International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-09, Issue-01, Apr 2023

Scalable Data Analytics Platform for ML in Smart Manufacturing Systems

[1]Prof. Vishal R. Shinde, [2]Miss Anushka Hemant Borole , [3]Miss Mrunali Mahesh Dhaske , [4]Miss Shreya Charansingh Yadav ,[5]Mr. Onkar Vijay Bhoir

[1]Asst. Prof. ,[2,3,4,5] UG Student ,[1,2,3,4,5] Computer Engg. Dept. Shivajirao S. Jondhle College of Engineering and Technology , Asangaon , Maharashtra , India.

[1]mailme.vishalshinde@gmail.com ,[2]boroleanu2002@gmail.com ,[3]mrunalidhaske@gmail.com, [4]shreyayadav1712@gmail.com , [5]bhoironkar28@gmail.com

Abstract - It is challenging for the manufacturing industry to meet changing customer demands. As a result, products must be manufactured efficiently, with minimal interruptions, and with low resource consumption. To accomplish this, current technologies must manage and analyse the data produced by industrial equipment. RAMI model architecture uses ML and large data techniques. It is essential that Industrie 4.0 standards and regulations are followed because the big data revolution in production is still in its early stages. This article outlines the specifications for creating a scalable analytics platform, to evaluate industrial data in accordance with Industrie 4.0 guidelines. In light of the criteria, the suggested standard big data architecture for commercial applications that use ML is contrasted with similar studies. The envisaged architecture has now been put into practice in the Lab Big Data at the SmartFactoryOWL and the scalability and performance of SmartFactoryOWL have been assessed using the concurrent computation of an industrial PCA model. The results show that the proposed architecture is exponentially scalable, suitable to ML scenarios, and will improve the automation and control processes of the production systems [1].

Keywords- Big data, Industrie 4.0, ML, Industrial automation .

I. INTRODUCTION

In order to realise the idea of the "Smart Factory," Industrie 4.0 has emerged as a digital transformation of industrial manufacturing [5]. It combines the production and IoT worlds. The so-called IIoT is a collection of interconnected intelligent sensors, actuators, and computing power embedded in machines, workpieces, transport tools, and products in smart production lines. Information interchange and how to leverage it to enhance business models and services are at the heart of Industrie 4.0. The data gathered during production operations must be managed, processed, and analysed for this reason. By analysing industrial data, finding patterns in it, and generating insights to make wise judgements and predictions, ML technologies play a significant role in industrial automation [7].

Big data as well as ML technologies are hence critical to industrial automation. Having a standardised, resilient architecture that combines BDA and ML technologies for effective industrial data analytics is a crucial step along this path. Many architectures for BDA in Industrie 4.0 have been developed, even though in the era of big data is still in its early stages, notably in the manufacturing industry [8]. However, these architectures either don't have enough physical structures to account for every stage of the data lifecycle. This paper offers a BDA that takes advantage of the potential of big data technologies to apply ML algorithms in the

smart manufacturing business, based on a literature assessment of the Industrie 4.0 requirements and standards. To offer data analytics platforms for automated production systems, numerous studies have been carried out. Examples of such systems are AWS IoT [2], which was created to integrate Amazon Web Services for IoT applications, and the IoT solution developed by General Electric and NTT Docomo and documented in [3]. These systems, however, are commercial and were made more for data presentation than for uses of ML, including predictive maintenance.The suggested ML architectures include many cloud-based services. The purpose of this study is to present a reference architecture for scalable data analytics in manufacturing technologies that adheres with RAMI 4.0 and corrects the problems faced in the previously stated research.

II. AIMS AND OBJECTIVE

a) Aim

The project's goal is to provide a solution that is linearly scalable, responsive to ML use cases, and will aid in enhancing the production systems' industrial automation

 DOI : 10.5281/zenodo.7863785

processes. It centres on the four letters Scalability, Availability, Performance, and Security (S, A, P, S).

b) Objective

• The major goal is to use contemporary technology to handle and analyse the massive volume of data produced by industrial equipment.

• Using ML and big data techniques in accordance with Industrie 4.0 specifications.

• Mainly concentrates on data integration, distributed control, and remote collaboration.

