International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 04 | Apr 2021 www.irjet.net p-ISSN: 2395-0072

Review Paper on Data Acquisition and Compression with Leakage Prevention

Sheetal A. Paturde[1], Prof. S. B. Rathod[2]

[1]M.E.Student , Computer Science and Engineering, Sipna College of Engineering and Technology, Amravati, Maharashtra, India.
[2]Prof, Computer Science and Engineering, Sipna College of Engineering and Technology Amravati, Maharashtra, India.

Abstract - *Big data-based acquisition and storage system (ASS) plays an important role in the design of industrial data platform. Data acquisition is a process of gathering, filtering and cleaning data before the data is put in data any other storage solution. Many big data frameworks have been integrated compression and serialization method. The impact of multiple compression and serialization methods on big data platform performance and tries to choose optimal compression method for the industrial data platform .They are manage big amount data storage. Now many schemes have been recently advanced for storing data on multiple clouds. Distributing data over different cloud storage providers (CSPs) automatically provides users with a certain degree of information leakage control, for no single point of attack can leak all the information .However, unplanned distribution of data chunks can lead to high information disclosure even while using multiple clouds .Big amount of data handling is very crucial to maintain so it is necessary to perform active compression methodology on this.*

***Key Words*: Data acquisition, Serialization, Data compression, Multicloud storage, Data storage.**

1. INTRODUCTION

Data acquisition and storage system (ASS) is important part of the design of industrial data platform. Industrial data platform is the core component of industrial data storage, computation and analysis for the management of intelligent plant. 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 divided in 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 .For such organizations play central role after the data acquisition. 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. This study focuses on the acquisition layer. we are discussing data acquisition on industrial data platform, big data platform include three modules, which are data acquisition module, data storage module and data computation module . 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. There are many forms of industrial data distribution acquires a lot of data stored in relational databases in the industry, and the real-time requirement of these data is not high. 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 monitory benefits, they are trying to get access to the data with or without authentication. In this case, some time data leakage problem occouring for this reason we are going to studied data leakage prevention in this paper. The data leakage and detection its impact on organizations. Data leakage is the unauthorized transmission of data or information from within an organization to an external destination or recipient. Data leakage is defined as the accidental or intentional distribution of private or sensitive data to an unauthorized entity and the illegal transfer of valuable/sensitive data by an entity to unauthorized entities. Some of the data are leaked and found in an unauthorized place. Data leakage poses a serious issue for companies as the number of incidents and the cost to those experiencing them continue to increase data loss directly and indirectly. 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. 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 where the organizations are in responsibility to overcome this problem. The data loss / leakage prevention solutions detect and prevent unauthorized attempts to copy or send sensitive data both intentionally and unintentionally.

Data leakage prevention helps the organization to prevent the leakage of sensitive data or information namely, personal identifiable information, financial information. 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. Well Several methods have been currently developed for storing data on multiple cloud environment .Distributing data over different cloud storage providers (CSPs) automatically provides users with a certain degree of information leakage control, for no single point of attack can leak all the information. However, unplanned distribution of data chunks can lead to high information disclosure even while using multiple clouds. An important information leakage problem caused by unplanned data distribution in multi cloud storage services. an important information leakage problem caused by unplanned data distribution in multicloud storage services Data security plays an important role in cloud in which lot of data is get shuffled and became unsecured while sharing to other users.

2. Literature Review

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 plat- form. 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] 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 monitory 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] 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.[7] 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]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] In fact, the data deduplication 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 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.

3. Design Goal

- To apply encryption over the data in storage
- Data sharing with encrypted format
- Lossless data compression technique implementation
- Apply data acquisition in cloud

4. Proposed Work

In this we are proposed a dynamic execution module in which the system will work on dynamic selection of the compression technique which make the user more useful and built up with more aspect so that this system will helps the user to perform efficient analytics, serialization and compression based on inputted data work. In the proposed system the cloud optimization done with prevention of leakage. In which the whole data is get stored in encrypted form. It will get access to the owner only.

5. Conclusion

This is to be conclude that the implemented system is more efficient as compared to the given paper in which the compression technique and prevention technique is get implemented. In this the data leakage is add feature included in proposed so that the implemented method is more effective. Distributing data on multiple clouds provides users

with a certain degree of information However, unplanned distribution of data chunk scan lead to avoidable information leakage. We show that distributing data chunks in a round robin way can leak user's data as high as 80% of the total information with the proposed mechanism will be helpful in lot extend to increase the data security in the cloud and improve the technique of data leakage. In this the leakage data can be workout more effectively as compared to existing system and leakage control in that no single cloud provider is privacy to all the user's data.

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"

[3] 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

[4] Hao Zhuang, Member, IEEE, Rameez Rahman,Pan Hui, Member, IEEE, and Karl Aberer, Member, IEEE "Optimizing Information Leakage in MulticloudStorage Services" IEEE TRANSACTIONS ON CLOUD COMPUTING, VOL. 14, NO. 8, JANUARY 2016.

[5] Bijayalaxmi Purohit, Pawan Prakash Singh" Data leakage analysis on cloud computing" Vol. 3, Issue 3, May-Jun 2013.

[6] R.Senthil kumar1, S.Kaviya priya2 , M.Manimegalai3, P.Logesh4"Privacy Preservation With Data Leakage Detection & Monitoring" *Vol 6 Issue 3 March -2018*

[7] 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.