

Framework and key technologies for big data based on manufacturing

Shan Ren^{1, a}, Xin Zhao^{2, b}

¹ Engineering College, Honghe University, Mengzi, Yunnan, China

² College of Life Science and Technology, Honghe University, Mengzi, Yunnan, China

^a282952654@qq.com, ^b353435088@qq.com

Keywords: manufacturing, big data, framework, key technologies

Abstract. Data type and amount in human society is growing in amazing speed which is caused by emerging new services such as cloud computing, internet of things (IoT) and social network, the era of big data has come. Data is a fundamental resource from simple dealing object. In order to fully understand the connotation of big data, this paper expounds the conception of big data, combines with the application requirements of manufacturing big data, the five layer stack type big data processing framework is proposed, and the key technologies of manufacturing big data applications are discussed and analyzed.

Introduction

Along with the rapid development of mobile Internet, internet of things and cloud computing, the prelude of the mobile cloud is opened. Data is continuous increase and accumulate at an unprecedented speed, and the era of big data has come. Big data has been already caused people's extensive concern in economic, academia, manufacturing, medical, agriculture and even government.

Nature [1] published a special issue on "Big Data". Given the unprecedented amount of data that has been produced, collected, and stored in the coming years, one of the technology industry's great challenges is how to benefit from it. At the same time introduced a series of technical problems and challenges arise with large-scale datasets from the aspects of Internet, high performance computing, environmental sciences and biological medicine. Science [2] published a special issue on "Dealing with data". Its core idea is effective organization and utilization of the big data and further making the science and technology promote the development of social. Davos World Economic Forum [3] issue a report related to big data named "Big Data, Big Impact: New possibilities for international development". This report discussed under the new way of data that has been produced, how to produce better social benefits by using the amount of data. The reports focus on the integration and utilization of personal mobile devices data with other types of data, meanwhile, also gave a detailed analysis of the key technologies and application areas of big data. The "Global Pulse" [4] initiative of United Nations published a report about "Big Data for development: Challenges & Opportunities". Report mainly explained the huge opportunities as well as challenges in an era of data deluge to the countries of the world especially to the developing countries. McKinsey [5] published article "How big data can improve manufacturing". In-depth analysis of the issues in how big data and advanced analytics to make biopharmaceuticals, chemical industry and discrete manufacturing more rationalization. Discussed on the problem of how the manufacturers based on the process that taking advantage of advanced analytics can reduce process flaws, saving time and money.

The development of big data has received worldwide attention. How to effectively manage, analysis and use the heterogeneous big data, and mine the potential and valuable knowledge of the big data, become a common concern topic of the whole world. It is also the most important significance of big data research.

Concept of big data

Big data is an abstract conception and not yet have a generally definition. Different definition of big data often illustrate and inductive from its characteristics. There are several representative definitions of big data.

3V definition [6]: three characteristics of big data are volume, variety and velocity. Volume relates to size of the data such as terabytes (TB), petabytes (PB), zettabytes (ZB), etc. Variety means the types of data. In addition, difference sources will produce big data such as sensors, devices, social networks, the web, mobile phones, etc. For example, data could be web logs, RFID sensor readings, unstructured social networking data, streamed video and audio. Velocity means how frequently the data is generated. For example, every second, minute, hour, day, week, month and year. Processing frequency may also differ from the user requirements. Some data need to be processed real-time and some may only be processed when needed.

4V definition: On the basis of 3V added a new feature. International Data Corporation (IDC) regard as big data should have value [7] and IBM consider big data must have veracity [8].

Big data from Wikipedia [9]: Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time.

Framework of four layer stack type data processing based on product lifecycle

In the process of manufacturing enterprise informationization, more and more sensing devices are used in production lines and products, and then a large amount of data are monitored and captured during production and usage process. These data can be sent to the service platform in real time by mobile network, and stored in the enterprise database. Through analysing and mining these data, the potential and valuable information and knowledge can be discovered. The information and knowledge are useful for product design, production, sale, usage, maintenance, recycling and even the whole lifecycle decision-making.

Based on the above consideration, a framework of five layer stack type data processing based on product lifecycle is proposed as seen in Fig. 1.

