

The Industrial Internet Platform: Trends and Challenges

Wang Chen^{1,2}, Song Liang², Li Shaokun³

1. National Engineering Laboratory for Big Data System Software, Tsinghua University, Beijing 100084, China

2. Innovation Center for Industrial Big Data of Sichuan, Chengdu 610041, China

3. Sichuan Provincial Economy and Information Commission, Chengdu 610041, China

Abstract: The fusion of manufacturing industries, next-generation Internet, and information technologies is catalyzing the development of the industrial Internet (II). The importance of developing the industry Internet platform (IIP) is recognized not only by leading international manufacturers, but also by people involved in framing China's manufacturing strategies. In this paper, the development trends of IIP are explained, and the challenges in building the client ecosystem, developer ecosystem, and data ecosystem of IIP are analyzed. Furthermore, the technical challenges of IIP in modeling and analyzing industry big data are discussed in depth.

Keywords: industry Internet platform; industrial big data; data analytics

1 Status of industrial development at home and abroad

In the past few decades, the influence of the Internet has remained mainly in the consumer sector. China's "Internet plus" strategy has made great progress in this sector, where electronic commerce and other technologies and applications have seen their places at the forefront worldwide. Beijing Baidu Netcom Science and Technology Co., Ltd.; Alibaba (China) Network Technology Co., Ltd.; and Tencent Computer System Co., Ltd. have been listed among the world's top 10 Internet companies. For some time, owing to the great influence of the consumer-oriented Internet in China, the impact of the Internet on industry did not receive as much attention as it deserved in China. After the publication of Germany's Industry 4.0 strategy, Chinese people began to pay close attention to the urgency and value of integrating informatization and industrialization. This expansion of the Internet into the manufacturing industry is called the industrial Internet (II).

The II is the only viable way to build a modern industrial ecosystem and achieve advanced intelligent development. It is the foundation of "smart" factories, and the way of realizing the potential of equipment, technical processes, and materials. It is also the key to improving production efficiency, optimizing resource allocation, creating differentiated products, and implementing value-added services [1]. The industrial Internet platform (IIP) is the hub of all connections of production factors, and the core of industrial resource allocation. It presents new features of manufacturing industry, such as cross-domain, cross-region, and full life cycle.

The United States leads the world in transferring industrial knowledge and experience into software and its platformization. By strengthening the platform-based network, data, software application, and system integration, decoupling the business of manufacturing process and product based on services, and promoting new forms of manufacturing such as networked equipment, platformization of business, and virtualization of resources, IIPs represented by GE Predix have been spawned.

Received date: April 3, 2018; **Revised date:** April 8, 2018

Corresponding author: Wang Chen, National Engineering Laboratory of Big Data System Software, Tsinghua University, Chief Engineer. Major research fields include big data for industry, industrial Internet, database, and big data system. E-mail: wang_chen@tsinghua.edu.cn

Funding program: CAE Advisory Project "Research on Development Strategy of 'Internet Plus' Action Plan" (2016-ZD-03); Project of National Natural Science Foundation of China "Big Data Analysis Theory and Key Technologies for Industrial Applications" (U1509213)

Chinese version: Strategic Study of CAE 2018, 20(2): 015–019

Cited item: Wang Chen et al. The Industrial Internet Platform: Trends and Challenges. *Strategic Study of CAE*, <https://doi.org/10.15302/J-SSCAE-2018.02.003>

Germany has an impressive heritage of industry, with manufacturing leading the Internet as the mainstream development path. Germany advocates using the Internet to promote the optimization of efficiency in techniques, processes, and equipment, and to promote the restructuring and continuous promotion of the value chain system. The open cloud platform MindSpere (developed by Siemens) is a typical example. The development path of China's II has instigated many innovative business models with diverse characteristics. Among the domestic industrial Internet platforms, there have been collaborative manufacturing platforms such as casicloud.com, product lifecycle management service platforms represented by iRootech Technology, and user-customized platforms represented by Haier.

2 Ecological challenges faced by IIPs

The II ecosystem must solve various problems: who will use it? (i.e., business ecosystem), who will provide the services? (i.e., developer ecosystem), and how will each link achieve cross-border integration? (i.e., data ecology).

