DATA ACQUISITION AND STORAGE SYSTEM FOR CORPORATE DATABASE USING BIG DATA

[1]Vivek M. Kumbhare, [2]Sachin A. Murab, [3]Parag D. Thakare, [4]Aniruddha A. Kolpyakwar

[1]Student of M.E.(CSE) IIIrd semester, [2]Head Of Department, [3]Assistant Professor, [4]Assistant Professor
[1,2,3,4]Department of Computer Engineering,
[1,2,3,4] Jagadambha College of Engineering & Technology, Yavatmal, India.

Abstract : Big data based data acquisition as well as database system plays an important role in the design of corporate data platform. Mostly big data frameworks have been integrated data compression and data serialization method. These methods cannot meet the importance of corporate production information management for requiring time-consuming and mass storage. Based on the existing big data frameworks, we propose an enhanced corporate big data platform in order to reduce the data processing time while requiring fewer data storage space. Specifically, focuses on evaluating the impact of multiple compression and serialization methods on the big data platform performance and tries to choose optimal compression and serialization method for the corporate data platform. Compared to the methods integrated in the past years, the experimental results showed the data compression time of the platform has been reduced by 73.9% with a less than 96% the size of data compressed, furthermore, the data serialization time has been reduced by 80.8%. With the increasing amount of data, it takes less time to compare with benchmark methods.
Big amount of data handling is very crucial to maintain so it is necessary to perform active compression methodology on this.

IndexTerms - Corporate data, Big data, Data compression & Data serialization.

I. INTRODUCTION

Big Data is also data but with a huge size. Big Data is a term used to describe a collection of data that is huge in size and yet growing exponentially with time. In short such data is so large and complex that none of the traditional data management tools are able to store it or process it efficiently. For example, social media. Big data analysis of corporate is considered as a necessary aspect for further improvement in order to improve the profit margin of corporate production and operation, and represents the next frontier of innovation, competition and productivity. Nowadays, corporate data platform is the core component of corporate data storage, computation and analysis for the management of intelligent plant. With the increasing number of intelligent equipments used in intelligent plant, however, intelligent plant can acquire a large quantity of data of Radio Frequency Identification (RFID) and intelligent equipments, thus providing rich data sets for manufacturing corporate. The current trend in corporate systems is to use different big data engines as a means of processing a large quantity of data that cannot be processed by ordinary infrastructures. Corporate infrastructure faces a large number of problems, including challenges such as defining different efficient architecture settings for different applications and defining specific models for corporate analysis. We aim at designing corporate data with higher performance and higher data compression ratio.

II. LITERATURE SURVEY

[1].Advanced process & heterogeneity process are the two issues which can be eliminated by using this paper proposed system. Stream analysis techniques are mostly used by researchers.
It is much useful to enable generalization of the model as well as visual process views with existing research tools.
Advanced Message Queueing Protocol used for real time sql stream to supports large amount of data.
Stream analysis techniques will gives data analysis representation features.

[2].In this paper reviewed the networking architecture and services for big data applications.
Advanced big data analyzing technologies, insights can be acquired to enable better decision making for critical development areas such as natural disaster prediction, health care & economic productivity.
Data is most important fact for both academic, corporate as well as for personal life.So its necessary to store,analyse & secure that data.
Thats why for privacy & security purpose new encryption method is required such as homomorphic encryption.

[3]. Materials data are currently stored in a large number of different ways, dependent on factors such as the area of research, data source, acquisition technique, and type of processing.
Machine learning has significantly accelerated materials discovery across a variety of application areas .Discuss the current & potential future applications for materials informatics in the corporate.
Data are vastly more useful when they are structured in a way that allows systematic search and analysis of the content, which is inherently more difficult with greater diversity of data. This means that along with importability, an ideal infrastructure should be able to structure uploaded data in a standardized way to as great an extent as possible.

[4].Time reduction is most important thing in world of data. Macy's merchandise pricing optimization application provides a classic example of reducing the cycle time for complex and large-scale analytical calculations from hours or even days to minutes or seconds.

Data had to be stored in enterprise warehouses or marts before analysis. Data sources were relatively small and structured, and came from internal sources.

The challenge is adapting operational and decision processes to take advantage of what the new technologies available in the market.

Analytics helps to optimize key processes, functions and roles. It can be leveraged to aggregate both internal and external data.

