

월마트 사례를 통해 다시 보는 데이터 웨어하우징

Data Warehousing Revisited: A Wal-Mart Case

박진수 (Jinsoo Park)

서울대학교 경영대학/경영전문대학원

최민정 (Minjung Choi)

서울대학교 경영대학

요약

데이터 웨어하우징(Data Warehousing, DW) 개념이 탄생한 1970년대는 미국 유통 대기업 월마트가 폭증한 데이터베이스를 좀 더 효율적으로 보고자 했던 시기와 맞닿아 있다. 그러나 지금까지 월마트의 데이터 관리에 관한 논의는 대부분 생산관리, 재고관리 등 운영 측면이 강조되어 왔고 데이터 자체에 대한 관심은 상대적으로 적었다고 할 수 있다. 본 연구는 데이터베이스에 관한 기술적인 개념을 먼저 아우른 뒤 월마트의 데이터 웨어하우징 프로젝트 사례를 구축 전(준비)-구축(실행)-구축 후(평가) 단계로 나누어 분석하였다. 본 연구는 월마트가 데이터 웨어하우징을 진행하면서 조직에서 지속적으로 발생하는 문제를 어떻게 해결하였는지, 그리고 데이터 관리를 통해 어떻게 전사적인 공헌을 이룩하였는지에 초점이 맞추어져 있다. 산업을 막론하고 전 세계적으로 가장 큰 데이터베이스를 가지고 있는 월마트의 사례를 현재의 관점에서 다시 살펴봄으로써 최근 재조명되고 있는 (빅)데이터 이슈와도 연결점을 찾을 수 있을 것이라 기대한다.

키워드 : 데이터 웨어하우징, 데이터 관리, 빅 데이터

I. Introduction

The concept of data warehousing was born in the 1970~1980s at the same time when Wal-Mart's growing database gave birth to its need to view records in a more effective way. As the world's largest retail warehouse, Wal-Mart represents the world's largest data warehouse. It is estimated to be more than 2.5 petabytes as of 2010 (*The Economist*, 2010). The company has been leading in the use of data-

base management techniques in catering to business needs and it is non-arguable that behind its success was always "data" above marketing, finance, or operations. A large portion of the discussion around Wal-Mart's database management heretofore has only been process-centric, however, that is, inventory management or sales forecast, not its data *per se*. The current study largely involves technical concepts around database and takes a deep look at Wal-Mart's chronicle of pre-implementation, implementation, and post-implementation of the data warehouse. The highlight underlies how Wal-Mart acted against solving endless problems and was able

† The study was partly supported by Brain Korea 21 fund.

to gear their data management to business contribution. By revisiting their data warehousing project case from a current point of view, we expect their lessons to be still available and applicable. With the renewed interest in (big) data and business intelligence, Wal-Mart's high respect for data and technology holds a lot more meaning than a couple of decades ago.

II. Research Background

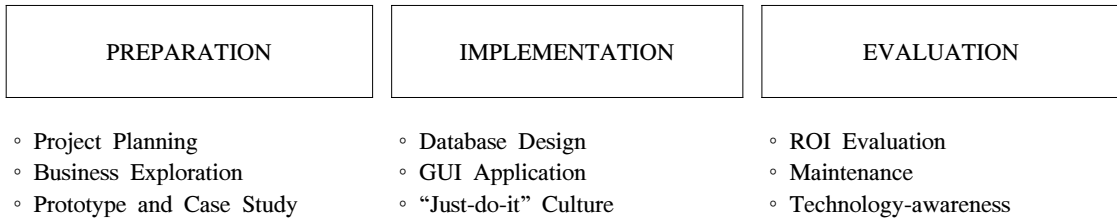
In prior literature, we have seen abundant research that stresses the importance of the alignment of business and technology. Reich and Benbasat (2000) investigated factors that influence the social dimension of alignment between business and information technology (IT) objectives: shared domain knowledge between business and IT executives, IT implementation success, communication between business and IT executives, and connections between business and IT planning processes. Luftman (2003) tried to assess a company's IT-business alignment by developing six criteria, that is, communications maturity, competency/value measurements maturity, governance maturity, partnership maturity, technology scope maturity, and skills maturity. According to the survey commissioned by the Society for Information Management (SIM) and analysis of responses collected from its members, IT and business alignment ranked No. 1 among the top five management concerns (Luftman, 2005). In classifying the expected roles of the Chief Information Officer (CIO), Lee, Park, and Choi (2010) found that the emphasis has been primarily moved from the technology competence to the structured relationship with non-technology managers. It is worth noticing that most of the studies have largely focused on the communication between IT and busi-