The construction of the II ecosystem is carried out based on the business of industrial enterprises from respective fields. Users' adherence to the platform depends on the platform's value to their businesses. Users of IIP can be divided into three categories: internal users of enterprises, enterprise, and end users in the industrial chain, and other enterprises in the same industry. For the internal users of enterprises, the value of the platform is mostly in improving the quality and efficiency of the business itself, such as the improvement of research and development efficiency, energy consumption reduction, and shortening of turn-over cycle. This is achieved by a series of data or model-driven analyses and optimization software applications on the platform. For enterprise users and end users in the industrial chain, the value of the platform is reflected by various collaborative and application-oriented services. Through the application of R&D cooperation, supply chain coordination optimization, manufacturing ability outsourcing, and centralized procurement of materials, the efficiency of upstream suppliers can be improved, and the costs of procurement and production can be effectively reduced. For downstream users, the value is more reflected in product-based services in operation and maintenance and product-centric value-added services, such as financial services, content services, and other services. For other enterprises in the same industry, especially small and medium-sized enterprises lacking information technology capabilities and large-scale construction funds, IIPs are more likely to play a role as providers of capabilities in industrial software, manufacturing resources, and industry knowledge, for example. It can be inferred from such a perspective that the challenge for the construction of a user ecosystem lies more in the capability of industrial software-as-a-service (SaaS).

The other factor in the IIP is the content provider of the plat-

form, who can be an internal business department or software service provider, or a third party independent developer. To construct the developer ecosystem, focus needs to be placed on the integrity of the platform's function, the ease of development, and the stability of the system. Further, the developer ecosystem must form a positive feedback effect with the user ecosystem of the platform, through which the platform gathers more and more users and emerging business demands, and provides more market opportunities for developers. From the perspective of personnel skills in developing industrial applications and models, industrial know-how is the core, but it is usually a great challenge for industrial talent to grasp information technology. Therefore, competition for the scarce existing developer resources has become the most important task for the establishment of the ecosystem of the IIP.

The construction of the data ecosystem is the key to the vitality of the IIP. The term data-driven refers to treating data as a natural resource. As it is a resource, we must find ways to exploit and make use of it. Therefore, it is necessary to determine what data the platform builders already have, what data they will have in the future, and what data they need to have in the future. The first level of data ecosystem construction is data source collation. Some of this data comes from existing IT systems, some from machines, some from upstream and downstream of the supply chain, and some from the Internet. The second level of data ecosystem construction is the opening of data sources. Big data in industrial enterprises has multiple sources, meaning that data will be collected from differing systems. It is necessary to consider how to synchronize and exchange data among systems, and how to correlate and integrate data based on industry domain models, including correlating and integrating data with business systems. The third level of data ecosystem construction is the consolidation of data sets. The issue is how to extract, store, and manage data from sources, multiple available data sets and corresponding data resource directories, and standardized interfaces based on application-oriented features. The next layer is data governance. For industrial data, it is possible that devices, the environment, transmission, and other factors will lead to data loss, duplication, errors, and other problems. Therefore, it is necessary to understand the quality of data by using data portrait tools, and then correct the data with the help of data cleaning tools.

3 Technical challenges faced by IIPs

Infrastructure-as-a-service (IaaS) technology has reached maturity, and has a short iteration cycle. In just over ten years, Amazon Web Services, Microsoft Azure, Alibaba Cloud, Tencent Cloud, and HUAWAI Cloud, etc., have emerged. At present, the developers of most examples of the platform as a service (PaaS) have insufficient knowledge of industrial know-how and professional technology. In general, targeted research and de-

velopment of PaaS has not been undertaken to narrow the wide gap in specialty between manufacturing and consumption [2]. Therefore, building industrial SaaS directly on the existing general cloud-computing platform will lead to problems such as high development costs and poor usability. Further, it will more easily lead to enterprises' misjudgment of the strategic direction and implementation path of the development of II. The research, development, and construction of PaaS should focus on the professional direction of the industrial fields. The two core components of the industrial PaaS layer in the IIP architecture (Fig. 1 [3]) are industrial big data systems, and industrial data modeling and analysis. Because these two technologies conform to the characteristics of industrial data and their application, they present significant challenges for the overall technical system of the IIP [4,5].

