Winter 12-6-2015

The Formation Of The Big Data Industry And Related Strategies

Shari S.C. Shang

Sheng-Chi Chen

THE FORMATION OF THE BIG DATA INDUSTRY AND RELATED STRATEGIES

Shari S. C. Shang, National Chengchi University, Taiwan, sshang@nccu.edu.tw
Sheng-Chi Chen, National Chengchi University, Taiwan, 102356503@nccu.edu.tw

ABSTRACT

Big Data, as moving into the post-cloud era in year 2015, is changing the way software is applied by industries. Meanwhile, the Open Data is transforming the sources of value for software development in which comprehensive digital data value creation is set to be the mainstream in IT application strategies. Big Data applications can enhance the knowledge level of technology applications as well as drive value growth for products and services. The key issue affecting Big Data development is the question of how to leverage mechanisms for joint creation by the general public to identify the value that Big Data can provide. This paper proposes that an effective Big Data industry ecosystem should encompass data owners, application innovators, technology leaders, and open API platforms. Using the methods of design science, the paper evaluates the way in which Big Data creates value in the industrial development process, and argues that it is important to attract mass participation and to satisfy the needs of both industrial and social development when putting forward a value proposition. Moving further, the next step is value creation initiatives, in which the first priority should be to promote the digitalization of industry seeking to create digitalized industries that can contribute to the gradual optimization of the industrial ecosystem as a whole.

Keywords: Big Data, Design science, Government policy, Industry ecosystem, Industry development.

INTRODUCTION

Disruptive innovation is transforming the world at a very rapid pace, and at the same time it is also changing consumer behavior; the continuing evolution of new types of information technology is changing how the "information society" operates. With the development of cloud computing applications, huge quantities of computing operations and data are converging in cloud-based data centers. As the volume of data in these cloud data centers builds up, the need for data analysis becomes ever more pressing. Industry has begun to recognize the potential business opportunities from data applications and the added value that they can create. Government agencies and business enterprises are both faced with the challenge of how to use technology effectively to extract from the data clues that can help in policy implementation or business development.

Big Data has yet to take the form of a fully-fledged industry, as such. The significance of Big Data, and its potential scope of use, is still being interpreted from the perspective of the individual enterprises and people that use it. Currently, no consensus has been reached regarding Big Data in academia. Most of the literature on Big Data restricts itself to examination of data mining technologies and applications, and some studies simply treat Big Data as being synonymous with data mining. [4] The aim of the present study is to help readers develop a clearer picture of the current state of development of Big Data, and of how it should be defined; the study analyses the roles played by individual actors in the Big Data ecosystem during this period prior to the formation of a "Big Data Industry." It is anticipated that this will help future researchers in this field to clarify potential issues, and will facilitate a more comprehensive exploration of the value generated by Big Data.

RELATED WORKS

Definition and Characteristics of Big Data

With demand for new types of application emerging all the time, data has come to be viewed as a commodity. Commercial needs have led business enterprises to start investing resources in finding data, and in using analytical technology to develop new products and services, thereby helping to support business decision-making and enhance operational efficiency, leading to the emergence of the "data economy." [8] The term "Big Data" is used to refer to large volumes of data, and more specifically the use of information technology to rapidly access, convert, process and analyze large quantities of data, so that data is transformed into useful information that can be utilized to understand phenomena, forecast trends and provide a reference for decision-making. [1] Market research firm IDC defines "Big Data" as "a new generation of technologies and architectures designed to extract value economically from very large volumes of a wide variety of data by enabling high-velocity capture, discovery and/or analysis." IDC anticipated strong growth in the global Big Data market over the period 2012-2016, forecasting a compound annual average growth rate (CAGR) for the global Big Data technologies and services market during this period of 31.7%, with total sales revenue forecast to reach US$23.8 billion by 2016. [6] International research and consulting firm Gartner defines "Big Data" as "high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision-making." The most widely adopted definitions of "Big Data" emphasize its "3V" characteristics: volume, velocity and variety. [1][2] Other definitions incorporate additional characteristics such as variability, veracity and value, etc. [10][14][16]

"Big Data" thus generally refers to taking quantities of data that are so large they could not be processed manually, and accessing, managing, processing and collating the data within a reasonable space of time so that it is transformed into information that can be interpreted by human users. However, some definitions emphasize not so much the sheer quantity of

data involved, but rather the question of whether or not value can be effectively extracted from the data. Overall, it is difficult to say precisely just how big data has to be in order to qualify as "Big Data." [7][10][12]