ness while practitioners on the other hand report the lack of it.

Wal-Mart is all the more a valuable example. The case showed several factors that contributed to the success of Wal-Mart's database management. Here, database management denotes not only the narrow concept of database but also the environments where data is processed and needed (Davenport and Snabe, 2011). Many a discussion around Wal-Mart's database management, however, has been only process-centric, that is, inventory management or sales forecast, not its data *per se*. The company put considerable efforts into identifying business needs prior to database design and actual implementation. They were also very smart at using technology and open to technology development departments' requests to connect their recognized business needs and IT capacity. The evaluation was completely company-wide, not confined within a certain level or a department because Wal-Mart knew by heart that it was the biggest issue. They strictly collected users' response and reflected their needs afterwards. The ROI report was used to assess the value of the previous system as well as to develop an advanced system. Wal-Mart's "not-to-stop" culture played a key role in continuously modifying errors and making gradual improvements despite countless obstacles, which finally let the initial design last until three years later.

III. Research Approach

Yin (1981) has provided extensive explanations on how a case study ought to be carried out. We adopted his suggestions in constructing and organizing the artifacts that consist of the case scenario for this current study to be "conducted systematically" (Yin, 1981, p. 64). Following this section, we



〈Figure 1〉 Conceptual Framework

analyzed Wal-Mart’s data warehousing project with regards to three distinct but congruent stages: preparation, implementation, and evaluation. The conceptual framework of the current study is provided to aid in understanding in <Figure 1>.

Most of the numbers and figures served as a back-up to our descriptive narratives, which we would like to call a “triangulation” method. As we agree with Jick (1979) in that “qualitative and quantitative methods should be viewed as complementary rather than as rival camps,” quantitative data has been presented when necessary to balance against the qualitative nature of a case study.

The analytical dimension of this study mostly revolves around organization-level. We would like to object to the commonly-held belief that information technology project is to be prepared or performed by the IT department only. Because data warehousing project is inherently a company-wide plan as well as it is the organization’s responsibility to consider the consequences, we attempted to view the case subject as a whole, not a separate component.

IV. History of Data Warehousing and Wal-Mart’s Early Database Systems

The most formal definition of a data warehouse (DW) is “a subject-oriented, integrated, time-variant

and non-volatile collection of data in support of management’s decision making process” (Inmon, 1992). A more concise, still reputable definition is “a copy of transaction data specifically structured for query and analysis” (Kimball, 1996). Put it simply, data from older systems is copied into a new computer system dedicated entirely to analyzing that data. Normally, the data warehouse stores a substantial amount of historical data. Users of this system are able to continuously ask or query it to retrieve data for analysis. The intent of data analysis is to better understand what is happening or what has happened within an organization. The value of better understanding is better decision making, which brings a tangible value to almost all companies in almost all industries.

However, building a data warehouse is not as simple as understanding its concept. Like every other technological implementation, a data warehouse should align with the needs and goal of the organization. Ordinary project deliverables, milestones, timelines, and project management apply to its construction. However, the previous manager of Wal-Mart Paul Westerman emphasizes the difference between a data warehouse application implementation and a typical functional application implementation from a technological standpoint. The key is that data warehouse development is focused solely on information delivery whereas classical applications are focused on functional efficiency (Westerman, 2001).