3.1 Key technical challenges in industrial big data systems

3.1.1 The collection technology of polymorphic data

The industrial hardware and software system is a complex, hybrid, and closed system. Differences in data format, interface protocol, and other standard systems can instigate considerable technical barriers. Even the basic specifications or generated information of the same type of equipment produced by the same factory often varies, owing to different components and materials. Therefore, whether it is data collection and analysis of the industrial system, or structured decomposition of collected data

by data storage systems, there are considerable potential application challenges. Owing to the different standards from economic entities (or the closed designs of the protocols), it is sometimes impossible to collect the data of a piece of equipment. Further, even if the data are available, in the actual implementation process of an industrial big data project, several months and a large amount of human investment are needed to organize data formats and fields [5]. The bigger challenge comes from the proliferation of complex, unstructured data that are diverse and highly variable. Because industrial systems have high standards of independence, sensitivity, and safety, the data are usually only compatible with (or processed by) specific software, which results in more difficulties in the overall structuration of information. Therefore, such challenges can only be overcome by a systematic approach, and through the comprehensive application of big data management technologies, such as industrial standardization, data model intelligent identification, or matching.

3.1.2 High throughput data access technology

Large amounts of machine data, especially time series data, are connected to industrial big data systems. In general, large manufacturing enterprises can connect and access hundreds of thousands of devices at the same time, and the data throughput rate can easily reach an astonishing magnitude of a million data points per second to ten million data points per second. In the face of such application requirements, big data platforms must have data access capabilities similar to those of real-time

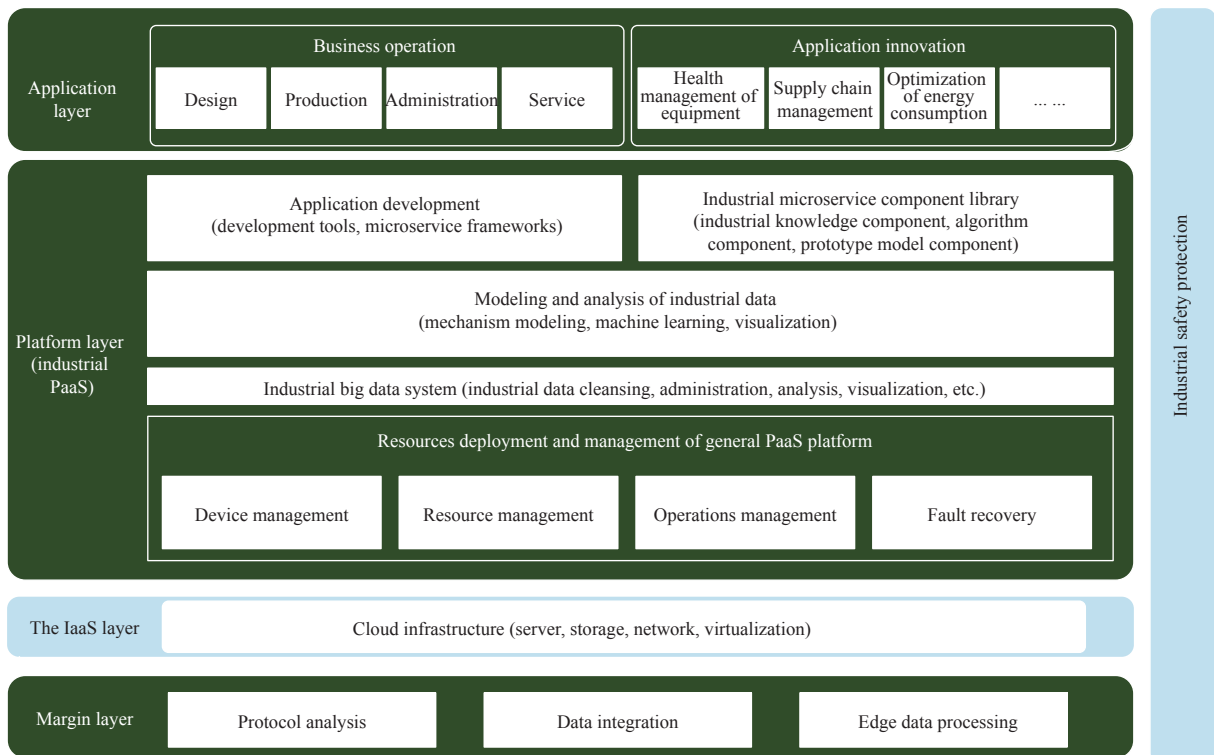


Fig. 1. IIP architecture.