Government Policy

Currently, Big Data technology is evolving rapidly all over the world, and the speed at which innovative new applications are being developed is picking up. Governments have also gradually begun to realize the important role that Big Data can play in promoting economic development, improving public services, and helping to safeguard national security. [8] The U.S. is currently at the forefront of global Big Data development. In March 2012, the U.S. Federal Government launched the Big Data Research and Development Initiative, which aims to strengthen the allocation of resources to basic technical research and public sector applications for Big Data, integrating these efforts with the promotion of Open Data, and incorporating relevant government agencies into the implementation of the Initiative. The British government has also recognized the major benefits that Big Data can provide with respect to policy implementation, and in 2013 allocated £189 million to support Big Data technology R&D; the areas where the British government saw considerable potential for Big Data application development included real-time information management for the public sector, improving citizens' experience with public services, enhancing the quality of healthcare provision, strengthening compliance with tax legislation, improving welfare provision and preventing fraud and inappropriate use. [5] In March 2014, the Australian government announced the adoption of The Australian Public Service Big Data Strategy, incorporating six basic principles and an action plan; the government expressed the hope that Big Data could be used to enhance the efficiency of government administration, and promote collaboration and innovation; the Strategy notes that Big Data and the related data analysis involve not just the adoption of new technology solutions, but also the transformation of attitudes and organizational culture. [13]

Taiwan is also facing the challenges and opportunities that the era of Big Data presents. Although Taiwan has considerable experience in hardware-software integration, Big Data applications and implementation are still just getting off the ground, and there is a shortage of inter-disciplinary teams (and related experience). There are still many issues that Big Data could help to address to which it has not yet been applied, and Taiwanese industry, universities and research institutes are currently working decide where resources should be allocated. The Taiwanese government has opted for a model that involves collaboration between government and the university sector to establish the Big Data application environment, aiming to cultivate the data analysts and other specialist human talent needed to accelerate the development of Big Data in Taiwan. At the same time, by exploiting the innovation and other capabilities of Taiwan's universities and research institutes for the purposes of situational and demand-based analysis, the government is able to arrange for in-depth statistical analysis of government data, generating research results that have real value as a reference for government policy implementation, thereby helping the government to achieve its policy goals.

Industry Development

Leading international corporation Amazon analyzes user habit data to forecast customers' purchasing behavior, and makes use of the EMR (Elastic MapReduce) platform to provide Big Data collection, storage and analysis functions, thereby helping technology developments to acquire a better understanding of consumer characteristics. Google provides tools such as Dataflow and BigQuery, giving technology developers comprehensive data processing channels and the ability to store and analyze Big Data on cloud-based platforms.

As regards corporate applications, companies in some industries have already begun experimenting with Big Data as a means of identifying new business opportunities and improving processes. By and large, these firms have seen an increase in gross profits, and a substantial improvement in productivity. Firms in the retail sector have been able to use social networking to learn more about consumer preferences, making it possible for them to recommend appropriate products to potential consumers; at the same time, retailers can offer discounts for specific products that consumers mention online, and transmit this information to those customers that need it. In the manufacturing sector, comparison of historical data with current equipment status can be used to identify trends and modes in equipment deterioration due to aging, making it possible to predict several weeks in advance when equipment is likely to fail or need to be shut down, and facilitating preventive inspection and diagnostics to avoid breakdowns and stoppages. In the healthcare sector, data relating to past trends can be used to support preventive diagnosis, and integrated services can be developed that combine wearable devices with cloud-based services to permit more timely analysis of medical data. [8][9]

The Big Data movement has already begun to spread from other countries to Taiwan. Leading Taiwanese companies in the hi-tech manufacturing, telecommunications, finance and retailing sectors are already undertaking proof-of-concept verification, seeking to position themselves at the forefront of the adoption of the latest Big Data technologies. It is clear from the pattern of adoption in industry and from the initiatives launched by governments around the world that the development of Big Data presents opportunities for transformation in many different industry sectors. The three key elements in the development of Big Data are data, technology and applications, and Big Data derives its value from three key roles: those played by data owners, application innovators, and technology leaders. [16] Leading international corporations such as Google and Amazon have leveraged their own experience with cloud-based services to begin providing Big Data services for external clients, making use

of Open API technology to enable third parties to access Big Data, analyze it, and use it create added value. [15] Taiwan currently lacks the environment and mechanisms needed to stimulate effective collaboration in the Big Data field; the three key roles are developing independently of one another, and there are problems with a lack of speed and limited resources, which have hindered the development of application services in Taiwan. Of course, it is even more challenging for Taiwanese companies to develop open platforms similar to those of Google or Amazon. Until such time as Big Data develops into a fully-fledged industry, the main emphasis should be on exploring particular aspects and inter-related issues, encouraging industry to move in the direction of cross-industry integration that takes the value proposition as the core concept, as well as on examining how the general public can be got involved in the joint creation of value from Big Data.

