

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Knowledge Management Based on Big Data Processing

Li Baoan

Beijing Key Laboratory of Internet Culture and Digital Dissemination Research,
Beijing Information Science and Technology University, Beijing, China

Abstract: At present, many large enterprises, like oil industry accumulated a large amount of data with a range of potential value of knowledge in their value activities over the years. How to help them to put these data into wealth are common problems faced by IT industry and academia. This study analyzed the five key problems of big data processing and knowledge management in-depth and then explained the composition and technical characteristics of knowledge management system based on big data processing. It explored the new approach of knowledge management which can adapt to the ever-change demands of enterprises.

Key words: Big data, knowledge management, cloud computing, IOT (Internet of Things)

INTRODUCTION

Big data is becoming more and more recognized and chase by the people, this is because the big data have broad research and application prospect. A survey shows that only 2009 years, Google through the big data business's contribution to the national economy is for 54 billion dollars and it is just great data contains huge economic benefit of the tip of the iceberg. In May 2011, McKinsey company issued a big data research report, "the big data: the next frontier, competitiveness, innovation and productivity", this study fully illustrates the big data research status and will give the society brings the value (Zhou *et al.*, 2012). Big data research is the social development and technological progress of urgent need.

Oil industry is an old traditional one which has distributed and numerous oil fields and subordinate enterprises. It has created enormous amount of wealth, in meanwhile and produced more and more enterprise's data. However, as an important resource, the oil exploration and drilling is becoming more and more difficult, the knowledge hide in the big data of enterprise is urgently in need of discovering and utilizing. At the era of big data, almost to a "Data is the Wealth Itself".

As cloud computing and IOT (Internet of Things) 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 study 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. This study introduced an approach to knowledge management based on big data processing.

KEY PROBLEMS IN BIG DATA PROCESSING

Data storage: Data is divided into structured data and unstructured data. Many data are belonging 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? These problems are all must face and solve in the actual applications (Kanagal *et al.*, 2011).

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) (Wang *et al.*, 2011). Most of the time, it needs to encapsulate data query tools, for business personnel to provide Web pages which are easy to use.

It 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 h day⁻¹. 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 (Liu and Li, 2009). 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.

Big data analysis: Big data processing is extremely complex (Li and Zhang, 2011), 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.

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 study 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 (Baoan, 2013). 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.

RESEARCH AND APPLICATION ON KNOWLEDGE MANAGEMENT

Knowledge management system: Knowledge management system is to build the organization or enterprise core platform and provide the basis of the knowledge management of the three core functions: multidimensional perfect knowledge base, dynamic intuitive knowledge map and intelligent fast knowledge search. Also provides the knowledge answer, expert map, knowledge dissertation, personal knowledge zone, evaluation integral mechanism, etc. Various kinds of interactive applications, at the same time satisfy the organization internal dominant and recessive knowledge management needs.

Knowledge management system has the following technical features: standard open data interface and intellectual diversity compatible, agile and extensible development platform, open system integration interface and terminal support, multi-channel application, active personalization, etc.

Knowledge management system based on B/S (Browser/Server) is designed by three layer framework, SOA (Service Oriented Architecture). To follow the

standard XML data, using the component technology, distributed cluster deployment model, support large-scale application of knowledge management. Distributed deployment can be used in the knowledge management application, knowledge retrieval, knowledge mining module and load balancing according to the increase of data and the concurrent user access.

The composition and technical characteristics of knowledge management system was showed in Fig. 1.

Sources and types of knowledge: Knowledge management information can be from the Internet or enterprise Internet, from the enterprise's business application system (such as OA system, ERP system, etc.), also can be come from staff's desktop data, or from the proprietary electronic resources CD, periodicals and various printing material, etc. Some knowledge could be obtained by context-aware systems (Li, 2012a). The types of knowledge information include structured data which were estimated about 20% of the whole sources of knowledge and a large number of unstructured or semi-structured data which were estimated about 80% of the sources of knowledge.