• To show the platform's viability and efficiency in intelligent manufacturing systems.

• To cover data definition, storage, processing, and retrieval, it incorporates a distributed data warehouse.

• Based on actual and historical data, it employs a data analytics methodology to develop empirical decision-making models.

III. LITERATURE SURVEY

Paper 1: The Proposal of a Lambda Architecture Adoption for Real Time Predictive Maintenance:

The IoT technology has advanced recently, and applications in the maintenance area are expected to follow. The paper proposes a maintenance platform based on lambda architecture that analyzes sensing data, detects anomalies, extracts a new detection rule in real time, and the cloud orders maintenance automatically. The system analyzes the data collected in batch in detail and updates the learning model of edge nodes to improve analysis accuracy. Through the collection and analysis of sensor data, this paper is able to visualize the status of factories, facilities, and products, as well as monitor production effectiveness, reflect production plans, control logistics, and replace defective products to optimize the supply chain, coordinate exterior systems, and speed up manufacturing , maintenance operations ,data integration, distributed control. Therefore, IoT platforms were successful in developing and operating IoT applications. In particular, the current issues include the insufficient analysis of real-time scenarios, the high price of collecting real-time data, and the difficulty in configuring failure detection rules. Thus increasing risk in deployment as well as operation. Current issues are specifically the high cost of collecting sensing data and configuring failure detection, as well as the insufficient analysis of real-time situation. [8].

Paper 2: The Data analysis & data mining in process industry: Role of ML:

In last few decades, knowledge discovery and decision-making/support have benefited greatly from the data mining and analysis in the processing sector. Information extraction, data pattern discovery, and forecasting are all made possible by ML, which serves as the computational engine for data analysis & data mining. This article offers a summary of the data mining and ML-based data analysis techniques that have been used in the process sector over the years. To determine the state-of-the-art in data mining and data analytics, 10 supervised learning algorithms, 8 unsupervised learning algorithms, and the status of the application for semi-supervised algorithms are employed. A variety of views are highlighted for future study on the use of data analysis & data mining in the process sector, and these perspectives are discussed [7].

Paper 3: A survey of IOT and BDA integrated solutions for industrie 4.0:

Massive amounts of industrial data are being produced as a output of the rapid expansion of the IIoT. Utilizing this data is necessary to support organisational and business objectives. Big data technologies must be quickly adopted in order to enable data analytics in industrial automation. In order to produce business insights from industrial data, this study examines the relationships between IIoT and big data technologies.Another source of needs for cloud-based services is use case scenario value-based services with an emphasis on monitoring system and preventative maintenance services. To determine how well these platforms fit the needs generated from the use case, a survey of a few chosen cloud-based platforms is undertaken. The findings indicate that current general cloud platforms could adopt more IIoT platforms and applications while current industrial cloud platforms should diversify their offering with big data frameworks. In addition, concerns about using the public cloud for IIoT applications are examined, and an architecture for combining cloud-based IIoT and big data solutions is offered[1].

IV. EXISTING SYSTEM

Big data approach is currently being used in manufacturing for preventative maintenance. Data sources, data transmission, big data processing, and visual presentation are the four levels that make up the architecture. Although Hadoop is utilised for offline data distribution and computation, Apache Storm is used for actual processing.The architecture puts a lot of emphasis on batch and real-time processing, as well as how to use these elements for preventative maintenance. However, this study does not cover elements like data integration, data storage, or streaming technology. The current system seeks to establish a reference architecture for scalable data analytics in manufacturing systems that complies with RAMI 4.0 and overcomes the problems in the research stated above.

Disadvantages Of The Existing System:

• Such services are commercial and were created more for data visualisation as for machine learning uses like predictive maintenance.

 DOI : 10.5281/zenodo.7863785

• Either architectures do not adhere to the Industry 4.0 standards or do not have solid frameworks to handle every stage of the data lifecycle.

• The RAMI 4.0's layers are not all covered by the lambda architectural approach (RAMI 4.0).

