Big Data: What's the Big Deal?
Deloitte Technology, Media, Telecom Center of Excellence

Deloitte Technology, Media, Telecom Center of Excellence (TCOE) is the first TMT Center of Excellence established in China by an international renowned accounting and consulting firm. TCOE is also Deloitte's third regional TMT Center of Excellence following North America and Europe.

TCOE's vision and objectives are: leverage Deloitte's years of TMT service experiences, act in global perspective but at local basis to set up a think tank and research team. TCOE consists of strong global industry experts, research consultants and professional local research groups in China. TCOE will provide proactive thought leadership to service providers, upstream and downstream enterprises, industry authorities and regulators in China TMT industry.
Contents

I. Big data drives big changes ......................................................... 1

II. The transition stages of big data application ......................... 3

III. Who will lead big data? ....................................................... 7

IV. The bottleneck on future big data development .................. 10

V. Authors ........................................................................ 12

VI. Contact details for Deloitte China Practice ....................... 13
I. Big data drives big changes

Big data is undoubtedly the most visible technology affecting the future development of diverse industries. Since 2011, an increasing number of business managers have realized that big data is an unavoidable issue for future development. Relating researches in 2009 were very limited across the globe, however, by year-end of 2012, about 90% of the Fortune Global 500 companies had implemented big data projects.¹ Market research firm IDC released a research that shows the big data market is expected to grow to USD16.9 billion in 2015.² The business data of all companies at present doubles in every 1.2 year.³ The explosive data information keeps reminding us that the future will be greatly changed by big data.

Why does big data come into the focus of all people? What kind of fundamental changes will big data bring about? With these questions, Deloitte interviewed professor Du Xiaoyong, Big Data Expert Committee Member of China Computer Federation and Dean of the School of Information, Renmin University of China.

In his opinion, there will be three fundamental changes by big data: first, with the help of big data, people can get rid of the dependence on algorithm and models because the data itself can help people to get close to the truth; secondly, big data weakens the causal relationship. As big data can dig out the relevance among various factors, we don’t have to know why we can utilize the results of the factors just because they are related. In a modern society of intricate information, the above application will significantly enhance the efficiency; thirdly, compared with previous database technologies, big data is able to handle semi-structured and non-structured data, which rapidly expanses the data scope that can be analysed by computers.

Professor Du further elaborates his points with some examples:

First, computer science was highly dependent on models and algorithms before the emergence of big data. If you want to obtain a precise conclusion, you have to develop models to describe issues, arrange the logic, understand the cause and effect, and design sophisticated algorithms to output conclusions closer to the reality. Therefore, the best solution of a problem depends on whether the modelling is rational, in which the trade-off among varied algorithms becomes the key to success.

However, the emerged big data completely changed people’s dependence on modeling and algorithms. For instance, assuming that there are two alternative algorithms, respectively A and B. When running in small amount of data, algorithm A produces more favorable results than B. In other words, algorithm A, for its part, can output better results. But people find that as data amount increases, the results of algorithm B running in large data are superior to that produced by algorithm A running in small amount data. The finding offers landmark enlightenments for computer science and computer’s derivative disciplines: when data gets bigger and bigger, the data itself (not the algorithms and models used by data research) ensures the effectiveness of data analysis results. Even without the precise algorithm, a conclusion that close to the reality can be produced as long as the date is sufficient enough. Thus, data is hailed as a new kind of productivity.

Secondly, if data is sufficient enough, we can arrive at a conclusion without understanding the specific causality. For instance, when Google helping users translate texts, it does not set up various grammar and

¹ Source: Smartplanet, 21 February 2012
³ Source: http://knowwpcarey.com/article.cfm?cid=25&aid=1171
translation rules, but make use of the word-using habit of all users collected in Google’s database to compare and recommend. Google will check up the writing habits of all its users, and then recommend the most common and frequently used translations to users. In this process, the computer may not know the problem’s logic, but as the data that records user behaviour increases, the computer will be able to provide the most reliable results without knowing the problem’s logic. Mass data and pertinent analytical tools have provided a complete new access for understanding the world.

