

A Scenario Analysis Of Big Data Technology Portfolio Planning

Wei-Hsiu Weng¹ and Woo-Tsong Lin²

Department of Management Information Systems, National Chengchi University, Taiwan

Abstract

“Big Data” refers to the technology, product and service applied in big, immediate and manifold structured and unstructured information. It helps companies store, transform, transmit and analyze huge amounts of information. It also provides advanced business analytics, develops business intelligence and leads to gains in business values.

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

Based on the Scenario Analysis approach, a technology planning strategy for Taiwanese organizations 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 two possible scenarios. The result provides a reference for organizations and vendors interested in incorporating the emerging Big Data technologies.

Keywords: Big Data, Scenario Analysis, Technology Portfolio, Strategy

1. Introduction

Big Data is an emerging terminology 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 key global vendors today include IBM, SPS, Oracle, Teradata and EMC. Solutions currently offered by these BDA vendors include Data Warehouse, Data Mining, Analytics, Data Organization, Data Management, Decision Support and Automation Interface, and so on. Innovative technologies and solutions in this field are currently under rapid development (Mukherjee, 2012).

Currently, major IT firms worldwide are exploring possible business opportunity in the Big Data generated market. The Taiwanese IT vendors are strong players worldwide in the manufacturing and integration of IT devices and services. To assist the Taiwanese IT vendors moving forward to the emerging Big Data market, this research aims to explore possible strategic plans for effective allocation of resources. To achieve this objective, a systematic approach of scenario analysis followed by technology strategy planning is conducted.

2. Literature Review

Big Data technology is used to store, convert, transmit and analyze large quantities of dynamic, diversified data (which may be structured or unstructured data) for the purpose of commercial benefit (Borkar et al., 2012). Big data technology applications need to be able to undertake real-time, high-complexity analysis of vast amounts of data, to help business enterprises perform decision-making 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). Recent development trends in the field of big data technology are discussed as follows.

2.1 NoSQL 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 in 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 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 in its evolution, characterized by slow but steady growth. By contrast, NoSQL market is still very much in the growth stage, and be expected to play an increasingly important role in big data technology in the future.

2.2 Hadoop big data technology

Hadoop is a big data technology that is attracting growing interest from 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; online auction site eBay has been using Hadoop to process unstructured data and reduce the burden of database storage requirements. Hadoop is a software platform that enables users to rapidly write and process large quantities of data. Designed by Doug Cutting, Hadoop was originally an Apache research project, but has since been adopted for commercial applications involving Petabyte-scale big data.

The Hadoop platform architecture comprises three main elements: the Hadoop Distributed File System (HDFS), the Hadoop MapReduce model, and the HBase database system (Kobielus, et. al., 2011). The operational model of the Hadoop computing framework basically involves coordinating the use of MapReduce distributed computing algorithms with the HDFS distributed file system, which makes it possible to transform ordinary commercial servers into distributed computing and storage groups, creating the capability to store and process huge volumes of data. Supporting this capability is HBase, which is a distributed database system capable of coping with Petabyte-scale data volumes.

One of most significant features of the HDFS database system is its use of streaming data access, which facilitates access of Petabyte-scale data. What is more, this system can be used with ordinary-specification hardware. 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. HBase is a data storage system that uses a Java-based distributed architecture.

Unlike conventional database systems that use a row-oriented storage method, HBase (like Google's BigTable) uses a column-oriented approach to storage. The advantage of using the column-oriented method is that the column in which each item of data is stored does not

have to be fixed; this helps to ease the complexity of having to add new columns, which is a problem for row-oriented methods.

Besides HDFS, MapReduce and HBase, which constitute the core components of the Hadoop platform, there are a number of other tools which help to make the functionality of the Hadoop platform more comprehensive. For example, Hive allows users to access large volumes of data using SQL language, which they may be more familiar with; Pig is a simple programming language that can be used to write MapReduce programs; ZooKeeper provides the configuration data needed by individual servers, to facilitate monitoring of the Hadoop operating environment; Mahout is a function library that helps developers to write MapReduce algorithms.