V. COMPARATIVE STUDY

Table 5.1 Comparative Analysis of Existing System

Sr. No.	Author	Project Title	Publication	Technology	Purpose
1.	Y. Yamato, H. Kumazaki, and Y. Fukumoto	The Proposal of the Lambda Architecture Adoption for the Real Time Predictive Maintenance	IEEE, 2019	IOT	Maintenance platform in which edge nodes determine the data from sensors, detect ambiguities, find new rule for detection in real time.
2.	Z. Ge, Z. Song, S. X. Ding, and B. Huang	The Data analysis & data mining in process industry: Role of ML	IEEE, 2017	Data analysis & data mining	A computer tool for data pattern identification, information extraction, prediction.
3.	K. Al-Gumaei and et al	A survey of IOT and BDA integrated solutions for industrie 4.0	IEEE, 2018	Industrial Internet Of Things (IIOT)	Real worth services approach that emphasises condition monitoring and preventive maintenance.

VI. PROBLEM STATEMENT

The manufacturing sector is continuously struggling to meet the shifting demands of its customers. Products must therefore be produced using effective procedures, few disruptions, and little resource usage. To achieve this, it is necessary to handle and analyse the large data generated by industrial equipment using modern technologies. Because in era of big data in manufacturing is still in its early stages, a reference architecture that integrates big data and ML technologies and complies with Industrie 4.0 standards and regulations is necessary.

VII. PROPOSED SYSTEM

The proposed system in this study is built on layers which includes the asset layer, the integration layer, the communication layer, the information layer, and the functional layer and finally the business layer. If not properly safeguarded, manufacturing equipment and systems are exposed to cyber threats due to their increased interconnection. This calls for simultaneous protection of all layers of the proposed architecture, from field devices on the asset layer to corporate management systems on the business layer. Training data is loaded from data storage and prepared for model learning for each ML use case. Using the training data, select and fit a suitable model is the next stage.

The model is then assessed using test data. These actions can be accomplished using a batch processing framework and are referred to as model building. Data generated in the asset layer is inconsistent in terms of usage and format; as a result, it requires integration using standard information models .The information layer, data is described by means of semantics and becomes information.

VIII. ALGORITHM

1. **Standardize the data:** Calculate the mean and standard deviation for each variable, then scale the data to have a resultant average of zero and a resultant standard deviation of one.

```
from sklearn.preprocessing import StandardScaler
scale = StandardScaler()
data = scale.fit_transform(X_data)
print(data)
```

2. **Compute the covariance matrix:** Calculate the covariance matrix S of the standardized data, which is given by

   S = (1 / (m-1)) X^T X

   where X^T is the transpose of X.

3. **Sort the eigenvectors:** The eigenvectors should be arranged in decreasing order of the associated eigenvalues.

   (A – λ I) X = 0

4. **Select the no. of principal components:** Based on the proportion of variable that is explained by each principal component, Choose how many.principal components, or k, to keep.

```
from sklearn.decomposition import PCA
pca = PCA( n_components = 2)
pca_values = pca.fit_transform(data)
```

5. **Transform the data:** Project the standardized data onto the k principal components by computing the matrix product

   Z = X V_k

   where V,_k is the matrix that contain the first k no. of eigenvectors.

 DOI : 10.5281/zenodo.7863785

International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-09, Issue-01, Apr 2023

6. **Determine the results:** Key patterns in the data can be captured by the principal components, which can be seen as new variables.

IX. MATHEMATICAL MODEL

When using data analytics, it is frequently necessary to create system models from historical data and then apply those models to the evaluation or process data.The principle component analysis (PCA) matrix, which has several industrial uses including dimensional reduction and condition monitoring techniques, is an illustration of such models. SVD for the given covariance matrix is used to create the PCA matrix using historical measurement vectors xk collected at time occurrences k = 1 through n.

$$\sum_x = \frac{1}{n-1}\sum_{k=1}^{n} 1(x_k - \mu_x)\,(x_k - \mu_x)^T \quad \ldots(1)$$

with

$$\mu_x = \frac{1}{n}\sum_{k=1}^{n} 1x_k \quad \ldots(2)$$

To ensure that the condition nC n is true, a huge proportion of historical measurement vectors are examined in big data applications. When compared to the work required to calculate the covariance matrix, the computational cost for the SVD is negligible in certain application instances[1].