RESEARCH METHOD

Big Data is already a major focus of discussion in the market. While many enterprises are aiming to use Big Data to solve organizational problems and identify emerging opportunities, they are often unable to verify whether their products and services will meet consumers' real needs, whether the new technologies involved can solve problems rapidly (and not be subject to replacement within the near future), and whether data access involves questions of data design and licensing mechanisms. [8] Where products or services are developed without following appropriate procedures or without in-depth discussion, there is a risk that such products or services may not be able to satisfy the market's needs, and that this will not be discovered until after a considerable amount of time and resources has already been wasted unnecessarily. The present study therefore makes use of the design science research methodology proposed by Peffer et al.[11] to examine the question of whether Big Data value creation in relation to government applications and industrial development is adequate to satisfy the value proposition requirements of industrial and social development, and to permit the development of a comprehensive industry eco-system. As shown in figure 1 the research process includes six steps: identify problem and motivate, defining the objectives of a solution, design and development, design and development, demonstration, evaluation, and communication with practices to carry out an explicitly applicable solution to a problem, the rigor of an artifact.

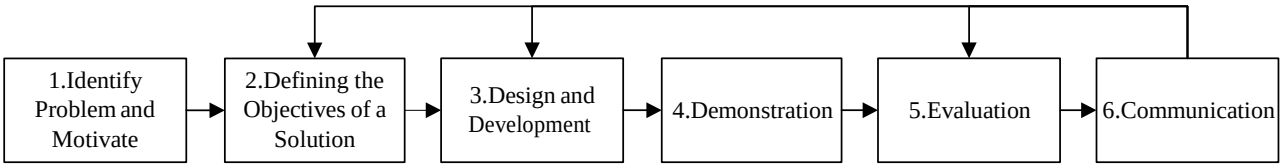

Figure 1. Research Process

Design science is of importance in a discipline oriented to the creation of successful artifacts. [11] It is regarded that eco-system is an artifact in this study. A topic, defined with social value and business strategy, will be resulted from team-built and co-creation by the process of industry analysis, customer recognition, and value proposition. There developed an artifact that comes to a formed topic by business viability, human values desirability, technology feasibility, and design innovation. As described in table 1, an artifact needs to be examined by the six steps of design science research methodology.

Table 1. Artifact Examined by Design Science Research Methodology

	Step	Description
1	Identify problem and motivate	Propose an issue and value proposition through analysis of innovative trends and observation of social needs.
2	Defining the objectives of a solution	Define design challenges and look for a consensus from stakeholders.
3	Design and development	Joint design with partners to build up a prototype service system.
4	Demonstration	Collaborated with target users to collect feedback and improvement.
5	Evaluation	Implement service system and analyze user behavior continuously
6	Communication	Diffuse and promote its importance to relevant professionals.

RESEARCH DESIGN

Role Definition

The value of Big Data lies in the development of cross-device, cross-system and cross-service smart data collection and analysis technologies, and in the ability to use mobile networks to link devices together and speed up the generation of different types of data by business enterprises. The development of the Big Data industry should begin with the forging of consensus regarding the needs of industrial and social development, with the putting forward of value propositions that can serve as the basis for value creation initiatives; the first priority should be to promote the digitalization of industry, before moving on from this towards the creation of digitalized industries, and the gradual formation of a comprehensive industry ecosystem. The present study proposes a model whereby the Big Data ecosystem as depicted in figure 2 is comprised of three key roles – data owners, application innovators, and technology leaders – with open platforms being used to provide data and tools.