databases. However, as the big data platform requires long-term and highly reliable storage of data, efficient data compression coding methods and low cost distributed extension capabilities are also major challenges for industrial big data systems. In addition, in order to meet the requirements of multi-condition complex query and high-performance response, in the process of data access, it is necessary to design a complete data organization system and index structure, implement effective auxiliary preprocessing and calculation, and optimize the efficiency of the read-write system for high-throughput data access [5].

3.1.3 Low quality data-analysis techniques

One of the key issues that big data technology needs to solve is the extraction of high-value information from a significant amount of low-value data, that is, making up for the deficiencies of low density information by utilizing a large scale of data. As industrial hardware and software systems correspond to strict mechanism models, the definition of each variable has a definite physical meaning. Low-quality data will lead to a change in the functional relationship between variables, which will negatively affect the analysis of industrial big data. In fact, there have been several problems with the data quality of information systems in manufacturing enterprises owing to historical, technological, and human factors. For example, the material in the enterprise resource planning (ERP) system has the problem that one item corresponds to multiple labels. Furthermore, the quality of data in the Internet of Things is far from ideal. The proportion of data quality problems such as invalid and renamed working condition data, wrong or irregular time scales, and time sequence disorder may be up to 30% in practical applications. As these problems negatively affect the results and evaluation of data analyses, the primary data must be effectively managed before analysis.

In industrial applications, owing to various reasons (such as time/space constraints and poor physical environmental conditions), a large amount of key information has not been measured, or sufficiently and accurately measured. This requires analysis algorithms to work based on imperfect and inaccurate data. To solve this problem, soft measurement technologies based on industrial big data analysis can be applied. Here, they can establish the correlation relationship model between indicators through big data analysis and use the measurement parameters of easy and measurable measurements to calculate the hard to obtain process measurement values, thus improving the overall data basis of the supplementary production process [5].

3.1.4 Management technology of multi-modal data

A large number of multisource heterogeneous data are produced in industry. The frequently seen types are structured business data, time-series data generated by device detection, and unstructured engineering data, among others. Each type of data requires efficient storage solutions and heterogeneous data

targeted storage models. However, the existing big data technology cannot meet all these requirements. Taking unstructured engineering data as an example, the various mass computer-aided design (CAD) files, test simulation files, pictures, documents, and other small files need to be organized and arranged efficiently, flexibly, and quickly for different characteristics such as product lifecycle, project name, and bill of materials (BOM) of hierarchical organization. At the same time, batch management and modeling, and quality control of data, are also required. Meeting these requirements, for both distributed file systems and object storage systems, is technically difficult. In addition, from the perspectives of convenient use and reducing the difficulty of development and learning, the internal processing differences of heterogeneous data should be completely transparent, to ensure that the data model and query interface exposed to users are as uniform as possible. For example, in IoT data analysis, a large amount of static data (such as sensor deployment information), need to be processed, and such operations often require cross-library mapping between time series data and structured data to establish a relationship; therefore, it is necessary to provide integrated query collaborative optimization based on multi-modal industrial big data [5].

3.2 Key technical challenges in industrial data modeling and analysis

3.2.1 Integration technology of strongly correlated data

Industrial big data analysis relies more on the integrity of the database than simply the amount of the data. Because of the existence of information silos in industrial production, the sources of industrial big data are characterized by spatial dispersion and time asynchrony. The application of industrial big data requires the integration of data at three levels: physical information, industrial chain, and cross-industry [5].

Physical information integration: The design and development stage mainly manages digital products, while the manufacturing service stage mainly manages physical products. Therefore, the whole-life cycle management needs to integrate digital and physical products to build a system integrating industrial information and physical products.

Industrial chain integration: Against the background of the big data of the Internet, cloud manufacturing production modes with the goal of production resource reorganization and optimization have been rapidly developed. The intelligent industry chain must open up the business scope and boundary of traditional enterprises and achieve data-driven collaborative business integration.