[5]. Proposed a novel big data architecture for end users using QoE improvement.

The data storage plane stores a wide variety of data collected originating from different data sources. Finally, simulation results demonstrate the high QoE performance of the proposed deep-learning-based algorithm. At any time, people are exchanging data and information from one plece to another place through some network which gives meaningful information & provides the foundation of services to the end users.

[6]. This paper proposed fault diagnosis system method which gives accuracy in their field.

Modern manufacturing systems, machines are more automatic & efficient with their quality & reliability. Machinery fault diagnosis systems is one of the fact for development of industries & internet .By some developed applications and knowledge based system data can be set properly as well as accurately.

[7]. Big data is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data- processing application software. Data with many cases (rows) offer greater statistical power, while data with higher complexity (more attributes or columns) may lead to a higher false discovery rate. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy and data source.
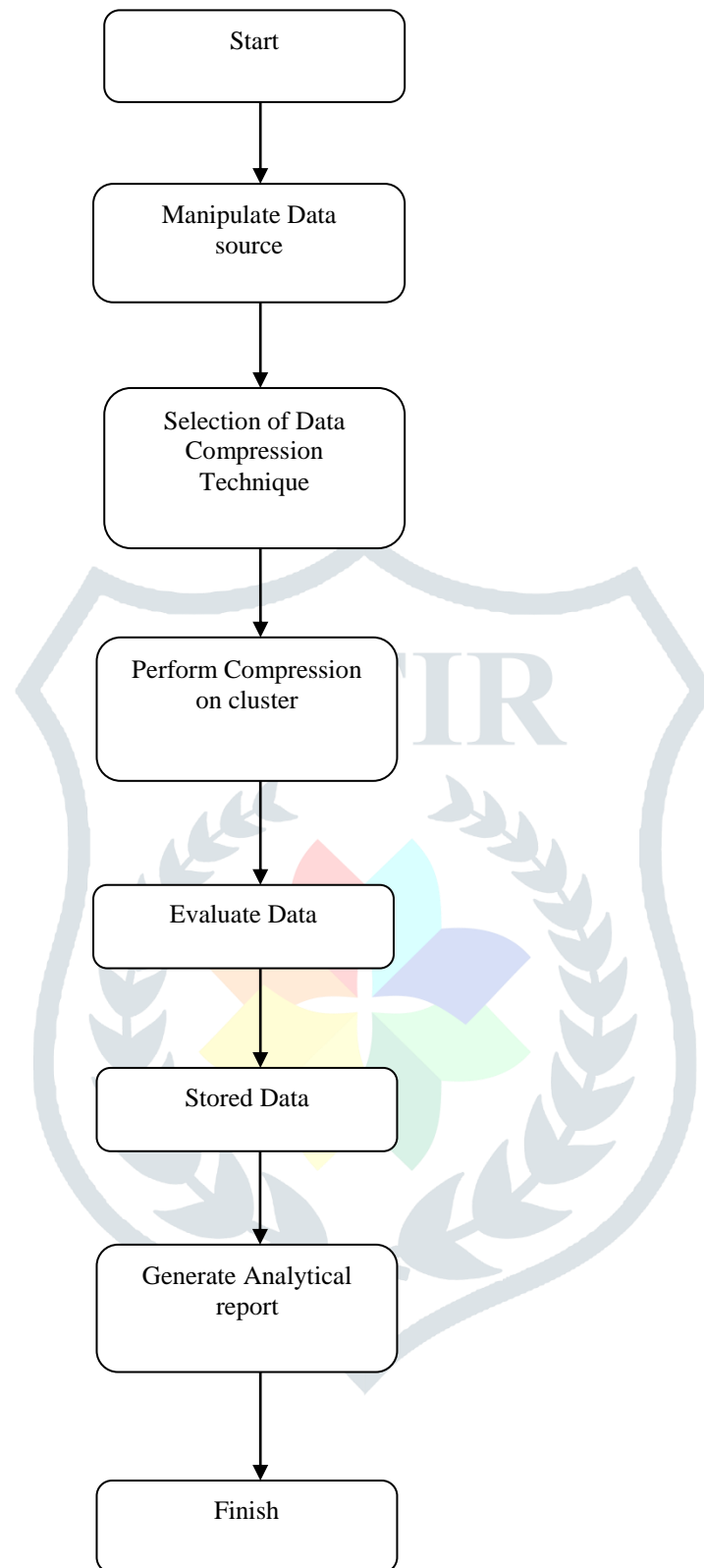
[8]. For meaningful information data must be processed by using some kind of tools like analytics & algorithms which are much advanced. Beacause of privacy & security matter there are number of visible as well as invisible factors with their components used to manage factory.

