Journal of Information Systems Engineering and Management
2025, 10(16s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Industry 4.0 Data Processing Requirements: Use Cases for Big Data Applications.

Rajesh Lomte[1], Prof. Narender Kumar[2]
[1]*Ph.D. Scholar, Department of Computer Science and Engineering, NIILM University, Kaithal, Haryana, India*
rajulomte1@gmail.com
[2]*Professor, Department of Computer Science and Engineering, NIILM University, Kaithal, Haryana, India*
drnarem.2016@gmail.com

ARTICLE INFO

Received: 08 Dec 2024
Revised: 29 Jan 2025
Accepted: 10 Feb 2025

ABSTRACT

By combining big data analytics, IoT, artificial intelligence (AI), and cyber-physical systems, Industry 4.0 has completely changed manufacturing and industrial processes. Efficient data processing has emerged as a crucial challenge as industrial systems produce enormous volumes of data in real-time. With an emphasis on real-time analytics, data storage architectures, scalability, security, and interoperability, this study examines the data processing needs crucial for Industry 4.0 applications. We highlight important use cases, such as supply chain optimization, quality control, predictive maintenance, and smart manufacturing, where big data applications spur innovation. Additionally, we examine how distributed processing frameworks, cloud computing, and edge computing are used to handle industrial data. In order to guarantee effective, scalable, and safe data management, the study also identifies potential solutions to problems like data integration, latency, and cybersecurity threats. This article offers useful insights for industries hoping to fully utilize big data in Industry 4.0 by addressing these aspects.

Keywords: Industry 4.0, Big Data, Data Processing, IoT, Smart Manufacturing, Predictive Maintenance, Edge Computing, Cybersecurity.

Introduction

Significant advancements in the practical aspects of manufacturing and production have been made possible by the fourth industrial revolution. The invention of steam engines marked the beginning of the industrial revolution. Computing technology have created enormous volumes of data in our day and age.

With the advent of Industry 4.0, numerous sensors like vision and vibration have been introduced, which can be used to analyze machine health and other related data such as online shopping patterns. Industry 4.0 requirements require companies to be transformed, as evidenced by the data that is evident. This is crucial for companies. Four dimensions are employed to depict Big Data. Volume, velocity, capacity and variety are represented by Vs.

Let's understand these four pillars of Big Data in more detail:

1. Volume: The amount of data generated. In the last ten years, the internet has generated more than half the data since its onset.
2. Velocity: The rate at which data is Generated. The amount of Data that is generated per second (Throughput).
3. Variety: The different types of data that are generated such as (videos, log files, etc.) and how accurate/interpretable they are.
4. Veracity: The different types of data (quality of data) from different sources (structured/semi-structured).

Industry 4.0 can use Bigdata for Predictive maintenance, Production planning, and Material Requirements Planning.

Big Data to the Rescue

Organizations are using Big Data to enter Industry 4.0 because of its affordability and the ability to integrate data with commodity hardware. Despite being established, Hadoop was not yet finished until Yahoo Inc. The availability of analytics, scalability, and accuracy have been made possible by Big Data. Both horizontal and vertical scalability are possible in Big Data's distributed architecture.

You can extract the generated data to predict a pattern that is not possible to detect before. The availability of big data helped to soften the attack of the Chackes network. The organization often effectively smoothed the attack and analyzed the files in the network magazine.

Information aggregation can develop a panel of panel that can provide valuable information. Big data can handle terabytes/petabytes. The organization uses big data to improve supply chain management and provide large networks. Big data helps to make important decisions. Because the understanding that it creates can help the entire organization to improve the process. Computer equipment has been greatly developed in the design over the past decade, which has been more insightful for large data analysis.

Big Data Architecture

Data ingestion platforms, which prepare structured and semi-structured data for analysis, as well as Mapper/Reducer architecture, are part of the larger picture. Mapper/reducer architecture is an aspect of big data analysis. The extracted data is divided into several key/value pairs in a smaller file format, which are then sorted and combined with the reducer phase. In the distributed/pseudo-Distributed mode, the operation is carried out on a single node.

Data can be obtained from websites, sensors, or other sources for Bigdata analysis. Inpache Kafka utilizes real-time big data streaming to facilitate the ingestion of data by T. Pipelines that stream big data can process data in real time for analysis purposes.. A file system for big data is a Hadoop distributed File System (HDFS).