The first task that the computer evolution faced in the data management history was automation of mundane tasks, such as data entry. With the advancing technology, the focus has been moved to controlling and making basic decisions using these automated computers. Inventory management is a good example. A third step entailed a fixed type of data presentation. These systems were often called *executive information systems* (EIS). All of these three systems were time-critical *on-line transaction processing* (OLTP) systems that required a very fast, usually under a second, response time.

Wal-Mart created an EIS in 1988 on the IBM platform. The EIS effectively joined data from three functional systems, that is, purchase order system, supplier system, and article system. This application was using the data with each independent application from each source of the corresponding systems. This was the inception of the data warehousing concept, but there was a critical problem: they did not provide enough analytical flexibility.

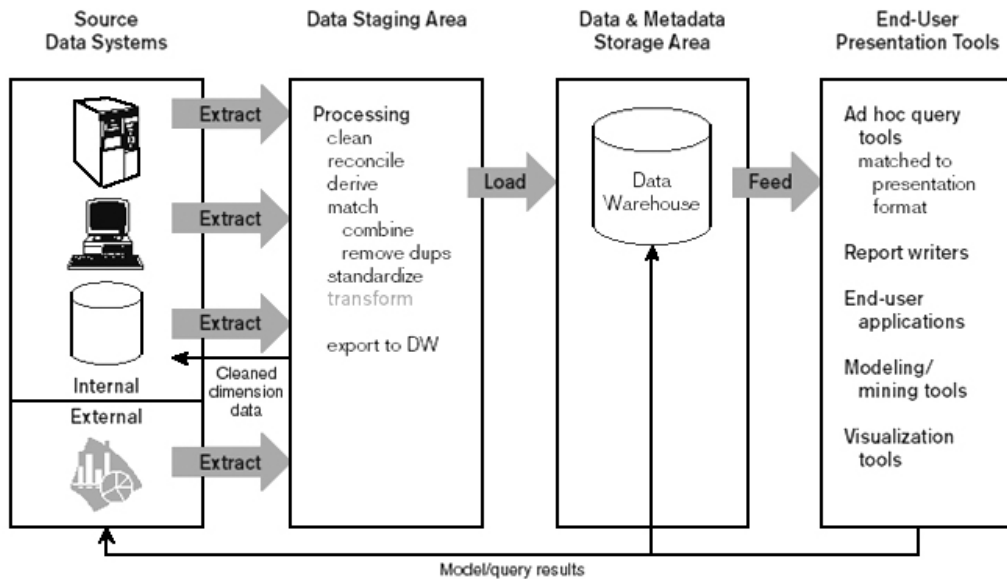
It was very structured but required the business executives to drill down through the data. Technology departments at Wal-Mart (and other companies at the time) thus began to allow the more technically oriented business people to access the data “directly,” using a computer access method called *structured query language* (SQL). The development of this language was a tremendous step forward toward data warehousing in terms of simplicity and flexibility which had been lacking.

Wal-Mart’s advanced users began downloading data from the operational system into their personal computers using SQL. The spreadsheets were used to create complex graphs which enabled executives to see obvious trends more quickly than before. This is what many companies still do today. However, another problem arose. Business executives

needed to be trained properly to understand how to retrieve the data effectively. This was important because the SQL data retrieval process could cause conflict with the operational and controlling applications. For example, it was not uncommon for the EIS to interfere with the order entry system when the executive wanted to review outstanding purchase orders for a specific article. This request might require all records in the entire purchase order database to be read, causing the order entry application to wait or compete for the same data. The result was a slower response time for the order entry application, which was by no means acceptable for the order entry department. They were quick to complain when the response time was over one or two seconds.

To solve this contention problem with the other applications, data in these systems was copied into an altogether different database, even onto a different dedicated computer in 1990 at Wal-Mart. The users were thus allowed complete access to the data without impacting the automation and controlling applications. Working within a separate database also gave the chance to “clean” the data. This was essentially the birth of actual data warehousing, as there is one very clear difference between an EIS and a data warehouse: the latter is completely separate from functional systems (For better understanding, refer to <Figure 2>). This means that data is copied from the functional OLTP systems, cleaned, transformed, and duplicated into the data warehouse database. The data in that database can then be designed to provide unrestricted read access solely for informational purposes. Early on, there were two main benefits of data warehousing:

1. there was no contention with the functional systems, and
2. historic data could be accumulated.



