INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING AND APPLICATIONS
VOLUME XVI, ISSUE VII, JULY 22, WWW.IJCEA.COM, ISSN 2321-3469

MACHINE LEARNING IN SMART PRODUCTION SYSTEMS WITH SCALABLE ANALYTICS PLATFORM

Mr. Sk. Alimoon[1], K.Karuna[2], V.Prakash[3], P. Lakshmi Latha[4], S. Bala chandrudu[5]

[1]Associate Professor, [2,3,4,5] Scholar, Krishna Chaitanya Institute of Technology &Sciences , Markapur,A.P, India

ABSTRACT:

The manufacturing sector faces significant challenges in meeting the consumer's shifting needs. Therefore, manufacturing procedures must be efficient, rarely interrupted, and resource-efficient. To do this, massive amounts of data generated by industrial machines must be managed and assessed using contemporary technology. Because the big data era in the manufacturing sector is still in its infancy, there is a need for a reference architecture that incorporates big data and machine learning technologies and is compliant with Industrie 4.0 standards and specifications. In this article, the requirements for developing a scalable analytics platform for industrial data are established using the Industrie 4.0 standards and literature.Based on these requirements, a reference large data architecture for business machine learning applications is proposed, and it is compared to similar publications. Finally, the parallel processing of an industrial PCA model in the Lab Big Data at the SmartFactoryOWL has been used to evaluate the performance and scalability of the suggested architecture. The results show that the proposed structure is linearly scaleable, adaptable to machine learning use cases, and would improve the industrial automation processes of the production systems.

Keywords — Big Data, Machine Learning, Industry 4.0, Industrial Automation.

1. Introduction

Thanks to Industry 4.0, a digital revolution in industrial manufacturing that integrates the production and Internet of Things (IoT) worlds, the idea of the "Smart Factory" has become a reality. An interconnected system of intelligent sensors, actuators, and computer devices, collectively known as the "Industrial Internet of Things," are built into smart production lines' machinery, workpieces, transport systems, and products (IIoT).

The fundamental ideas of Industry 4.0 centre on information sharing and how it may be used to enhance business models and services. Data acquired during industrial processes must be organised, processed, and evaluated in order to do this. Machine learning technologies play a crucial role in industrial automation through the analysis of industrial data, the finding of patterns within it, and the creation of insights to make educated decisions and predictions. Therefore, big data and machine learning are essential for industrial automation. A critical step on this path is having a standardised, reliable architecture that blends big data and machine learning technologies for efficient industrial data analytics.Even though the big data era is still in its infancy, several concepts for big data analytics in Industrie 4.0, particularly in the manufacturing industry, have been put forth. However, these designs either don't follow Industrie 4.0 standards and specifications or don't have true structures to cover every stage of the data lifecycle. Based on a literature review of the requirements and standards for Industry 4.0, this article presents a big data architecture that makes advantage of the capacity of big data technologies to deploy machine learning techniques in the smart manufacturing business. Several studies have been done to provide data analytics platforms for automated production systems.

2. Literature Survey

Recently developed IoT technologies are expected to have applications in the maintenance industry. Furthermore, due to poor real-time condition analysis, high costs associated with collecting sensor data, and the configuration of failure detection algorithms, IoT servicing applications are not yet widely used in Japan. In this paper, we propose a maintenance platform based on the concept of lambda architecture, where edge nodes examine sensing data, identify anomaly, and obtain a new detection principle in real time, while a cloud automatically commands maintenance, also thoroughly assessments all data collected through batch processing, and upgrades learning models of edge nodes to improve analysis accuracy. Recent advancements have been made in cloud and Internet of Things (IoT) technology. Despite the fact that there are a wide variety of IoT applications for the cloud, manufacturing and maintenance are thought to be the most likely application areas because they are also the subject of Industrie 4.0. We are able to see the conditions of factories, facilities, and goods by collecting and analysing sensor data. Additionally, we have the ability to present production schedules, manage logistics, and replace faulty items in order to enhance supply chains, cooperate with external systems, and expedite production and maintenance processes. Therefore, IoT platforms want to be able to efficiently construct and manage IoT applications. The present IoT systems, however, are insufficient to speed up maintenance activities because they are primarily built to view the statuses of the things using sensor data that has been acquired. The high expense of gathering sensing data and setting up fault detection criteria, as well as the inadequate analysis of real-time scenario, are current problems. Deployment and operation are now riskier.

Data mining and analytics have made significant contributions to knowledge discovery and decision-making/support in the process industry over the past few decades. Fundamental tools for data pattern discovery, information extraction, and prediction are provided by machine learning, which serves as a computational engine for data mining and analytics. This article provides an overview of the machine learning-based data mining and analytics tools that have been applied in the process industry throughout the years. The state-of-the-art in data mining and analytics is examined using eight unsupervised learning, ten supervised learning, and the implementation status of semi-supervised learning algorithms.

A variety of views are highlighted for future study on data mining and analytics in the process industry, and these perspectives are discussed.