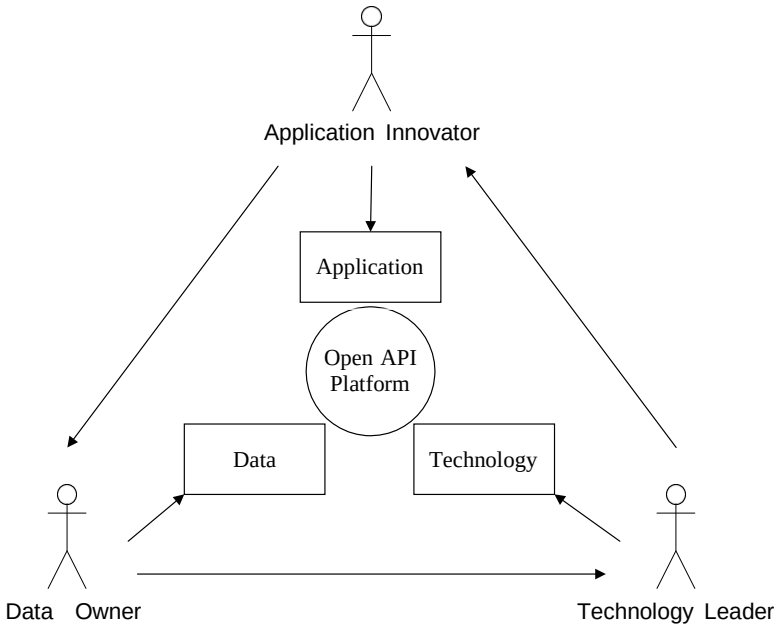

Figure 2. Big Data Ecosystem

- Application innovators: These individuals are able to spot opportunities before others notice them, and come up with unique ideas for creating value through data extraction. Those filling this role may not be directly involved with data or technology, but this means that their thinking is not constrained by concerns about practical feasibility.

- Data owners: Data owners either own, or are in a position to collect, large amounts of data, and are legally entitled to access these data, either for their own use or by licensing a third party to extract value from the data. This role does not necessarily involve actually working with the data directly, and the data owner may not possess suitable technology for deriving value from the data; the role may not require creative thinking, and the data owner may not actually know whether the data is capable of providing value or not.

- Technology leader: This will normally be a consulting firm, technology provider, analytical service provider or data specialist that possesses specialist, practical technology for implementing complex analysis. Those fulfilling this role may not actually possess data themselves, and may not be able to think of creative new uses for data.

- Open API platforms: The function of an open application programming interface (API) platform is to provide a data sharing market and data analysis tools, effectively integrating the three key elements of data, technology and applications. Open API platforms provide important opportunities for application innovators, particularly in relation to social media. They can benefit from the assistance of application innovators in developing value-added services for specific market segments, and there is the potential for cross-sectoral integration in the future, with the sharing of data in different fields being used to drive further growth in innovative Big Data applications.

Promoting Strategies

The key to a breakthrough in the development of Big Data applications lies in the formation of consensus regarding value, so that the value proposition of each application is capable of bringing about the formation of a unique industry ecosystem. In terms of Open Data, there is a need to deal with issues relating to data security, personal data, privacy, licensing, incentives for openness, etc. As regards technology adoption, the key issue here is the widespread shortage of data analysis technology and inter-disciplinary data science talent. The present study proposes four strategies for putting the development of the Big Data ecosystem on a sound basis:

1. Building up the Big Data application service environment, and realizing open APIs for data and analytical tools: Promoting an Open Data regulatory framework and incentive mechanisms, and establishing a Big Data application service platform along with guidelines for bringing data and analytical tools onto the platform, together with data application development model and date analysis model guidelines.

2. Cultivation of Big Data human talent: Using the Big Data application service platform to provide data sources and support application testing and community development. Besides focusing on helping data owners to rapidly develop the ability to utilize data in effective ways, there is also a need to encourage collaboration between industry and the university sector to

bring about the launch of Big Data related courses in universities, the holding of themed application development competitions, and collaborative development of data analysis tools and technologies.

3. Using themed services as model projects: Specific industries suited to the introduction of Big Data applications should be selected, and an effort made to identify latent data application needs; model projects could then be launched, such as Industry 4.0, Bank 3.0 or Retail 4.0, described in table 2.

4. An across-the-board expansion of innovative Big Data application services: Through the formation of innovation teams and open innovation collaboration mechanisms, and the holding of seminars and competition, it should be possible to bring together data owners, application innovators and technology leaders to develop inter-disciplinary Big Data service concepts and to develop related applications.

Table 2. Industry-specific Application Models

Sector	Promotion Recommendations
Industry 4.0 (Smart Manufacturing)	• Introduction of Cyber-Physical Systems (CPS) – which integrate computing, communications and control functions – into the production process, to create model "smart factories" • Moving from centralized to dispersed production models, and using cloud-based data centers to share data with other regions, permitting real-time monitoring of factory inventory levels and production status, and integrating the upstream and downstream segments of the supply chain
Bank 3.0 (Financial Service Evolution)	• Emphasizing virtual banking and digital banking, with the development of virtual services that are not dependent on physical infrastructure • Using emerging technologies and applications to drive reform in the financial sector, bringing about the emergence of a steady stream of new personalized, mobile services; the reduced importance of physical branches will rewrite the "rules of the game" in the financial sector
Retail 4.0 (Retailing Sector Transformation)	• Driving a continued evolution from online shopping e-commerce models towards closer integration of physical and virtual channels • Integrating indoor navigation based on high-precision location based services (LBS) with mobile advertising to provide personalized marketing and enhance the efficiency of service and marketing provision • Transforming physical stores into places where customers can experience new products and services, thereby boosting sales

