

The Big Data-Driven Industrial CPS Real-Time Processing System Based on Storm

Junyu Ye ^{a)}, Lichen Zhang ^{b)}

College of Computer Science, Guangdong University of Technology, Guangzhou, 510006.China.

^{a)} 364773360@qq.com

^{b)} lchzhang@gdut.edu.cn

Abstract. In industry 4.0, CPS will play a huge role in the industrial production and supply management [1-3]. CPS makes the industrial system more intelligent, making the subsequent sales businesses more data-oriented and intelligent. However, the industry environment produces large of data during the production process. How to deal with these huge data quickly and analyze meaningful value message becomes the core issue of CPS. In this paper, a real - time processing system of CPS for industrial management system is proposed. The architecture is based on the characteristics of industrial management system and designed for solving the big stream data processing.

Key words: Big Data, Real-time processing, CPS, Storm, Industrial system, Intelligent.

INTRODUCTION

The Cyber-Physical System(CPS) is an integrated system for computing, communication and physical systems [4-5]. In an industrial environment, the optimization of industrial system and intelligent processing architecture are related to the performance of the whole CPS. With the Big data-driven, industrial CPS not only necessary to store data collected by industrial sensors but also needs real-time data mining and analysis. That requires a high reliability and fault-tolerant real-time processing system. By this real-time processing system, we can monitor the whole process of industrial production in real time, improve the efficiency of production and finally realize the intellectualization of industrial management system [6].

CORE TECHNOLOGY OF BIG DATA PROCESSING PLATFORM

The Batch System: Hadoop

Hadoop is the top distributed open source project of the Apache software foundation. The core of Hadoop system is the HDFS [7] distributed file system and MapReduce [8] big data processing technology. HDFS stores large amounts of data through a cluster of cheap PC units and supports fault recovery and easy expansion and MapReduce takes advantage of cheap PC units to cluster, batch large amounts of data, support distributed parallel computing and failover. With this to two core technologies, Hadoop is a software framework that enables distributed processing of large amounts of data and it can processed in a reliable, efficient, and scalable way.

Stream Processing System: Storm

Storm is an open source, big data processing system [9]. Unlike other systems, it is intended for distributed real-time unlimited continuous data flow and is not related to development languages. Storm is not just a traditional big data analysis system, it can also be used to build a complex event processing (CEP) system. The CEP system is functionally classified as computing and testing, and both this two classes can be implemented in the Storm through user-defined algorithms. CEP can be used to identify meaningful events in the flood of events and then process them in real time.

Cloud Computing Storage Model: HBase

In Cloud computing, it is necessary to analyze and process the distributed and massive data so that the data management technology must be able to manage a large amount of data efficiently [10]. HBase is a distributed storage system based on Bigtable open source implementation, which can be used to manage unstructured and semi-structured data types. In the industrial CPS system, the sensor, as the data acquisition component of the underlying system, has the characteristics of wide distribution and much data collection. And the different types of sensors cause data to be mostly unstructured data. Therefore, HBase is needed as a distributed storage system for industrial environment.

THE BIG DATA-DRIVEN INDUSTRIAL CPS REAL-TIME PROCESSING SYSTEM BASED ON STORM

We propose a big data industrial processing system based on Storm. The overall system architecture is shown below.