Figure 2 showed the sources and types of knowledge in knowledge management system.

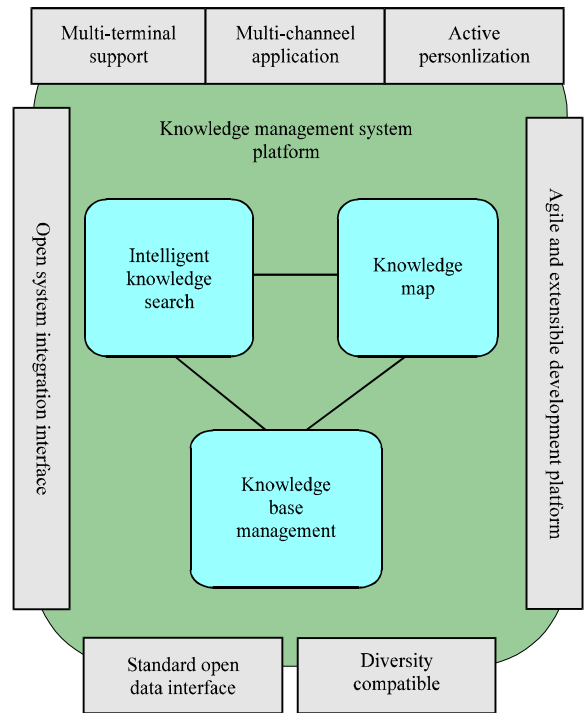


Fig. 1: Knowledge management system



Fig. 2: Sources and types of knowledge

CONCLUSION

It will produce a lot of valuable data in the enterprise long-term production and operation process. For these data collection, analysis and reuse, will effectively promote the enterprise management; make the business enterprise to obtain competitive advantages and continuous profitability. Currently, with the progress of IOT, cloud computing and big data processing technology, these will provide the possibility to make data into wealth for the enterprise. In this study, it makes the big data processing and knowledge management related issues on further research and gives an application of knowledge management to be used in oil or other data-intensive enterprise based on big data. Believe that with increasingly perfect and popularization of the research and application of big data processing technologies, it will be for the enterprise and the society to create more value and promote the information technology increasingly progress (Li, 2012b).

ACKNOWLEDGMENTS

The study was supported by the specialty Construction and Comprehensive Reform Pilot Project under Beijing Municipality (Grant No. 71M1310830), Funding Project for Natural Science Foundation of China (Grant No. 61070119) and the Project of Construction of Innovative Teams and Teacher Career Development for Universities and Colleges under Beijing Municipality (Grant No. IDHT20130519).

REFERENCES

- Baoan, L., 2013. Service oriented enterprise based on value-aware service engineering. *Inform. Technol. J.*, 12: 478-481.
- Kanagal, B., J. Li and A. Deshpande, 2011. Sensitivity analysis and explanations for robust query evaluation in probabilistic databases. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, June 2011, Greece, pp: 841-852.
- Li, B. and W. Zhang, 2011. End-to-end resources planning based on internet of service. *Proceedings of the International Conference on Web Information Systems and Mining*, September 24-25, 2011, Taiyuan, China, pp: 19-26.
- Li, B.A., 2012a. Research on context aware service based on IOT. *Int. J. Digital Content Technol. Appl.*, 6: 35-41.
- Li, G.J., 2012b. The scientific value of big data research. *Commun. China Comput. Fed.*, 8: 8-15.
- Liu, J.B. and J.Z. Li, 2009. Improvement of a duplicate code detection technology. *J. Beijing Inform. Sci. Technol. Univ.*, 24: 44-49.
- Wang, J.S., Y.M. Shi and Y.S. Zhang, 2011. Key technology of distributed search engine based on Hadoop. *J. Beijing Inform. Sci. Technol. Univ.*, 26: 25-28.
- Zhou, X.F., J.H. Lu, C.P. Li and X.Y. Du, 2012. Big data challenge from the data management perspective. *Commun. China Comput. Fed.*, 8: 16-20.