

Research of Big Data Based on the Views of Technology and Application

Zan Mo, Yanfei Li

School of Management, Guangdong University of Technology, Guangzhou, China
Email: 649319529@qq.com

Received 1 April 2015; accepted 19 April 2015; published 22 April 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In the era of big data, large amounts of data affect our work, life and study, even national economic development. It provides a new way of thinking and approaches to analyze and solve problems, which gradually becomes a hot research. Based on describing the concept and characteristics of big data, this paper describes the development of technologies in big data analysis and storage and analyses the trends and different values in commercial applications, manufacturing, biomedical science and other applications. At last, the authors sum up the existent challenges of big data applications and put forward the view that we should deal with big data challenges correctly.

Keywords

Big Data, Big Data Technology, Application, Data Analysis

1. Introduction

Before big data appear, database has become an important processing platform because of the data processing convenience. But when database is faced with non-relational or large-scale data, there is a difficulty dealing with them. Big data not only enhance the related computing services technologies but also change the traditional mode of many industries. The latest report released by Markets and Markets shows that [1], from 2013 to 2018, the annual compound growth rate of the global market for big data will be 26 percent, from \$14.87 billion in 2013 to \$46.34 billion.

Big data are the hottest words in the IT industry, followed by data warehouse, data analysis and data mining. The commercial value of the using big data gradually becomes the focus profits of different professionals. Big data help people acquire knowledge from the massive, complex data, and become another focus after integrated circuit and Internet information technology. IBM, Amazon, Microsoft and other large companies are constantly committed to develop and utilize big data, triggering the development boom of big data.

2. Development of Big Data

2.1. Definitions and Characteristics

The original concept of the idea of big data is from the world of computer science and econometrics [2]. McKinsey & Company is the first company to refer big data. In June 2011, McKinsey issued a report on “big data”, which carried out a detailed analysis of the impact, key technologies and application. From that on, big data caused different industries’ concerns.

There are many kinds of definitions of big data. Wikipedia points out that 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 [3]. McKinsey believes that big data is one data set whose size exceeds the typical database software acquisition and storage, management and analysis. In Victor Meyer-Schonberg’s “BIG DATA” [4], big data means using the method of all the data but not random analysis (sampling). IDC (International Data Corporation) is defined as to meet 4V (Variety, Velocity, Volume, Value) index called big data.

The characteristics of big data are submitted by Victor Meyer-Schonberg in “BIG DATA”. There are four characteristics, including volume, velocity, variety, value [4].

Volume refers to the huge amount of data. With the development of data storage and network technology, data storage expands from TB to ZB. Only in 2011, 1.8ZB (1.8 trillion GB) of data are created. Warehouse management server will increase 10-fold to 50-fold to cater to the growth of big data. Velocity refers to the mobility of data streams. It is difficult to deal with data in a traditional way because data run fast. Through cloud computing, it can achieve fast data processing. Variety refers to relational and non-relational data generated by a variety of ways. With the development of mobile networks, people are more widespread to use real-time data. The quantity of semi-structured and non-relational data is also increasing. Value reflects the value of the significance of big data applications, which has a scarcity value, uncertainty and diversity.

All in all, although the definition of big data contains different concerns and technologies, but there is a consensus point, big data refers not only too large amounts of data, but also including a large amount of data processing techniques. “4V” characteristics show a large number of data. Volume, velocity and variety are aim to realize the value of big data. Data collection, storage, analyze, is prepare for dig out the value of data. Big data emphasizes complexity in data analysis, and it pays more attention to data processing efficiency and the data value.

2.2. Development Trend

From an economic development perspective, many large companies focus on big data seriously. IDC’s report claimed that global data will increase by 50 times over the next decade, shown in [Figure 1](#). Oracle President, Mark Hurd, said that now it’s the era of big data explosion, and data grew at an alarming rate. At present, the amount of data around the world is million trillion. Data increase 8 times from 2005 to 2011. In 2020, the expected amount of data could reach 35 million trillion.

The development trend of big data published in “2012 Hadoop and Big Data Technology Conference” showed the top three topics are: data resources, big data privacy issues and integration of big data and cloud computing. The magazine editor of *Wired*, Chris, has asserted that the data have made the traditional scientific method obsolete. Although this statement is a bit extreme, but big data indeed has changed our lives, our way of thinking. Big Data is widely used. Now many large companies use big data to streamline processes and create efficiencies, such as Microsoft, Apple, Oracle, Amazon, Google, FaceBook and Twitter. They are experienced in dealing with big data sets [5].