Foundation Layer. It is responsible for real-time monitor the lifecycle data (e.g. engineering design, process design, machining, assembly, maintenance, testing and after-sales service), and transmit these data to upper layer. At the same time, the multi-source and heterogeneous data of product lifecycle should be identified and integrated into standard event log format, and providing original and useable data for data mining and analysis. Through wireless sensor network which composed with radio frequency identification, GPS, sensors and distributed servers, the lifecycle data of products are sent to the foundation layer, and stored and preprocessed in there. A highly automated, scale-out storage and compute platform should be established to realize big data application. Cloud model can be used to store and aggregate the big data and resource. In the cloud model, resource allocation is achieved by multiple cloud data node, which makes the data closer to the users. Therefore, the response time of the system is shortened and the productivity of the enterprise is increased.

Management layer. To support the deep analysis of multi-source and heterogeneous big data, a management platform of big data should be established. By this platform, the structured data and un-structured data of lifecycle can be unified a standard schema. Then, meet the requirements of real-time data transmission, query, calculation, large-scale data statistics and update, and provide technical support for lifecycle management. Parallel and distributed data management is the function of the big data management layer. NoSQL, Hadoop, MPP and Esper are used to manage the lifecycle big data.

Analysis Layer. In order to apply the big data, big data analysis must be done. Data mining and machine learning algorithms and models are referred to analyze and interpret big data, and help customers to acquire valuable data information. The advanced big data analysis theories and methods are cluster, association rules, regression, classification and prediction, etc.

Application layer. This layer aims to provide decision-making for enterprises and provide query, analysis, statistics, forecast and other services for end-users. From the point of whole lifecycle of products, the big data analysis results can be used to optimize system operation parameters, estimate cost, maintenance service management, service-oriented manufacturing and predictive maintenance, etc. Finally, these results can provide data support and knowledge support for product innovation design that based on big data.

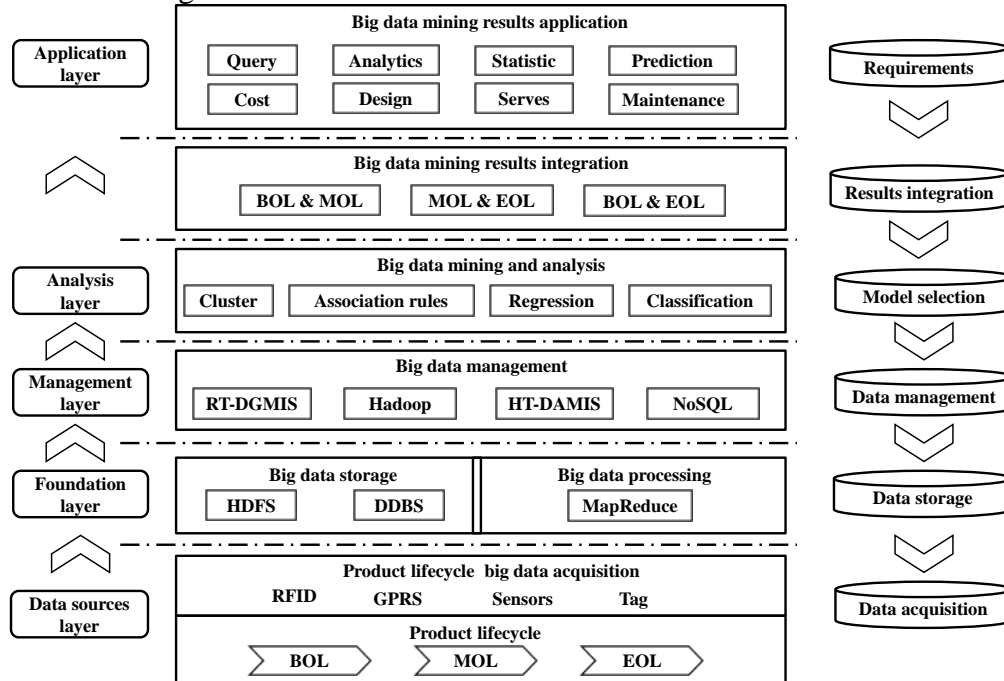


Fig.1 Framework of four layers stack type manufacturing big data processing based on product lifecycle

Key technologies of big data processing based on product lifecycle

Distributed computing framework MapReduce. MapReduce programming model is one of the most typical batch processing technologies, and is widely used in the field of data mining, data analysis, machine learning, etc. Mapreduce is running in the Key/Value pairs. Map function, written by the user, takes an input pair and produces a set of intermediate key/value pairs. The reduce function, also written by the user, accepts an intermediate key and a set of values for that key, and synthesize the results as a final output. The overall flow of a MapReduce operation is seen in literature [10].

