

## A Big Data Technology Foresight Study with Scenario Planning Approach

Wei-Hsiu Weng<sup>1\*</sup> and Woo-Tsong Lin<sup>2</sup>

Department of Management Information Systems, National Chengchi University, Taiwan

weng@iii.org.tw<sup>1</sup>, lin@mis.nccu.edu.tw<sup>2</sup>

*\*Corresponding Author*

Received 12 Aug 2013; received in revised form 12 Dec 2013; accepted 27 Dec 2013

---

### Abstract

While the importance and value of Big Data are gradually being recognized by most enterprises worldwide, Big Data related technologies, and the priority of adopting these technologies have so far not been clearly recognized. To fill this gap, this paper focuses on the technology planning strategy of organizations that have an interest in developing or adopting Big Data related technologies.

Based on the scenario analysis approach, a technology planning strategy is proposed. In this analysis, thirty Big Data related technologies are classified into six strategic clusters, and the importance and risk factors of these clusters are then evaluated under four possible scenarios. The main research findings include the discovery that most NoSQL technologies are rated high to medium in importance and high risk in all four scenarios, and that scenario changes will have less impact on Cloud Analytics, Embedded Analytics, as well as Big Data Visualization technologies. These results provide a reference for organizations and vendors interested in incorporating emerging Big Data related technologies.

*Keywords: Big Data, scenario analysis, technology foresight, strategy*

---

### 1. Introduction

Big Data technology is used to store, convert, transmit and analyze large quantities of dynamic, diversified data (which may be structured or unstructured) for the purpose of commercial benefit (Borkar et al., 2012). The applications of big data technology need to be able to undertake real-time, high-complexity analysis of vast amounts of data, to help business enterprises make decisions within the shortest possible timeframe (Bryant, 2008). With the rapid pace of development in cloud computing, both public cloud and private cloud data centers are continuing to accumulate enormous volumes of data; as a result, big data technology applications are becoming ever more important (Agrawal, 2011).

Big Data is an emerging terminology used to represent the fast growing data size

encountered in organizations and societies (Bollier, 2010; Brown et al., 2011). Big Data Analytics (BDA) refers to a technology and framework for quickly storing, converting, transferring and analyzing massive amounts of constantly updated, huge, varied, structured and unstructured data for commercial gain (Russom, 2011). BDA has now evolved from large database storage systems to cloud technology in order to analyze and process data in a way that is more economical, more effective and easier for the customer to manipulate (Baer, 2011). The leading global vendors today include EMC, IBM, Oracle, SAS, SAP and Teradata. Solutions currently offered by these BDA vendors include data warehousing, data mining, analytics, data organization, data management, decision support, automation interface, to name only a few. Innovative technologies and solu-

tions in this field are currently under rapid development (Mukherjee, 2012).

Presently, major IT firms worldwide are exploring possible business opportunities in the Big Data generated market. However, what is the scope of Big Data technologies? And what are the possible outlooks in terms of the importance as well as the risks of these technologies? These key questions need to be answered before one can have confidence in the accuracy of technology strategy planning. To assist IT vendors moving forward in the emerging Big Data market, this research aims to explore possible planning strategies for adopting or developing Big Data related technologies. To achieve this objective, a systematic approach of scenario analysis followed by technology strategy planning is conducted.

## **2. Literature Review**

### **2.1 Evolution of Big Data Technology**

Chen et al. (2012) describe the evolution of Big Data technology. These researchers use business intelligence and analytics (BI & A) as a unified term, and treat big data analytics as a related field. They argue that the evolution of Big Data technology is characterized by BI & A 1.0, BI & A 2.0 and BI & A 3.0. Data management and warehousing is considered the foundation of BI & A 1.0. BI & A 2.0 systems require the integration of scalable techniques in text mining, web mining, social network analysis, and spatial-temporal analysis with those existing DBMS-based BI & A 1.0 systems. BI & A 3.0 integrate Big Data technology with mobile applications, such as mobile BI, mobile and sensor-based content, location-aware analysis, person-centered analysis, context-relevant analysis and mobile visualization and HCI.