X. SYSTEM ARCHITECTURE

Description: Due to the exponential development in data that is generated by IIoT and industrial systems, the industry is compelled to use new technologies to handle big data. Therefore, a reference architecture that combines big data frameworks with ML solutions is needed for industrial automation. The primary contribution of this work is the suggestion of a reference architecture model for big data and ML in industrial automation that is compatible with the RAMI 4.0. The conceptual architecture has been developed using the SmartFactoryOWL, and its performance and scalability have been tested using actual industrial data. The implemented platform has proven to be adaptable, linearly scalable, and compatible with industrial analytics requirements [1].

1.Asset Layer(AL): This layer includes components such as machines, personnel, products, and engineering systems. The main data sources in the production industry are these elements.

2.Integration Layer(IL): This layer is where the changeover from the real world to the virtual one takes place.

3.Communication Layer(CL): Between the integration layer and the information layer, this layer facilitates communication. Information can be exchanged using Ethernet, Bluetooth, or Wi-Fi interfaces, or it can be delivered and received using protocols like TCP/IP, HTTP, and FTP.

4.Information Layer(IL): Data is semantically described and transformed into information at the information layer. On this layer, data is saved for later access and examination.

5.Functional Layer(FL): The functional layer (FL) is divided into the visualisation and analysis sublayers. The visualisation layer presents the prediction findings and gives data scientists the option to annotate predictions with expert knowledge to make the analysis process easier.

6.Business Layer(BL): Business needs, use case descriptions, and obstacles to be solved are all stated on this layer.The study findings from production data analytics aid stakeholders in making wise decisions and optimising their operations.

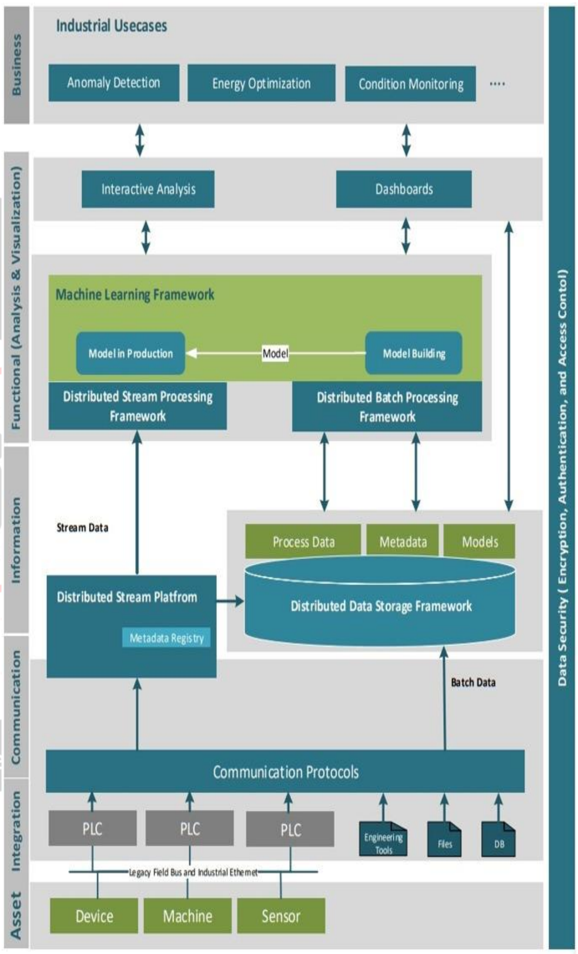

Fig.10.1: The System Architecture

XI. ADVANTAGES

- The foundation of transformation is data analytics and digital trust, with a focus on people and culture to drive it.
- It helps in creating strong digital connections with customers who have more influence.
- Significant change is needed for powerful enterprise-wide data analytics capabilities.
- Globalization is accelerating thanks to Industry 4.0, but it still has a very local flavour.

 DOI : 10.5281/zenodo.7863785

- Moving towards cloud to cut infrastructure expenditures.
- The unified data platform, which centralises all data and provides predictive analytics, has improved the efficiency of data management.