Distributed data base system (DDDBS). Structured data is the most important data in product lifecycle. At present, more and more structured data are collected in enterprise database, as time goes on, two problems are appeared: (1) historical data and current data are stored in the same database that affected the data processing performance of system; (2) cannot solve the problem of data value-add of historical data. Aiming at these problems, two kinds of methods are proposed: data generation management information system based on real-time and data application management information system based on historical time (HT-DAMIS). RT-DGMIS is used to manage the data that change frequently [11]. However, HT-DAMIS is responsible for accumulation of the data that is produced from RT-DGMIS so that easily to query, analysis and mining historical data. DDDBS should be used for managing structured big data of MMP. The objective is to rapidly query historical data, to update real-time data, and to conduct data analysis and data mining. The architecture of DDDBS is shown in Fig. 2.

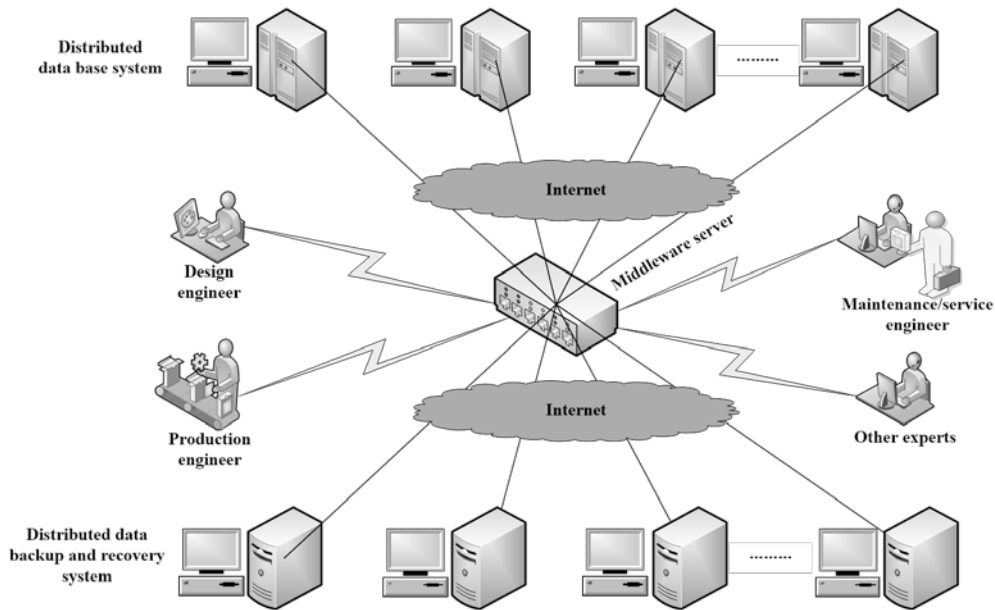


Fig.2 Architecture of distributed data base system

Hadoop distributed file system (HDFS). Un-structured was the data without format, spatial and temporal constraints, and a great amount of un-structured data are produced and stored by different lifecycle actors and application systems. Recently, more and more manufacturing enterprises pay more attention to the problem of statistical analysis and data mining for un-structured data. Because of relational database management system (RDBMS) cannot meet the applications demand of big data in flexibility and high scalability (especially the scale out), it is unsuitable to management un-structured data. Same as structured big data, distributed approach are used for managing and storing MMP un-structured big data, such as Hadoop Distributed File System (HDFS) and not only Structured Query Language (NoSQL) data management system. They store data on multiple storage nodes, which unified management and distribution resources, and provision of file system of access interface for customers. The architecture of HDFS is shown in Fig. 3.

