International Core Journal of Engineering Volume 6 Issue 6, 2020
ISSN: 2414-1895 DOI: 10.6919/ICJE.202006_6(6).0023

Architecture Design and Research of Smart Factory Data Platform Based on Big Data

Ying Lan[1,a],Weijie Zhou[1,b]

[1]School of Shanghai Maritime University, Shanghai 200000, China.

[a]1072298135@qq.com

Abstract

Based on the analysis of the research status of smart factories at home and abroad, the research on the architecture technology of smart factory data platform based on big data is carried out to provide technical reference for the operation analysis, prediction, decision-making regulation and digital twin information physical fusion of intelligent production. The definition and connotation of smart factory, as well as the source and characteristics of smart factory big data are discussed. The popular open source big data calculation engine of Hadoop, Spark, Storm is used to propose the data source layer, data transmission layer, data storage layer, resource management layer, and processing analysis The technical framework of the smart factory big data platform composed of the layer and the business application layer can effectively solve the requirements and difficulties of the multi-source complexity and real-time nature of smart factory big data. The proposed data platform technology architecture will have important reference value for the realization of smart manufacturing and smart factories.

Keywords

Big Data; Smart Factory; Smart Manufacturing; Digital Twins.

1. Introduction

Today, with the rapid development and maturity of emerging technologies such as the Internet of Things technology, cloud computing technology, and big data technology, the manufacturing industry will face a transition from traditional manufacturing to new production modes of informatization, automation, and intelligence. In recent years, it has also attracted great attention from all over the world.

The developed countries in Europe and America led by the United States and Germany have formulated a series of manufacturing plans under the new situation. The United States has successively launched a series of industrial strategic planning and industrial Internet concepts since 2009. Represented by the United States General Electric (GE) company, and in 2012 released the "Industrial Internet: Breaking Wisdom and Machines" industrial policy report; 2013 Hanover At the Industrial Fair, the German government proposed the concept of "Industry 4.0" and introduced it into its production control system and industrial software.

In response to the above challenges, the Chinese government proposed the "Made in China 2025 Plan" in 2015 to promote the in-depth integration and innovation of new-generation information technologies such as cloud computing and big data and modern manufacturing. In terms of academic research, Ji Xu et al. Proposed cloud computing and big data cloud manufacturing architecture and key technologies for the high-peak materials industry; Zhang Jie partially eliminated the big data-driven "association + prediction + regulation" smart workshop analysis and decision-making new model; Wang Jianmin Analyze the source of industrial big data and summarize the key technologies

ISSN: 2414-1895 DOI: 10.6919/ICJE.202006_6(6).0023

of big data management analysis; Tao Fei and others proposed a digital twin five-dimensional model based on big data transmission and connection and its 10 major application fields.

Based on the above background, this paper conducts research on the problem of the smart factory data platform, in-depth discussion of the smart factory data platform technology framework based on big data, in order to mine valuable information from massive data to know the operation and optimization of the smart factory.

2. Smart factory

2.1 Definition of smart factory

Compared with the traditional manufacturing industry, intelligent manufacturing is its transformation and upgrading, as well as intelligent activities such as analysis, reasoning, judgment and decision-making in the manufacturing process, with human as the core position while achieving human-machine integration; realizing intelligent manufacturing The key is the establishment and implementation of smart factories. The definition of a smart factory was first proposed in Germany ’s “Industry 4.0”, and uses emerging technologies such as the Internet of Things, cloud computing, and big data to connect and gradually integrate machines, devices, personnel, and software programs through sensors and networks. In order to efficiently monitor, collect, process and analyze data, realize intelligent processing, information management and services, and build efficient, energy-saving and green humanized factories.

2.2 Implementation of Smart Factory

The first point of realizing a smart factory is to use the Internet of Things to connect "dark data", that is, unused data, and convert it into a useful information system, so as to quickly respond to consumer demand changes and market mutations, and achieve agile demand Driven manufacturing model.

