Advanced Engineering Forum *Online: 2012-09-26*
ISSN: 2234-991X, Vols. 6-7, pp 1036-1040
doi:10.4028/www.scientific.net/AEF.6-7.1036

Research and Application of Massive Data Processing in Oil Services

LI Baoan

Computer School, Beijing Information Science and Technology University, Beijing, China 100101

liba@bistu.edu.cn

Keywords: Massive Data Processing, Oil Services, Cloud Computing, IOT (Internet of Things)

Abstract. Big data problem has caused widespread concern from industry to academia in recent years. As the amount of data produced by various industries and sectors of rapid growth, increasing demands on data processing and analysis capabilities, how to face the challenges of data, discover new opportunities, the issue has received wide attention. As a traditional industry, the oil drilling or refinery enterprise is facing the operational status of the system to produce large amounts of data. This text introduced an approach to massive data processing for oil enterprise based on cloud computing and Internet of Things.

Introduction

To this day, it has no doubt about the coming of the era of big data, especially in the Internet, telecommunications, finance, and so on, almost to a "Data is the Business Itself". New Internet technology, networking, social networks bring convenience to people at the same time, also had a large amount of data. How to effectively store and query the data, how to get useful information from the huge amounts of data through data mining or other methods, for a good user experience, enhance competitiveness of enterprise, all these problems have brought many challenges to oil drilling or refinery enterprises. IDC research shows that, the digital realm already exceeded 1.8 trillion GB data, and enterprise data is to increase year by year at a rate of 55%. Now, just two days can create from data generated since the dawn of civilization to the 2003 total [1]. Big data has become an important characteristic of the times.

Data analysis will be perceptual judgment into quantitative analysis, in the hope of improving customer experience play an important role. The data is also the best standard to measure performance [2]. As cloud computing and cloud storage promotion, more and more data can be collected and used. The problem faced by the enterprise is no longer the lack of data, but how to the appearance of data through, the meaning of the analysis. Data can be success in the work to bring benefits only through conscious processing and analysis. The increased amount of data for oil enterprise provides accurate grasp the user group and individual network behavior model foundation. If the data can be made full use, it can provide personalized, precision and intelligent and production operating and personality service than existing old production form more cost-effective new business model. At the same time, oil drilling or refinery enterprise can also through the grasp of the data, looking for more and better can increase oil profit, lower operating cost ways and means. In the oil systems inside, traditional business models are individual business system, between each other is not shared storage and content. As cloud computing, the introduction of the business system gradually formed shared resource pool mode, the requirements of data storage, processing and the unity of the show. The data fusion brought before independent system can't show some information. Now we introduced an approach to massive data processing based on cloud computing and Internet of Things.

The Key Problems in Big Data Processing

Data Storage. Data is divided into structured data and unstructured data. Many data are belong to repeat storage data. How to reduce the number of the invalid copies of the system? How to effectively use the data compression technology to reduce the space and the processing time? How to ensure the

security of data storage? How to aim at the different types of data, to get high efficiency and convenient targeted store?

Data Inquires. It is not enough only store data, it is more important to use data. High performance inquire of data is one of the basic demand. The different use of inquires will make a lot of changes about the data analysis and business scope, so there must have suitable work tools. In many cases, SQL queries could support the rapid inquiry, but the traditional support SQL relational database existing spreading problem, make similar Hadoop data tools cannot provide the index finished rapid inquiry (Hadoop is a suitable tool for some cases need deep analysis of inquires) [3]. Most of the time, we need to encapsulate data query tools, for business personnel to provide Web pages which are easy to use.

We developed the distributed data mining analysis engine, distributed search engine and the corresponding cloud service components which can make the custom query massive data request for SaaS (Software as a Service) way, thus provide the services, effective processing needs for structured and unstructured data queries.

Real-time Data Processing. At present the data processing models are mostly batch processing. Considering the business efficiency cannot be influenced, data analysis/processing most in the midnight, the next day to see results, but this way cannot meet the needs of the business. For data real-time demand is largely the needs of the development of the business, not only business manager hope to see real-time business running situation, but also the users don't want to wait until the next day to enjoy services. The development of mobile Internet of your past midnight is no longer its business, and is likely to be the peak of the business. Therefore, Internet needs to provide services in 24 hours. Real-time processing becomes the key of fast uninterrupted processing in business system requirements. Not only that, real-time data processing was also able to minimize "batch waterfall", or even completely ruled out [4]. The fault on night of the operation causes bottlenecks, in no one found in time of treatment can lead to a more serious delay or even accidents. Real-time process in the first time to find the system problems, and to solve them in accordance with established strategies.