<Figure 2> Generic Two-Level Data Warehousing Architecture (Hoffer *et al.*, 2002)

In the first place, a query would not interfere with the operational functions no matter how much data was required. Secondly, the functional systems typically maintained only a limited amount of data (mostly because of performance considerations) whereas the data warehouse could maintain historic information. This led to the development of applications to show the graphical representation of historic trends or patterns-without the fear of affecting the OLTP applications.

Another interesting differentiation was that only the OLTP systems concerned response-time requirements while there was no such requirement for a data warehouse. Because business executives did not operate under the same procedural pressures as those using the OLTPs, they could afford to wait for an answer. It was accuracy that they wanted so the speed was secondary. The final difference between functional systems and data warehouses was that the former simply displayed data values in

their database while the latter was used to analyze data from one database or many databases. This type of processing is now known as *on-line analytical processing* (OLAP). The above explanation is summarized in <Table 1>.

V. Wal-Mart's Success Drivers-Preparation

There is plenty of literature that strongly recommends business-technology alignment. The question is "How?" Wal-Mart was very keen at aligning its business goal with the technology. People often like to call Wal-Mart a data company, rather a retailer (McKendrick, 2011), but the company was truly business-oriented in viewing and utilizing its advanced techniques. Its preparation and implementation of data warehouses were grounded upon one common theme, contributing to business, as demonstrated by its project planning, business exploration, and case study phases.

〈Table 1〉 Comparison between OLTP and Data Warehouse Databases (Teorey, 1998)

	OLTP (transactional solutions)	Data Warehouse
Optimized for	Transaction performance	Query performance
Structured for	Data integrity	Ease in querying
Users	Thousands of users	Relatively fewer users (typically under 100)
Size	Generally small (MB up to several GB)	Large (from hundreds of GB up to several TB)
Data	Current data	Historical data
Normalization	Normalized data (many tables, few columns per table)	Denormalized data (few tables, many columns per table)
Update frequency	Real-time (Continuous updates)	Periodically (Batch updates)
Query	Simple to complex queries	Usually very complex queries

5.1 Project Planning

At the beginning of a data warehouse project, technical managers at Wal-Mart were trained to understand what will be happening in the project and, where possible, to plan and allocate time for it. It was based on the belief that the more successful data warehouse can be constructed after business benefits have been illustrated and active corporate sponsorship has begun. A data warehouse implementation can stagnate if it lacks the support from the company so Wal-Mart assures the focus of the data warehouse to be aligned with the business, not to be based solely on a technical perspective.

Wal-Mart also knew that misconceived alignments can also cause problems. Forcing the data warehouse to align with a company goal that does not really fit well can become an issue during the implementation. Sometimes this misalignment causes the data warehouse to be used for functional business processing and not for providing information. Westerman puts a stress. "Keep in mind that the data warehouse is a historical representation of the company's actions and, in most cases, data will be *added* but *not updated*" (Westerman, 2001, p. 34, emphasis added).

Effective communication was another important factor during a data warehouse implementation. At Wal-Mart, the combination of responsible leadership and effective communication was the key for rapid development. The manager of the data warehouse project was chosen as a hands-on person with high contact to the team. He or she held a status meeting every morning during the critical times and weekly afterward. Impromptu meetings were also standard practice when a technical obstacle stopped development progress. Often these meetings would immediately change the work schedule and led the entire team to focus on solving the immediate problem.

There is one story about the communication between business and technology departments (Westerman, 2001). When a data warehouse was initially to be built at Wal-Mart, the executives did not appreciate the value of logical data modeling for the database. Rather, it was perceived as an unnecessary waste of time. The technology team leader and his colleague convinced Wal-Mart management to allow using time to logically model the data before creating a physical model. They were given only one week. During that week, they created and proposed four new logical/physical designs. One of the

designs added the daily data, in nearly the same size database. Then they were given another week to explore the possibility of maintaining daily data. They concluded after a lot of analysis¹⁾ that it was possible to implement daily sales history on the computer Wal-Mart had been using. That database design was adopted without change for the next three years. Shortly after this success, the company created the logical database design group. The story not only exemplifies the company's urgent need to access huge data at the time as well as organizational openness to solving the problem.