3. Technologies

Big Data provides a new method to traditional data analysis, which has a variety of technologies, including Hadoop and MapReduce, cloud computing, grid computing and so on. This paper sorts out the following technologies.

3.1. Hadoop and MapReduce

In the related technologies, more representative one is Hadoop, which is represented by non-relational data

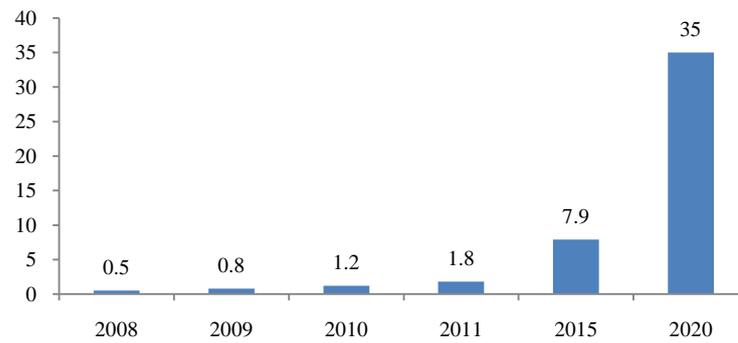


Figure 1. The forecast of global data growth (unit: ZB).

analysis techniques. By the virtue of processing for non-structural, massively parallel processing, easy using and other advantages, Hadoop becomes a mainstream technology. MapReduce is a model proposed for parallel processing and generating big data by Google in 2004 [6], which is a linear, scalable programming model. Hadoop is an open source realization of MapReduce. With its open source and easy using, Hadoop has become the first choice for big data processing. It not only create targeted marketing applications, make full use of transaction data, but also improve accuracy and timeliness of fraud detection. Many Internet companies, including Facebook, Google, eBay and Yahoo, have developed a large scale applications based on Hadoop. MapReduce and Hadoop can significantly improve the efficiency of big data processing.

3.2. Big Data Acquisition Engine

In addition to the requirements of efficiency and speed, big data collection also requires security. A general data acquisition engine which combines rule engine and finite state automaton together, helps to verify the security and correctness of the big data acquisition flow [7]. When adding a new collection node, the rule engine will automatically make the whole system more flexible and scalable. At the same time, it ensures the state transition, and improves safety and clear logic. Big data acquisition, integrated with JESS rule engine, not only can control the state transitions and match, but also to monitor the unusual status and location errors. Rules engine can clearly show the errors and details which are matching wrong, ensure the safety and accuracy of the data acquisition.

3.3. MFA (Mean Field Analysis)

Big data processing system requires some related components to use in parallel multiple instances of the same task, so as to achieve the desired level of performance applications. In order to enable administrators and developers to maintain the growth rate of the data, these systems' reliability assessment is critical. A set of methods for approximate inference of probabilistic models, based on MFA, can solve the performance evaluation system problem of big data [8]. Through behavioral modeling to assess the performance of data structure, MFA can calculate the related basic performance in a limited time. In addition, MFA can set up and evaluate in a shorter time, because it does not depend on the number of instances. In the process of assessing the performance of big data, MFA technology is very effective.

3.4. Other Technologies

In addition to the above-mentioned techniques, M2M (Machine To Machine) technology is an important one. M2M platform can expand the number of data producers and data consumers flexibility, accomplish new services in a very short period of time, re-use and combine data from different sources. Existing studies have shown that automatically creating M2M decision support system has much room for development [9]. There are also grid computing, cloud computing and other technologies in big data analysis and processing. Big data technology is not a single technology, but mix with a variety of other techniques, so as to play the biggest role in the storage and analysis.

4. Applications

The arrival of big data change many applications, including business, traditional manufacturing, biomedical field and other applications. Big data brings opportunities to the enterprises. Previously untapped data resources can be stored and processed. The new data collection techniques and advanced data mining tools provide an unprecedented opportunity. This paper analyzed from business applications, manufacturing, biomedical industries and other industries.

4.1. Business Applications

Business studies show that timely and effective use of data-driven knowledge is a competitive advantage. Combined with cloud computing, using LDA to extract themes can provide usefulness for unstructured data, can help companies to export the competitive advantage [10]. Through a lot of structured and unstructured data, enterprise use big data intelligent analysis technology to identify fraud risk, trends and patterns. Big Data can not only help companies solve problems, but also to prevent crime. One example is the Griffins companies [11]. The fraud detection and prevention solutions of big data left traces of interaction, improved data visualization, and increased opportunities to identify fraudulent activity.