In an article by Waller and Fawcett (2013), data science, predictive analytics and big data are collectively referred to as DPB. They argue that data science is the application of quantitative and qualitative methods to solve relevant problems and

predict outcomes. Data scientists need deep domain knowledge and a broad set of analytical skills. Predictive analytics is a subset of data science. Although predictive analytics is related to many long-standing quantitative approaches, it stands as distinct from each. Predictive analytics attempts to quickly and inexpensively approximate relationships between variables while using deductive mathematical methods to draw conclusions. Chiang et al. (2012) argue that the current state of the analytics software industry makes it difficult and cumbersome to conduct analyses without a deep perspective of the underlying systems and technologies. They advocate that BI & A should be an interdisciplinary area that integrates data management, database systems, data warehousing, data mining, natural language processing, text mining, network analysis, social networking, optimization, and statistical analysis.

### **2.2 Features of Big Data Technology**

Big Data has three features: volume, velocity and variety (McAfee and Brynjolfsson, 2012). Most discussion in the past has focused on how to store the volume of data. However, velocity and variety are crucial in competitive differentiation. Variety refers to the variety of data formats. Data can be structured data that can be sorted or non-structured data such as pictures, music, videos, essays and discussions. Compared to structured data, non-structured data provides a better reflection of reality for making important decisions. The New York City government for example integrated its previously separate criminal record, surveillance and traffic control systems into a single crime fighting system (Garicano, 2009). Big data analysis technology was used to identify patterns of criminal behavior and optimize police assignments, effectively reducing the local crime rate. The other feature is velocity. In a business environment where every second counts, businesses must collect and analyze data in a timely manner so as to make crucial decisions faster than their competitors. Examples include re-

al-time patient bio-monitoring in the medical industry, process improvement information in electronics manufacturing, and web view/click-through data in advertising and marketing. By processing huge volumes of constantly changing information that must be processed immediately, businesses can convert a mass of seemingly useless data into a product with economic value.

### 2.3 Progress of Big Data Technology

Big Data technology can be divided into two broad categories: Advanced SQL technology, which is oriented towards the use of relational databases, and NoSQL technology, the emphasis of which is on non-relational databases (Baer et al., 2011). Advanced SQL is specifically designed to provide real-time analysis results with large quantities of structured data. However, as the scale of data collection grows ever larger, and as the different categories of data that need to be processed become ever more complex, non-structured data is presenting business enterprises with new challenges, especially in terms of data storage and analysis (Borkar et al., 2012).

NoSQL non-relational database systems (Adrian, 2012) offer enhanced performance and extensibility, making them ideally suited to processing large amounts of non-structured, highly variable data. There are four main types of NoSQL database: key-value databases, in-memory databases, graphics databases, and document databases. The Advanced SQL database platform segment is currently going through a period of market consolidation, indicating that this segment is entering the mature phase of its evolution, characterized by slow but steady growth. By contrast, the NoSQL market is still very much in the growth stage, but is expected to play an increasingly important role in the big data technology of the future.

Hadoop is a form of big data technology that is attracting growing interest from various enterprises; more specifically, it is a form of key-value database technology (Baer, 2011). Leading U.S. retailer

Wal-Mart is using Hadoop to analyze sales data and identify new business opportunities. As well, online auction site eBay has been using Hadoop to process unstructured data and reduce the burden of database storage requirements.

MapReduce is a parallel computing method that permits the processing of very large volumes of data within a clustered computer framework. MapReduce breaks down tasks into two stages – Map and Reduce – in order to achieve a distributed computing effect.

