Design and Application of Distributed Computing Environment Based on Data Drive

Zhenguo Zhu¹, Chao Bi², Rong Zheng³ and Song Zheng¹,*

¹College of Electrical Engineering and Automation, Fuzhou University, Fuzhou, China
²Research Institute of Fujian Histrion Group Co., Ltd., Fuzhou, China
³IAP (Fujian) Technology Co., Ltd., Fuzhou, China

*Corresponding author email: s.zheng@fzu.edu.cn

Abstract. This paper proposes a distributed computing environment based on data-driven mechanism, which can be deployed on different levels of Industrial Internet, designs and implements an Industrial Internet architecture with high reusability and rapid deployment. This architecture has the consistency of the computing environment in the cloud and terminal control layer, which is beneficial to the realization of remote management and collaborative control of Industrial Internet. Data driven technology enables heterogeneous industrial equipment to achieve interconnection and interworking at a low cost. The multi-engine computing environment constructed by it is not only an effective method to solve the application design and development of complex systems, but also a key technology to realize edge computing and fog computing. This paper designed a corresponding test bed platform, test results show that the technology can not only provide connectivity of heterogeneous data source and efficient technical support, and to ensure that the distributed industrial equipment real-time and synchronization of data transmission, but also has the strong edge computing capabilities to cope with the constantly generated data stream, and reducing the pressure of the cloud server storage, computing.

Keywords: Industrial cloud computing; Data engine; Edge computing; Heterogeneous interconnection; Configuration environment.

1. Introduction

The new production model brought by Industrial Internet will be the integration and integration of many production factors. All participants and production resources and production materials are highly interactive. The unique needs and characteristics of customers and products are likely to be integrated into all stages of design, production and operation, engineers can optimize the configuration, intelligent supervision and dynamic reconstruction of the production process according to the specific situation. The establishment of Industrial Internet not only involves the internal network of the factory, but also involves the external network of the factory, and involves the intelligent joint control of massive heterogeneous equipment. However, due to the lack of unified standards for Industrial Internet, equipment between different manufacturers cannot be interconnected. In addition, factories, equipment, network, users and products in the network cannot establish corresponding trust mechanism and security protection measures due to their independent closed systems. This requires us to build a more open, interconnected and interoperable system platform to meet the needs of standardization and generalization brought by the proliferation of networking and collaboration of devices and systems. Xiao et al. (2014) designed a user interoperability framework (UIF) that enables device users to interoperate...
with heterogeneous devices with different contexts with consistent syntax and semantics; Figay et al. (2012) present a federative interoperability framework for technical enterprise applications; and Jardim-Goncalves et al. (2012) explained the fundamentals of sustainable interoperability and analysed its application prospects. This paper believes that promoting the deep integration of information technology and manufacturing technology, promoting the digitalization, networking, intelligent transformation and development of industrial equipment and production requires building a unified distributed computing environment in the cloud and terminal control layer, which is the basis and key to the development of the industrial Internet, and the best way to minimize the cost of building enterprise information infrastructure.

The operation of any industrial control algorithm is inseparable from the support of the computing environment. Moreover, there are many computing environments with different underlying structures in industrial complex systems, and the existence of these heterogeneous forms is the main obstacle to system interconnection (Reveliotis (2015)). For the design of the industrial Internet computing environment, the main disadvantages of related studies are that they mainly consider the interoperability between heterogeneous systems, the computational efficiency of complex application algorithms, and the operational collaboration between edge computing and cloud computing. Few studies discuss the consistency of the cloud-edge computing environment technology system, or talk about how to solve the contradiction between the diversity of industrial applications and the unity of the virtual machine computing environment to meet the changing application needs of the industrial Internet. In the final analysis, these scientific issues discuss the scalability and high reuse value of the platform architecture. In addition, the processing of complex data structures has high requirements for visualization technology. If you continue to use computer software to directly program the diversity of visualization applications, the software development costs will remain high, which to some extent will also limit the application benefits of the Industrial Internet computing environment. The construction of visualization technology framework can effectively simplify the complexity of industrial cloud computing, and visually present the process data and result data of industrial cloud computing to users for analysis and decision-making.