2.3 Major Vendors of Big Data Technology

Recognizing the potential business opportunities presented by big data, leading international corporations such as IBM, Oracle, Microsoft, SAP/Sybase, Teradata and EMC have all been moving aggressively into this new field, working to provide enterprises with practical commercial applications for the effective storage and analysis of large volumes of data (IDC, 2012). IBM, which has for some years now been focusing heavily on the business intelligence market, recently acquired data warehousing firm Netezza, and will be incorporating Netezza's technology into its Business Analytics Optimization (BAO) solutions as a means of enhancing their competitiveness. Similarly, Teradata has acquired Aster Data, and EMC has purchased Greenplum (Cohen et. al., 2009). This wave of M&A activity reflects leading IT companies' determination to develop the big data market.

BDA vendors have accelerated their expansion and acquisitions. HP for example acquired 3PAR in September, 2010. In November 2010, EMC acquired Isilon. In December 2010, Dell acquired Compellent. In March 2011, Teradata acquired Asterdata. In June 2011, Oracle acquired PillarData. In September 2011, HDS acquired Blue Arc. In April 2012, EMC acquired Pivotal as well.

The wave of acquisitions can be divided into two types. One is system manufacturer's acquisition of a storage system manufacturer to provide customers with a one-stop solution. Oracle's acquisition of PillarData and IBM's acquisition of XIV were both of this type. The consolidation enabled large vendors to acquire complementary technologies that make their own system more complete. The other type is buy-outs between storage system manufacturers. If a storage system manufacturer was big enough and was not acquired by a system manufacturer, it began acquiring other smaller storage manufacturers with innovative technologies in order to flesh out their own product lines and technology. EMC's acquisition of Isilon and HDS' acquisition of Blue Arc were both of this type. Whether it was a system manufacturer acquiring a storage system manufacturer or storage manufacturers acquiring each other, in both cases, vendors used their acquisition strategy to acquire key technologies.

Continued innovations in information technology means that, large vendors will continue to make new acquisitions in order to make their products more complete and competitive.

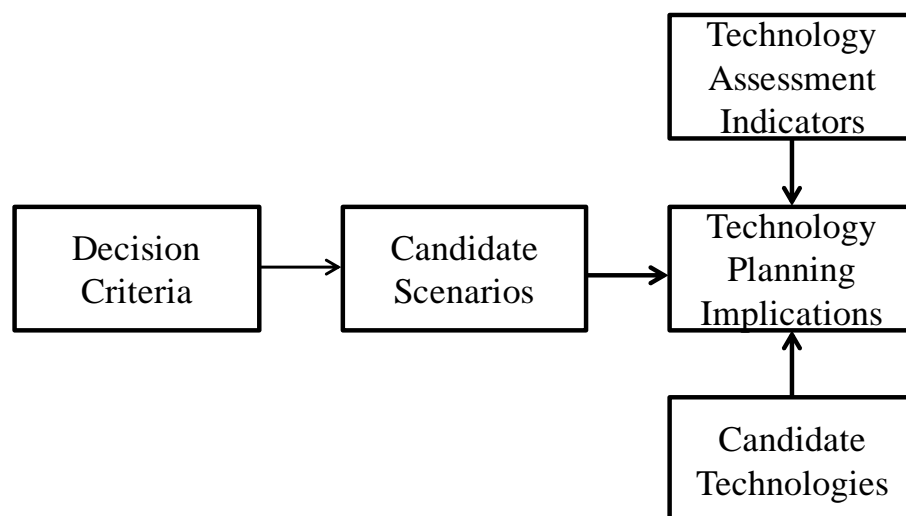
3. Research Methodology

Scenario Analysis (SRI, 1996) has been used in various domains for analyzing and forecasting technology development trends. Many versions and variations of the SRI scenario analysis methodology have been proposed (Mietzner & Reger, 2005). The technology portfolio planning process (Yu, 2006) is a systematic procedure for the strategic decision of finding a 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 help this process and to develop strategic action plans and policies or to 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 of technology portfolio.

Figure 1. Research Framework of Technology Strategy Planning



4. Application of the Process and the Results

4.1 Decision Criteria

To identify decision criteria, an expert panel is formed with eleven experts from the IT industry and academics. Expert panel discussions are conducted for decision factors from the social, technological, economical and political perspectives. Possible decision factors are discussed, such as the market outlook for a technology, and the competence of the industry to acquire this technology. The final set of indicators are summarized in the following table.