## 3. Research Method

### 3.1 Scenario Analysis

Scenario Analysis (SRI, 1996) has been used in various domains for analyzing and forecasting trends in the development of technology. Many versions and variations of the SRI scenario analysis methods have been proposed (Mietzner and Reger, 2005). The technology portfolio planning process (Yu, 2006) is a systematic procedure used to assist in the strategic decision necessary to find the cluster set of resource allocations among available technologies that best fits the goal of an organization. Scenario planning is a key technique used by futurists to develop future models in order to facilitate this process and to develop strategic action plans and policies, as well as create a vision for the future (Erdogan et al., 2009).

The major steps of the technology strategy planning process are as follows (Bishop et al., 2007).

- (1) Identify decision criteria, which are the motivational forces for the resource allocation decision.
- (2) Propose possible future scenarios by exploring combinations of significant impact variables.
- (3) Compose a set of technology alternatives and classify them into clusters.
- (4) Generate a set of technology assessment indicators from mutually exclusive dimensions.

- (5) Find the best plan for a technology portfolio.

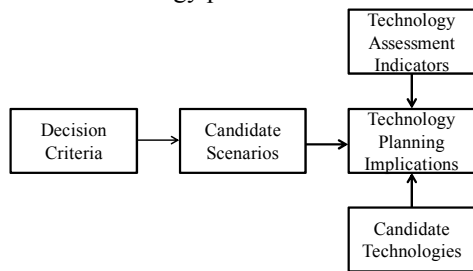


Figure 1: Research Framework of a Technology Foresight study

### 3.2 Expert Panel

To conduct the technology foresight study, an expert panel was formed with eleven domain experts selected from both the IT industry and the academic world. This expert panel consisted of the following members:

- (1) Three consultant managers of publicly listed IT services firms.
- (2) Four CEO and VP level executives from independent software vendors.
- (3) Two CIOs of publicly listed IT device manufacturers.
- (4) One professor of Management Information Systems.
- (5) One professor of Industrial Engineering and Management.

A facilitator led the expert panel discussion sessions by following the steps in Figure 1 above. Activities in these sessions included open discussions, anonymous voting, as well as the administration of surveys.

## 4. Results

### 4.1 Decision Criteria

To identify decision making criteria, expert panel discussions were conducted concerning decision making factors from the social, political, economic and technological perspectives. Possible decision factors were discussed, such as the market outlook for a technology, as well as the competence of the industry to acquire this technology. The final set of indicators is summarized in Table 1.

Table 1: Major Decision Factors

Decision factors	Issues
Social factors	1. Availability of Big Data for quality of life improvement for people
Technological factors	1. Entrance barrier level of Big Data technology 2. R&D strength of the industry
Economic factors	1. Strategic benefit of the enterprises 2. New business opportunity for the industry
Political factors	1. Strength of Big Data industry promoting policies of government

### 4.2 Candidate Scenarios

There are many different scenario alternatives which organizations may select for big data technology trends. Impact variables which are most likely to affect the scenario development were identified by the expert panel. Through evaluations from different combinations of these variables, final choices of scenarios were then determined. After the Expert Panel discussions, the scenarios were labeled and elaborated upon. The results are illustrated in Table 2.

Table 2: Candidate Scenarios

Scenario Code	Global IT Spending Outlook	Big Data Technology Breakthrough	Final Scenario Choice and Naming
00	High	High	Big Demand
10	Low	High	Cautiously Optimistic
01	High	Low	Slow Progress
11	Low	Low	Pessimistic

A detailed description of the scenarios is as follows.

#### 4.2.1 Scenario 00: Big Demand

In the Big Demand scenario, the foreseen global economic situation is strong, and the worldwide IT spending outlook is in good shape. At the same time, with the progress of continuous research in both industry and academia, the development of Big Data technology is experiencing a major breakthrough.

#### 4.2.2 Scenario 10: Cautiously Optimistic

In the Cautiously Optimistic scenario, the global economic outcome is in a downturn, possibly due to slow recovery from previous global financial turmoil, or encounters with new financial crises. However, the advancement of Big Data technology is not likely to stagnate, since leading vendors have invested significantly in R&D, and the volume and growing velocity of global data continue to progress.