The first step in the implementation of a smart factory is to connect the workshop systems; because before the emergence of the Internet of Things technology, many existing enterprise equipment has been deployed and there is no built-in connection function, there is currently no universal communication standard to support traditional equipment and objects The Internet is interoperable, but on the open platform of the Internet of Things, an ecosystem solution that connects traditional devices and intermediates has been successfully listed, and organizations such as OPAF and IIC and Intel ’s giant manufacturers are stepping up the development of Internet of Things standards.

In order to apply advanced analysis to explore the great value of data, the second part of the smart factory implementation is to collect, store, preprocess, and analyze data; most existing equipment can generate massive data, but due to the excessive number, these data cannot be used in data centers. Quick analysis; In order to extract the most important information, in terminal analysis or fog computing strategies, algorithms are deployed at streaming data collection points, gateways, clouds, or anywhere in between to implement efficient performance calculations, thereby assessing the required data and not Required data.

3. Big data

3.1 Big data concept

In the "Big Data Era" written by Schonberg, big data refers to not using shortcuts such as random analysis (sampling survey), but using all data for analysis and processing; it has a large number, high speed, diversification, Multiple attributes such as value.

Big data requires different technologies at different levels; the data collection layer needs to use ETL tools to extract the data from the distributed and heterogeneous data sources to the temporary intermediate layer, complete the data preprocessing work such as cleaning and conversion, and load it into the data warehouse or market; The data storage layer needs to use distributed file systems, data warehouses, relational / non-relational databases, etc. to achieve storage and management of various data; the data processing and analysis layer needs to combine machine learning and mining algorithms,

ISSN: 2414-1895 DOI: 10.6919/ICJE.202006_6(6).0023

and use the computing framework engine to achieve large-scale Data analysis and processing, and visualize the analysis results. Its two core key technologies include distributed storage and distributed processing, so as to effectively process large-scale data within a tolerable time.

3.2 Big data sources for smart factories

The big data sources of smart factories mainly include: enterprise informatization data, industrial Internet of Things data, and cross-border data, as shown in Figure 1.

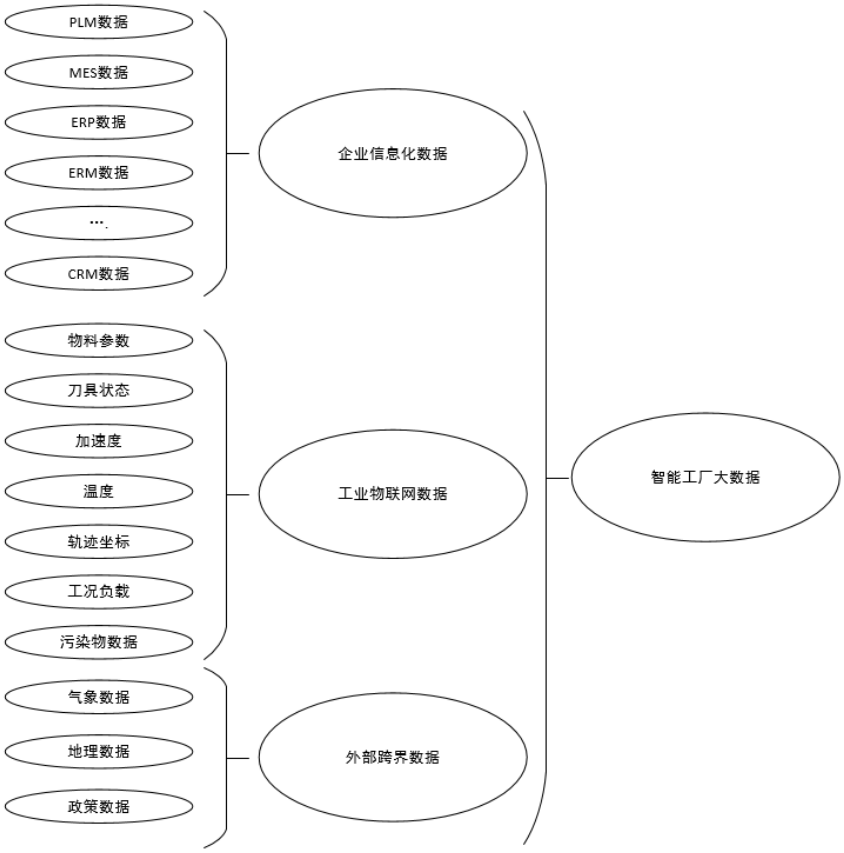

Figure 1. Intelligent factory big data source

As a traditional industrial data asset database, enterprise informatization data mainly comes from PLM (product full life cycle management system), MES (manufacturing execution system), ERM (enterprise resource management system), PQM (product quality management system), CRM (customer Relation management system), etc., this part of the data from raw materials storage, processing, inspection to factory circulation, throughout the product life cycle and the entire value chain, often has a high value density.

