Construction of Cloud Storage Platform based on the MongoDB and its Application in the Oilfield Production

Chao Yang\(^1\)*, Shaowei Huang\(^1\), Lining Zhao\(^1\), Bing Xiao\(^1\), Lei Ren\(^2\)
and Yuchao Zhang\(^3\)

\(^1\)Engineering Technology Research Institute of Huabei Oilfield Company, 062552 Renqiu, Hebei Province, China
\(^2\)Daqing Oilfield Highway and Bridge Engineering Co. Ltd. Daqing, Heilongjiang Province, China
\(^3\)Yangtze University, Jinzhou, Hubei Province, China

ABSTRACT

With the rapid development of the digital constructions in Huabei oilfield, the area that the automatic monitoring systems must cover becomes larger and larger over times. In addition, the systems built in different local region face many problems, such as the un-unified data format, decentral storage, difficult to share, and low data utilization. Based on MongoDB, a cloud storage platform is built, aiming at synchronizing all the real-time production data from the different automatic monitoring systems to the unified storage format. Then, the secondary development of unified production data on this platform can easily realize the oil/water well on-line analysis, 3D-simulation, real-time monitoring, report, etc. The built cloud platform will satisfy the demand of the integration and optimization for production engineering.

INTRODUCTION

An set of automatic monitoring systems has been developed in Huabei Oilfield since 2000. The digital management system is initially well suited for its work. The automatic system plays a key role in the following three main issues: the simplification of the oilfield surface engineering, improving information level and reducing labor intensity. But with the extension of the IoT in the production of oil and gas of CNPC (PetroChina Co., Ltd.) and the increasing demand of the fine management, many problems are exposed, such as information island, the unstandardized data, inconsistent process [1, 2]. It is difficult to realize the deep application in the oilfield business and coordinate the production.

With the application of the IoT technology, especially with the a large number of the sensors are used in the oilfield, a lot of the semi-structured and unstructured data was produced[3, 4]. But the original setup which uses relational databases exists triggered many insurmountable problems, such as mass data storage, access and maintenance, the management between the semi-structured data and unstructured data, the easy use, and the expansibility scalability of the databases[5].
NoSQL

NoSQL is a new mode of data storage and queries relative to traditional relational databases[5]. Using the Cloud storage platform, NoSQL solves the deficiency of the relational databases in the mass data storage and provides a different mechanisms in the data storage and queries [6]. Main advantages of NoSQL are simply design, distributed storage, high expansibility, mass storage, quick queries and better usage. Relative to the table model of the relational databases, the manifestations in the data structures in NoSQL are column, document, key-value, graph, etc. NoSQL is better for the data management of the weak structured data, the semi-structured data and unstructured data, such as document, graph, JSON [7].

MongoDB Introduction

MongoDB is one of NoSQL technologies based on the distributed file storage, and it’s programmed in C++, C and JavaScript [8]. Compared with other storages, MongoDB has advantages, such as high performance, easy deployment, easy use, easy data storage and management.

MongoDB is a production between relational databases and NoSQL, which is the most similar with the relational databases within NoSQLs. MongoDB uses the set concept set instead of the tables, supports loose dynamic structure through documents. MongoDB has strong data model performance ability. It supports the Boole, digital, string, data, object basic data types (booleans, strings, ...) and special data types, such as embedded documents, object ids, binary data, and code.

The query language of MongoDB is powerful, it can almost instead of the single table query in the relational databases and can exchange the atomicity of the field. The index mechanism is rich. It can support all index, composite index, TTL index, building index in the embedded field and array. The query optimizer of MongoDB is combined with index and can generate the most efficient query plan. Beside the index rule of MongoDB is basically as same as the relational databases. The most optimized skills of the relational databases also can be used in MongoDB.

MongoDB supports master/slave model and back-up copy model, the two models can be exchanged depending on the actual demand which is used in data back-up, disaster recovery and guarantee the high use of the cluster.

CLOUD STORAGE PLATFORM

The construction of Cloud storage Platform Based on MongoDB in the oilfield production mainly includes the data synchronization software from the monitoring center to the dispatching center, real-time databases based on MongoDB, the primary production databases based on Oracle, time synchronization service, other applications and so on.