Use Cases of Industry 4.0

It was Industry 4.0 that coined the phrase "smart factories." Big data can be used for equipment monitoring. Data can be used to forecast whether the equipment will be available soon. Additionally, you can use big data (employee fatigue analysis) and the vision sensor. Future sales can be predicted using the production plan and a sizable data set (the last ten years' worth of data). You can make a production graph using these data to help speed up production and lower stock losses. You can identify the data template (seasonality, trend) that will support your decision-making.

The big data frame can be used to analyze website user reviews. An essential tool for recommending products to customers is mood analysis. Customer reviews, sales, and other features can also be used to create effective advisory engines. This kind of analysis employs a flow of vast amounts of data. The best product for the product can also be suggested by the price engine. In order to give customers, the best deal, competitors' websites may be shut down. The effective price of airline tickets is a crucial choice to make because the cost varies depending on the season and for each supplier. Accurate pricing can optimize the price in addition to customer satisfaction.

The telecom sector used big data analysis to forecast customer leaks. Data from client visits, surveys, and payment methods are all examples of client data that can be used. Industry profits, production, and margins are all enhanced by big data [5]. Large-scale data It is crucial to many engineering domains, applications, and artificial intelligence. On the basis of data on data. By 2022, Grand View Research, Inc., USA, will have $72.8 billion. When using a large data set to evaluate a diagnosis, health care is crucial.

Therefore, big data plays a high role in the medical information system [2]. Some factors affect the integrity of the version with the origin and uncertainty of the data. Dirty and noisy data [7]. This algorithm is machine learning, deep education and Natural language (NLG) generation Medical, Smart City, Industrial 4.0, etc.

The template is designed through deep training. Enhancement of computation and optimization [8]. Data science for cost optimization, work and product settings, etc. As in the 1970s, the big data evaluation's hierarchical structure was computed in 1980. In [9], mass data sets, 1990s data analysis, statistical training, business analysis in the 2000s, and big data analysis in 2010 were examined. However, businesses are prepared for the large-scale data analysis technology and are willing to acknowledge certain issues [10].

Activity and manual data for diagnosis and prediction in the medical system are part of the monitoring and management concept of the system (SHMM) [9]. [11] discusses cloud computing and the Internet of Things for healthcare. In the medical field, large data applications offer numerous benefits. Medical services and sectors are supported by big data technology. For advanced patient care, increased workload, disease detection, cost savings, precise treatment, and improved medical diagnosis, this calls for the quality, surgery, and performance of data in medical systems.

Table 1 Importance of V's in Big data

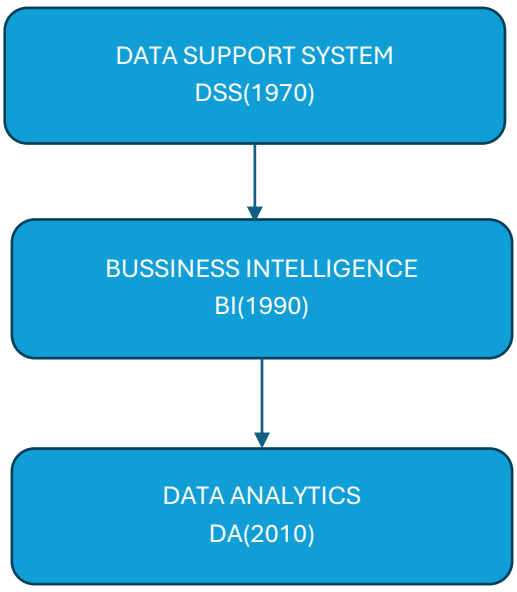
Volume	Variety	Velocity	More v's
Megabytes	Structured Data : Tables	Batch	Visualization
Gigabytes	Semi-Structured Data : Records	Real-Time	Value
Terabytes	Un-Structured Data: Records	Stream	Veracity
	Partial data : Transactions	Near Real	Virality
	Spatial data: Transactions	Time and so on	Viscosity and so on
	Sensor Data: Transactions		
	Image and so on : Files, etc.-Records		

Big data in business organization

Business analysts and big data are interchangeable terms. (Cassandra, Spark, Rapid Miner, SAS, Knime, Orange, Weka, Hadoop, HPC, STOOBLE, QUOBLE, etc. Modified Business analysts use the reporting and decision-making support system to forecast the benefits of operation. When a company has a lot of data, cloud analysis and big data tools like Hadoop, Cassandra, and others have helped it. The Fuzzy Hybrid Core Means algorithm (HKFCM) uses clustering to preserve volume data.