Thirdly, big data can maximize the use of human behavior data recorded through internet to make analysis, thanks to its ability to handle multiple data structures. Before the emergence of big data, all data processed by computers shall be structured and recorded in corresponding database. But big data greatly reduces the requirements on data structure. Multi-dimensional information left by people on internet, ranging from social network, geographic location, behavior habits and preferences etc. will be processed on a real-time basis, outlining the rich and complete characteristics of an individual.

Figure 1: Disparity between traditional data and big data

<table>
<thead>
<tr>
<th>Traditional data</th>
<th>Big data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000 MB~ GB</td>
<td>PB ( Petabytes ) ~EB ( 1,000PB )</td>
</tr>
<tr>
<td>Integrated</td>
<td>Distributed</td>
</tr>
<tr>
<td>Structured</td>
<td>Semi-structured or non-structured data</td>
</tr>
<tr>
<td>Have complicated known correlations</td>
<td>Correlations known are little and rare</td>
</tr>
</tbody>
</table>
The substantial changes brought by big data unlimitedly extended the breadth and depth of management. Business executives began to consider how to improve management efficiency with big data. Although the concept of big data was widely known in recent two years, its commercial application mode has started to upgrade on a gradual basis. Deloitte believes that in the last two years, the application of big data has experienced the uplift from logical judgement to systematic thinking.

At the early stage of application, big data was mainly used to make simple logic judgment. For instance, when many people, of one accord, searched the keyword "influenza virus" on Google, the large amount of region-concentrated searches made Google a platform forecasting the spread of influenza virus ahead of many disease control and prevention centres. This, however, is just the most initial exploration of big data.

Deloitte analysed the application mode of big data and found out that the application of big data has switched from the original stage of logical thinking to the stage of systematic thinking. Meanwhile, big data experienced three transitions in problem-solving strategies:

The first is the transition from simple correlation identification to systematic survival relationship. At the first stage of application, big data served as a useful tool for analysing data correlations. When many relations in reality interweave together, big data can explore the possible identified relationship in chaos and indicate a direction for managing such chaos. While at the second stage of application, simple correlation identification is just the beginning, and the key to success is making use of these relations with the thinking of systematic survival.

For example, when searching micro-blog posts about the movie So Young, Letv stumbles across that the number of posts discussing the movie Tiny Times is eight times more than that of the former. This clearly indicates Tiny Times' box-office appeal. Letv decisively chose to be the marketing agent of Tiny Times. However, the direct deduction drawn from data is the basic application of big data. High attention degree on micro-blogs is a guarantee for the investors' returns. Of course, whether a film can attract more audience is influenced by complex factors, including competitiveness of other films of the same release period, the marketing efforts and promotion approaches by different media, and even the weather and school holidays, etc.

The on-going evolution not only enables big data to find out the correlations with existing factors, but also to integrate them and perceive such correlations among essential factors with the thinking of systematic development.

The second is the transition from problem-finding to interactive growth. At the first stage of application, big data is used to identify problems through deep data mining to the record and description of a variety of variables in a complex system and their relations, so as to improve the problem-solving efficiency. However, at
the second stage, the application focuses more on the deep cognition and understanding of users, through which, products that truly meet users' needs are created and grow together with users.

Case 1: In-depth analysis on user demand by big data made the House of Cards a success before its birth

Netflix, the world’s largest online paid video and flat rate DVD-by-mail service provider, utilizes big data to analyse the users’ habits. By using search techniques, Netflix compared and observed users’ watching habits and found out a totally unrelated coincidence: the audience who like watching House of Cards produced by BBC in 1990 are the fans of the famous director David Finch and Oscar-winner Kevin Spacey. Netflix believed that a TV drama combining such three elements may greatly increase the probability of success. Hence, Netflix invited David Finch to reproduce House of Cards and Kevin Spacey to star in this teleplay. Netflix spent USD200 million to subscribe two seasons of new House of Cards before the release of any trailers or sample clips.