The exponential rise of the "industrial internet of things" is producing enormous amounts of industrial data. To support organisational and business goals, this data must be used. Fast adoption of big data technology is necessary to enable data analytics in industrial automation. This study explores the connections between IIoT and big data technologies to get business insights from industrial data. A source of requirements for cloud-based solutions is the Industrie 4.0 use case scenario value-based services, with a focus on condition monitoring and preventative maintenance services. A survey of a few selected cloud-based platforms is conducted to ascertain how well these platforms match the needs derived from the use case.The results show that existing general cloud platforms may embrace additional IIoT platforms and applications, while existing industrial cloud computing platforms should broaden their product offering with big data frameworks. An architecture for fusing cloud-based IIoT and big data solutions is also provided, and issues regarding using the public cloud for IIoT applications are discussed.

Infrastructure, business, defence, and transportation are just a few of the industries that employ cyber-physical systems (CPS). In this work, a cyber-physical production system (CPPS) architecture is designed in order to provide an adaptive and autonomous system in the machining line. With this CPPS, it is also able to monitor tool wear and breakdown as well as surface roughness. A few preliminary examples are given. The phrase "cyber-physical systems" was originally used in the US in 2006 to refer to how computer interacts with a physical process, frequently through a feedback loop where the computation effects the physical process and vice versa. It covers a broad range of subjects, including military, aviation, transportation, and infrastructure.Although a CPS need not always include the Internet, the rapidly growing disciplines of computer science, information technology, and networking have given rise to a new class of CPSs known as IoT systems. This might lead to the manufacturing industry's fourth industrial revolution (also called Industry 4.0 in Germany). The first industrial revolution started in 1764 with the invention of the mechanical loom, water power, and steam power devices. The second industrial revolution began with the construction of the first Ford assembly line in 1913. The third industrial revolution began with the invention of the first PLC in 1968 and the widespread deployment of industrial robots. The Cyber-Physical Production System will define the fourth industrial revolution.

Distributed machine learning has typically been approached from a data parallel viewpoint, where vast data are divided across numerous workers and an algorithm is executed concurrently over several data subsets, in order to speed up and/or ensure correctness. A related subject that has received comparably less attention is how to guarantee efficient and accurate model simultaneous execution of ML algorithms when parameters of an ML programme are distributed to several workers and undergo concurrent iterative modifications. In our opinion, the problems that model and data parallelisms provide for theoretical analysis, algorithmic optimization, and system design are rather separate. By recognising and utilising the dynamic structural elements of the model, we create the STRADS model-parallelism system in this work, which provides a programming abstraction for parameter update scheduling of ML programs. STRADS improves distributed ML's memory efficiency and enables a flexible trade-off between scheduling efficiency and adherence to inherent model dependencies. We compare the performance of model-parallel algorithms on STRADS to popular Lasso, topic modelling, and matrix factorization implementations.

3. SYSTEM ARCHITECTURE

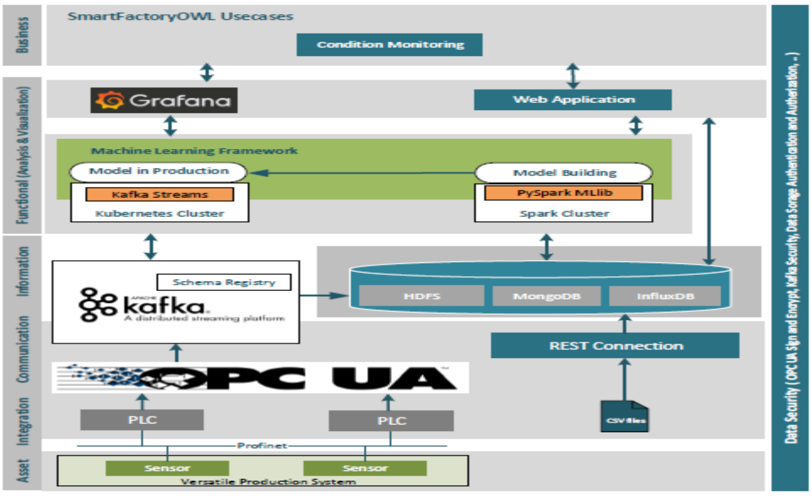

Fig. 1 System Architecture

4. IMPLEMENTATION

4.1 Modules Description

4.1.1 User : The user can initially register. For further chats, he required a functional user email and phone following enrollment. The administrator can activate the user after registration. Once the admin has activated the user, they may log onto our system. The user can preprocess the data. According to this theory, data from the Asset layer, Integration layer, Communication layer, Information layer, Functional layer, and Business layer are important in Industry 4.0. All of the user-added data must first be added to the database server. The PCA method's results may then be checked.

4.1.2 Admin: Admin can sign in using his login information. After logging in, he may start the users. The only programmes that let the enabled user log in are ours. The admin can change the details of each tier's attributes in accordance with Industry 4.0 regulations. The pca algorithm can be run by the user-posted data admin. A small number of columns are selected from each layer using the PCA technique.

4.1.3 Data Preprocess: Data pre-processing ensures the availability of high-quality data derived from the source dataset. It is an important step since a substantial portion of the data generated and stored in sqlite databases is deficient in high-quality data. By using the read frame function of the Django pandas io module, all models object querysets will be converted into pandas dataframes. since the input for machine learning techniques is pandas dataframes. At the information layer, data is semantically characterised and converted into information. Semantics may be kept in metadata registries or established information models for particular areas of the information business, such as OPC UA Companion Specifications. Data is stored for subsequent access and review on this layer.