The configuration, encryption, data storage, data reconstruction, and access steps are all completed by this method. It is simple to find new data for business profits with these tools and analysts in memory. The industry's assistance forecasts the outcomes using trained data or education. Big business data It is possible to undervalue the organization, which has developed into an analysis in the decision support system. Big data is crucial because it compares technology, goods, services, clients, vendors, and feedback from coworkers. In addition to suggesting the access and storage of large amounts of diverse data, the term "big data" also suggests a process by which industries use and commercialize data in a variety of ways [28].

Large-scale data processing techniques work well in corporate settings and can handle massive volumes of data. However, it includes the data's duality, noise data, and values that are not between data. It comes in a variety of formats and was supplied in response to the request that was received. client. As a result, the industry benefits from new products, reduced time, and cost optimization. Recognizing market conditions based on consumer demand, development process, and mood analysis for online business control.



Big data in industrial processing

Large volumes of data produced by a variety of sources and methods are referred to as big data. Including the significance of ,industrial data storage involves keeping data outside the purview of a variety of data databases. Analysis A decision-making support system, prediction modeling, future activities, and business operations of customers, industries, or corporate systems are just a few of the many benefits that big data provides. The end user is promoted by big data analysts and service analysts who move by approach based on the end user interface based on analytical mechanisms, according to Analytic Analytics Analytics (EIS).

One data analysis technique used in the aforementioned industry and other sectors is called "big data." Large-scale industrial data storage (BID) is supported by volume cloud computing in 4. -ITS, which also processes it for computation. Benefits from the government include communication, manufacturing, retail trade, agriculture, sports, fitness, consumer goods, insurance, construction, transportation, banks, financial services, and tourism. Through the decision-making system, the production industry must also manage the resources of long-term profitability identification . Large-scale data analysis is necessary in the production sector to maximize the advantages of various operations and optimize resources.

Novel Industrial Methods 4.0 uses a variety of data creation sources for processing in industry. Large data in Raznoye is used by other industries, including small and medium-sized, medium-scale, large-scale, and very-large. 87% are used by the telecommunications sector, 76% by the financial services sector, and 60% by the medical services sector . By combining production resources, processes, and products, big data analysis (BDA) is a useful tool for sustainable intellectual production that boosts profitability. Numerous industries, including on-demand services, resource combination, rapid elasticity, and economic efficiency, can benefit from big data.

The organizational deployment of large-scale data technology affects a number of factors, including human resources, technology resources, management support, acceptance, safety, confidentiality, complexity, and normative behavior. The corporate technology or environmental circumstances determine whether or not large data is adopted . Big data can also be used for fraud detection, new product and service recommendations, cost reduction, and quicker and better decision-making. Robotics and automation systems are developed using big data, which transforms business.

Large volumes of data are handled efficiently by machine learning in the majority of IoT industries for precise forecasting and decision-making . Technology tools for intellectual production include the Internet of Things (IoT) and industrial Internet, cloud computing (CBT), cyber physics systems (CPS), data production (DM), and artificial intelligence (AI) . IoT has the potential to improve processing and operation efficiency by 30% . The author of [5talked about improvement factors, including the capacity of different industries, including the ability to manage and analyze big data, finance, education, production, wholesale, and retail trade.

Challenges and opportunities in big data

Though there aren't many big data, many authors pose a range of issues. Unbalanced, big data availability, big data management, big data aggregation, big data analysis, and big data purification . The management of excellent data, excellent data and system functions, and machine training all discuss this issue. Large data industrial analysis faces a variety of challenges that require the use of creative production systems. It is made up of long-term investments in IT infrastructure, rules and industries, protocol systems, low vision and dedication, and risk associations with new technologies . It is also challenging to automate industry automation.

Future medical systems will require a variety of tasks, such as data type standardization, integrated data schema, data quality parameters, and data, data, data. The tasks include data representation, data compression, data reduction, data life cycle management, data confidentiality, aggregate scale, analysis tools, etc. MREDHANISM, potential for growth and consumption, and collaboration among multidisciplinary systems [18]. The intricacy of data, computational complexity, and system complexity are also challenging.

Big data includes intellectual data analysis, data storage, data exchange, data confidentiality, data technology, and talent, but there are also challenges in locating crucial data. Big data applications, information barriers, and limited data standardization make data integration inadequately supported by data exchange. The author outlined some of the issues with big data as a crucial component in : The data of the data, the data of the data, the data of the semantic generation, and the data of the data in the absence of data.