Also, as House of Cards is online broadcasted, by using powerful database monitoring system, Netflix is easy to identify where the users press pause key, how many users stop watching it after several chapters and how many users playback and replay it. These precise data analyses can provide guidance for producing sequels.

In early 2013, Ted Sarandos, CCO and Vice President of Netflix, said that the House of Cards originally produced by Netflix enjoys the highest audience in countries covered by Netflix services.

Before producing the TV drama House of Cards, Netflix analysed the interest and possible taste of the users who have common characteristics. Based on the predicated taste, Netflix tailored the TV production, which achieved great success. This production process fully explains how big data helps enterprises have an in-depth understanding of user needs through interacting with them and create perfect products that satisfy their needs.

The third is the transition from existing state analysis to future ecological reconstruction. At the first stage of application, big data is used for the static analysis of the current state of the users. For instance, the big data can record the behaviours, preferences, geographic location and other real-time information of the users. The enterprises may make use of such data to deliver more accurate promotion information at a proper time and location so as to better target the users. In contrast, at the second stage, the big data helps enterprises reconstruct the user needs and create new ecological environment, in which both enterprises and users may achieve organic growth together with a new business mode.
Case 2, PatientsLikeMe explores new diagnose and treatment models for chronic diseases with the help of big data

No one likes to post his most private information on Facebook or Foursquare, but PatientsLikeMe, a social network being funded by the Robert Wood Johnson Foundation with USD1.9 million, has become an exception.

PatientsLikeMe is an exclusive social network for patients with chronic diseases. On this open platform, patients can measure their treating progress and researchers can obtain the medical data. Up to date, PatientsLikeMe has nearly 200,000 patients on the platform building and sharing their medical records.

PatientsLikeMe Research and Development Director Paul Wicks said on the TED Conference: "The gap between diagnose and medical care is pending to be filled. Once you have the right measuring standards and approaches, you will do things amazing."

There are thousands of diseases but approached for patients to conduct self-measurement is so rare. Paul Wicks took an instance and explained that patients with multiple sclerosis can use a form consisting of 7 questions to conduct self-measurement. Up to date, 30,000 users of PatientsLikeMe have participated in filling the forms. Paul Wicks hopes to create similar standard self-measurement tools for other diseases.

David Knowles, a 59-year-old property manager who lives on St. Croix in the U.S. Virgin Islands, was searching online for information about a new treatment for multiple sclerosis when he came across PatientsLikeMe. Knowles, who has had MS for 10 years and had participated in several patient-community groups, was immediately intrigued. "There's no other site with data like PLM has," says Knowles. "You can click on a symptom and say, 'Well, there's 850 people with this symptom, and this is what they're using to treat it.'"

Knowles was particularly interested in a drug called Tysabri. One of his doctors had recommended it, but Knowles was concerned about negative side effects (brain infection, anxiety, and fatigue). On PatientsLikeMe, Knowles found data from hundreds of patients taking Tysabri. After reviewing their results, he decided the risks outweighed the rewards for him, and he went to his doctor with a list of other treatments he wanted to discuss instead. "I feel like I'm in charge of my medical care now," he says. "Of course, I still listen to my neurologists, but now it's more of a team approach." Knowles lauds PatientsLikeMe for its "wealth of information about treatments from people who are actually using them. And you can follow patient responses over time -- after three months, six months, a year."

PatientsLikeMe's profit model is as follows: upon the explicit authorization of users, the website will sell users' information to pharmaceutical companies which can leverage those huge and detailed records of user information to research the mechanism of drug action for different patients and acquire sufficient information for R&D of new drugs.

Source: Most Innovative Companies-2010
http://article.yeeyan.org/view/49269/99965
Through the in-depth analysis, PatientsLikeMe aggregates data of large amounts of users to provide valuable data to patients, medical care and pharmaceutical industry, which is reshaping the manner of seeing doctors, diagnose and treatment solutions, drug R&D and sales plan of pharmaceutical industry while benefiting all three parties. Big data is restructuring this eco-environment and has created new business models.