Mass Data Analysis. Mass data processing is extremely complex, the main problems include: (1) the effective mass of data cleaning and import. Data may appear in any of the exceptions, in the treatment of the former must clean, or is likely to cause the system in dealing with half collapse. Large amounts of data processing all have time limits, data achieve to TB levels must be parallel processing, or cannot complete file transfer; (2) when single machine can't meet the needs of computing power and storage capacity requirement, must implement the distributed processing. Parallel data processing algorithm have more differences from serial algorithm, programming ability to have higher demand; (3) online analysis (online analytical processing, OLAP) and batch data calculation is auxiliary of the relationship between each other, data and the complex relationship, generate reports could take several hours, and import data warehouse, using OLAP multidimensional analysis may be a few minutes to get results; (4) need to use a sampling in less loss to sharply reduce processing precision case data; (5) to enhance processing speed, should optimize the hardware.

The Effective Management of the Data. Since the data has been in a core position, many businesses started with data as the center, to re-examine the business system, hoping to get the benefits of large data. But big data are not throwing into the warehouse; instead need more fine management means, so that to be able to operate data effectively. The concrete measures include: (1) considering large data security; (2) to reconsider data interpretation, analysis and prediction ability; (3) establish a data driven business work mode, will change from leadership as the guidance to data as the guidance; (4) to solve the contradiction of data and process, will make the apart of process and data; (5) the construction of business from the centre with application to the centre with data. The enterprise' upward expansion ability is very important facing different database and analyzing environment. Hadoop is quickly adopted in enterprises by the reason which has the easy function to outside expansion. The key is to use low cost server cluster for large-scale parallel processing; it requires less professional skills than other data management mode, so that to reduce more personnel requirements and can realize more economic smooth expansion.

A Solution of Refinery Production Scheduling System Based on Cloud Computing and IOT

The Information Flow on Scheduling Business. Take advantage of production management information system, the petrochemical enterprises scheduling work-related data, analysis, storage, statistics, publications, and other various management links organically together [5]. Systems to enterprise scheduling business as the core, in accordance with the information flow in a production run of logistics processes, design their systems function modules, enabling enterprises to the logistics base in each production unit for real-time monitoring, scheduling, reached the flexible, real-time regulation to optimize production purposes. The information flow of production scheduling business shows in Fig. 1.