4.1.4 Principal Component Analysis (PCA): In order to evaluate or analyse current process data when employing data analytics, it is usually required to first build system models using past data. Principal component analysis (PCA) is a technique for identifying different patterns in a dataset by reducing variances. To make discovery and analysis easier, data sets are cleaned up. The Principal Component Analysis method is based on a number of mathematical ideas, such as Variance and Convariance Eigen Vectors and Eigen. A variety of virtual machines have been developed on our big data cluster in order to build the necessary clusters for the PCA's parallel computation.

5 RESULTS

Fig. 2 Home page

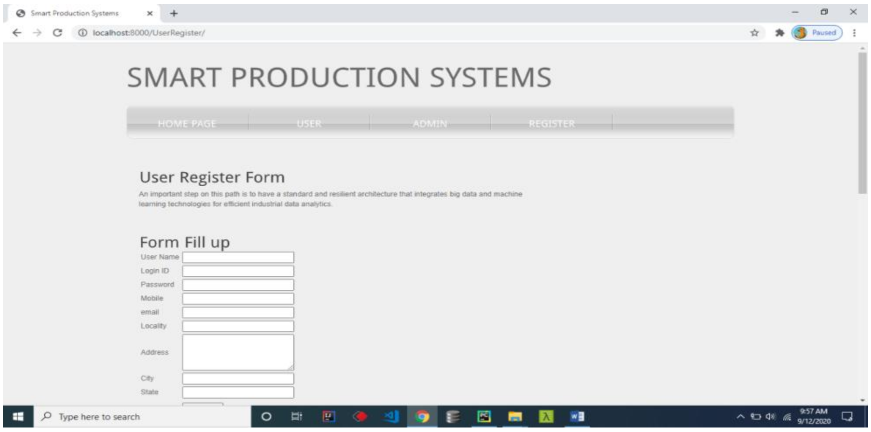

Fig. 3 User Register

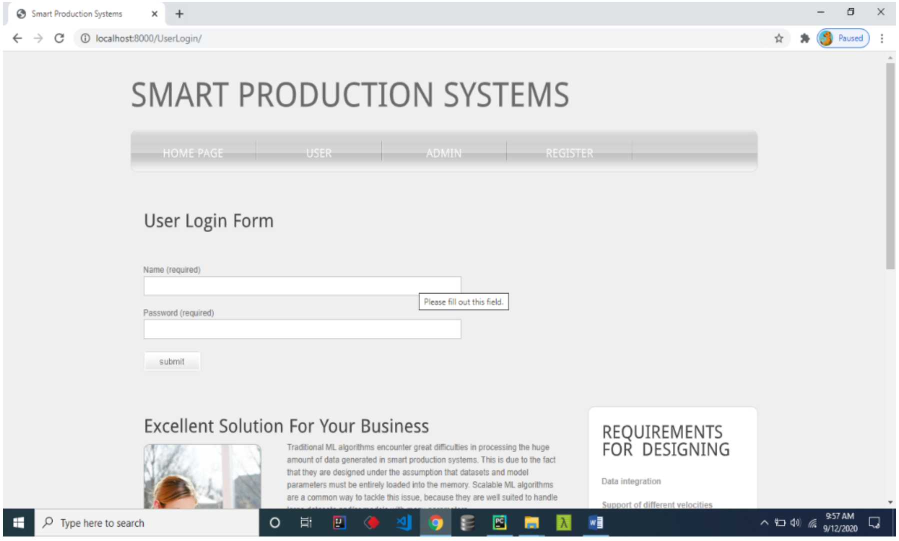

Fig.4 User Login

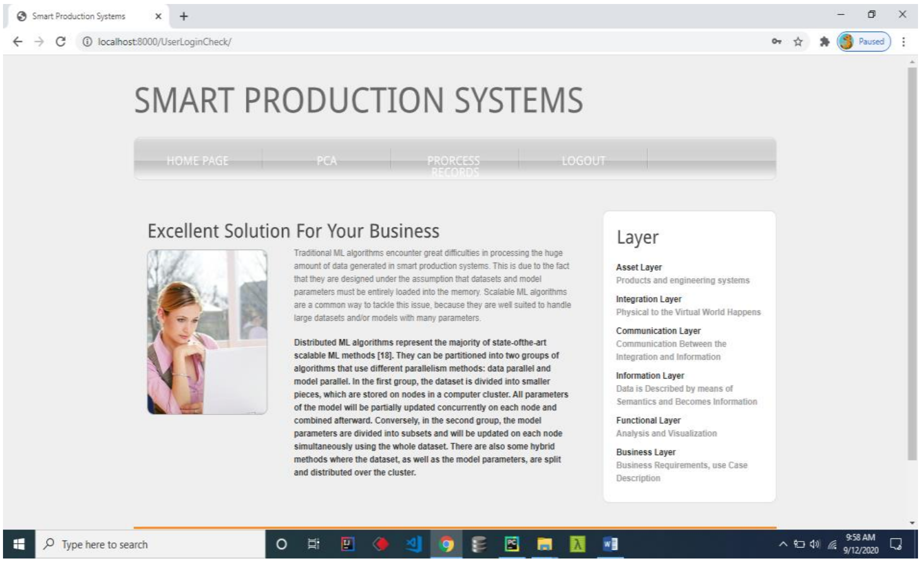

Fig. 5 User Home Page

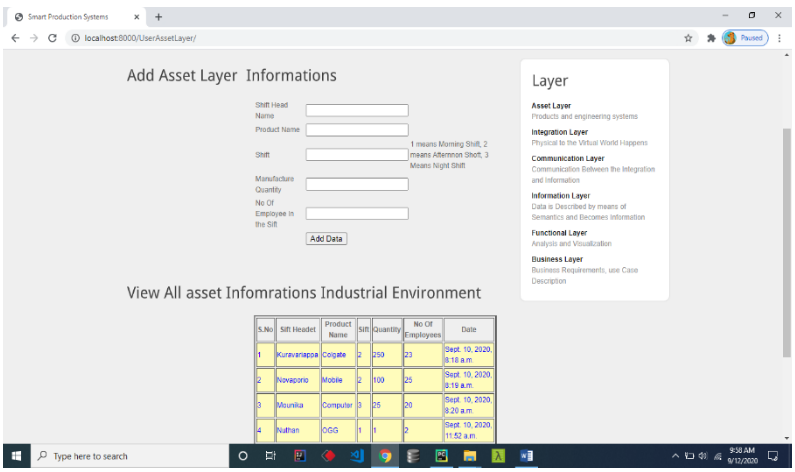

Fig. 6 Asset Layer Information

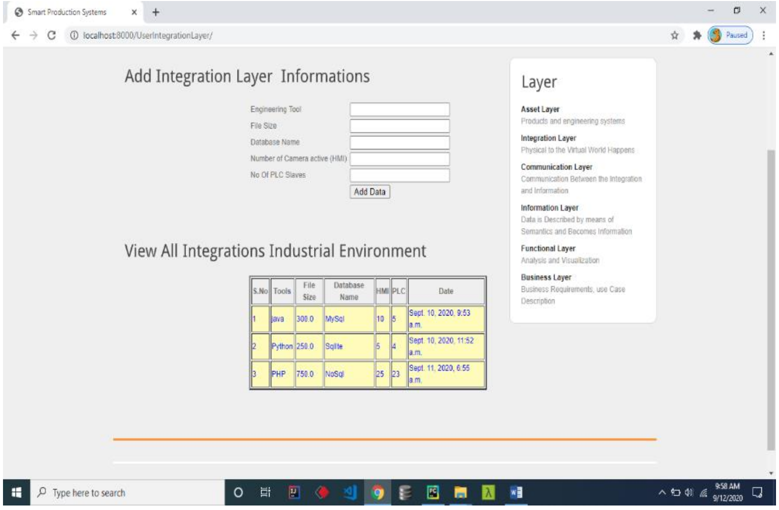

Fig. 7 Integration layer information

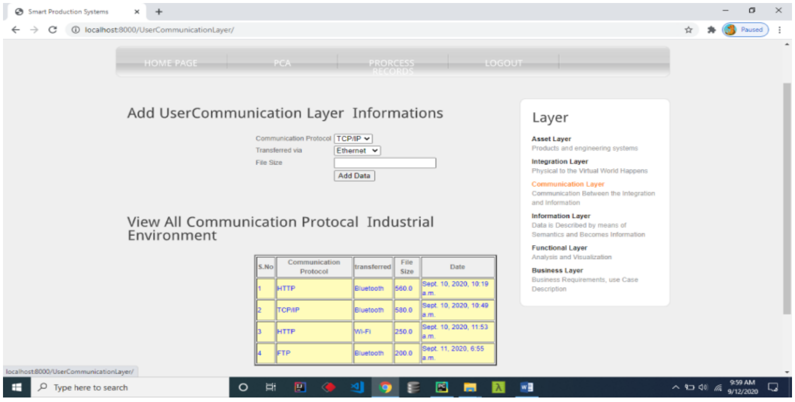

Fig. 8 Communication Layer Information

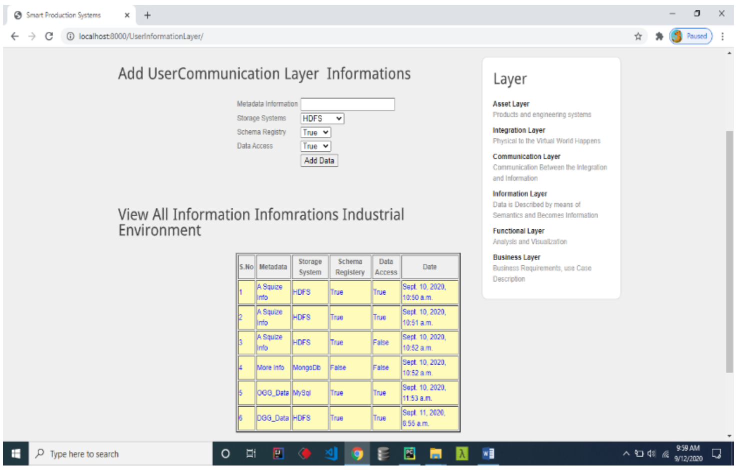

Fig. 9 Information Layer

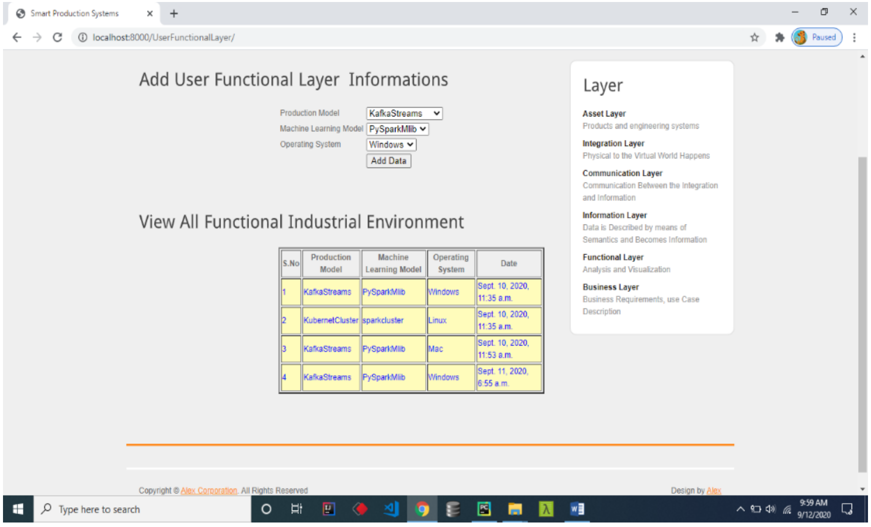

Fig. 10 Functional layer Information

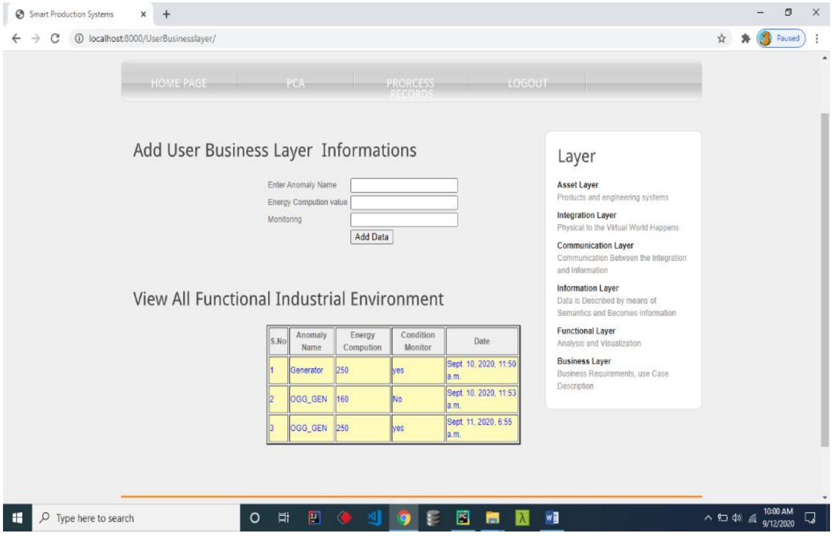

Fig. 11 Bussiness layer Information

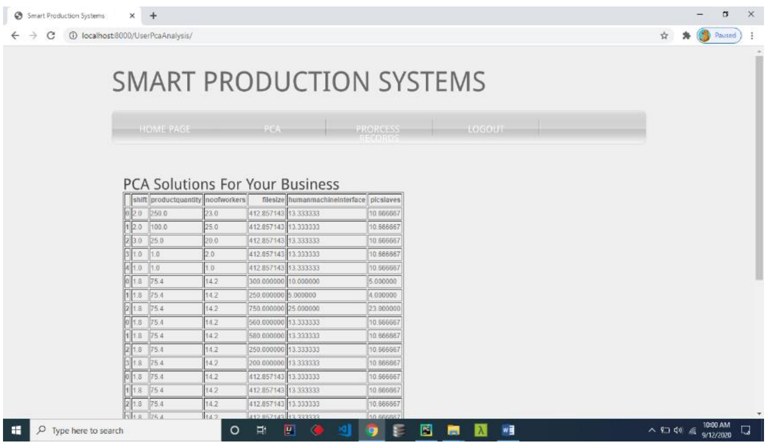

Fig. 12 PCA Data

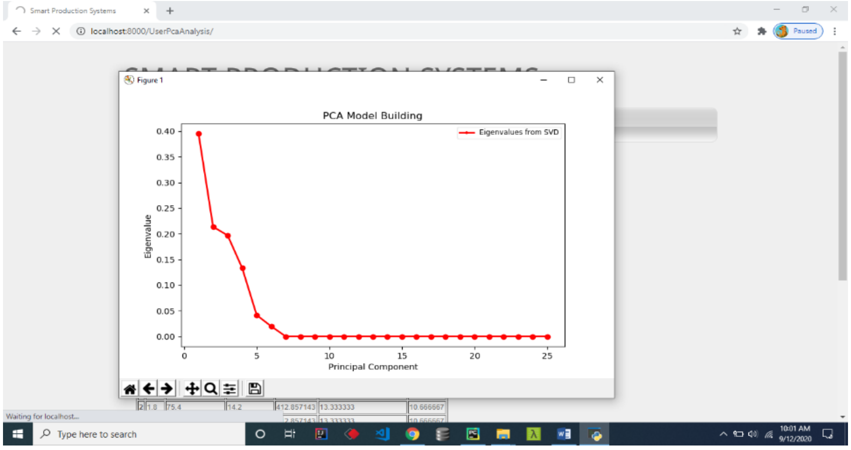

Fig. 13 PCA Graph

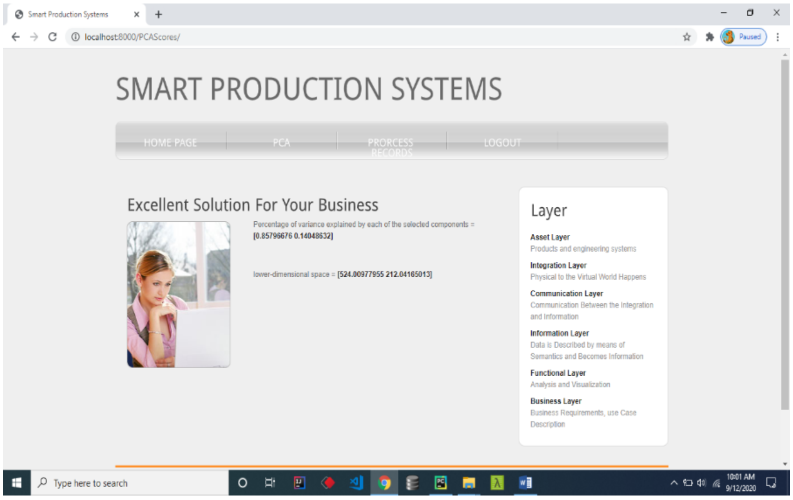

Fig. 14 PCA Results

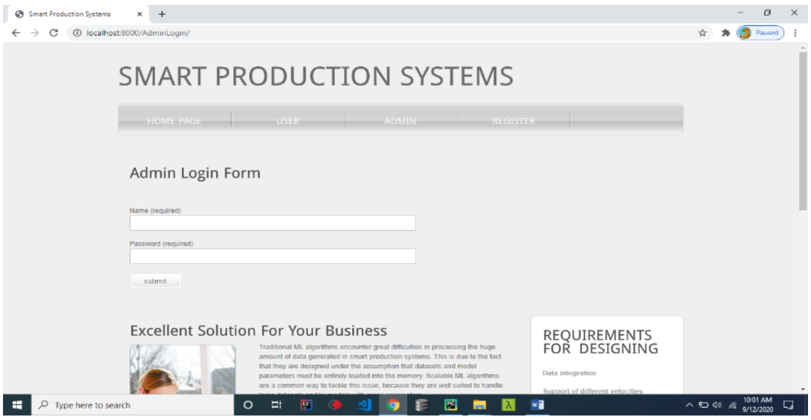

Fig. 15 Admin Login page

Fig. 15 Admin Home Page and Activating Users