Figure 1: Transition of big data application stages - from logic judgment to systematic thinking
There is no doubt that big data will bring about more profound changes to our times. Both the old IT powers like IBM and CISCO and the up-and-coming IT players focusing on big data in the Hadoop ecosystem have occupied major sections on big data’s industry chain in just a few years. Who will lead big data technology in future? Can China’s manufacturers seize a place before the explosive growth of big data? Deloitte interviewed the big data expert John Ho, former senior vice president of Asialinfo and developed a corporate footprint diagram of the development of big data.

Chart 2: Corporate footprint diagram at different information processing sections of big data

Big Data can be divided into six sections by information processing: data acquisition, data cleansing, data storage and management, data analysis, data visualization, and industrial applications. Various companies have started to take up places at every section.
1. **Data acquisition**: Traditional IT companies like Google and CISCO have already started deploying data acquisition. In China, Taobao, Tencent, Baidu and other companies have collected and stored large amount of data on user habits and consumer behavior.

2. Deloitte anticipated that there would be more specialized data collection companies to design industry data acquisition systems to meet with the specific needs of various industries.

3. **Data cleansing**: How to filter out useful data, complete data cleansing and pass it to the next section after numerous unsystematic and disorderly data are collected is a process that becomes increasingly demanding along with the more specialized labor division of big data. In addition to the old IT powers like Intel, Teradata, Informatica and some other professional data processing companies have shown greater vitality. In China, Audaque and other similar manufacturers have begun to emerge. Deloitte anticipates that there will be a large number of companies specialized in data cleansing in future.

4. **Data storage and management**: Data storage and management are two segments of data processing, which are closely related. The data management approach determines data storage formats which further limit the depth and breadth of data analysis. Usually, it is more effective for one manufacturer to design these two segments due to high correlation. From the perspective of market occupation, the old data storage providers like IBM and Oracle have obvious established advantages. They have further developed their original storage business and easily occupied larger market share. Emerging companies like Apache Software Foundation bring together the wisdom of industry specialization via open-source strategy and become leaders of the development of big data.

5. **Data analysis**: Traditional data processing companies SAS and SPSS have obvious advantages regarding data analysis. However, Hadoop, a data analysis company based on open source software infrastructure has achieved explosive growth in recent years. For example, Cloudera which established in 2008 is committed to helping companies manage and analyze data based on open-source Hadoop products. With the ability to help customers complete customized data analysis, Cloudera has a large number of well-known business users, such as Expedia, JP Morgan Chase and other companies. It is estimated that its market value has reached 700 million $^5$ in just five years.

6. **Data interpretation**: Data results from big data analysis will be restored to specific industry issues. Data analysis companies like SAP and SAS added industry knowledge to their existing businesses and become leaders of this section. Meanwhile, the professional data restoration companies like wibidata emerging with the development of big data have begun to flourish.

7. **Data visualization**: Big data really starts to help manage practices in this section. Qualify and compute the conclusion derived from big data via analyzing and visualizing data and apply it to specific industry, which require the professionals with industry expertise to combine industry practice with the deduction from big data to develop solutions that can change the state quo of industry.

---

$^5$ Source: Most Innovative Companies- 2010

http://article.yeeyan.org/view/49269/99965
### Figure 2: Big data information processing sections and challenges