4.2. Manufacturing

As a traditional industry, manufacturing is also under attack due to the advent of the big data. Big data may push the next revolution in manufacturing—forecast manufacturing. In order to become more competitive, manufacturers need to accept emerging technologies, such as advanced analysis and physical network [12], to improve their efficiency and productivity based on a systematic approach. Big data can help reduce defects and control costs during automated production. By tracking into every detail of the product for each part, from manufacturer to store installation, data allows manufacturers to track for better solutions. Monitoring defect rates and on-time delivery can also help suppliers to select and evaluate the performance. Tools which are being developed to process and manage big data generated by sensors and other equipment will change the product invention, manufacturing, transportation and services from a fundamental way [13].

4.3. Biomedical Industry

Big Data is changing the biomedical industry with bringing benefits for human being. According to the McKinsey Global Institute, effective use of big data will help the US health care sector save \$300 billion per year in savings, reduce spending by 8% [14]. Big Data analysis can be applied to echocardiography, angiography, and magnetic resonance imaging or computed tomography to form cardiac imaging [15]. Big data imaging may also provide new insights about the disease, treatment and interventions. Due to the large amount of new data generation techniques and computing power of “big data” bioinformatics analysis, immune genomics provides new ways to understand disease etiology, immune function and regulation, as well as a more comprehensive knowledge of genetic variation [16]. Over time, the development of biomedical technology is changing the personal health, so that patients can get more beneficial control for themselves.

4.4. Other Applications

In addition to the above-mentioned applications, big data is also available in the study of history and geography. Historians have been making sense of reams of data for centuries. Now computational tools, along with a proliferation of digital source materials [17], are opening up new ways of understanding history. In geography, visual analysis system—Exploratory Data Analysis Environment [18], a software applied to analyze complex data sets of Earth system simulation, presents a visual interactive analysis tool that can be used to transform data into perspective, thereby enhancing the earth system processes.

Overall, in business applications, big data are changing the way of exporting competitive companies' advantage, identifying fraud risks and analyzing trends and patterns. In traditional manufacturing industries, big data can control the quality of products, reduce defects and product costs and track every aspect of the product. In the biomedical industry, the development of big data changes the way of treatment, so that more patients can get a better solution. In the study of history and geography, big data also appeared in the historical archives research and geography climate research. With further research and the attention of big data, big data will become in-

creasingly widespread.

5. Conclusions

In the past decade, the development of the Internet has produced a lot of data. History shows that the availability of new and timely data does not always bring expected results. It also has shortcomings, such as data privacy, lack of professional, data leakage of personal information and other challenges.

The biggest challenge of big data is data privacy. It issues across the entire life cycle of big data: collection, combination, analysis and use. Development of big data puts forward a series of challenges that requires careful balance between threats and opportunities [19]. In this regard, the organization should classify information and control it, such as formulate specified retention period according to the provisions they face. Second challenge is shortage of professional. With the participation of big data market, more and more enterprises have an important opportunity to involve in big data. But the companies lack sufficient qualified staff and necessary analysis and research skills [20]. Therefore, big data are not only a technical phenomenon. It may also bring significant change. Companies need to face the risk correctly so as to meet the challenges of big data.

Based on the definition and characteristics of big data, this paper concludes the development trend of big data. As a hot research, the company has interests in it and gets some results gradually. The analysis of big data processing technologies is also deeper. People use Hadoop and MapReduce technology to enjoy the big data applications in commercial, manufacturing, biomedical industry and other aspects of the benefits. At the same time, there are also some challenges like data privacy, lack of professional and data leakage. Despite the great convenience to people's lives brought by big data, it is a double-edged sword yet. If the way of data management doesn't change, it will bring great challenges. Therefore, business managers need to carefully comprehend and correctly respond to the challenges.

Fund

This study is funded by National Natural Science Foundation of China: 71171062.