With the widespread use of intelligent equipment and perception networks such as CNC machine tools, industrial robots, sensors, RFID, etc., the Industrial Internet of Things data includes not only real-time production data such as material parameters, tool status, and working load, but also operating environments such as pollutants and harmful gases. Real-time data to realize the unified management of all production data and process data.

With the continuous integration of the Internet and industrial manufacturing, in addition to the internal big data from the manufacturing industry, it also includes external "cross-border" data that affects the production and operation of enterprises, such as meteorological data, geographic information data, policy data, economic data, Expose regulatory data, etc.

3.3 Big data features of smart factories

Smart factory data is diversified, diversified, and heterogeneous, and the processing scenes are complex and changeable, so its data presents typical big data 3V characteristics: mass, diversity, and high speed; the specific manifestations are as follows: (1) mass, Taking production in the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202006_6(6).0023

metallurgical industry as an example, it is a complex multi-stage process production process that combines multiple processes, gaps and continuous operations. It has the characteristics of multi-stage production, multi-stage transportation, multi-stage storage, ground and space cross transportation, and its logistics measurement takes you , Logistics information, such as operation points, is a complex mixture of multi-dimensional processes. (2) Diversity, taking the metallurgical industry as an example, that is, there are equipment information and operating parameters of equipment such as slab continuous casting, round billet continuous casting machine, desulfurization device, etc. The carbon content of steel, temperature, humidity and other data also include Programmable logic controller (PLC) control programs and other semi-structured data (3) High speed, PLCs, sensors and other equipment in smart factories continuously sample the production process in a very short time window, and the resulting data flow is in accordance with time Sequences flood into the database like a tide.

4. Technical architecture of smart factory data platform

Big data processing includes complex batch data processing, interactive query and mining based on historical data, real-time data stream processing, post-graph data processing, and the actual application of smart factories simultaneously exist in the above several scenarios, so the current popular Hadoop, Spark

The Storm open source computing engine is also the three most important distributed frameworks of the Apache Software Foundation.

Based on the above, a smart factory data platform architecture as shown in Figure 2 (Ambari and Zookeeper management coverage) is formed. From bottom to top, it is divided into data source layer, data transmission layer, data storage layer, resource management layer, processing analysis layer and business application layer. Data flows from bottom to top; the following is a detailed introduction of each layer.

