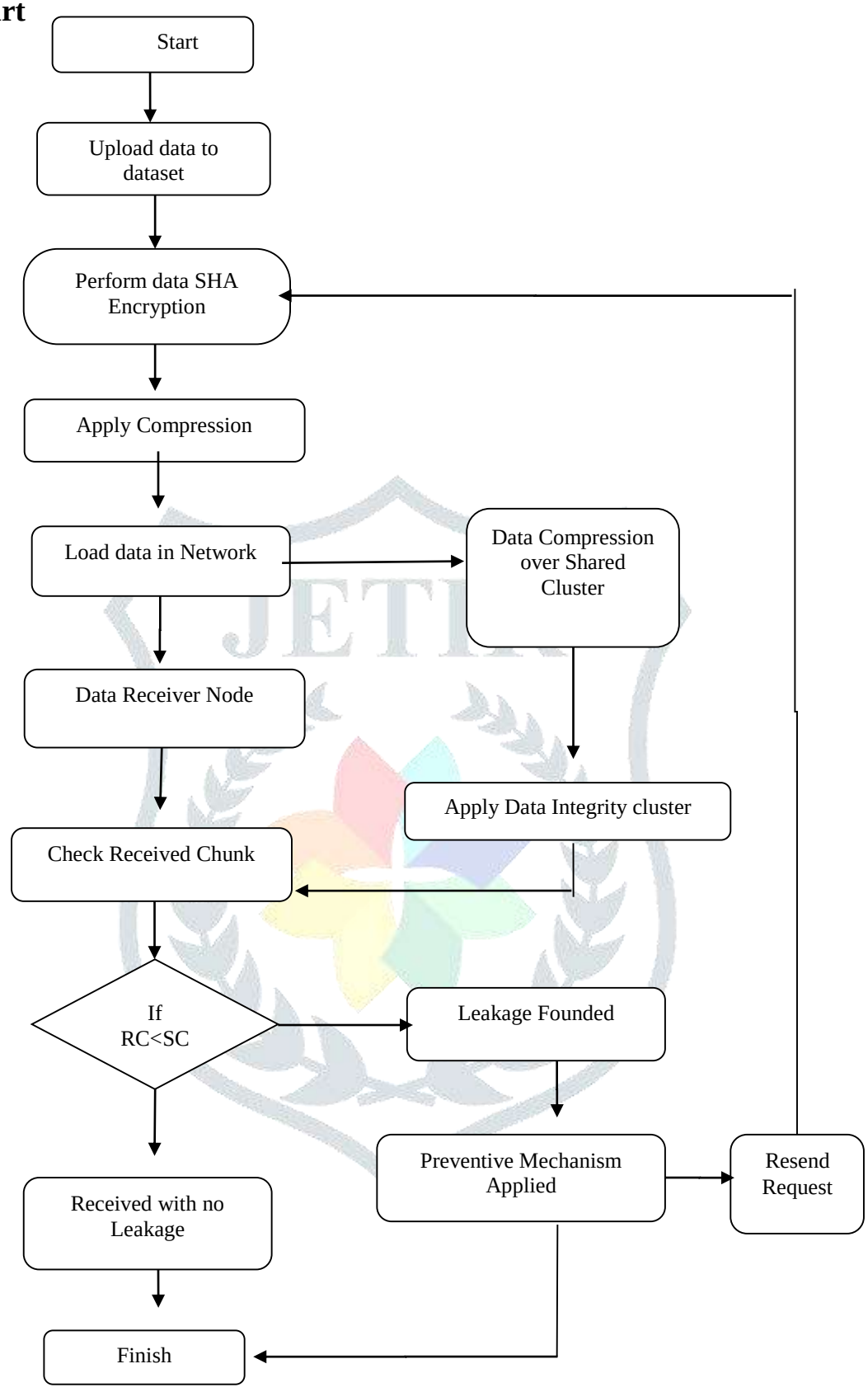

Fig. 1 Flowchart for the proposed system

5. Proposed Work

The proposed system helps to provide the control over our data leakage and assure more cloud security in the environment. While using multiple clouds there was the chance of data leakage and this system helps to design a fundamental to prevent the issue related to data leakage. A mechanism was design for stopping the leakage and provides the information by sending the request to get actual intimation of data leakage or provide a file which contains the issue of data leakage point. The system actually helps to determine as an auto responder in leak file. The implemented compression mechanism helps to prevent the leakage over file sharing all over the network. The system also helps to analysis the data leakage and helps to finalize the content of leak file. It also to perform proper efficient analytics and allow serialization with compression based input data.

6. Objectives

1.To store the data uploaded in highly secured environment

2. To provide an effective data sharing technique for further implementation

3. To implement an effective compression techniques

4. To prevent the data leakage over the network

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] 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

[3] P. Lade, R. Ghosh, and S. Srinivasan, ``Manufacturing analytics and industrial Internet of Things," *IEEE Intell. Syst.*, vol. 32, no. 3, pp. 74_79 May/Jun. 2017.

[4] H. Chen, Y. Hu, P. Lee, and Y. Tang, "Nccloud: A network-coding-based storage system in a cloud-of-clouds," 2013.

[5] B. Hauer ,Data and Information Leakage Prevention Within the Scope of Information Security", december 16, 2015.

[6] T. G. Papaioannou, N. Bonvin, and K. Aberer, "Scalia: an adaptive scheme for efficient multi-cloud storage," in Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis. IEEE Computer Society Press, 2012, p. 20.

[7] 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

[8] Nina Pearl Doe 1, Sumaila Alfa 2, V. Suganya, "Efficient method to prevent cloud" Volume 16, Issue 3, Ver. III (May-Jun. 2014)

[9] Zan Mo, Yanfei Li,"Research of Big Data Based on the Views of Technology and Application" American Journal of Industrial and Business Management, 2015, 5, 192-197

[10] K. Manoj Kumar+ G. Shubhang+ G. Rajesh Chandra "DATA LEAKAGE DETECTION SYSTEM FOR CLOUD-BASED STORAGE SYSTEMS" (IJAER) 2014, Vol. No. 8, Issue No. V, November.

[11] M. Hart, P. Manadhata, and R. Johnson, "Text classification for data loss prevention," in Privacy Enhancing Technologies, 2011, pp. 18–37

[12] Y. Shapira, B. Shapira, and A. Shabtai, "Content-based data leakage detection using extended fingerprinting," arXiv preprintarXiv:1302.2028, 2013