The Cloud platform servers in the dispatching center use open-source, stable and safe CentOS6 operating system [9, 10]. The virtual machines are installed and configured and the applications are divided legitimately [11, 12]. All the working of the application platforms are supported. See Figure 1.
All the data of the oil-wells, water-wells, stations and pipelines is acquired, stored and displayed to the SCADA systems. The data is transmitted from all the real-time databases to MongoDB cluster at the specified frequency by the data synchronization software. All the real-time production data is stored in MongoDB cluster. The primary database store the basic information, time data. The NTP Server provided the time synchronization service for all the servers and SCADA systems. The data from the data unified access interface is back in the API way. The production command system, production operation management and other professional analysis software are provided with the data service.

MongoDB cluster is composed of 6 servers including 3 data slicing storage servers and 3 configuration and routing servers. Every data slicing storage server is designed 3 slicing to store data. Every copy of the slicing configuration is deployed in the different servers to make sure accessibility. Each configuration and routing server is configured the mongos service to response return data. To improve reading rate, index in the time field of relative tables is created. If the load of the cluster increase, the new slicing can be added to balance the load. See Figure 2.
THE APPLICATIONS OF THE CLOUD STORAGE PLATFORM

The big data analysis helps the organization to work better, based the data service of the production command system and back-end data. Through the construction of the working flow, every department can be effectively linked, auto-circulated, uploaded information and published data timely. The office efficiency and agility of management can also be improved. The information is more valuable. See Figure 3.
Oil/Water Well On-line Analysis Module

Oil/water well on-line analysis module: the intelligent analysis system as the key function is one of the most important manifestations of the IoT oilfield. By reading the data from the cloud storage platform, many analysis functions are realized.

(1) One well one scheme
Through building the “one well one scheme” early-warning module, the boundary data of every well is set according to the working conditions for preset and diagnosis model. The down-hole dynamometer card from the surface dynamometer card is derived. The fault diagnosis of the well, early-warning and alarm is realized by using pattern recognition technology. See Figure 4.

![Figure 4. One well one scheme.](image)

(2) Trend prediction
Based on the big-data, the future changing trend is predicted with the history variation law. Based on the comprehensive analysis model of the business logic, the advanced early-warning can be realized.

(3) Data mining
By using big-data technology, wrong data can be distinguished automatically. The exact result of the data analysis can be checked out with the algorithms of Shingling and Simhash. All the wrong dynamometer card, wrong data and missing acquisition can be recognized.

(4) Show the recessive knowledge, automate the explicit knowledge
By setting the factor weights, the experience of the professional technical personnel can be changed to the predicting and early-warming model. The information of the balance degree and system efficiency can be displayed to provide the whole reference information.

3D-simulation Model

3D-simulation monitoring software is popular recent years[13]. The problems in the original SCADA and DCS are the following:

a) The monitoring systems are independent, low integrations and lacks of connection.

b) Every system is used 2-dimensional plane, no vertical. The real status of the pipe and station can not be reacted directly. The workers have to know all the production flow of the station.
With the sensors in the spot and the Cloud storage server, the real working conditions of the station are displayed by using three-dimensional animation (3DMax). The function of the around view and scene walkthrough, the situation of the station can be learned. The oil-zone, water-zone, gas-zone and the position of the different devices can be displayed directly and clearly by the zone browsing. All the flows and ranges of the oil, gas and water can be displayed by the technological process. The production flow, the operational principles and inside structures of important devices can be showed in detail by the production training. The underground pipeline network can be displayed the device static data, monitored the real-time data, configured the production alarm parameters. The alarm linkage, risk warming and escape line can be showed with this software. See Figure 5.

![Figure 5. 3D-simulation.](image)

**Real-time Monitoring**

With the real-time data which is from the cloud storage platform [14], the data of all the SCADA can be displayed in the plant monitoring center and published on the web. Using Component Art UI and High chart, different ways of display are realized, such as GIS-Simulation, pie, line chart; display of the real-time data in all region and the history data analysis of oil/water wells [15].