<table>
<thead>
<tr>
<th>Data acquisition</th>
<th>Company name</th>
<th>Content</th>
<th>Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taobao, Tencent, Baidu, Google, Cisco</td>
<td>Collect various user daily data on the network</td>
<td>How to become industry-specialized</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data cleansing</th>
<th>Company name</th>
<th>Content</th>
<th>Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teradata, Intel, Informatica, Audaque</td>
<td>Select and transmit useful data</td>
<td>Data selection model and transmission speed</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data storage and management</th>
<th>Company name</th>
<th>Content</th>
<th>Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM, EMC2, Oracle, Greenplum</td>
<td>Store and manage the collected data</td>
<td>Open source community and commercial big data vendors must develop management and analysis tools which are easy for traditional IT people and business people to operate</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data analysis</th>
<th>Company name</th>
<th>Content</th>
<th>Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAS, SPSS, Greenplum, Cloudera, The Apache Software Foundation</td>
<td>Conduct customized analysis on data</td>
<td>How to conduct customized analysis on data according to the needs of enterprises</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data visualization</th>
<th>Company name</th>
<th>Content</th>
<th>Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAP, SAS, wibidata, Microstrategy</td>
<td>Visualize the data from analysis and restore the industrial issues reflected in the big data conclusion</td>
<td>Analysts need to have capabilities in statistics, computer and math, as well as industry knowledge and business acumen, to combine the results of big data and business practice</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Industry application</th>
<th>Applied industry</th>
<th>Content</th>
<th>Challenge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business intelligence, public sector, marketing</td>
<td>Apply the resulting data that could reflect industry trends to the industry, and guide the development of larger consumer market</td>
<td>How to design the big data analysis results into an executable business implementation plan</td>
<td></td>
</tr>
</tbody>
</table>
IV. The bottleneck on future big data development

By analysing big data's industry chain, we find that major companies have started to occupy positions at each production section along the industry chain of big data. As high-performance computers and the storage and management process of mass data continue to improve, problems that can be solved by technologies won't be a problem in the end.

Deloitte thinks three sections really matter in restricting the development and application of big data:

The first is the trade-offs among the legitimacy of data acquisition & extraction, data privacy protection and application. The extraction of private data from a group of users by any company or organization should be known to the users, and business application of users' privacy data should be accepted by them. However, China or even the world at present is lagging behind the developing speed of big data in terms of a number of management issues, such as how to protect user privacy; how to develop business rules; how to punish the violation of user privacy and how to enact laws and regulations.

In Deloitte's perspective, many big data services in future may walk in the grey belt at their early developing stages. Once the business operation begins to take shape and impact large number of consumers and companies, can related laws and regulations as well as market norms be created on an accelerated basis.

It can be predicted that the application of big data is infinite at technical level; however, due to the constraints of data acquisition, the data available for serving people and business is far less than the amount big data can collect and process in theory. Constraints on data acquisition source will significantly restrict the business application of big data.

Secondly, giving play to the synergy of big data requires a balanced competition and cooperation among companies at varied sections of the industry chain. Big data requires the companies in its ecosystem to make more cooperation. Without a macro view of the whole industry chain, a singular company cannot understand the correlations among data at each industry chain section according to the separate data it holds. Therefore, it is limited for the company to estimate and affect consumers.

As for some industries e.g. banking and insurance, in which the information is apparently asymmetric, there is a strong demand for sharing data among companies. For instance, both of the two industries in general need to build up a shared database for the whole industry, enabling their members to learn about a single user's credit record, eliminate the information asymmetry between the underwriter and the consumer, and smooth the deal process. However, under many circumstance, the companies that need shared information are both competitors and co-operators to each other. Before sharing the data, a company has to make trade-offs so as to avoid losing its competitive strength when sharing such information. In addition, when many merchants collaborate together, they are likely to form a seller alliance that may damage consumers' interests and affect the fairness of competition.
The most imaginative developing direction for big data is to integrate data of various industries and provide an all-round data stereogram in order to understand and reshape user demands in a systematic perspective. However, the sharing of cross-industry data requires the trade-offs among interest relations of too many companies. If without independent third-party organizations to coordinate such relations among all participants and develop data sharing and application rules, the space for big data to come into play will be greatly restricted. The lack of authoritative and independent third-party organizations will be a major factor constraining the most potential of big data.

Thirdly, it is the interpretation and application of big data conclusions. Although big data can reveal the possible relevance among variables at data analytical level, how to reify such relevance into the industry practice? How to develop executable solutions to apply the conclusion of big data? Facing these questions, executors are required to be able to interpret big data and know the correlations among various industry development factors. This section is based on the development of big data, and also relates to other factors at management and execution level.