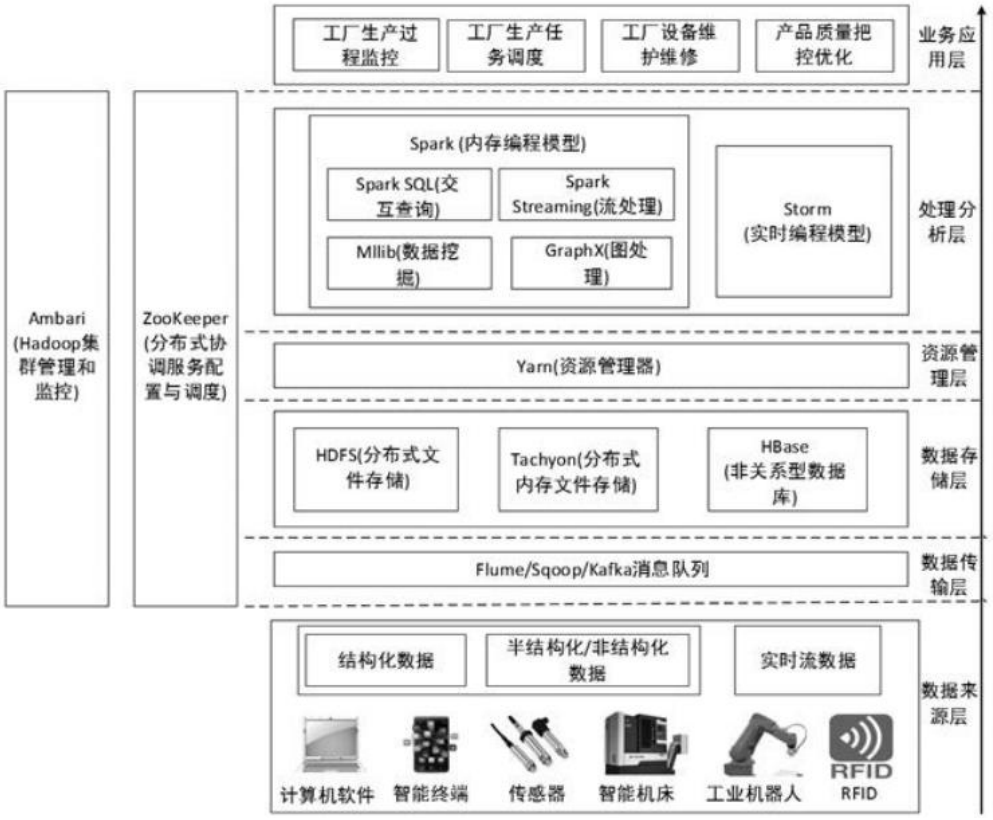

Figure 2. Technical architecture of smart factory data platform

ISSN: 2414-1895 DOI: 10.6919/ICJE.202006_6(6).0023

4.1 Data source layer

The data source layer is the data provider, which is mainly responsible for real-time and reliable production collection and obtaining how far heterogeneous data is within the scope of the smart factory; because the industrial production line is running at high speed, industrial equipment generates data types of multi-source heterogeneous data; because the industrial production line is at high speed In operation, the types of data generated by industrial equipment are mostly non-institutional data, and the real-time requirements for data are also higher. At this level, the factory's physical manufacturing resources, including intelligent machine tools, industrial robots, computer software, and intelligent terminals, collect data by installing and configuring industrial sensors, RFIS tag QR codes, and barcodes on manufacturing resources, and through wired Internet, wireless Basic network facilities such as the Internet, GSM network, infrared, and Bluetooth connect these manufacturing resources, and perform data transmission and exchange in accordance with the Internet of Things protocol.

4.2 Data transmission layer

Data transmission is located between the data source layer and the data storage layer. It is the first step for the data from various data sources in the smart factory to enter the big data system. It is responsible for importing data from external data sources into the persistence layer of HDFS and HBase; commonly used Technologies include: Flume, Sqoop, Kafka, etc.

Flume and Sqoop are mainly responsible for the collection and transmission of various static data, allowing users to extract data from relational data into Hadoop. Once the final analysis results are generated, Sqoop can transfer these results to the meeting data storage. For example, the CNC machine tool workshop management system is responsible for planning and allocating resources at one stage of the manufacturing unit, usually one to several hours, so as to process mass production; its information data includes the location of the manufacturing unit, the routing between them, and the tools required for mass manufacturing. , Bill of materials parts library and unit operation status, etc.

Since manufacturing real-time data streaming requires a very high-availability input pipeline, limited use of Kafka is responsible for the transmission and import of real-time streaming data. For example, the sensors built in the intelligent machine tool monitor the transmission of real-time data such as the tool's speed, acceleration, trajectory coordinates, temperature, etc .; the network monitoring of the intelligent workshop real-time data, rapid and continuous arrival, must be collected and calculated in real time, and the response time is seconds Level or even subtle level.

4.3 Data storage layer

The data storage layer is responsible for storing massive uninstitutionalized and semi-structured loose data, realizing random and real-time read and write access to large data. The Tai layer includes Hadoop's core component distributed file storage system HDFS, and the Spark ecosystem's distributed memory file storage Tachyon, Hadoop's real-time query framework HBase.

As the foundation of the HDFS ecosystem, HDFS is suitable for running on cheap computer clusters, storing large files in the form of stream data written once and read many times.