**Report Module**

Thanks to the real-time databases and history databases, electronic reports can be obtained from the network. The oil and gas wells production data management system (A2) production report is presented automatically, and the report data can be stored uniformly. The work efficiency is improved, the query is used easily, the oil well production state can be mastered and analyzed. This technology is first time used in the automatically generating the oil and gas wells production data management system(A2) production report.
CONCLUSIONS

In this paper, we report a cloud storage platform based on MongoDB. Then we describe how apply this platform is applied to the oil/gas production in Huabei oilfield. The data and information collected from different SCADA systems are stored uniformly. Thanks to on this platform, the data can be shared extensively. The uniform storage, usage and management of data make it easily to properly realize, among others, the oil/water well on-line analysis, 3D-simulation, real-time monitoring and, reporting and etc. Finally, this solution makes easy to implement the integration and optimization of the production engineering.

REFERENCES

1. Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari , and Moussa Ayyash. 2015. “Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications,” J. IEEE Communications Surveys & Tutorials, 17(4): 2347-2376.
2. Qingchen Zhang, Chunsheng Zhu, Laurence T. Yang, ZhiKui Chen, Liang Zhao, and Peng Li. 2017. “An Incremental CFS Algorithm for Clustering Large Data in Industrial Internet of Things,” J. IEEE Transactions on Industrial Informatics, 13(3): 1193-1201.
3. Charith Perera, Rajiv Ranjan, Lizhe Wang, Samee U. Khan, and Albert Y. Zomaya. 2015. “Big Data Privacy in the Internet of Things Era,” J. IT Professional, 17(3): 32-39.
4. Qingchen Zhang, Laurence T. Yang, ZhiKui Chen, and Peng Li. 2018. “High-order Possibilistic C-Means Algorithms Based on Tensor Decompositions for Big Data in IoT,” Information Fusion, 39: 72-80.
5. Stonebraker M. 2010. “SQL Databases vs NoSQL Databases,” J. Communication of the ACM, 53(4): 10-11.
6. Qingchen Zhang, ZhiKui Chen, Ailing Lv, Liang Zhao, Fangyi Liu, and Jian Zou. 2013. “A Universal Storage Architecture for Big Data in Cloud Environment,” C. IEEE International Conference on Internet of Things: 476-480.
7. Aviv Ron, Alexandra Shulman-Peleg, and Anton Puzanov. 2016. “Analysis and Mitigation of NoSQL Injections,” J. IEEE Security & Privacy, 14(2): 30-39.
8. Yong-Shin Kang, Il-Ha Park, Jongtae Rhee, and Yong-Han Lee. 2016. “MongoDB-Based Repository Design for IoT-Generated RFID/Sensor Big Data,” J. IEEE Sensors Journal, 16(2): 485-497.
9. Qingchen Zhang, ZhiKui Chen, and Laurence T. Yang. 2015. “A Node Scheduling Model Based on Markov Chain Prediction for Big Data,” J. International Journal of Communication Systems, 28(9): 1610-1619.
10. Qi Zhang, Lu Cheng, and Raouf Boutaba. 2010. “Cloud Computing: State-of-the-art and Research Challenges,” J. Journal of Internet Services and Applications, 1(1): 7-18.
11. Qingchen Zhang, ZhiKui Chen, and Yonglin Leng. 2015. “Distributed Fuzzy C-Means Algorithms for Big Sensor Data Based on Cloud Computing,” J. International Journal of Sensor Networks, 18(1-2): 32-39.
12. Qingchen Zhang, Laurence T. Yang, ZhiKui Chen, Peng Li, and M. Jamal Deen 2017. “Privacy-preserving Double-Projection Deep Computation Model with Crowdsourcing on Cloud for Big Data Feature Learning,” J. IEEE Internet of Things Journal, DOI: 10.1109/JIOT.2017.2732735.
13. Yanming Li, and Chengliang Liu. 2012. “Integrating Field Data and 3D Simulation for Tower Crane Activity Monitoring and Alarming” J. Automation in Construction, 27: 111-119.
14. Amr Osman, Mohamed El-RefaeY, and Ayman ElNaggar. 2013. “Towards Real-Time Analytics in the Cloud,” C. IEEE Ninth World Congress on Services: 428-435.
15. Qingmin Li, and Minglei. 2007. “Simulation Method for the Applications of Ferromagnetic Materials in Suppressing High-Frequency Transients Within GIS,” J. IEEE Transactions on Power Delivery, 22(3): 1628-1632.