#### 4.2.3 Scenario 01: Slow Progress

In the Slow Progress scenario, the foreseen global economical situation is

strong, and the worldwide IT spending outlook is in good shape. However, the progress of academic and industrial Big Data technology research and development is slow. As a result, potential users may relocate their resources to other areas with more promising technologies.

#### 4.3 Candidate Technologies

To assess the possible Big Data Analytics technologies for the proposed scenarios, another technology expert panel of ten members was formed. This panel differed from the previous panel. The purpose of a different expert panel was to assure independence between technology planning activities. Big Data technology data were collected by interviewing these panel members, as well as from secondary data which included vendor propositions and research literature. The final list of the most promising Big Data Analytics technologies is exhibited in the following table.

Table 3: Candidate Big Data Analytics Technology

Cluster	Technology
Data Warehouse (DW)	DW1: Central enterprise data warehouse DW2: Data warehouse appliance DW3: Data marts for analytics DW4: Analytics processed within the EDW DW5: Extract, Transform, Load (ETL)
NoSQL BDA (NS)	NS1: MapReduce NS2: Hadoop NS3: NoSQL or non-indexed DBM NS4: Column oriented storage engine NS5: Text mining
Advanced SQL BDA (AS)	AS1: Complex SQL AS2: Distributed SQL AS3: OLAP AS4: Advanced SQL Appliance AS5: SQL Accelerator
Cloud Analytics (CA)	CA1: Public cloud analytics CA2: Private cloud analytics CA3: Social analytics CA4: Software as a service (SaaS) CA5: Internet of Things (IoT)
Embedded Analytics (EA)	EA1: Predictive analytics EA2: Complex event processing (CEP)

Cluster	Technology
	EA3: In-memory database EA4: In-database analytics EA5: In-line analytics
Big Data Visualization (DV)	DV1: Advanced data visualization DV2: Real-time reports DV3: Dashboards DV4: Visual discovery DV5: Infographics
Total items	30

#### 4.4 Technology Assessment Indicators

The expert panel on technology then applied the scenario analysis approach to assess the candidate Big Data technologies of the six major clusters in two dimensions: importance and risk. These two dimensions are quantified by selected indicators summarized in Table 4.

For the surveying process of the risk dimension, a three point scale was used in the beginning. However, this yielded a result with many ratings squeezed together. A

nine point scale was then surveyed by the same expert panel and produced a dispersed and readable result. Therefore, although the final result is presented in a three level scale of Low, Medium and High, the survey adopted a nine point scale. These scores were then converted to three level indicators with Low Level for 1~3 points, Medium Level for 4~6 points, and High Level for 7~9 points. The final set of Technology Assessment Indicators is shown in Table 4.

Table 4: Technology Assessment Indicators

Dimensions	Indicators	Low Level	Medium Level	High Level
Importance	Global market size	< US\$1B	US\$1B~US\$10B	> US\$10B
	Enterprise adoption ratio	< 10%	10%~60%	> 60%
Risk	Entrance barrier	1~3 points	4~6 points	7~9 points
	Strength of industry	1~3 points	4~6 points	7~9 points

#### 4.5 Technology Planning Implications

Based on the important indicators and risk indicators in Table 4, the expert panel assessed the Big Data technologies compiled in Table 3 with respect to the four scenarios. The assessment results are exhibited in figures 2-5 and discussed as follows.

##### 4.5.1 Technology Planning Implications for Scenario 00: Big Demand

For the Big Demand scenario, the assessment outcome is depicted in Figure 2. In this scenario, the Advanced SQL tech-

nologies would be of high importance and low or medium risk in general. This is mainly because the Advanced SQL technologies, based on the development of Relational DBMS, are relatively mature and have a large base of users worldwide. Also note the NoSQL technologies are positioned in both high importance and high risk. Though NoSQL technologies are viewed as the next big opportunity for the IT industry, these technologies are new to most enterprises and the adoption of them is considered highly risky.