XII. DESIGN DETAILS

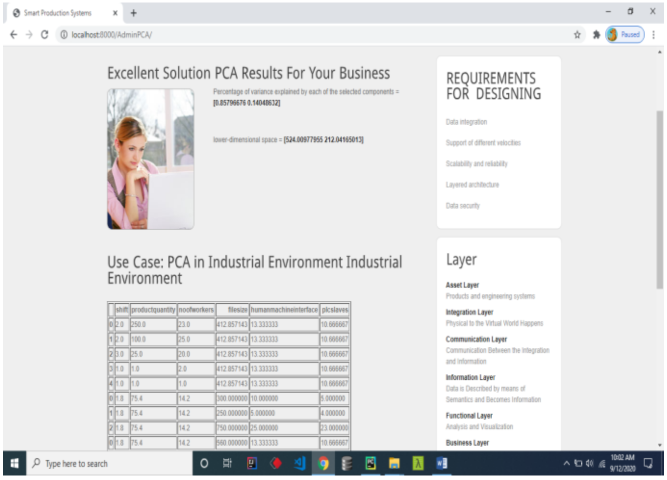

Fig 12.1: Result

PCA is all about finding the directions of maximum variance in high-dimensional data and project it onto a smaller dimensional subspace while retaining most of the information. In this paper it is attempted to build the PCA algorithm for a smart manufacturing system. In this study the computed values of covariance matrix and eigenvalues are shown as a result.

The "core" of a PCA is represented by the eigenvectors and eigenvalues of a covariance (or correlation) matrix.The new feature space's directions are determined by the eigenvectors (principal components), and its magnitude is determined by the eigenvalues. In other words, the data variance along the new feature axes is explained by the eigenvalues. The eigen decomposition of the covariance matrix, a d*d matrix whose elements each indicate the covariance between two features, is the traditional method for doing PCA. The objective is to enhance computing efficiency while keeping the less dimensions of the information by projecting a d-dimensional dataset onto a (k)-dimensional subspace (where k<d).

XIII. CONCLUSION

Thus, we have tried to implement the paper "Scalable Data Analytics Platform for ML in Smart Manufacturing Systems",Khaled Al-Gumaei , Arthur Muller , Jan Nicolas Weskamp , Claudio Santo Longo , Florian Pethig , and Stefan Windmann, IEEE,2019 and the conclusion is as follows:

The industry is forced to employ new technology to handle big data due to the exponential growth in data volume generated from IIoT and industrial systems. Consequently, a RAMI 4.0 is required for industrial automation that blends big data frameworks with ML solutions. This paper's is primarily responsible for the proposal of a reference architecture that fits with the RAMI 4.0 for big data and ML in industrial automation. The conceptual architecture was developed using the SmartFactoryOWL, and its performance and scalability were tested using actual industrial data. The implemented platform has proven to be linearly scaleable, adaptable, and compatible with industrial analytics requirements.

REFERENCE

[1] "Scalable Data Analytics Platform for ML in Smart Manufacturing Systems",Khaled Al-Gumaei , J. N. Weskamp , Arthur Muller , Claudio S. Longo , Florian Pethig , and Stefan Windmann, IEEE,2019

[2]"AWS IoT Platform website," [Accessed April-2019]. [Online]. Available: https://aws.amazon.com/iot/how-it-works/?nc1=h ls

[3] "NTT Docomo press release website," [Accessed April-2019]. [Online]. Available: fhttps://www.nttdocomo.co.jp/english/info/media center/pr/2015/0708 00.htmlg

[4] J. Herwan, H. Sawade, R. Oleg ,S. Kano, and N. Kasashima, "Cyber- The Physical System Architecture for Machining Production Line," 2018 IEEE ICPS, 2018.

[5] Industrie 4.0, "Industrie 4.0," [Accessed 28-February-2018].

[6] K. Al-Gumaei and et al., "A survey of IOT and BDA integrated solutions for industrie 4.0," ETFA, September 2018.

[7] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "The Data analysis & data mining in the process industry: The role of ML," IEEE Access, 2017.

[8] E. Trunzer and et al., "A flexible architecture for the mining of data from heterogeneous data sources in APS," 2017 IEEE ICIT, March 2017.

[9] Y. Yamato, H. Kumazaki, and Y. Fukumoto, "The Proposal of a Lambda Architecture Adoption for the Real Time Predictive Maintenance", 2016.

 DOI : 10.5281/zenodo.7863785