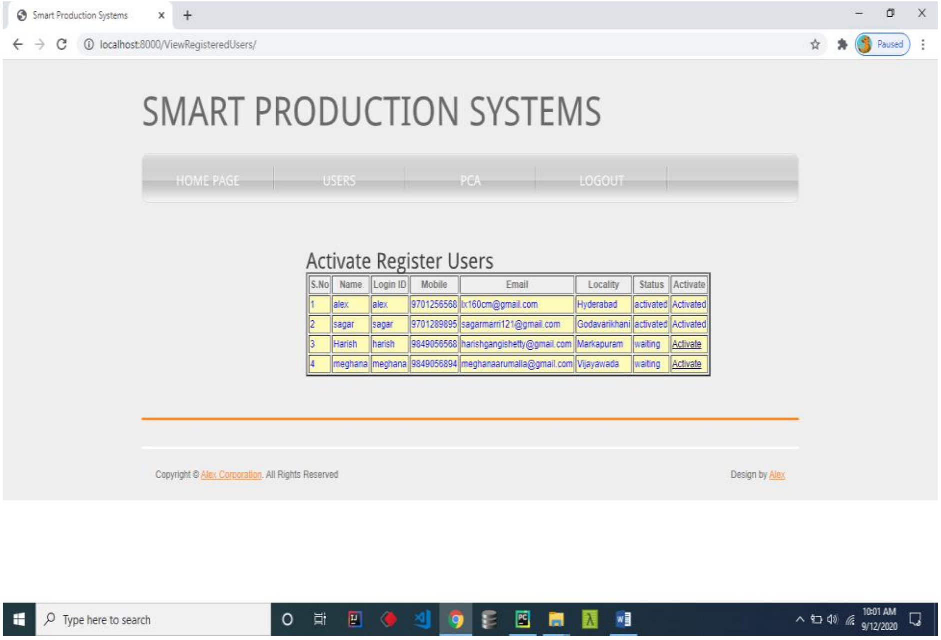

Fig. 16 Admin Home Page

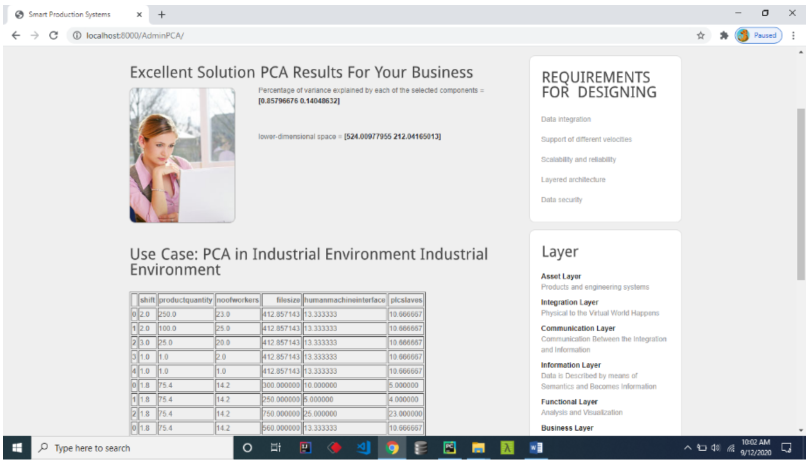

Fig. 17 Admin Side results

5. CONCLUSION

Since the amount of data created by IIoT and industrial systems is increasing at an exponential rate, the industry is compelled to use new technologies to manage big data. Machine learning may be used to monitor, forecast, diagnose, and improve production operations. Therefore, a reference architecture that combines big data frameworks with machine learning solutions is needed for industrial automation. The Industry 4.0 standards and associated material served as the basis for the principles for developing such an architecture used in this study. The primary contribution of this work is the suggestion of a reference architecture for big data and machine learning in industrial automation that is compatible with the Reference Architecture Model for Industry 4.0. The conceptual architecture has been developed using the Smart Factory OWL, and its performance and scalability have been tested using actual industrial data. The developed platform has shown to be adaptable, linearly scaleable, and compatible with industrial analytics requirements.

The conceptual architecture has been developed using the Smart Factory OWL, and its performance and scalability have been tested using actual industrial data. The developed platform has shown to be adaptable, linearly scaleable, and compatible with industrial analytics requirements. Future research will assess how well the created platform performs in comparison to other technologies, such as Apache Storm vs. Kafka Streams on Kubernetes. It is