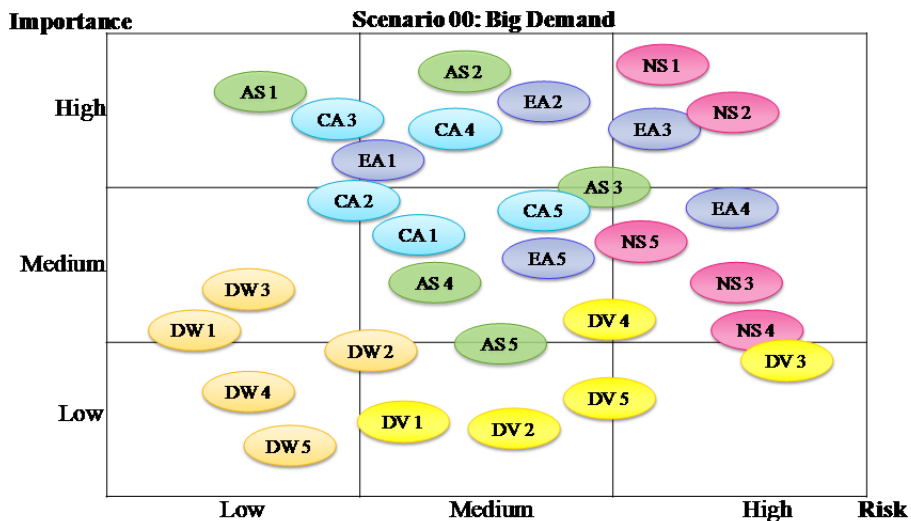


Figure 2: Technology Assessment for Scenario 00: Big Demand

#### 4.5.2 Technology Planning Implications for Scenario 10: Cautiously Optimistic

For the Cautiously Optimistic scenario, the assessment outcome is depicted in Figure 3. In this scenario, the economic outlook is not as good as in the Big Demand scenario. Compared with the results of Scenario 00 in figure 2, most Big Data

technologies have lower importance ratings except the NoSQL (NS) technologies. This implies that when the global IT spending outlook is weak, the progress of many Big Data technologies will be affected. However, vendors will continue the development of NoSQL technologies and await possible new opportunities.

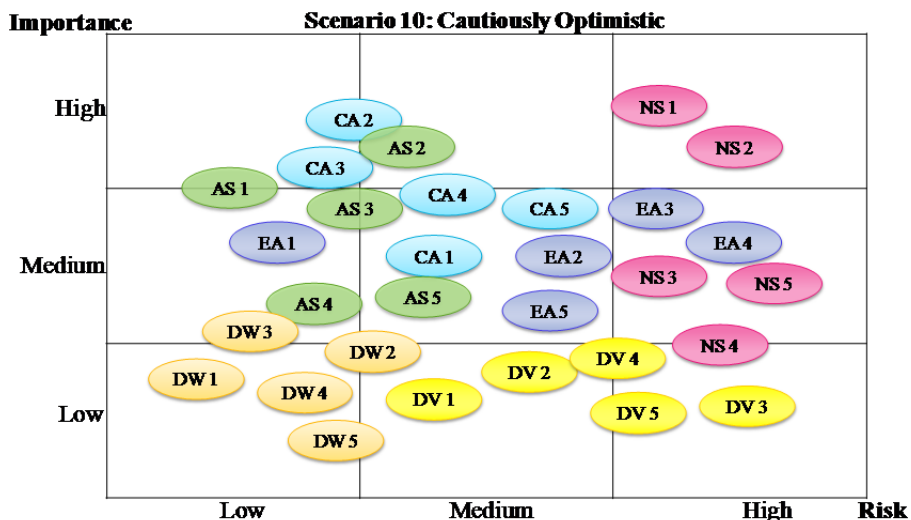


Figure 3: Technology Assessment for Scenario 10: Cautiously Optimistic

#### 4.5.3 Technology Planning Implications for Scenario 01: Slow Progress

For the Slow Progress scenario, the assessment outcome is depicted in Figure 4. In this scenario, the risk of Advanced SQL (AS) and Cloud Analytics (CA) technologies would increase compared with the

previous two scenarios. The Advanced SQL technologies, based on the development of Relational DBMS, would have decreased importance. In general, the Cloud Analytics would also have lower importance, due to the slow advancement of technology development.