The main contribution of this paper is to propose a distributed computing environment based on a data-driven mechanism to build a cloud-edge big data processing environment for Industrial Internet, which can effectively solve the problems of high cost and low efficiency of industrial Internet application development, and can adapt to diversified industrial application needs.

2. Requirements Analysis for Consistency in Distributed Computing Environments

Industrial Internet is actually a very large scale distributed computer system, which is essentially a huge multi-dimensional data network formed under the distributed connection of many nodes and the action of ubiquitous computing. All industrial elements, including people, computers and industrial equipment, are extensively linked together to improve the productivity and reduce costs of the entire industrial society. The factors that determine the production efficiency and final form of future industrial society are the scale and depth of connection between different industrial factors. In June 2013, GE released the first industrial Internet platform, Predix, which is used to manage the data generated by the operation of large industrial machines in the industrial cloud computing environment and mine the data value. Leading enterprises in developed countries have also adopted the Industrial Internet platform as an important direction of strategic layout, and gradually built the platform layout capability that they control. In just a few years, hundreds of Industrial Internet platforms have emerged in different fields. According to IoT Analytics, the number of industrial Internet platforms in the industry has exceeded 150, mainly from traditional IT enterprises and manufacturing enterprises, including Predix, Mindsphere, COSMOPlat, supOS, Fiibeacon and INDICS (Xiao (2018); Peter (2012); White paper on industrial Internet platform (2019)). Situation in face of the rapid expansion of the number of platform, the communication protocol and data format differences between different networks, industrial production field exist a large number of machinery and equipment of heterogeneity, and the diversity of all kinds of data sources also widening, causing heterogeneous and complex computing environment, lead to the complexity of industrial interconnection increasing, also greatly limited the industrial efficiency and scale of the Internet connection.
It can be seen that most of the Industrial Internet platforms emerging in the industry are still a kind of "application" platform closely related to the type of application business. They cannot use software as a pure tool like the traditional consumer Internet, and the application cost remains high. The ultimate goal of Industrial Internet platform is to realize ubiquitous connection and optimal configuration of industrial resources, reduce the complexity of Industrial Internet applications, improve the efficiency of industrial application design, development and deployment, and reduce risks and costs. For such a complex large system, the complexity of system research can be minimized only when all research objects are described and expressed in a unified language specification and all data processing contents are driven by a standardized computing environment (Lin (2017); Derhamy (2018)). Aiming at the existing problem of the above requirements and the development method, this paper proposes a general industrial Internet platform technology system based on the multi-engine distributed computing environment built by data engine technology, which forms a unified computing environment in the cloud computing layer and the edge layer, providing an effective way to promote the rapid development of industrial Internet technology.

3. Technological Innovation of Data Engine Concept
The biggest difference between industrial Internet and traditional Internet is that the former needs a specific cloud control platform, cloud control combined with cloud computing and remote control, and find suitable virtual control resources and control strategy resources for information transmission, management and decision making of distributed or parallel subsystems. These subsystems can access the cloud platform data center through the data communication link for integrated management and control. The establishment of the data link ecology requires a built-in computing environment suitable for industrial big data processing, that is, the data structure can be interpreted and driven according to the standard of the unified standard, and the data relationship of the control configuration is reconstructed.

In order to realize the technology innovation, IAP (Fujian) Technology Co., LTD, and Industrial Automation Control Technology and Information Processing Key Laboratory of Fujian Province University jointly proposed the data engine technology, data engine is a computing unit based on unified data communication specification and standard data processing method, and can be used as an independent agent, used to deal with a set of relatively simple control tasks, and the complex application tasks can be decomposed into multiple sub-problems to be processed by different combinations of computing agents, forming a huge multi-agent distributed parallel computing environment to meet the needs of industrial cloud computing.