feasible to build integration concepts for devices and data thanks to the current design. For example, changing the data format or adding or deleting devices from different manufacturers won't affect the components in higher levels. There is still dependency between the systems at higher levels. For instance, altering the producers' and consumers' components is necessary when transitioning from Kafka to another message broker system. Therefore, in order for these components to be entirely separated, it is required to enhance the integration between them. This may be achieved by creating consistent interfaces between layers, such as REST interfaces.

8. REFERENCES

[1] Industrie 4.0, “Industrie 4.0,” [Accessed 28-February-2018]. [Online]. Available: http://www.plattform-i40.de
[2] ISO/IEC, “Information technology- internet of things reference architecture (iotra),” ISO/IEC CD 30141, 2016.
[3] Z. Ge, Z. Song, S. X. Ding, and B. Huang, “Data mining and analytics in the process industry: The role of machine learning,” IEEE Access, 2017.
[4] E. Trunzer and et al., “A flexible architecture for data mining from heterogeneous data sources in automated production systems,” 2017 IEEE International Conference on Industrial Technology (ICIT), March 2017.
[5] K. Al-Gumaei and et al., “A survey of internet of things and big data integrated solutions for industrie 4.0,” Emerging Technologies and Factory Automation (ETFA), September 2018.
[6] P. Chapman, J. Clinton, and et al., “CRISP-DM 1.0,” The CRISP-DM Consortium, August 2000.
[7] “AWS IoT Platform website,” [Accessed April-2019]. [Online]. Available: https://aws.amazon.com/iot/how-it-works/?nc1=h ls