CONCLUSION

Currently, the Big Data industry is still in the early stages of its development, and the roles that need playing within the industry are still somewhat unclear. The industry has yet to achieve economies of scale, and there is a lack of strong commercial incentives for the development of new applications. Although many companies have expressed optimism about the benefits that Big Data applications can provide for industry, large corporations are still dragging their feet when it comes to the development of Big Data applications. Big Data application development can help to strengthen corporate competitiveness and enhance firms' operational efficiency; in promoting Big Data, specific application fields need to be taken as the starting point, so as to gradually build up data, R&D and testing and analysis models, along with forecasting methods.

Close collaboration with the general public – to benefit from the "wisdom of the crowd" – also has an important role to play in supporting the development of the Big Data industry. Business enterprises and innovation teams should be encouraged to make use of Big Data in developing innovative services and business models; innovation competitions, guidance provision and human talent cultivation initiatives can be used to help industry strengthen its data application capabilities. Taking the needs of industry as the starting point, concepts development by government agencies, universities and research institutes can be verified and examined by open discussion, with business enterprises providing feedback regarding usability and the feasibility of commercialization. Above all, enhancing the quality and value of data in both the public and private sectors is of key importance for ensuring the sound, healthy development of the Big Data industry.

REFERENCES

[1] Amir G. & Murtaza H. (2015) 'Beyond the hype: Big data concepts, methods, and analytics', *International Journal of Information Management*, Vol. 35, No. 2, pp. 137-144.
[2] Big data definition, *Gartner, Inc*, available at http://www.gartner.com/it-glossary/big-data/ (accessed 20 August, 2015).
[3] Hevner A. R., March S. T. & Park J. (2004) 'Design science in information research', *MIS Quarterly*, Vol. 28, No.1, pp. 75-105.
[4] Hsinchun C., Roger H. L. C. & Veda C. S. (2012) 'Business intelligence and analytics: From big data to big impact', *MIS Quarterly*, Vol. 36, No. 4, pp. 1165-1188.
[5] Gang-Hoon K., Silvana T. & Ji-Hyong C. (2014) 'Big data applications in the government sector', *Communications of*

the ACM, Vol. 57, No. 3, pp. 78-85.
[6] John G. & David R. (2011) 'Extracting value from chaos', *IDC IVIEW*, available at http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf (accessed 20 August, 2015).
[7] Jianqing F., Fang H. & Han L. (2014) 'Challenges of big data analysis', *National Science Review*, Vol. 1, No. 2, pp. 293-314.
[8] Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C. & Byers AH. (2011) 'Big data: The next frontier for innovation, competition, and productivity', *McKinsey Global Institute*, available at http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation. (accessed 20 August, 2015)
[9] McAfee, A. & Brynjolfsson, E. (2012) 'Big data: The management revolution', *Harvard Business Review*, pp. 61-68.
[10] Min C., Shiwen M. & Yunhao L. (2014) 'Big data: A survey', *Mobile Networks and Applications*, Vol. 19, No. 2, pp. 171-209.
[11] Peffers, K., Tuunanen, T., Rothenberger, M. A. & Chatterjee, S. (2007) 'A design science research methodology for information systems research', *Journal of Management Information Systems*, Vol. 24, No.3, pp. 45-77.
[12] Qi F., Jun T. & Yufang X. (2015) 'The survey of big data', *Proceedings of International Conference on Computer Science and Electronic Technology*, pp. 403-407, Dec. 27-28, Shenzhen, China.
[13] Roche S. (2014) 'The Australian public service big data strategy', *Department of Finance and Deregulation*, Australian Government, available at http://www.finance.gov.au/sites/default/files/Big-Data-Strategy_0.pdf. (accessed 20 August, 2015).
[14] Sagiroglu S. & Sinanc D. (2013) 'Big data: A review', *Proceedings of International Conference on Collaboration Technologies and Systems (CTS)*, pp. 42-47, May 20-24, San Diego, USA.
[15] Xinhua E. Jing H., Yasong W. & Lianru L. (2013) 'Big data-as-a-service: Definition and architecture', *Proceedings of 15th IEEE International Conference on Communication Technology (ICCT)*, pp. 738-742, Nov. 17-19, Guilin, China.
[16] Yuri Demchenko, C. & Membrey, P. (2014) 'Defining architecture components of the big data ecosystems', *Proceedings of the International Conference on Collaboration Technologies and Systems (CTS)*, pp. 104-112, May 19-23, Minnesota, USA.