Using machine learning, the Matoury approach was chosen to guarantee the best industrial and health outcomes. This was observed. During experimental analysis, this meta substitution is used for object selection, clustering, classification, and related rules. Filtering big data For many applications, data filtration offers the best results at many stages. the outcome of the application system's commercial success. Big data is present. The industrial database system is a significant approach.

The Matoury approach was selected to ensure the best industrial and health outcomes through machine learning. This was noted. This meta substitution is applied to object selection, clustering, classification, and associated rules in experimental analysis. Big data filtering Data filtration provides the best outcomes at many stages for a variety of applications. the result of the commercial success of the application system . There is big data. One important strategy is the industrial database system .

However, prior to handling the application, other system errors should be permitted and analytically assessed in every system. Error measurement, crash data (such as redundant data, strip values, error code, contradiction errors, etc.), discharge, and members are all included in data filtration . Tasks are a means of meeting and guaranteeing the complexity of the large data's features (speed, diversity, and volume).

This assesses how reliability analysis is used to ascertain the complexity and relationship between the features of big data . At the same time, it's critical to use a limited set of software and hardware support requirements to manage data collection, integration, and storage . Big data is a contemporary and significant field that centres on the author's numerous objectives, findings, and potential directions for future research. The author's contributions to big data and multi-technology industry are detailed in this document. This indicates a lot of unresolved issues, opportunities, and problems that will need to be resolved in the future. Numerous topics pertaining to analysis, processing, and distribution in diverse industries were covered in the article.

Applications in contemporary businesses and the medical field were the focus of this article. Focus is placed on machine learning, in-depth learning, classification, PCA, sampling, AHP, and preliminary data processing. By working on industrial data, medical data, and other applications, this approach has been refined for optimal service. However, there is room for improvement in terms of the current and examined parameters, such as accuracy, sensitivity, specificity, and so on. For more effective results, you can include additional crucial measurement parameters in the suggested tasks, such as voice prediction values (NPV), FLD (FLD), false negative speed (FPR), and O detection speed (FPR). The execution system will become more complex as a result of these outcomes. Time becomes more complex as accuracy, sensitivity, and specificity increase.

Conclusion

Production-level BDA (Big Data Analytics) Prior to putting BDA into practice, it was also known as pre-diagnosis (PD) and quality and process management system (QPC) improvement. Future enhancements will focus on big data values like performance indicators (PI), transparency, and decision support systems (DSS). In sustainable intellectual production, where data and services are integrated, data sources must be successfully integrated to guarantee business outcomes. A data purification model to identify and alter anomalous data is a crucial requirement of the outcome. A significant problem in the realm of big data-based industrial and medical applications is the identification, removal, and release of dirty data.

This review demonstrates that, depending on all applications, data processing, discharge, and purification techniques are crucial in large data sets. Another significant outcome and completion of the majority of studies that comprise an article state is the development of an optimized model for detecting and removing emissions in large-scale applications.

Big data filtering, big data processing, and big data analysis for organizational technology improvement are the three main areas of this article that require more research. Large data sets can cause issues such as discharge, missing and deleting, removal, and replacement when compared to data cleansing.

References

[1] Geng D, Zhang C, Xia C, Xia X, Liu Q, Fu X. Big data-based improved data acquisition and storage system for designing industrial data platform. IEEE Access. 2019;7:44574–82. https://doi.org/10.1109/ACCESS.2019.2909060