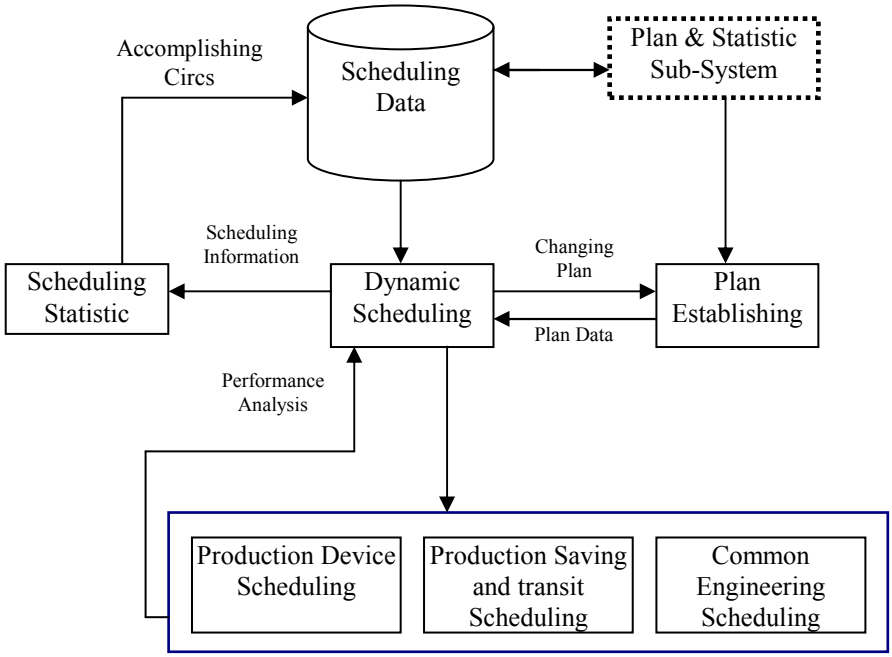

Figure 1 The Information Flow of Production Scheduling Business

The Solution of Production Scheduling Management System Based on Cloud Computing and IOT. The production scheduling management system is a very important business in oil production enterprise. The material balance, consonance plan, emergency management, production cost management are the core services of production scheduling management. The production data coming from each measurement point in DCS, PLC, and DDZ digital instruments are inputted into the scheduling system in the real-time foundation database. The data information from the real-time foundation database can be directly used in the dynamic scheduling management module to handle a variety of business; also available to extract data from a relational database. The specific forms of schedule data reports, documents and information can be gotten through the relation configuration [6].

The related system interfaces with other system have been predesigned, and the component libraries such as foundation component library and oil enterprise professional component library have been used. It speeds up the realizing of enterprise management system and makes it easier for users in system combination and system expansion. It improves system configuration flexibility. Adhering to the overall planning, step-by-step implementation of principles, and the methodology based on SOA (Service Oriented Architecture) and component oriented technology [7,8], it bring the enterprise's information work to save money and time, maximize to meet user's requirements, to reduce costs, optimize production and increase their work efficiency.

Fig. 2 shows the solution of production scheduling management system based on cloud computing and IOT. It is mainly divided into four parts, like as Data Acquisition, Data Storage & Data Transformation, Data Analyzing & Processing, and Data Services.

Data Acquisition offers the origin data sources. It mainly came from DCS, PLC, TGS, DDZ, LIMS and other sensors or electronic instruments. The non-structured data from Internet or other channel are another important data sources. The data from other ERP systems are the third kind of sources.

We used distributed relation database and real-time database to realize the production Data Storage and Data Transformation.

Data Analyzing & Processing is the most important part of massive data processing. It was consisted of Data Analyzing and Processing algorithms, Distributed Search Engine, Distributed Data Mining Analysis Engine, and Cloud Service Components.

Data Services offer various personalized data services such as Dynamic Scheduling, Integration Query, Quality Management, Measure Management, Oil Storage &Transit Management, Workshop Management, and other customers or enterprise services.