[8] “NTT Docomo press release website,” [Accessed April-2019]. [Online]. Available: fhttps://www.nttdocomo.co.jp/english/info/media center/pr/ 2015/0708 00.htmlg
[9] Y. Yamato, H. Kumazaki, and Y. Fukumoto, “Proposal of Lambda Architecture Adoption for Real Time Predictive Maintenance,” 2016 Fourth International Symposium on Computing and Networking, 2016.
[10] J. Herwan, S. Kano, R. Oleg, H. Sawade, and N. Kasashima, “Cyber- Physical System Architecture for Machining Production Line,” 2018 IEEE Industrial Cyber-Physical Systems (ICPS), 2018.
[11] DIN SPEC 91345, “Reference architecture model industrie 4.0 (rami4.0),” DIN Std. DIN SPEC 91 345, April 2016.
[12] M. O. G¨okalp and et al., “Big data for industry 4.0: A conceptual framework,” 2016 International Conference on Computational Science and Computational Intelligence (CSCI), December 2016.
[13] J. Wan and et al., “A manufacturing big data solution for active preventive maintenance,” IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, August 2017.
[14] Big Data Working Group, “Big data taxonomy,” October 2014.
[15] J. Ellingwood, “An introduction to big data concepts and terminology,” 9 2016, [Accessed July-2018]. [Online]. Available: https://www.digitalocean.com/community/tutorials/ an-introduction-to-big-data-concepts-and-terminology
[16] Qaware.de, “Big data landscape 2018,” 2018, [Accessed July-2018]. [Online]. Available: http://www.qaware.de/fileadmin/user upload/QAware-Big-Data-Landscape-2018.pdf
[17] J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, “A survey of machine learning for big data processing,” EURASIP Journal on Advances in Signal Processing, May 2016.