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 Taiwanese industry
Economical factors	1. Strategic benefit of Taiwanese enterprises 2. Business opportunity for Taiwanese industry

4.2 Candidate Scenarios

There are many different scenario alternatives organizations may select for big data technology trends. Impact variables which are most likely to affect the scenario development are identified by the expert panel. By evaluations from different combinations of these variables, final choices of scenarios are then determined. After the Expert Panel discussions, two scenarios are selected as most likely to take place in the next five to ten years. The two scenarios are named and elaborated. The result is 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
01	High	Low	
10	Low	High	Cautiously Optimistic
11	Low	Low	

4.2.1 Scenario 00: Big Demand

In the Big Demand scenario, the foreseen global economical situation is strong and worldwide IT spending outlook is in good shape. At the same time, with the progress of

continuous researches both in industries and academics, the Big Data related technologies are also experiencing major breakthroughs.

4.2.1 Scenario 10: Cautiously Optimistic

In the Cautiously Optimistic scenario, the global economical outcome is in downturn, possibly due to slow recovery from previous global financial turmoil, or encounter new financial crisis. However, the advancement of the 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.

4.3 Candidate Technologies

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

Table 1: Candidate Big Data Analytics Technologies

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) 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 technology expert panel 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 the following table. Point scale of 1 to 9 are surveyed from the panel. These scores are 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 are shown in the following table.

Table 3: Technology Assessment Indicators

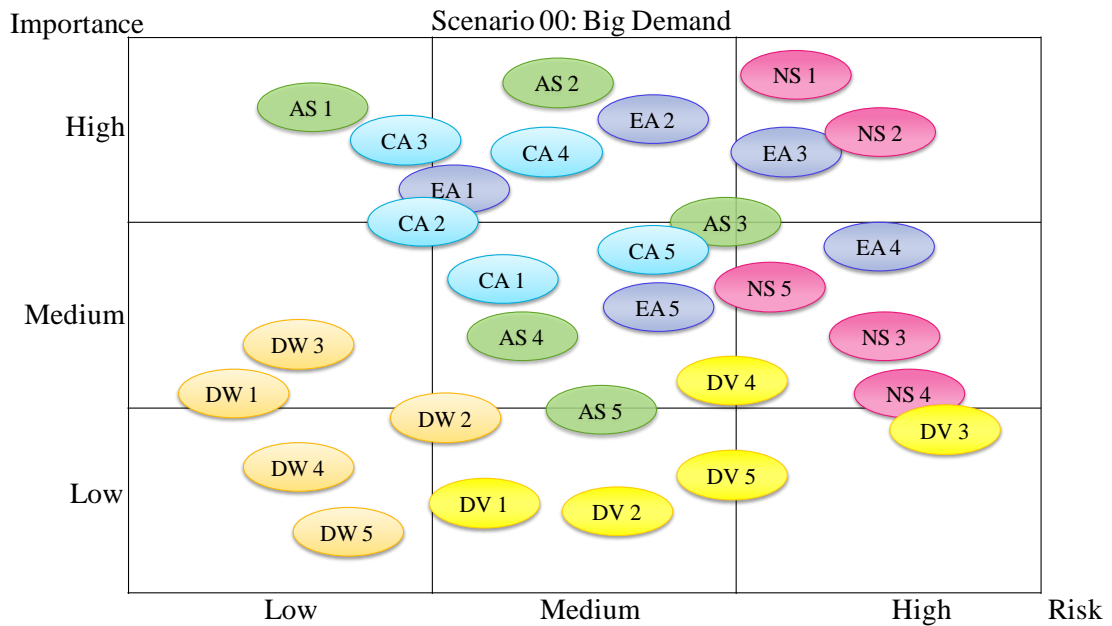
Dimensions	Indicators	Low Level	Medium Level	High Level
Importance	Global market size	< US\$1B	US\$1B~US\$10B	> US\$10B
	Enterprise adoption ratio	< 10%	10%~60%	> 60%
Risk	Entrance barrier	1~3 points	4~6 points	7~9 points
	Strength of industry	1~3 points	4~6 points	7~9 points

4.5 Technology Planning Implications

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 technologies would be of high importance and low 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 next big opportunity for IT industry, these technologies are new to most Taiwanese enterprises and the adoption of them is considered highly risky.