[2] "Technology _ Grand View Research_ Big Data Market Research. Report 2015 to 2022 by Grand View Research, Inc.".
Heureux AL, Member GS. Machine learning with Big Data : Challenges and Approaches. IEEE Access. 2017;5:7776– 97. https://doi.org/10.1109/ACCESS.2017.2696365
[4] 4. Hussain S, et al. Implications of deep learning for the automation of design patterns organization. J Parallel Distrib Comput. 2018;117:256–66. https://doi.org/10.1016/j.jpdc.2017.06.022
[5] Tsui KL, Zhao Y, Wang D. Big data opportunities: System health monitoring and management. IEEE Access. 2019;7:68853– 67. https://doi.org/10.1109/ACCESS.2019.2917891
[6] Ghasemaghaei M. Are firms ready to use big data analytics to create value? The role of structural and psychological readiness. Enterp Inf Syst. 2019;13(5):650–74. https://doi.org/10.1080/17517575.2019.1576228
[7] Dang LM, Piran J, Han D, Min K, Moon H. "A Survey on Internet of Things and Cloud Computing for Healthcare," pp. 1–49, 2019, https://doi.org/10.3390/electronics8070768
[8] Rathee G, Sharma A, Saini H, Kumar R, Iqbal R. A hybrid framework for multimedia data processing in IoT-healthcare using blockchain technology. Multimed Tools Appl. 2020;79:15–6. https://doi.org/10.1007/s11042-019-07835-3
[9] Mathivanan Durai, R. B. Dravidapriyaa, S.P. Prakash, Wanjale, K. H., M. Kamarunisha, & M. Karthiga. (2025). Student Interest Performance Prediction Based On Improved Decision Support Vector Regression Using Machine Learning. International Journal of Computational and Experimental Science and Engineering, 11(1). https://doi.org/10.22399/ijcesen.999
[10] Kurumbalapitiya D. *Data acquisition*. 2005.
[11] Ma Y, et al. Remote sensing big data computing: Challenges and opportunities. Futur Gener Comput Syst. 2015;51:47–60. https://doi.org/10.1016/j.future.2014.10.029
[12] Agrawal D, Das S, Abbadi AE. Big data and cloud computing: current state and future opportunities. 14th Int Conf Extending Database Technol. 2011;530–3. https://doi.org/10.1145/1951365.1951432
[13] Fan J, Han F, Liu H. "Challenges of Big Data analysis," *Natl. Sci. Rev*, vol. 1, no. 2, pp. 293–314, 2014, https://doi.org/10.1093/ nsr/nwt032
[14] Chen M, Mao S, Liu Y. Big data: a survey. Mob Networks Appl. 2014;19(2):171–209. https://doi.org/10.1007/ s11036-013-0489-0
[15] Sukumar SR, Natarajan R, Ferrell RK. Quality of Big Data in health care. Int J Health Care Qual Assur. 2015;28(6):621–34. https://doi.org/10.1108/IJHCQA-07-2014-0080
[16] Rabhi L, Falih N, Afraites A, Bouikhalene B. "Big Data Approach and its applications in Various Fields: Review," *Procedia Comput. Sci*, vol. 155, no. 2018, pp. 599–605, 2019, https://doi.org/10.1016/j.procs.2019.08.084
[17] Rahul K, Banyal RK, Goswami P. "Analysis and processing aspects of data in big data applications," vol. 0529, no. May, 2020, https://doi.org/10.1080/09720529.2020.1721869
[18] K. Wanjale, A. A. Deshmukh, J. C. Vanikar, A. P. Adsul and S. P. Bendale, "Detecting Human Eye Blinks through OpenCV," 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), Pune, India, 2024, pp. 1-5, doi: 10.1109/I2CT61223.2024.10543678
[19] Steckel T, et al. Big Data Analysis of Manufacturing processes. J Phys Conf Ser. 2015;659:012055. https://doi. org/10.1088/1742-6596/659/1/012055
[20] Krishnan R, Samaranayake VA, Jagannathan S. A hierarchical Dimension Reduction Approach for Big Data with application to Fault Diagnostics. Big Data Res. 2019;18:100121. https://doi.org/10.1016/j.bdr.2019.100121
[21] O'Donovan P, Leahy K, Bruton K, O'Sullivan DTJ. An industrial big data pipeline for data-driven analytics maintenance applications in large-scale smart manufacturing facilities. J Big Data. 2015;2(1):1–26. https://doi.org/10.1186/s40537-015-0034-z
[22] Shah D, Wang J, He QP. "Feature engineering in big data analytics for IoT-enabled smart manufacturing – comparison between deep learning and statistical learning," vol. 141, 2020, https://doi.org/10.1016/j.compchemeng.2020.106970
[23] García S, Ramírez-gallego S, Luengo J, Benítez JM, Herrera F. "Big data preprocessing : methods and prospects," pp. 1–22, 2016, https://doi.org/10.1186/s41044-016-0014-0

[24] Bonde M, Bossen C, Danholt P. Data-work and friction: investigating the practices of repurposing healthcare data. Health Inf J. 2019;25(3):558–66. https://doi.org/10.1177/1460458219856462
[25] Bossen C, Pine KH, Cabitza F, Ellingsen G, Piras EM. Data work in healthcare: an introduction. Health Inf J. 2019;25(3):465– 74. https://doi.org/10.1177/1460458219864730
[26] Kaur P, Sharma M, Mittal M. ScienceDirect Big Data and Machine Learning based Secure Healthcare Framework. Procedia Computer Sci. 2018;132:1049–59. https://doi.org/10.1016/j.procs.2018.05.020