Cross-sector and cross-industry integration: In the industrial environment of Internet Plus, it is necessary to implement information integration of the upstream and downstream processes, peripheral ecosystem, and other wide areas. For example, a farm

machinery company in the United States needs to make integrated use of agricultural resource data, such as weather dynamics, water resources distribution, and seed conditions with its own agricultural machinery product data, to provide better, efficient, and accurate services for farm production. This method of integration is more sophisticated, and offers a higher level of integration of industrial big data. Therefore, it is necessary to comprehensively analyze the manufacturing process, BOM structure, operating environment, and other types of industrial semantic information.

3.2.2 Analysis techniques of businesses with strong mechanisms

The mechanism is the basic logic and operating principle of modern industry. The industrial production process is also a strictly controlled process of strong mechanisms; that is, a large number of professional theoretical models in related fields are used to formalize the dynamic processes of physical, chemical, and biological changes in the real world. In addition, there are a large number of closed-loop regulation or control mechanisms, which can be used as parameters to make the industrial production process infinitely approximate to or even exceed the design goals. Traditional data analysis follows the isolation principle of “one takes its own charge,” which rarely combines mechanism models (or even fully data-driven models), and rarely considers the role of closed-loop control/regulation mechanisms. The challenge of a strong mechanism model to data analysis technology is mainly shown in the following three aspects. First, the implementation of the organic combination of the mechanism model and data model. Specifically, how to embody the mechanism model in the data model (for example, the mechanism model provides key features for the analysis model, and the analysis model provides post-processing support for the mechanism model or multi-model fusion prediction), or how to input the data model as the mechanism model (namely, to provide parameter calibration). Second, the computational model fusion. The mechanism model generally needs compute-intensive operations (CPU multi-core collaboration or computing cluster parallelization) or memory-intensive (GPU parallelization) operations, and data analysis is usually an intensive throughput (I/O) operation (for example map-reduce or parameter server). The two computing modes meet different respective requirements, and the selection needs to be based on the analysis of the algorithm and even on the software. Third, the method of integrating knowledge and experience with domain experts can make up for the knowledge deficiency of current front-line staff, and realize the visualization of process logic. For example, in the case of physical change processes, emphasis needs to be placed on automation of knowl-

edge rather than on rediscovery. Knowledge in different fields should be systematized, and information retrieval and update should be optimized based on big data analysis technology. Further, for relatively clear and formalized domain knowledge, modeling should be carried out by relying on the space-time pattern description and analytic identification technology provided by the big data-modeling tool, and verification and quality improvement should be carried out through massive historical data to constantly refine expert knowledge to form artificial intelligence [5].

4 Conclusions

The IIP is a product of the deep integration of manufacturing industries and the new generation of information technology such as cloud computing, big data, and artificial intelligence; it requires the effort of a large number of researchers and industrial workers in practice while searching for a breakthrough. As a large manufacturing country, China has the largest amount of industrial equipment in the world, is producing a large number of industrial data, and has a rich Internet ecosystem and a large body of industrial and information talent. We should make full use of such conditions to innovate management ideas, reconstruct industrial ecology, and enhance the position of China’s manufacturing in the global industrial chain. Relying on the IIP, we will promote the in-depth application of the II and the transformation and upgrading of the manufacturing industry, and strive to transcend it in a new era of the industrial revolution.

References

- [1] The State Council of the PRC. Interpretation of guiding opinions on deepening “Internet + advanced manufacturing” to develop industrial Internet [R]. Beijing: The State Council of the PRC, 2017. Chinese.
- [2] National Industrial Information Security Research and Development Center. Comparative analysis of industrial Internet platform at home and abroad [R]. Beijing: National Industrial Information Security Research and Development Center, 2017. Chinese.
- [3] Industrial Internet Alliance. Industrial Internet architecture white paper [R]. Beijing: Industrial Internet Alliance, 2017. Chinese.
- [4] Industrial Big Data Special Unit of Industrial Internet Alliance. Industrial big data technology and application practice [M]. Beijing: Publishing House of Electronics Industry, 2017. Chinese.
- [5] Wang C, Guo Z H, Wang J M. Industrial big data and its technical challenges [J]. Telecommunications Network Technology, 2017(8): 1–4. Chinese.