References

- [1] Sina Technology, IT Industry (2015) <http://tech.sina.com.cn/it/2013-09-02/11088699338.shtml>
- [2] Church, A.H. and Dutta, S. (2013) The Promise of Big Data for OD: Old Wine in New Bottles or the Next Generation of Data-Driven Methods for Change. *OD Practitioner*, **45**, 23-31.
- [3] Wikipedia (2015) Big Data. <http://en.wikipedia.org/wiki/Bigdata>
- [4] Mayer-Schönberger, V. and Cukier, K. (2013) Big Data: A Revolution That Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt, Boston.
- [5] Gobble, M.M. (2013) Big Data: The Next Big Thing in Innovation. *Research-Technology Management*, **56**, 64-67. <http://dx.doi.org/10.5437/08956308X5601005>
- [6] Dean, J. and Ghemawat, S. (2008) MapReduce: Simplified Data Processing on Large Clusters. *Communications of the ACM*, **51**, 107-113. <http://dx.doi.org/10.1145/1327452.1327492>
- [7] Xu, X.B., Yang, Z.Q., Xiu, J.P. and Chen, L.I.U. (2013) A Big Data Acquisition Engine Based on Rule Engine. *The Journal of China Universities of Posts and Telecommunications*, **20**, 45-49. [http://dx.doi.org/10.1016/S1005-8885\(13\)60250-2](http://dx.doi.org/10.1016/S1005-8885(13)60250-2)
- [8] Castiglione, A., Gribaudo, M., Iacono, M. and Palmieri, F. (2014) Exploiting Mean Field Analysis to Model Performances of Big Data Architectures. *Future Generation Computer Systems*, **37**, 203-211. <http://dx.doi.org/10.1016/j.future.2013.07.016>
- [9] Renu, R.S., Mocko, G. and Koneru, A. (2013) Use of Big Data and Knowledge Discovery to Create Data Backbones for Decision Support Systems. *Procedia Computer Science*, **20**, 446-453. <http://dx.doi.org/10.1016/j.procs.2013.09.301>
- [10] Ribarsky, W., Wang, D.X. and Dou, W. (2014) Social Media Analytics for Competitive Advantage. *Computers & Graphics*, **38**, 328-331. <http://dx.doi.org/10.1016/j.cag.2013.11.003>
- [11] Hipgrave, S. (2013) Smarter Fraud Investigations with Big Data Analytics. *Network Security*, **2013**, 7-9. [http://dx.doi.org/10.1016/S1353-4858\(13\)70135-1](http://dx.doi.org/10.1016/S1353-4858(13)70135-1)
- [12] Lee, J., Lapira, E., Bagheri, B. and Kao, H.A. (2013) Recent Advances and Trends in Predictive Manufacturing Systems in Big Data Environment. *Manufacturing Letters*, **1**, 38-41. <http://dx.doi.org/10.1016/j.mfglet.2013.09.005>

-
- [13] Noor, A. (2013) Putting Big Data to Work. *Mechanical Engineering*, **135**, 32-37.
- [14] O'Driscoll, A., Daugelaite, J. and Sleator, R.D. (2013) 'Big Data', Hadoop and Cloud Computing in Genomics. *Journal of Biomedical Informatics*, **46**, 774-781. <http://dx.doi.org/10.1016/j.jbi.2013.07.001>
- [15] Narula, J. (2013) Are We Up to Speed? From Big Data to Rich Insights in CV Imaging for a Hyperconnected World. *JACC: Cardiovascular Imaging*, **6**, 1222-1224. <http://dx.doi.org/10.1016/j.jcmg.2013.09.007>
- [16] Mack, S.J. (2013) Human Immunology in the Era of Big Data. *Human Immunology*, **75**, 2-3. <http://dx.doi.org/10.1016/j.humimm.2013.12.002>
- [17] Hoffmann, L. (2013) Looking Back at Big Data. *Communications of the ACM*, **56**, 21-23. <http://dx.doi.org/10.1145/2436256.2436263>
- [18] Steed, C.A., Ricciuto, D.M., Shipman, G., Smith, B., Thornton, P.E., Wang, D., Williams, D.N., *et al.* (2013) Big Data Visual Analytics for Exploratory Earth System Simulation Analysis. *Computers & Geosciences*, **61**, 71-82. <http://dx.doi.org/10.1016/j.cageo.2013.07.025>
- [19] Cumbley, R. and Church, P. (2013) Is "Big Data" Creepy? *Computer Law & Security Review*, **29**, 601-609. <http://dx.doi.org/10.1016/j.clsr.2013.07.007>
- [20] Nunan, D. and Di Domenico, M. (2013) Market Research and the Ethics of Big Data. *International Journal of Market Research*, **55**, 505-520. <http://dx.doi.org/10.2501/IJMR-2013-015>