In addition, the data engine technology adopts the real-time sharing mechanism of unified memory. The memory sharing area is specially designed in each data engine structure, which provides a good interactive channel for data synchronization between different data engines and ensures a high degree of consistency of real-time data between different data engines. As long as each heterogeneous controller loads the data engine, various control algorithms can be transplanted and shared with each other in a variety of different types of computing platforms, thereby supporting online configuration and debugging.

This shows that Data engine technology has an absolute advantage in avoiding the "Information Isolated Island" dilemma in the information exchange level of the Industrial Internet, when a large number of data engines are deployed to the cloud platform and common gateway of the Industrial Internet, a computing platform of a common platform is formed, and tens of thousands of data engines will cooperate to complete the complex computing tasks of the Industrial Internet application. From the perspective of the law of scientific and technological development, industrial Internet is the product of the extension of the traditional Internet to the industrial field, and its development will inevitably move towards standardization.
The data engine also has the feature of computing visualization, because its core is a real-time database based on Shared memory, and all intermediate calculation results are the contents of the database. And, the data engine will establish an associated communication address mapping mechanism between the real-time value of the driving data and the components of the human-computer interaction system. As shown in figure 1, the operator only needs to encapsulate the control algorithm into a uniform standard paradigm. The component is downloaded to the virtual controller of the shared memory, and the configuration execution of the different control stations is converted into the data update calculation and management process of the real-time database in the data engine via the Internet communication channel. Thus, the data change process and computing relationship can be monitored in real time through the human-computer interaction interface, so as to realize the remote control of distributed or parallel subsystems. More importantly, this feature enables the system to achieve online dynamic optimization by continuously optimizing the graph configuration program without affecting the continuity and stability of its operation. Moreover, this process can be realized by adding, replacing and deleting configuration components or changing the connection relationship between components, which greatly reduces the complexity of remote operation and maintenance of industrial Internet.

4. Technical Architecture Design

How to build a large Industrial Internet distributed computing environment, how to provide stable data services for the application layer software is the focus of this paper. On the basis of data engine technology, we build a data-driven industrial Internet PaaS platform IAPcloud, which not only provides a series of unique database resources for industrial users, but also designs a set of easy-to-use control algorithm editing, compilation and debugging environment, namely "configuration element" engineering application environment. As shown in figure 2, the configuration data of the applied algorithm is stored in the distributed data engine. The configuration data of human-machine interaction (HMI) is also distributed in the corresponding data engine by means of dynamic cache.
configuration data of the device information management is stored in a dedicated device library, and the information of the entire industrial Internet device is maintained through a synchronization mechanism. It can be seen that all information processing methods in the whole process from algorithm configuration to algorithm execution under the framework of IAPcloud adopt consistent computing environment to drive execution, which is a data-driven industrial Internet PaaS platform. The superiority of the Industrial Internet PaaS platform must be due to its integrity from data collection, cloud platform to human-computer interaction (Chen (2018)). Since the data on the cloud platform mainly comes from industrial production equipment, and the data source of production equipment of most enterprises is “heterogeneous”, how to extract and collect various types of data, how to manage and apply massive data, realize the equipment layer The large number of heterogeneous devices and nodes in real-time and reliable interconnection and so on, all involve the issue of standardization (Andreas (2015)). Therefore, this paper proposes a universal gateway IAPbox, which contains dozens of industrial communication drivers, and has both edge computing, network disconnection and remote management functions. IAPbox can easily integrate the heterogeneous physical world with the homogeneous information world, the key lies in the embedded computing environment based on data engine (Song (2017)). Therefore, both the geographically distributed IAPbox and the distributed virtual controller on the cloud platform adopt the same computing environment, ensuring the synergy of the overall operating environment of the industrial Internet, which is of great benefit to the design, development and application function collaboration of industrial applications. Based on such characteristics, IAPcloud will have excellent performance in practical engineering applications (Song (2018)).