[18] L. Zhou, S. Pan, J. Wang, and A. V. Vasilakos, “Machine learning on big data: Opportunities and challenges,” Neurocomputing, 2017.
[19] S. Lee, J. K. Kim, X. Zheng, Q. Ho, G. A. Gibson, and E. P. Xing, “On model parallelization and scheduling strategies for distributed machine learning,” 2014.
[20] A. N. Richter, T. M. Khoshgoftaar, S. Landset, and T. Hasanin, “A multidimensional comparison of tssssoolkits for machine learning with big data,” Aug 2015.
[21] G. D. F. Morales and A. Bifet, “Samoa: scalable advanced massive online analysis.” Journal of Machine Learning Research, 2015.
[22] Big Data Value Association, “Big data challenges in smart manufacturing,” Big Data Value Association, 1 2018.
[23] H. Kagermann and et al., “Recommendations for implementing the strategic initiative industrie 4.0: Final report of the industrie 4.0 working group,” April 2013.
[24] A. H. R. Drath, “Industrie 4.0: Hit or hype?” EEE industrial electronics magazine, vol. 8, no. 2, pp. 56–58, 2014.
[25] P. ODonovan and et al., “An industrial big data pipeline for data-driven analytics maintenance applications in largescale smart manufacturing facilities,” Journal of Big Data - a SpringerOpen Journal, 2015.
[26] M. Kiran and et al., “Lambda architecture for cost-effective batch and speed big data processing,,” 2015 IEEE International Conference on Big Data (Big Data), 2015.
[27] IEC Standards, “Industrial communication networks - fieldbus specifications - part 1: Overview and guidance for the IEC 61158 and IEC 61784 series,” 2104.
[28] DIN SPEC 62541, “Opc unified architecture (iec 62541:2015),” DIN Std. DIN SPEC 62541, 2015.