5.2 Business Exploration

Getting started with building a data warehouse can be a challenge for many companies. Where do you begin? How should you proceed? Because the investment in a data warehouse is a substantial commitment, many companies will need to assure their management that they can successfully implement the data warehouse. You may require a business case study and/or a prototype (or proof of concept) to be built before implementing a full-sized data warehouse.

Wal-Mart's focus on business can be shown in a prototype system they built with a much smaller computer system than would be needed in the real implementation. After this was functioning, they built a business case based on the success of the prototype. Because the system would be extremely large and therefore more expensive, they also had to obtain business sponsorship. The prototype and the business case helped achieve added funding of the Wal-Mart board of directors. For Wal-Mart, this

ensured the data warehouse was clearly aligned with the business.

Wal-Mart built a prototype based on the needs of the business, and then they received support from the entire company. Technology group worked closely with the merchandising group to create the prototype and only when it was approved did they build the system. By doing so, they had a general idea of what was needed before the start, and the company learned to invest their resources in building a prototype. The previous project leader at Wal-Mart Westerman asks to "speak with the business people" and "prioritize the business needs and deliver a small, but valuable piece of the data warehouse" (Westerman, 2001). He details each of the processes.

1. Define the goals
2. Gather the business questions
3. Prioritize the business questions
4. Define the business question

The goals constitute the basic scope of the data warehouse. Then there will be questions that should be answered to help achieve those goals. Those questions will need to be prioritized. Finally the answer to each question must be documented. To accomplish these steps, Wal-Mart used a workshop approach. In each workshop, one group of business people gathered and expressed opinions in a two-day session. To choose the right mix of people, Wal-Mart had all the participants at the same management level to avoid uncooperative silence and encourage a healthy discussion. It also took explicit care of the number of people to maximize the synergy of the conversation. Each participant and his or her job function were selected carefully prior to workshop.

1) This will be recounted later from a database design point of view in the next chapter.

5.3 Prototype and Case Study

The business case study is a selling document that the IT department needed to communicate the information desires of the merchandising organization. If you have funding to build a data warehouse, you don't need to create this document. It is possible to go directly into the construction phase. However, the main reason companies build a business case study *a priori* is that they are not yet convinced of the value of information and therefore need it to be demonstrated and clearly documented.

As mentioned above, Wal-Mart did build a business case study, but it was based on the prototype, which was built with the merchandising group. By the time Wal-Mart first thought of a data warehouse in 1990, they did not have much expertise in building this type of a system. Wal-Mart needed a technology partner so they hired Teradata Corporation to build a prototype. After the prototype was completed, a business case study was constructed. This business case study contained the business user vision, potential payback, projected investment costs, and a resource plan. The document became the basis for the initial \$20 million investment that was needed for the first implementation (Westerman, 2001).

About a year later, they also built a business case study for "Sam's Club," another division within Wal-Mart, to evaluate and measure the return on investment (ROI) after implementation. It was determined over \$12,000 per query and the company was performing more than 1000 queries a day. That is to say, the calculated ROI²⁾ of the data warehouse was \$12 million a day (Westerman, 2001). This evaluation played as evidence for the value of

2) This will be laid out more in detail in the next chapter.

technology with Wal-Mart stores and let the company realize the quantified payback of data warehousing, thus boosting the development of new applications.

VI. Technical Construction and the Wal-Mart Culture-Implementation

6.1 Wal-Mart's "Just-do-it" Culture

Wal-Mart has invested over \$4 billion supporting the database warehousing technology and is doing more. At Wal-Mart, there is a "just do it" attitude throughout the entire company. This means that the employees take individual responsibility and action to resolve daily issues. They have a custom called "the sundown rule," which guides you must respond to the people who have contacted you by the end of the day. The founder of Wal-Mart, Sam Walton, recalls "most everything I've done I've copied from someone else" (Walton, 1992), which means that almost all of his ideas were from his competitors and improved on them.