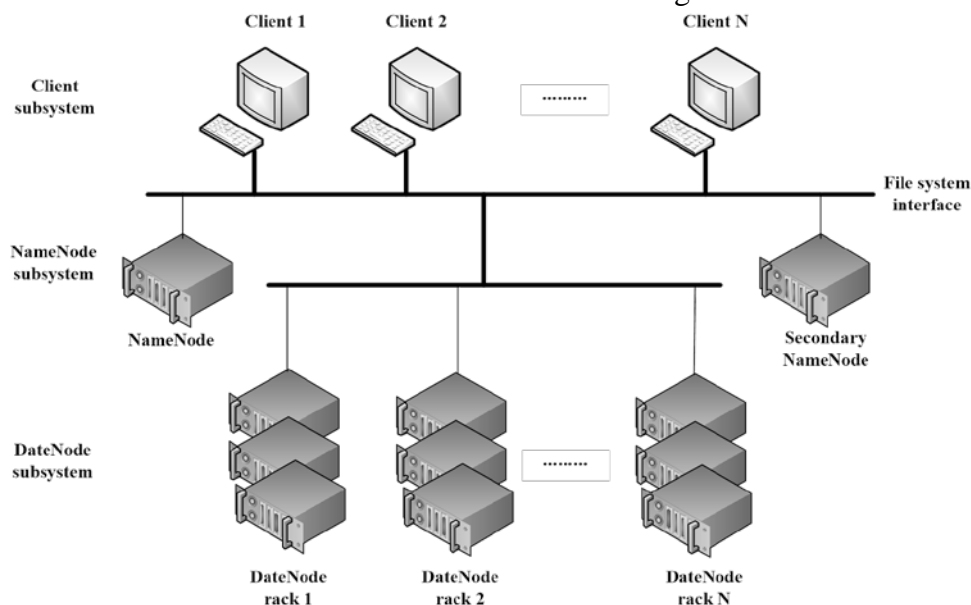


Fig.3 Architecture of hadoop distributed file system

Big data visualization technology. The graphical way is easier to understand and accept by user than pure text. Due to visualization technology can process and display the analysis results as an intuitive method, and it was widely used as the most effective means of explanation of big data. Tag Cloud, History Flow [12], Chart API and Data Driven Documents (D3) are the commonly used technology of visualization.

Conclusions

With the rapid development of cloud computing and internet of things, we are facing the explosive growth of data in the new century, and the era of big data has come. In this paper, a simple big data related research achievement in recent years are reviewed and summarized. Firstly, the various conceptions of big data are introduced. Secondly, combining the product lifecycle with big data technologies, the framework of five layer stack type big data processing is proposed, and the main functions of each layer are expounded. Finally, the key technologies of big data processing are discussed and analyzed.

Acknowledgement

Authors would like to acknowledge financial supports of Science and Technology Department Project of Yunnan Province in China (2013 FD 049), and the program of Hong he University (XJ15Y20).

References

- [1] Nature. Big Data. Information on <http://www.nature.com/nwes/specials/bigdata/index.html>
- [2] Science. Special online collection: Dealing with data. Information on <http://www.sciencemag.org/site/special/data/>
- [3] World Economic Forum. Big data, big impact: New possibilities for international development. Information on http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf
- [4] UN Global Pulse. Big Data for Development: Challenges & Opportunities. Information on <http://www.unglobalpulse.org/projects/BigDataforDevelopment>
- [5] McKinsey & Company, Information on http://www.mckinsey.com/insights/operations/how_big_data_can_improve_manufacturing
- [6] Grobelnik, M. Big-data computing: Creating revolutionary breakthroughs in commerce, science, and society. Information on http://videlectures.net/eswc2012_grobelnik_big_data/
- [7] Barwick, H. The “four Vs” of Big Data. Implementing Information Infrastructure Symposium. Information on http://www.computerworld.com.au/article/396198/iiis_four_vs_big_data/
- [8] IBM. What is big data? Information on <http://www-01.ibm.com/software/data/bigdata/>
- [9] Big data. Information on http://en.wikipedia.org/wiki/Big_data
- [10] Dean J, Ghemawat S. MapReduce: Simplified data processing on large clusters[C] //Proc of OSDI, Berkeley, CA: USENIX Association, (2004), p.137-150
- [11] Yang Julong. Big data technology solutions: foundation, design, development and practice. (Publishing house of electronics industry, China 2014)
- [12] Meng Xiaofeng and Ci Xiang. Big data management: conceptions, techniques and challenges. Journal of Computer Research and Development. Vol. 50(2013), p.146-169