A) Data engine and configuration element is extensible, and do not need to change the status of architecture, makes IAPcloud platform has nothing to do with any industry professional knowledge. Users can use their configuration components and computing environment to build specific industrial Internet applications, so as to make it a true deeply interconnection platform of industrial Internet and avoid falling into the predicament of "Information Isolated Island" again at a higher level.

B) Adopt the multi-function general gateway technology and multi-engine cloud platform virtual controller technology to ensure the consistency of the platform computing environment. It can realize the rapid deployment of industrial Internet for complex industrial systems.

C) The visual configuration technology provides a more flexible configuration environment for developers and users of industrial Internet. It does not require developers to program and develop related software functions instead of graphical configuration development, this also provides important conditions for the deep integration of industrial production management and control technology.

D) IAPcloud is based on a complete technical standards system that is compatible with most engineering applications in the industrial field and has good scalability.

E) The IAPcloud platform also provides a cloud simulation calculation engine that can use digital simulation technology to establish a corresponding digital model for the production process, provides industrial Internet-based parallel control for the establishment of a digital twin environment in an industrial Internet environment.

5. Test Bed Analysis and Evaluation

We believe that the ultimate goal of industrial Internet development is to achieve a high degree of standardization. The advantage of standardization is that many industrial resources in the Internet are highly reusable. Different applications can cooperate with each other, and the platform can quickly respond to changes in application requirements, thus improving the operational efficiency of industrial society to the greatest extent and reducing the cost of rapid deployment. This requires the deep integration of industrial systems IT and OT, the establishment of a good architecture at the beginning, and the connection of seemingly isolated software and hardware through a unified technical route to form a standard system. This technology route includes the design of good computing environment, the establishment of network communication environment, and the development of Industrial Internet application development framework and software tools. Only in this way can the complexity of heterogeneous environment be completely solved from the underlying architecture, that is, the problem of engineering application standardization, and the IAPcloud based on data engine technology has the
prerequisite for the establishment of industrial Internet standardization platform. In this paper, a set of industrial cloud computing test bed is used to verify the scientificity of the distributed computing environment concept proposed in this paper. The test bed consists of several different brands of PLC, general gateway equipment, lighting system and a set of servo motor control system. These PLCs can be used to simulate the actual equipment in the factory, and the multi-dimensional data of various terminal heterogeneous devices in the test bed will be uploaded to the cloud for collaborative computing through the Internet.

5.1. Real-time Performance of Heterogeneous Data Collection, Collaborative Analysis and Calculation

We embed the data engine real-time database directly in the data memory area of various actual control stations. Based on QNX control system of IPC (IPC - S2221) as controller, run the logic of the diagonal recursive neural network control algorithm, and the final control signal is generated and transmitted via shared memory to the output element corresponding to the PLC as the executing device, then output by the CS1W-DA08C analogy output module, transmitted to the voltage input terminal of the Delta servo drive for control. The feedback loop of the control loop is collected by the Omron CS1W-CT021 high-speed counting module to collect the high-speed pulse signal of the Omron C6B2-C optical electronic rotary encoder. The data of the speed is calculated by Omron's embedded data engine, and it is fed back to the IPC for the iterative operation of the neural network control algorithm to form a closed-loop control test bench for heterogeneous terminals.

Figure 3. Real-time analysis diagram of coordination and monitoring of heterogeneous terminals.

As shown in figure 3, due to the adoption of a data-driven standardized computing environment, the data engine can co-process all the transactions of the shared memory database, and autonomously stimulate the CPU process to process the update request of the relevant data, and perform high-speed scheduling, updating, calculation and management on the calculation timing of the algorithm data model. We randomly change the control model signal given by Omron PLC and increase the interference signal. The IPC controller can share the data changes in real time and quickly iteratively adaptively control the model signal curve. At present, the minimum iteration calculation period of the advanced control algorithm controller based on PLC and IPC in this paper is less than 20ms, which can realize the collaborative calculation of multi-dimensional heterogeneous data quickly.