The data warehouse was no exception. In fact, Wal-Mart was not the first to have the idea of building a data warehouse. Kmart, a fierce competitor at the time had just a few months earlier announced they were building a merchandising database with historic data down to the store-item-level information. Wal-Mart never had the goal of building the world's largest commercial database, either. They were just using leading-edge technology to solve the business problem they had. The struggle for solution was essential and timing, competition, and initiatives were important matters.

Wal-Mart has always adopted technology quickly when it could improve the business. For a central-

ized organization with an enormous size, they have a very impressive record of technology implementation. When they began building the data warehouse, they were in a very similar situation as other companies. Today, we still see companies in the same predicament that Wal-Mart was in at that time. The “just do it” attitude or the culture of taking leadership initiatives has helped the company survive the challenge and sustain its growth up to now.

6.2 Database Design

Wal-Mart’s approach to data modeling was a bit different from that of other companies. It was different from Wal-Mart’s approach today, too. At that time, it was focused on the ability to save, maintain, and retrieve the huge amount of data in the limited system capacity. The user community within Wal-Mart was asking for weekly sales data. Through many efforts put in the design and implementation, the technology team determined the number of days it would take for the average Wal-Mart product to sell within a week, i.e., 2.3. Put it differently, when an article was sold, it would sell 2.3 times in that week. This meant there could be many articles that would sell only one time in a week. Specifically, there were 4.7 days where no articles were sold in a week. Learning that only 5000 unique articles would sell each day, 20000 in a week, and 40000 in a month was important information given that the database was too big to fit in the machine. Using this knowledge, the technology team used the database compression extensions to reduce the volume even more, by compressing on 0. “Compressing on zero” means the database assumes a default value of zero and therefore does not need to store or allocate physical space for the value of zero. This is important be-

cause it requires as much space to save a large integer value like 1,234,567 as the value 0 unless you cannot save frequently repeated value.

6.3 GUI Application

The GUI team was responsible for building the graphical user interface (GUI) that the user community would use. They were responsible for getting the data out of the database and presenting the analysis to the user. The team started development of the application shortly after the technology team completed the physical design. They were to create an application using personal computers running on Microsoft Windows. This was a huge challenge because only a few experts knew how to do it. Another major problem to address was network problem. Specifically, they could not retrieve *sets* of records. As one had to have all of the primary key values to retrieve a record, it was possible to retrieve one record or many single records, but not a set of records.

Let’s take an example. For the point-of-sale (POS) table, the key values needed to access a record were article number, store number, and week number. If you knew the value of the keys (e.g., article number = 10001, store number = 10, week number = 9301) then the response time to get the record was very fast. However, if you wanted records for article 10001 and store 10, but not specifying the week key, you had no choice but to retrieve a set of sales records for ALL weeks. If you didn’t know this entire key, the database would scan the entire database to find the records that matched.

If it had been a small database, it would haven not be an issue but it was Wal-Mart, a retail giant with already 1 billion records by that time (now more than 100 billion). One technology team mem-

ber took responsibility to solve it. “Taking responsibility” is very different from being assigned the responsibility at Wal-Mart culture. This is the way the team worked. Each team member shared the task. It took a couple of weeks to solve this particular problem—making a Cartesian product to create the keys needed. A Cartesian product is in this case the joining of all of the possible week key values. There are only 52 weeks in a year, so if you join the week table with no constraints you have all the needed key values to directly retrieve the records. What was devised was a statement that would be executed with every SQL statement to force the Cartesian product, which seems a little strange to the eyes of the viewers today.³⁾

Other companies might have stopped the project and waited for the technical problem to be solved. The culture at Wal-Mart, however, demanded that the team quickly take responsibility for the problem, solve, and move on. The Wal-Mart way was not to stop but find a solution and continue. Getting the information into the hands of the business people was the entire focus and this is another drive that enabled fast establishment of a data warehouse, despite the high failure rate.