In this section, human factor is the key to success. In technical perspective, the executor shall be able to understand most of the data techniques, interpret the conclusion deduced by big data analysis; in industrial perspective, the executor shall be well-informed of the relations among various links of the production process, the potential relevance among factors, and can map big data conclusions onto specific execution; in management perspective, the executor shall develop executable solutions, make sure there is no conflicts between solutions and management procedures, and avoid making new problems while solving the old ones. Considering the conditions above, an executor is required to be an outstanding manager with systematic thinking and well versed of the technology, being able to perceive correlations between big data and the industry in an associated and complicated systematic perspective. The shortage of such talents will constrain the development of big data.
V. Authors

William Chou
Deloitte China TMT, Industry Leader
Deloitte China VC Program Leader
Venture Capital Association of IAC Vice Chairman
Email: wilchou@deloitte.com.cn

William Chou is the Managing Partner of Deloitte China TMT Industry, Education Industry and Deloitte China VC Program. He also serves as a mentor for the 7th AAMA Cradle Program.

William has more than 25 years of accounting, audit and consultancy experience. He is the leader of a cross-functional team which focuses on the market demands from industrial shifts. In addition, he is in charge of the research of Deloitte China’s operational strategy and the industry, as well as to provide professional services to clients. He participates in the business of Deloitte Global, having a seat in Deloitte Global TMT Execution Committee. He is also a member of Deloitte China Board of Directors.

Zhaoli Meng
Head of Deloitte China TMT Center of Excellence
Email: zhmeng@deloitte.com.cn

Dr. Zhaoli Meng graduated from the National University of Singapore and is the Head of Deloitte China TMT Center of Excellence. She has more than ten years of research experience in areas as Internet economy, electronic industry, new energy industry and corporate management. In recent years, she published dozens of research outputs on China Daily, Interfax, Harvard Business Review, PKU Business Review and CEIBS Business Review, etc. She is a co-author of Managerial Economics, which was selected by many reputed colleges as a MBA textbook. In addition, she is lecturing MBA courses in several universities, such as Shanghai International Studies University and Central University of Finance and Economics, etc. Before she joined Deloitte, Dr. Zhaoli Meng served as the Director of Research of the Accenture Center of Excellence in Greater China Region, the Lead Researcher of Samsung Economic Research Institute and the Assistant Professor in Renmin University of China.
Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee, and its network of member firms, each of which is a legally separate and independent entity. Please see www.deloitte.com/cn/en/about for a detailed description of the legal structure of Deloitte Touche Tohmatsu Limited and its member firms.

About Deloitte Global

Deloitte provides audit, tax, consulting, and financial advisory services to public and private clients spanning multiple industries. With a globally connected network of member firms in more than 150 countries, Deloitte brings world-class capabilities and high-quality service to clients, delivering the insights they need to address their most complex business challenges. Deloitte has in the region of 200,000 professionals, all committed to becoming the standard of excellence.

About Deloitte in Greater China

We are one of the leading professional services providers with 22 offices in Beijing, Hong Kong, Shanghai, Taipei, Chengdu, Chongqing, Dalian, Guangzhou, Hangzhou, Harbin, Hsinchu, Jinan, Kaohsiung, Macau, Nanjing, Shenzhen, Suzhou, Taichung, Tainan, Tianjin, Wuhan and Xiamen in Greater China. We have nearly 13,500 people working on a collaborative basis to serve clients, subject to local applicable laws.

About Deloitte China

The Deloitte brand first came to China in 1917 when a Deloitte office was opened in Shanghai. Now the Deloitte China network of firms, backed by the global Deloitte network, deliver a full range of audit, tax, consulting and financial advisory services to local, multinational and growth enterprise clients in China. We have considerable experience in China and have been a significant contributor to the development of China's accounting standards, taxation system and local professional accountants.

This publication contains general information only, and none of Deloitte Touche Tohmatsu Limited, its member firms, or their related entities (collectively the “Deloitte Network”) is by means of this publication, rendering professional advice or services. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser. No entity in the Deloitte Network shall be responsible for any loss whatsoever sustained by any person who relies on this publication.

©2013. For information, contact Deloitte Touche Tohmatsu Certified Public Accountants LLP.