5.2. Experiment on the Scalability of Configuration Environment

Infrastructures such as cloud platforms and IAPBox based on multiple data engines have characteristics that are independent of the specific application of the Industrial Internet, and their engineering applications are closely related to a consistent configuration environment (configuration tools and configuration components). On the basis of experiments in section 5.1, we added colour lights and
illumination sensors, connected with the motor control system through the gateway device, and formed a colour light display system varying with the motor speed. All devices are connected to the local gateway devices via LAN, and finally uploaded to the cloud server for collaborative monitoring in the industrial PaaS.

Figure 4. Scalability testing of the configuration environment.

As shown in figure 4, since a corresponding mapping mechanism is established between the data engine and the components of the human-computer interaction system, we only need to add and configure the newly added illuminance device to determine which data inside the device component enters the data sharing area. The shared area is synchronized with other controller nodes, and the shared components are directly involved in the configuration programming process of the control logic through the orderly connection of the graphic components. The real-time values of the shared components can be dynamically presented in the configuration screen, and the communication is performed. The data is transparent and easy to use. Based on this, engineers can decide the information sharing strategy more autonomously and select the specified data to exchange data between multiple controller nodes. This method can not only improve the flexibility of data sharing between controllers, but also reduce the impact of unnecessary shared data on network load and controller performance.

5.3. Abbreviations and Acronyms Advantage Analysis of Remote Fast Operation and Maintenance

At present, most of the remote operation stations studied can only provide some simple control strategy remote parameter modification and remote status monitoring functions. It is difficult to realize online modification and dynamic debugging of complex heterogeneous coordinated control system operation strategies, which cannot meet the remote control of configuration engineer’s demand. However, the visual configuration technology based on the standard computing environment can solve the appeal problem well and effectively reduce the complexity of the industrial Internet. Moreover, data synchronization (or exchange) based on shared memory ensures high temporal and spatial consistency of data between different data engines, providing technical conditions for the industrial users to algorithm configuration without disturbance update in the running state. That is the dynamic reconstruction technology of remote visual operation and maintenance, which is synonymous with real-time evolution, enabling it to change its functionality and/or topology online while the system is running. The benefit is that the system is updated with application consistency and service continuity, while at the same time minimizing interference to the application (Almeida (2014)).

As shown in figure 5, we studied online the Diagonal Recursive Neural Network (DRNN) control model proposed by Ku et al. (2002); Kazemy et al. (2007) in this experiment. When the learning error value was less than a certain small value, we dynamically reorganize the PID conventional control algorithm configuration logic of the motor speed control in the test bed into a neural network control algorithm.
that has been learned online. There was no large disturbance in this process, and only a small overshoot occurred at the moment of switching, which is the deviation caused by the alternating fluctuation of the error feedback value generated at the switching instant. It can also be seen from the weight curve of the neural network that the weight curve before reconstruction is under ideal industrial control conditions, and the curve is smooth. After the reconstruction, the weight updating process of the neural network control algorithm is adjusted in real time with the fluctuation of the actual feedback error.

Figure 5. Visual dynamic reconstruction analysis.

The reconstruction process only requires the engineer to transmit the configured algorithm logic online to the data engine of the corresponding control station, and the corresponding algorithm component is automatically invoked for collaborative computing. Relative to the input and output of the device, it is only a change in data. More importantly, it can also monitor the changes of intermediate data in each link in real time in the logic configuration, which greatly improves the real-time performance of remote monitoring, which is very important for practical engineering applications.