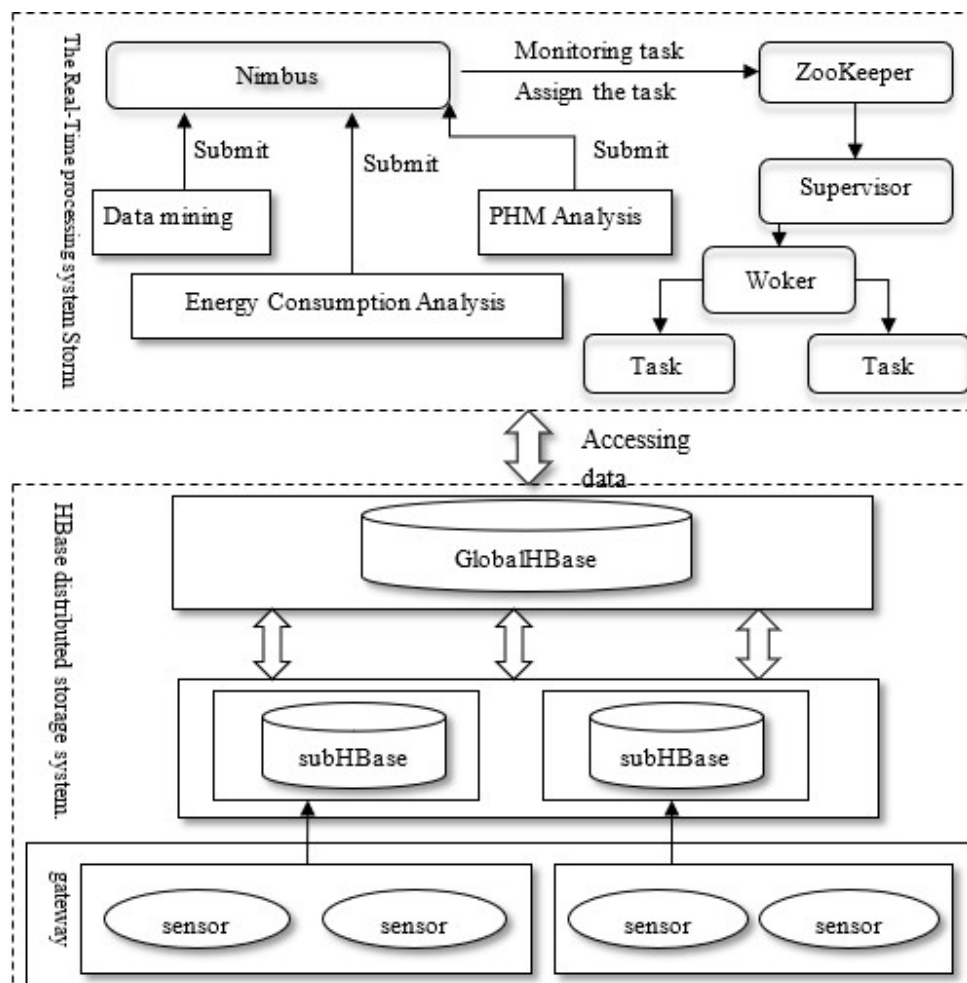


FIGURE 1. The big data industrial processing system architecture based on Storm

This processing architecture for real-time data processing will obtain the sensor collected data from the HBase distributed storage system and then submit corresponding task topology to the master node Nimbus daemon according to the different requirements (such as data mining, PHM [11] Analysis). After that Nimbus will assign the task to the node Supervisor which working in the cluster. The Supervisor starts or closes the worker process according to actual

requirements. Worker is the working process responsible for executing the topology in the supervisor cluster node, which is mainly used to generate the thread task for processing the topology, so that each topology can be processed in this Storm system.

About Zookeeper, its main function is to manage The Storm working progress node's error and supervisor node's error. If an failure exception occurs in the task topology execution scenario, the Storm will restore task execution failure on site and continue to perform and deal with the topology. This allows the subsystem to recover quickly from the error scene.

Figure2 shows an example about the Industrial PHM analysis topology process [12]. The whole topology process is divided into 4 parts:

- 1 Kestrel Spout: Kestrel is a data source. This is mainly responsible for obtaining the real-time status of industrial robots from the HBase distributed storage system.
- 2 PHMBolt: This part is mainly to carry out PHM data mining and analysis for industrial robot running state information obtained from Kestrel Spout data source.
- 3 Feedback Bolt: Feedback Bolt will feed back the PHM information from PHMBolt to industrial robots.
- 4 HBase Bolt: It used to store the PHM message into HBase.