[29] Q. Li and et al., “Scalable formalization of publish/subscribe messaging scheme based on message brokers,” International Workshop on Web Services and Formal Methods, 2008.
[30] S. Windmann, H. Trsek, and O. Niggemann, “Concepts to increase the IT Security in Industrial Automation Systems,” Automation, April 2016.
[31] CITRIX, “Citrix hypervisor,” [Accessed 18-April-2018]. [Online]. Available: https://xenserver.org/
[32] R. Samuel and Y. Cao, “Nonlinear process fault detection and identification using kernel pca and kernel density estimation,” Syst. Sci. Control Eng., vol. 4, 2016.
[33] M. Mansouri, M. Nounou, H. Nounou, and N. Karim, “Kernel pca based glrt for nonlinear fault detection of chemical processes,” J. Loss Prevention Process Ind., vol. 40, 2016.
[34] S. Windmann and O. Niggemann, “A self-configurable fault detection ystem for industrial ethernet networks,” at - Automatisierungstechnik, vol. 65, 06 2017.
[35] G. E. Plassman, “A Survey of Singular Value Decomposition Methods and Performance Comparison of Some Available Serial Codes,” 2005.
[36] J. J. Downs and E. F. Vogel, “A plant-wide industrial process control problem,” Computers & chemical engineering, vol. 17, no. 3, pp. 245–255, 1993.
[37] M Narasimha Rao, SK Althaf Hussain Basha, Shaik Abdul Aziz , “A Distributed and Collaborative Electricity Consumption and Low Cost Maintenance across Multi Cloud Environment”, International Journal For Recent Development In Science And Technology(IJRDST), Volume 04, Issue 07, Jul 2020,pp. 288-293, ISSN:2581–4575.
[38] PuttaVenkata Ravi Kumar , SK Althaf Hussain Basha, Mahammad Hafeez N S , “Distributed Approach for Detecting Spammer Across Twitter Through Clustering Techniques” , International Journal For Recent Development In Science And Technology(IJRDST), Volume 04, Issue 07, Jul 2020, pp.252-257,ISSN:2581–4575.
[39] SK Althaf Hussain Basha, B Sasidhar, “A Review on the Challenges of E-Commerce Security Issues, Privacy, Trust and Solutions”, International Conference on Consumer Dynamic and Marketing Strategies in Globalized Economic Era-Perspectives and Challenges, GRIET, Hyderabad, 2013.
[40] B Sasidhar , SK Althaf Hussain Basha, , “ The Effect of E-Commerce Applications on Marketers and Consumers: A Case Study ”, International Conference on Consumer Dynamic and Marketing Strategies in Globalized Economic Era-Perspectives and Challenges, GRIET, Hyderabad,2013.
[41] Jinka Sreedhar , SK Althaf Hussain Basha, Pammi Pavan Kumar, , “Innovative Techniques and Technologies in Translation in a Multilingual Context -2012”, Third International Conference on Translation, Technology and Globalization in Multilingual Context, ITA, NewDelhi.
[42] B Sasidhar, SkAlthaf Hussain Basha, A.Govardhan, , " Data Mining Techniques using in E – Learning Domain", Published in proceeding of National Conference on "Data Mining and Data Warehousing" (DMDW2009) atMRCET,Secunderabad,121-127,2009.
[43] Prasad, GNR, “ DIGITAL TECHNOLOGY AND ITS POSITIVE IMPACT ON THE ENVIRONMENT”, International Research Journal of Modernization in Engineering Technology and Science Volume:02/Issue:08/August-2020