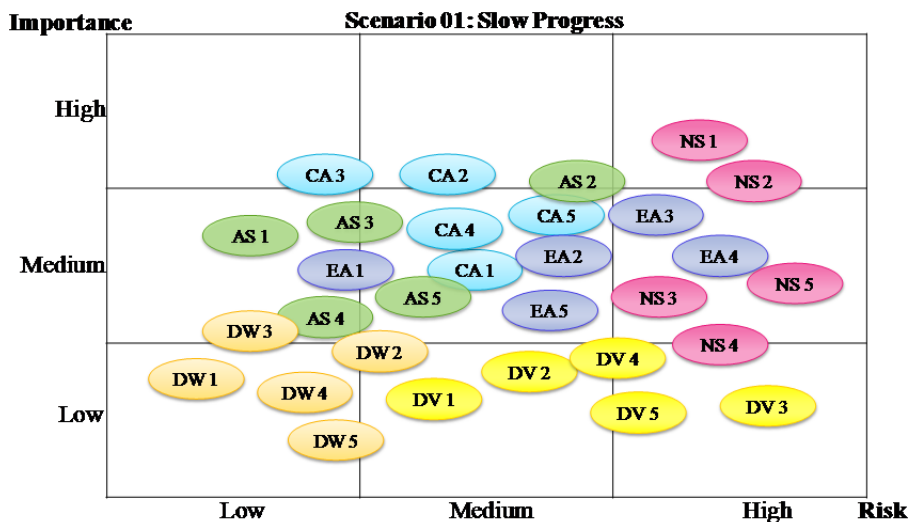


Figure 4: Technology Assessment for Scenario 01: Slow Progress

#### 4.5.4 Technology Planning Implications for Scenario 11: Pessimistic

For the Pessimistic scenario, the assessment outcome is depicted in Figure 5. In this scenario, the global IT spending outlook is not good. The advancement of Big Data related technologies is also slow-

ing down. The risk rating of most Big Data related technologies would rise and their importance rating would drop. An important observation is that the NoSQL (NS) technologies are the most robust against scenario changes.



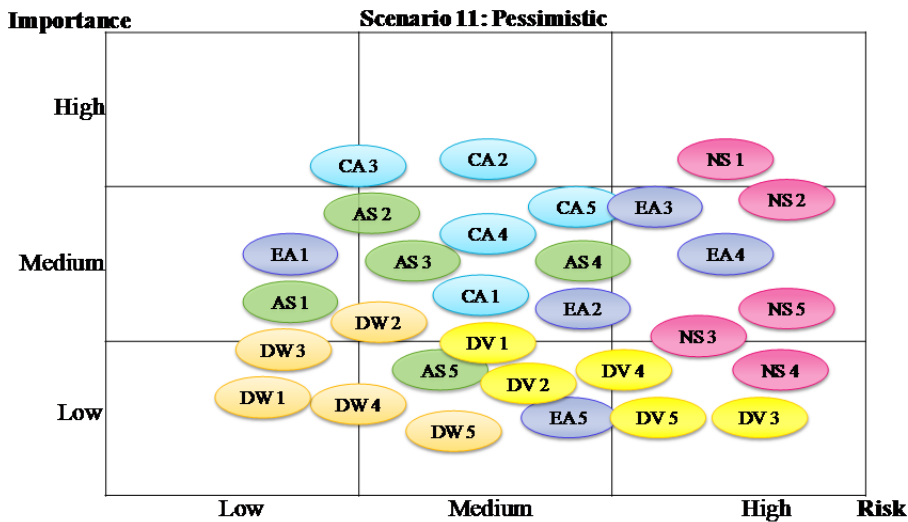


Figure 5: Technology Assessment for Scenario 11: Pessimistic

## 5. Conclusions

### 5.1 Research Findings

In this study, a systematic approach geared towards deriving foresight towards possible Big Data technology developments over the next five years was conducted. Highlights of the research findings are summarized in Table 5. Based on these

results, the strategic thinking of an organization toward developing or adopting Big Data technologies for competitive advantages can be initiated. For example, these findings suggest that NoSQL technologies should have a higher priority for organizations in the pursuit of new market opportunities.