Tachyon is a high-performance, high-fault-tolerance, memory-based open source distributed storage system that can be used under the cluster framework Spark and above the file storage system HDFS.

HBase is a highly reliable, high-performance, column-oriented NoSQL distributed database; it uses HDFS as its file storage system and Zookeeper as a collaborative service. Due to the complexity of the smart factory data and the diversity of sources, in the design of database tables, the data generated by different data sources is stored in different data tables; at the same time, in order to improve data access performance and achieve millisecond response, according to the business Requirements and storage requirements design each table row key.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202006_6(6).0023

4.4 Resource management

In order to realize the unified deployment and operation of the computer framework, they are all deployed on the Hadoop 2.0 resource management framework YARN; YARN, as a general resource management system, provides unified management and scheduling of resources for the data processing layer; Need to be elastically scalable, do not mix and match applications, and share the underlying storage to avoid data migration across the cluster.

4.5 Processing the analysis layer

The processing and analysis layer is responsible for the processing of big data, and uses various computing frameworks to write code models to realize the pretreatment of smart factory large protective gear, data association analysis, data mining, etc. to reveal and realize the timing evolution laws of manufacturing data, performance prediction of manufacturing equipment, and tasks Decision control, etc .; distributed parallel computing frameworks include memory computing framework Spark and stream computing framework Storm.

Spark is responsible for the offline batch processing of enterprise information data in the smart factory, and the task time span is generally monthly / day / hour level; Spark SQ is responsible for the interaction between the enterprise information data in the smart factory and the historical data in the Internet "cross-border" data Query, the component GraphX is responsible for the processing of structural data; MLib is responsible for data mining and provides various algorithm models, such as algorithms such as logistic regression and Bayesian in classification, K-Means and fuzzy K-means in clustering, and machine learning Neural networks, least squares, deep learning and other algorithms, the task time span is generally minutes / second level; Spark Streaming library is responsible for the processing and analysis of second-level data.

Storm is a free, open-source distributed real-time computing system that can process data stream data simply, efficiently, and reliably; due to its millisecond real-time response speed, Storm is mainly responsible for parallel computing of industrial IoT data in smart factories, such as machine tools Real-time analysis of tool running status, real-time alarm of workshop management system and other tasks.

4.6 Business Application Layer

Big data has been applied to all walks of life in human society. Having big data is not the purpose. It is the key to using advanced analysis to discover the huge value of data. Industrial big data with enterprise informatization data, industrial physical Internet of Things data and cross-border data integration contains many industrial production laws, just like a "gold mine" contains infinite wealth value; industrial big data monitors and produces industrial production processes It is widely used in task scheduling, fault diagnosis and prediction of industrial production equipment, network collaborative manufacturing, etc.

5. Conclusion

Based on the analysis of the big data sources, characteristics, and complex processing scenarios of smart factories, this paper designs and studies the technical architecture of the smart factory big data platform based on the current three distributed open source computing frameworks of Hadoop, Spark, and Storm, and proposes data Source layer, data transmission layer, data storage layer, resource management layer, processing analysis layer and business application layer. Follow-up will be based on improving enterprise productivity.

References

[1] Ji Xu, Zhong Ganji, Yu Yang, Li Zhongming. Key technologies and applications of cloud manufacturing in polymer materials industry [J]. Computer Integrated Manufacturing System, 2015, 21 (11): 3072-3078.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202006_6(6).0023

[2] Zhang Jie, Gao Liang, Qin Wei, Lv Youlong, Li Xinyu. Big data-driven intelligent workshop operation analysis and decision-making method system [J]. Computer Integrated Manufacturing System, 2016, 22 (05): 1220-1228.

[3] Wang Jianmin. Industrial big data technology [J]. Telecommunication network technology, 2016, 8 (8): 1-5.

[4] Tao Fei, Liu Weiran, Liu Jianhua, et al. Digital twin and its application exploration [J]. Computer Integrated Manufacturing System, 2018, 24 (01): 1-18.

[5] Guo An, Yu Dong, Hu Yi. Application Research of Information Physics Fusion Technology in Machine Tool Fault Diagnosis System [J]. Small Microcomputer System, 2017, 38 (04): 896-900.