VII. Data Warehouse and Aftermath-Evaluation

The data warehouse project is never finished at Wal-Mart, or under its “never-stop” culture. After rolling out the system, more and more people want-

3) For those who can read SQL, we provide the statement below. The detailed explanation is illustrated in Westerman (2001, p. 169).
 ((Article number = Store number or Article number NOT = Store number) or
 (Store number = Week number or Store number NOT = Week number)).

ed access to that system. However, they encountered many problems initially. Among them was that everyone was using the same user ID and password to log on and run queries. If everyone has the same ID, how do you know who is running which query? The problem was quickly fixed so that everyone had a different user ID but another issue was that they all still had the same password: walmart. This should frighten any security administrator today but it was a much later priority.

Again, in any other company, this is not what one could possibly imagine. The development would have been closed down until the problems had been resolved. The Wal-Mart culture did not allow stopping and the rollout continued. The company was not concerned about people having complete access to all the data. Detecting the problem and solving it was a much important issue.

7.1 Evaluation

People liked it so much that they asked for the system to be installed in supplier negotiating rooms, too. They used the system in front of the supplier to leverage the pricing agreements. This was a very powerful negotiating tool. One year later, the company decided to calculate the ROI.

Even though there were problems, the overall feedback for the system was positive. The only way to quantify results was to do an ROI analysis. There needed to be an assessment of the influence this data warehouse had on Wal-Mart. Some of the users were asked to what they had done with the system and how they used it. All successful users were asked to document their successes. Finally, there were audits of who was doing the most work on the system. Through the query statistics application, it was possible to know how many queries were

run and by whom. Building on the initial question, the users were asked:

- How were you able to improve the business?
- What analyses did you use?
- What was the financial benefit?
- Could you have done this without the system?
(When there was a financial benefit and it could not have been achieved without the system, it was asked to be further documented.)

The success stories were consolidated into a single document listing the person, his or her job, the analysis that was performed (charts and graphs), the conclusion that was made, the action that was taken, and the financial benefit. All of these were accumulated into an ROI document. All told, a handful of people were selected in the assessment, yet the total ROI for these people exceeded the cost of the initial implementation, which was about \$20 million. To call attention to the value of the system, they broke down the number even further, on a per-query basis. The company earned over \$12,000 per query. Moreover, they were running over 1000 queries per day. Of course, not all queries were from experienced power users. Regardless of the percentage of power users, that is a huge ROI. With an ROI of \$12000 per query, you can see why Wal-Mart continued its data warehouse development.

7.2 Maintenance

Wal-Mart was having performance problems after the ROI calculation was completed. The technology team was working to resolve these issues but it simply took time. They had added what they called *deferred queries* to enable a more stable environment for users. Deferred queries meant exactly

what the name implied: it allowed the user community to defer the execution of a query until later. That way, they would not have to stay connected to the bouncing network and database. Teradata Co. actually developed the code and Wal-Mart did the alpha, beta, and production testing simultaneously because they put the application directly into production when it was functional. Instead of simply allowing the user to press the Run Query button, the system was modified to allow the execution of the query without staying connected. The deferred queries application would run the query and save the results in the database. The users were allowed to see the status of the query and could retrieve the results after the query completed. Deferred queries also allowed user to run the queries overnight if they wished as well as a few other options. The important part of the deferring a query was that users did not have to leave it on their PC. They could submit the query and check it later, whenever they wanted. This is important because Wal-Mart suffered network problems. When users did not have to stay connected to the network for the duration of the query run time, the query did not fail as often.

The deferred query application is now a standard part of many off-the-shelf applications. You can buy the product and have his functionality immediately. Companies that adopt technology early will solve many technical problems before others adopt the technology. Almost everything that Wal-Mart built regarding data warehousing over the years is now available in an off-the-shelf product.