Table 5: Results and Implications

Big Data Technology Cluster	Result	Implication
Data Warehouse (DW)	Data Warehouse (DW) technologies remain in low importance and low risk corner in general in all scenario analyses.	These technologies are mature technologies classified in the BI & A 1.0 technology category. Data Warehouse (DW) software has been implemented in the production IT systems of many enterprises. Adoption is safe but opportunity is also limited.
NoSQL BDA (NS)	Most NoSQL (NS) technologies are rated high to medium importance and high risk in all four scenarios.	This implies the emerging NoSQL technologies are still in the early stage of development, and both the market potential and risk need to be carefully balanced for most enterprises.
Advanced SQL BDA (AS)	If the market demand for Big Data technologies is not strong, Advanced SQL (AS) technologies will have lower importance.	This reflects the fact that Advanced SQL is extended from Relational DBMS which is competed by other emerging Big Data technologies. When the global IT spending outlook is low, most enterprises will be more conservative on investing

Big Data Technology Cluster	Result	Implication
		further in this type of technologies.
Cloud Analytics (CA)	Cloud Analytics (CA) technologies are rated of similar importance but lower risk than NoSQL (NS) technologies in all scenarios.	This implies Cloud Analytics (CA) technologies may serve as safer candidates for organizations which are seeking new opportunities with Big Data technologies but are vulnerable to risk.
Embedded Analytics (EA)	In general, Embedded Analytics (EA) technologies are rated of similar importance but higher risk than Advanced SQL (AS) technologies.	This implies Advanced SQL (AS) technologies may serve as a safer choices than the Embedded Analytics (EA) technologies for organizations which plan to extend SQL based technologies in the future.
Big Data Visualization (DV)	Scenario changes will have less impact on the Big Data Visualization technologies (DV), which remain in the higher risk and lower importance corner.	Big Data Visualization technologies (DV) are technologies with high risk for now and in the near future. Their potential for development requires further observation and investigation.

On the other hand, vendors interested in exploring the market opportunities of Big Data technologies can use the analysis framework and outcome of this research as a reference for their strategic planning, thereby avoiding many unnecessary trial and error marketing efforts. In particular, with a clear picture of the Big Data technologies scenario analysis, vendors can better position themselves for the most suitable market sector in terms of importance and risk.

## 5.2 Future Research Suggestions

Further studies on Big Data technology planning could be conducted more thoroughly through various scenario investigations. In particular, the following questions are worthy of attention.

- (1) How are Big Data technology development trends influenced by the adoption of trends in other innovative technology areas, such as cloud computing (Agrawal, 2011) and mobile apps?
- (2) How are Big Data technology adoption trends related to IT industry competitiveness as well as the level of IT readiness of different countries (EIU, 2007)?

In addition, more decision making factors that affect the enterprise adoption of Big Data technology could be integrated and more insights could be obtained by incorporating decision techniques such as AHP (Saaty and Vargas, 2001) into the analytical framework and process of this research.