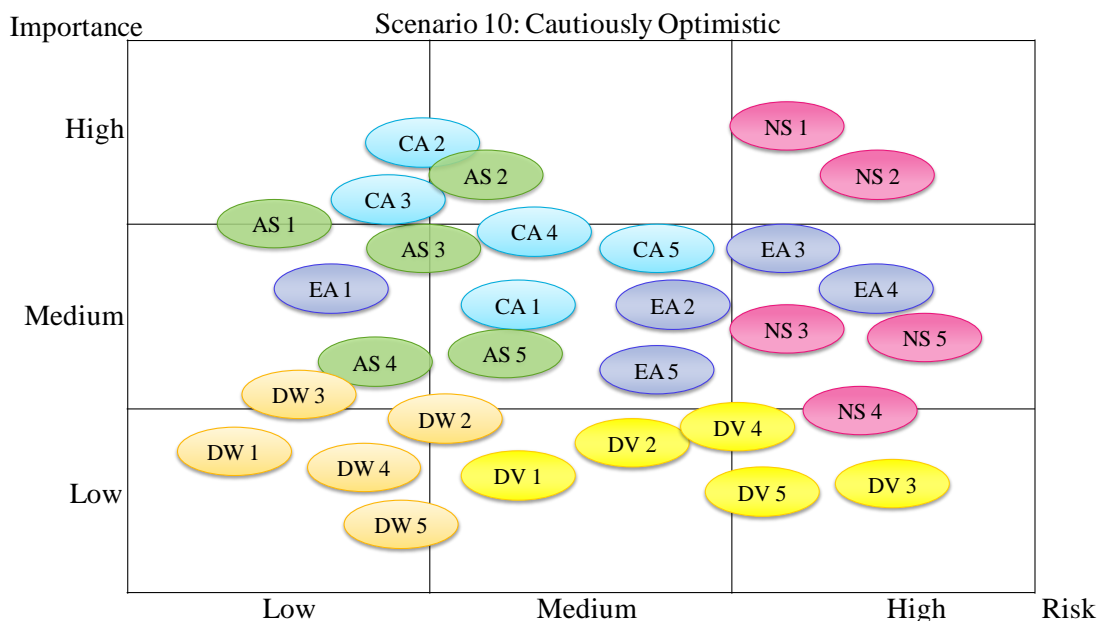
Figure 2. Technology assessment for Scenario 00: Big Demand



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 economical outlook is not as good as in the Big Demand scenario. However, the advancement of Big Data related technologies is not slowing down. Global major vendors and academic institutes will keep the Big Data R&D in progress, and await for the next cycle peak of market demand.

Figure 3. Technology assessment for Scenario 10: Cautiously Optimistic



5. Conclusion

In this study, a systematic approach of deriving the possible Big Data technology portfolio planning for the next five years is conducted. The highlight of findings are:

1. Depending on the future global IT spending outlook, Big Data technologies will have different importance and risk ratings.

2. NoSQL technologies are the Big Data technologies that will obtain high consideration with respect to both importance and risk.

3. If the market demand for Big Data technologies is not strong, Advanced SQL technologies will have lower risk.

4. Data Warehouse technologies are mature technologies. The risks are low in both scenario, and when the market is not booming, they will obtain higher importance considerations.

5. Scenario changes will have less impact on the Cloud Analytics, Embedded Analytics, as well as Big Data Visualization technologies. These are technologies with potential and worth of attention.

Based on these results, the strategic thinking of Taiwanese industry toward leveraging Big Data technologies for competitive advantages can be initiated. For example, these findings suggest that NoSQL technologies should have higher priority in the agenda of developing Big Data technologies of an organization.

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, and avoid 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.

References

- Adrian, M. (2012). Who's Who in NoSQL DBMSs, *Gartner Report G00228114*.
- Baer, T., Sheina, M., & Mukherjee, S. (2011). What is Big Data? The Big Architecture. *Ovum Report, OI00140-033*.
- Baer, T. (2011) 2012 Trends to Watch: Big Data. *Ovum Report, OI00140-041*.
- Bishop, P., Hines, A., & Collins, T. (2007). The current state of scenario development: an overview of techniques. *Foresight*, Vol. 9, No. 1, 2007, pp. 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>.
- 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.
- 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, pg. 540-555, <http://www.itcon.org/2009/35>.
- 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.
- Mietzner, D., & Reger, G. (2005). Advantages and disadvantages of scenario approaches for strategic foresight. *Int. J. Technology Intelligence and Planning*, Vol. 1, No. 2, 2005.
- Mukherjee, S. (2012). Deploying Big Data Systems. *Ovum Report, OI00140-035*.
- Russom, P. (2011). Big Data Analytics. *TDWI Research*, 4th Quarter, 2011.
- SRI. (1996). Integrated Technology Planning, *SRI International Report*, Menlo Park, California.
- Yu, O. (2006). *Technology Portfolio Planning and Management*. Springer Publisher, Berlin.