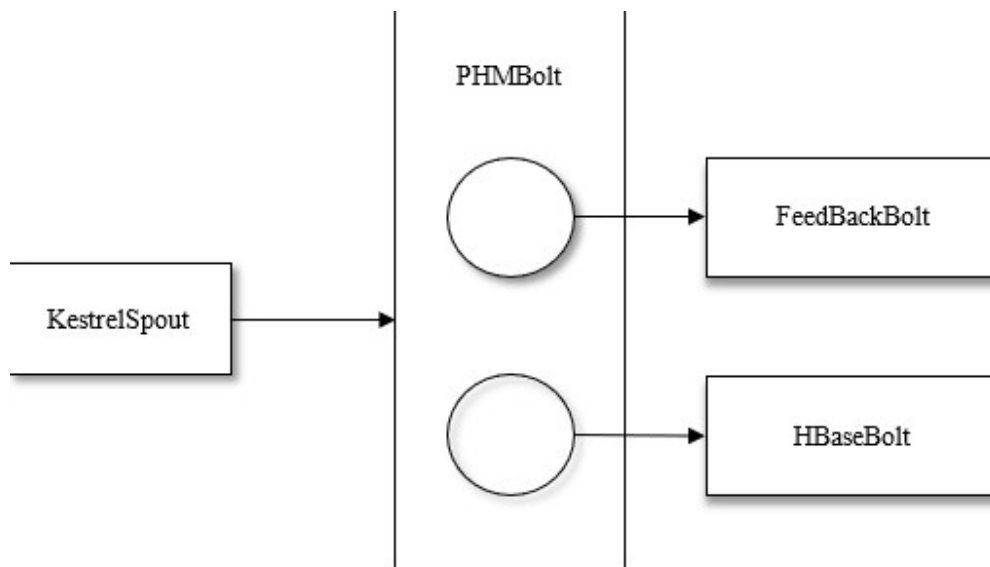


FIGURE 2. The PHM Topology Structure

In this case, First the Storm will read the sensor data which coming from HBase, after that it will generate the corresponding topology, the topology tasks are about the industrial PHM. Finally, the result feedback information will sent to the corresponding Industrial management system and stored in the HBase storage system.

CONCLUSION

The real-time processing system architecture of big data industry CPS based on Storm which proposed in this paper implemented a distributed processing platform with scalability, low latency, high reliability and fault tolerance. This architecture combines with industrial data characteristics to conduct data mining and analysis in real time, which can guarantee the real-time performance of industrial CPS system and suitable for industrial CPS production environment.

ACKNOWLEDGMENTS

This paper is supported by the National Natural Science Foundation of China(No.61572142) and Natural Science Foundation of Guangdong Province (No.2015A030313490).

REFERENCES

1. Shuai Yang. Industry 4.0 and industrial Internet: comparison, revelation and coping strategies. [J]. CONTEMPORARY FINANCE & ECONOMICS, 2015, 08:99-107.
2. Changhong Pei, Yan Yu. Germany "industry 4.0" and china-Germany manufacturing cooperation new development. [J]. Research on Financial and Economic Issues, 2014, 10:27-33.
3. Jieli Li, Ming Chen, Reference architecture of CSOCPS under "Industry 4.0"[J]. Modern Manufacturing Engineering, 2017, 08:147-161.
4. Zuopeng Li, Tianchi Zhang, Jing Zhang. Survey on the Research of Cyber-Physical Systems [J]. Computer Science, 2011, 09:25-31.
5. Yafei Hu, Fangmin Li, Xinhua Liu. CPS: Network System Framework and Key Technologies [J]. Journal of Computer Research and Development, 2010, S2:304-311.
6. Yingmai Gao. Industrial big data value mining path. [J]. China Industry Review, 2015, Z1:21-27
7. Qi Yu, Jie Ling. Research on cloud storage security technology based on HDFS [J]. Computer engineering and design, 2013, 08:2700-2705.
8. Jianjiang Li, Jian Cui, Dan Wang, Lin Yan, Yishuang Huang. Survey of MapReduce Parallel Programming Model [J]. Acta electronica sinica, 2011, 11:2635-2642.
9. Quinton Anderson. Storm Real-Time Processing Cookbook [M]. Beijing: China Machine Press, 2014.
10. Qiaoxing Li, Baohua Qiang, Chunyan Yang. The Storage Model of Big Data Basic-elements in HBase Database and Its Realization [J]. Journal of Guangdong University of Technology, 2014, 09(31):8-13.
11. Yu Peng, Datong Louisiana Peng. A review: Prognostics and health management [J]. Journal of Electronic measurement and instrument, 2010, 01:1-9.
12. <https://github.com/twitter-archive/kestrel>.