## References

- Adrian, M. (2012). Who's Who in NoSQL DBMSs, *Gartner Report G00228114*.
- Agrawal, D., Das, S., & Abbadi, A. (2011). Big Data and Cloud Computing: Current State and Future Opportunities. *ACM EDBT Conference*, March 22–24, 2011, Uppsala, Sweden.
- Baer, T. (2011). 2012 Trends to Watch: Big Data. *Ovum Report, OI00140-041*.
- Baer, T., Sheina, M., & Mukherjee, S. (2011). What is Big Data? The Big Architecture. *Ovum Report, OI00140-033*.
- Bishop, P., Hines, A., & Collins, T. (2007). The current state of scenario development: an overview of techniques. *Foresight*, 9(1), 5–25.
- Bollier, D. (2010). *The Promise and Peril of Big Data*, The Aspen Institute.
- Borkar, V., Carey, M., & Li, C. (2012). Inside “Big Data Management”: Ogres, Onions, or Parfaits? *ACM*

- EDBT/ICDT Joint Conference*, Berlin, Germany.
- Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of 'big data'? *McKinsey Quarterly*, McKinsey & Company, Oct. 2011.
- Bryant, R., Katz, R., & Lazowska, E. (2008). Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society. <http://www.cra.org/ccc/initiatives>.
- Chen, H., Chiang, R., & Storey, V. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4):1165–88.
- Chiang, R. H. L., Goes, P. & Stohr, E. A. (2012). Business Intelligence and Analytics Education, and Program Development: A Unique Opportunity for the Information Systems Discipline. *ACM Transactions on Management Information Systems*, 3(3), 12:1-12:13.
- Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J., & Welton, C. (2009). MAD Skills: New Analysis Practices for Big Data. *ACM VLDB conference*, August 24-28, 2009, Lyon, France.
- EIU. (2007). The means to compete - Benchmarking IT industry competitiveness. *The Economist*, July 2007.
- Erdogan, B., Abbott, C., Aouad, G., & Kazi, A., (2009). Construction IT in 2030: a scenario planning approach, *Journal of Information Technology in Construction (ITcon)*, Vol. 14, Special Issue Next Generation Construction IT: Technology Foresight, Future Studies, Roadmapping, and Scenario Planning, 540-555.
- Garicano, L. (2009). How Does Information Technology Help Police Reduce Crime? *TNIT Newsletter*, Number 3 - December 2009.
- IDC. (2012). Worldwide Big Data Technology and Services 2012–2015 Forecast. *IDC #233485, Volume: 1*, Tab: Markets.
- Kobielus, K., Moore, C., Hopkins, B., & Coyne, S. (2011). Enterprise Hadoop: The Emerging Core of Big Data. *Forrester Research*. October, 2011.
- McAfee, A., & Brynjolfsson, E. (2012). Big Data: The management revolution. *Harvard Business Review*, October, 1-9.
- Mietzner, D., & Reger, G. (2005). Advantages and disadvantages of scenario approaches for strategic foresight. *Int. J. Technology Intelligence and Planning*, 1(2).
- Mukherjee, S. (2012). Deploying Big Data Systems. *Ovum Report*, OI00140-035.
- Russom, P. (2011). Big Data Analytics. *TDWI Research*, 4<sup>th</sup> Quarter, 2011.
- Saaty, T. & Vargas, L. (2001). *Models, Methods, Concepts and Applications of the Analytic Hierarchy Process*. Boston, MA: Kluwer Academic Publishers.
- SRI. (1996). Integrated Technology Planning. *SRI International Report*, Menlo Park, California.
- Yu, O. (2006). *Technology Portfolio Planning and Management*. Berlin: Springer Publisher.
- Waller, M. A. & Fawcett, S. E. (2013). Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics*, 34(2): 77–84.

### **About Authors**

**Wei-Hsiu Weng** is a Ph.D. candidate in the Department of Management Information Systems, National Chengchi University, Taiwan (ROC). He received his M.S. degree in Computer Sciences from the University of Wisconsin at Madison, USA, and his B.S. degree in Chemical Engineering from National Cheng Kung University, Taiwan (ROC). His research interests are in the areas of Supply Chain Management, Technology Innovation Management and Software Industry Management.

**Woo-Tsong Lin** is a Professor in the Department of Management Information Systems, National Chengchi University, Taiwan (ROC). He received his D.Eng. in Industrial Engineering from the University of California at Berkeley, USA. His research interests are in the areas of Supply Chain Management, Electronic Commerce and Software Industry Management.