Data must be processed with advanced tools (analytics and algorithms) to reveal meaningful information. For example, to manage a factory one must consider both visible and invisible issues with various components.

Information generation algorithms must detect and address invisible issues such as machine degradation, component wear, etc. on the factory floor.

III. PROPOSED WORK

In this we are proposed a dynamic execution module in which the system will work on dynamic selection of the data compression technique as well as evaluation of analytics tools 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.

IV. PROPOSED MODEL

```
                    ┌──────────────┐
                    │    Start     │
                    └──────────────┘
                           │
                           ▼
                    ┌──────────────┐
                    │ Manipulate Data │
                    │    source    │
                    └──────────────┘
                           │
                           ▼
                    ┌──────────────┐
                    │ Selection of Data │
                    │  Compression │
                    │   Technique  │
                    └──────────────┘
                           │
                           ▼
                    ┌──────────────┐
                    │ Perform Compression │
                    │   on cluster │
                    └──────────────┘
                           │
                           ▼
                    ┌──────────────┐
                    │ Evaluate Data │
                    └──────────────┘
                           │
                           ▼
                    ┌──────────────┐
                    │  Stored Data │
                    └──────────────┘
                           │
                           ▼
                    ┌──────────────┐
                    │ Generate Analytical │
                    │    report    │
                    └──────────────┘
                           │
                           ▼
                    ┌──────────────┐
                    │    Finish    │
                    └──────────────┘
```

V. REQUIREMENTS

software requirements:

1. Eclipse
2. Apache Tomcat
3. My-SQL
4. JDK7.0

hardware requirements:

1. Ram : 4 GB
2. HDD: 500 GB
3. Processor : i3

VI. CONCLUSION

We starts the process of corporate data processing and the requirement of corporate data analysis . Here we analyse the corporate data by using frameworks with pictorial view, so we can easily analyse the corporate data. We comparing the optimal serialization methods with other high performance methods by using framework for optimize the existing frameworks method.Compared with Java defaults, Protobuf takes only some time to process the data. At same time, our data platform has many data analysis functions including machine-learning, Finite Element Analysis, Optimization and Knowledge Mapping.

REFERENCES

[1] C. Esposito, M. Ficco, F. Palmieri, and A. Castiglione, ''A knowledge- based platform for big data analytics based on publish/subscribe services and stream processing,'' Knowl.-Based Syst., vol. 79, pp. 3–17, May 2015.

[2] X. Yi, F. Liu, J. Liu, and H. Jin, ''Building a network highway for big data: Architecture and challenges,'' IEEE Netw., vol. 28, no. 4, pp. 5–13, Jul. 2014.

[3] B. Meredig, ''Corporate materials informatics: Analyzing large-scale data to solve applied problems in R&D, manufacturing, and supply chain,'' Current Opinion Solid State Mater. Sci., vol. 21, no. 3, pp. 159–166, 2017.

[4] P. Basanta-Val, ''An efficient corporate big-data engine,'' IEEE Trans. Ind. Informat., vol. 14, no. 4, pp. 1361–1369, Apr. 2018.

[5] X. He, K. Wang, H. Huang, and B. Liu, ''QoE-driven big data architecture for smart city,'' IEEE Commun. Mag., vol. 56, no. 2, pp. 88–93, Feb. 2018.

[6] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, ''An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data,'' IEEE Trans. Ind. Electron., vol. 63, no. 5, pp. 3137–3147, May 2016.

[7] Breur, Tom (July 2016). "Statistical Power Analysis and the contemporary "crisis" in social sciences". Journal of Marketing Analytics. (https://www.worldcat.org/issn/2050-3318).

[8] Lee, Jay; Bagheri, Behrad; Kao, Hung-An (2014). "Recent Advances and Trends of Cyber-Physical Systems and Big Data Analytics in Industrial Informatics" (https://www.researchgate.net/publication/266375284). IEEE Int. Conference on Industrial Informatics (INDIN) 2014