6. Conclusion

Industrial Internet is a large complex industrial system composed of people, machines and objects. All data sources of industrial applications are heterogeneous machines or people distributed in different regions. Its architecture must solve the problem of how to adapt to the distributed data environment and how to improve the ability and level of data processing complexity. Otherwise, it is difficult to form an efficient form of interoperation and collaboration between different fields, different application scenarios and different modelling types required by users. The main contribution of this paper is to propose a data-driven technology based on shared memory, which provides a unique computing environment for heterogeneous collaboration of industrial Internet. The industrial Internet computing environment constructed by the multi-engine mode can effectively deal with the problems of complex systems. The natural formation of computing space can reduce the complexity of industrial Internet applications and improve the consistency of distributed computing technology. IAPcloud based on this technology provides developers with a development environment based on graphical configuration, which effectively improves the development efficiency. With the help of visual computing and dynamic reconstruction technology, the platform can be quickly deployed, greatly lowering the threshold of industrial Internet applications.
References

[1] Almeida, J.P.A., Sinderen, M.V., Pires, L.F., Wegdam, M. (2004). Platform-independent Dynamic Reconfiguration of Distributed Applications. *Proceedings IEEE 10th International Workshop on Future Trends in Distributed Computing Systems (FTDCS 2004)*. IEEE Computer Society Press, 286-291.

[2] Andreas, M.H., and Vandikas, K. (2015). Computations on the edge in the internet of things. *Procedia Computer Science*, 52, 29-34.

[3] Chen, L.X., Pan, Z., Liang, G., and Jie, X. (2018). Adaptive fog configuration for the industrial internet of things. *IEEE Transactions on Industrial Informatics*, 1-1.

[4] Derhamy, H., (2018) Architectural Design Principles for Industrial Internet of Things. Industrial Electronics.

[5] Figay, N., Ghodous, P., Khalfallah, M., and Barhamgi, M. (2012). Interoperability framework for dynamic manufacturing networks. *Computer in Industry*, 63(8), 749-755.

[6] Jardim-Goncalves, R., Popplewell, K., & Grilo, A. (2012). Sustainable interoperability: the future of internet based industrial enterprises. *Computers in Industry*, 63(8), 731-738.

[7] Kazemy, A., Hosseini, S. A., & Farrokhi, M. (2007). Second Order Diagonal Recurrent Neural Network. *IEEE International Symposium on Industrial on Industrial Electronics*. IEEE.

[8] Ku, C. C., & Lee, K. Y. (2002). Diagonal recurrent neural networks for dynamic systems control. *IEEE Transactions on Neural Networks*, 6(1), 144-156.

[9] Lin, S.W., Miller, B., Durand, J., Bleakley, G., (2017). The Industrial Internet of Things Volume G1: Reference Architecture.

[10] Peter, C., Evans, Marco A., (2012). 11. Industrial Internet: Pushing the Boundaries of Minds and Machines.

[11] Reveliotis, S., (2015). Real-time management of complex resource allocation systems: necessity, achievements and further challenges. *IFAC-PapersOnLine*, 48(7), 50-57.

[12] Zheng, S., Zhang, Q., Zheng, R., Huang, B.Q., Song, Y.L., and Liu, C.R. (2017). Combining a multi-agent system and communication middleware for smart home control: a universal control platform architecture. *Sensors*, 17(9), 2135-.

[13] Zheng, S., Lin, Z.C., Zeng, Q.J., Zheng, R., Rong, Z., and Liu, C.R. (2018). iapcloud: a cloud control platform for heterogeneous robots. *IEEE Access*, 1-1.

[14] White paper on industrial Internet platform. Industrial Internet industry alliance. (2019).

[15] Xiao, G., Guo, J., Xu, L. D., and Gong, Z. (2014). User interoperability with heterogeneous lot devices through transformation. *IEEE Transactions on Industrial Informatics*, 10(2), 1486-1496.

[16] Xiao, L.L. (2018). Comparative study on industrial Internet platforms at home and abroad. *Information and Communication Technology*, v.12, 64(03), 29-33.