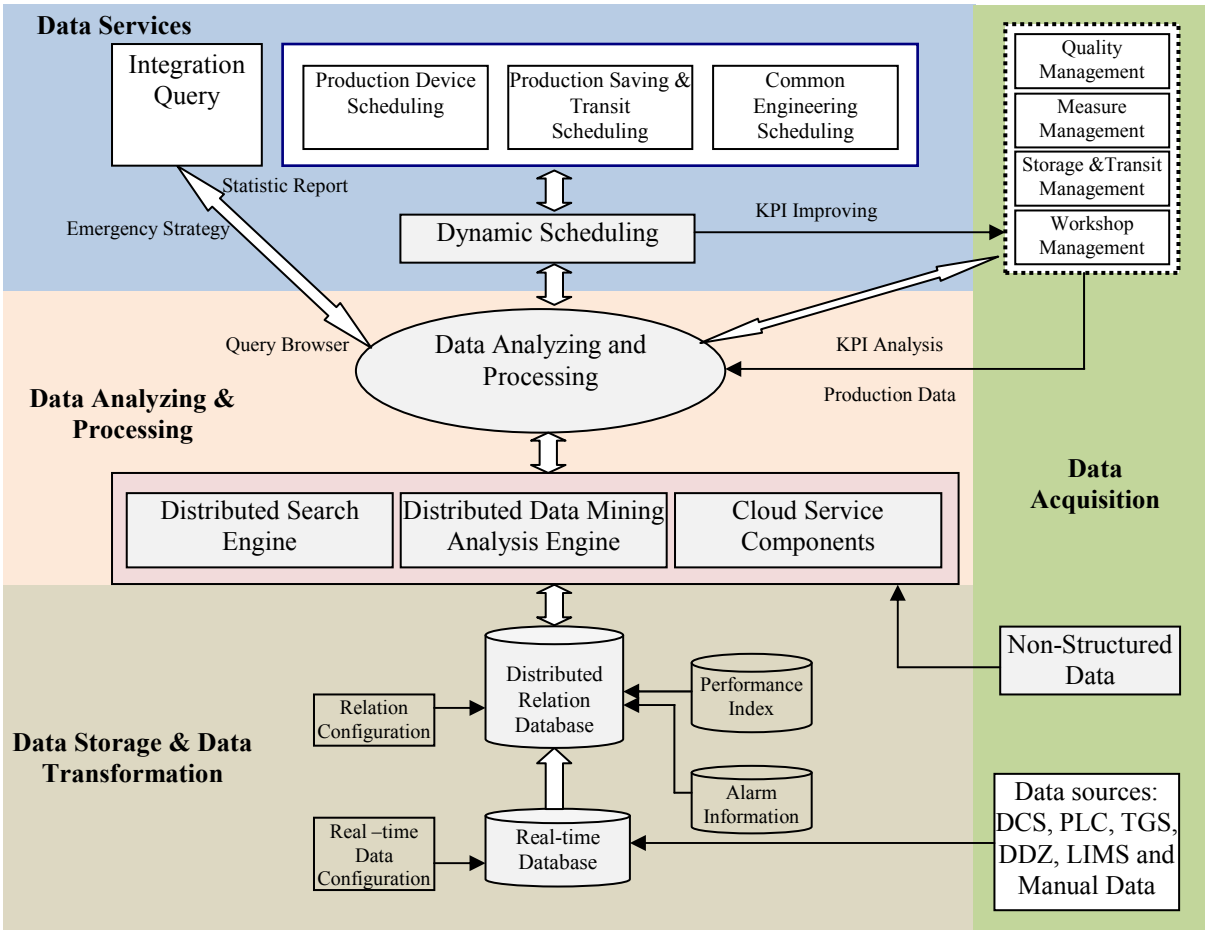

Figure 2 The Solution of Production Scheduling System Based on Cloud Computing and IOT

Conclusions

For the business of oil industry, the main task is to use the data center for the enterprise service. As the oil industry competition pressure increases, oil enterprise in the face of the competition is not only counterparts, and on the Internet industry competition [9]. This paper discusses how to effectively use the business system which produced a large number of data to improve oil the competition ability of the enterprise, and gave a typical big data application in enterprise production dynamic scheduling. Big data processing in the application of the oil industry has just started at present, there are also the many content needs further research [10]. We expect academia and industry collaborating closely and further promoting the use of data efficiency.

Acknowledgement

The work was supported by Graduate Teaching Reform Project of Beijing Information Science and Technology University (Grant No. YKJ201210), Funding Project for Academic Human Resources Development in Institutions of Higher Learning under the Jurisdiction of Beijing Municipality (Grant No. PHR201007131), and Opening Project of Beijing Key Laboratory of Internet Culture and Digital Dissemination Research (Grant No. 5026035410).

References

[1] Yuan Xiaoru, Yang Zhenkun, Data-Intensive Business, Communications of the CCF. vol.8, no.6, pp.6-7.

[2] Qian Yuming, Ding Yan, Fen Jun, Application of Massive Data Technologies in Telecommunication Services, Communications of the CCF. vol.8, no.6, pp.13-16.

[3] Wang Junsheng, Shi Yunmei, Zhang Yangsen, Research on Key Technology of Distributed Search Engine Based on Hadoop, Journal of Beijing Information Science and Technology University, vol.26, no.4, pp.25-28 (2011)

[4] Liu Jianbin, Li Jianzhong, Research on the Improvement of a Duplicate Code Detection Technology, Journal of Beijing Information Science and Technology University, vol.24, no.3, pp.44-49 (2009)

[5] Li Baoan, Zhang Wei, End-to-End Resources Planning Based on Internet of Service. Springer: Lecture Notes in Computer Science, v 6988 LNCS, PART 2, pp.19-26 (2011)

[6] Li Baoan, Research on the Production Scheduling Management System Based on SOA. Springer: Lecture Notes in Computer Science-Web Information Systems and Mining, vol.6318 LNCS, pp.286-294 (2010)

[7] Ron Barack, etc, SCA Java Component Implementation Specification, http://www.osoa.org/display/Main/Service+Component+Architecture+Specifications (2007)

[8] Matthew Adams, et,. Service Data Objects For Java Specification, http://www.osoa.org/display/Main/Service+Data+Objects+Specifications (2006)

[9] Richard Welke, Rudy Hirschheim, Andrew Schwarz. Service Oriented Architecture Maturity. IEEE Computer, no.2, pp.61-67 (2011)

[10] Yanbo Han, Xiaofei Xu and Keqing He: Service Computing for the Future Internet. Journal of Communication of China Computer Federation, vol.6, no.9, pp.10-11 (2010)