7.3 Technology as a Competitive Advantage

As information needs grow, in general, a company will view its ability to utilize this information

as a competitive advantage. The priority given to the previously neglected information center and the technology department will increase as well. This is such a predictable scenario when the business learns more about the significance of information. Surprisingly (or unsurprisingly?) this is rarely seen in reality. In contrast, Wal-Mart had the firm belief that the technology they have is a strategic advantage over their competitors. The company showed its support for technology by investing more in technology as opposed to that in some other area, such as inventory⁴). How did this affect the organizational structure? It changed the function of the information center. As it evolved to be able to suggest and document the needs of the business users, its function became more like that of a business analyst. A major advantage for both the technology department and the business department was that the information center would learn and teach more about the daily functions of the business people to enable better, information-driven application development.

The company had learned how to apply technology to the business value and extended this 20-year-old experience to every way of doing their functional jobs. Its data-driven decision making has affected many of the strategy undertakings, most importantly, pricing and location. For example, they calculated that elasticity of the marginal cost associated with an additional sale with respect to sales volume is approximately -0.2; in other words, a 10 percent increase in total sales volume decreases marginal cost by 2 percent (Basker and Van, 2007).

4) Wal-Mart was a bit slow to add inventory figures because they did not have them until after the replenishment system was fully implemented. Westerman (2001, p. 171) remembers it was after 1991 that Wal-Mart upgraded the system.

Also, Wal-Mart located its stores in places where it expects to be profitable, taking into consideration information about demand and cost factors as well as the competitive environment and how it is expected to evolve after Wal-Mart's entry (Holmes, 2006; Jia, 2005).

VIII. Discussion: Challenges Faced

The amount of data stored by businesses doubles every 12 to 18 months. The very biggest companies, those at or near the 100 terabytes, probably triple every three years (Babcok, 2009). Databases are not just getting bigger, however, but they are also becoming more real time. Data also is coming from increasingly complex sources. Businesses are thus destined to deal with the complexities of engineering databases that combine historical and real-time data from multiple sources. Designing and building the hundreds, even thousands, of tables that make up multi-terabyte databases and the queries used to extract useful knowledge can test the technical and management skills of any database administrator. But the advantages of big databases are obvious: Most of the largest are data warehouses for analytical tasks where more, and more-detailed, data means better insights. With real-time or near-real-time data, the value of those insights will increase exponentially. Dan Phillips, Wal-Mart's Vice President of information systems says, "We know how many 2.4-ounce tubes of toothpaste sold yesterday, and what was sold with them" (Babcok, 2009). Wal-Mart now refreshes sales data hourly, adding a billion rows of data per day, allowing more complex searches.

In the midst of this, earlier in 2011, Wal-Mart announced its intentions to acquire Kosmix, a start-

up company focused on social media. When this announcement happened, many analysts were surprised and were analyzing the key reasons for a giant retailer to acquire a startup company.

After acquisition, the Kosmix team operates as part of the newly formed company called Walmart Labs. The fundamental technology Kosmix was building is semantic analysis. By applying semantic analysis to social media, Kosmix understands the connections between people, topics, location and products.

With its ability to search and analyze connections in real-time data streams to deliver highly personalized insights to users, the giant knowledge base captures information and relationships about entities such as people, events, topics, products, locations and organizations. In a recent interview with the press (Geron, 2011; Roush, 2011), Anand Rajaraman, Head of WalmartLabs, says:

Wal-Mart is the biggest retailer in the world, but they are not the number-one player in e-commerce-Amazon is. About a year ago, Wal-Mart decided that e-commerce is a strategic priority. It's not like they had not been investing in e-commerce, but they said, 'It's time to go to the next level.' (...) The goal of WalmartLabs is to make sure that Wal-Mart is at the forefront of "e-commerce 2.0," so that we help define it rather than playing catch-up.

As he said, Wal-Mart is still seeking innovation as a frontier, not as a me-too company. He added in the same interview, "You'd be surprised. Wal-Mart has been one of the most innovative companies." By analyzing social media activity, they are expecting to make recommendations about products, events or any other activity that the user is in-

terested in.

While the idea sounds great, doing this in reality is a huge problem-it has always been such, but especially since there are thousands of data pieces flowing from live data sources including tweets, Facebook posts and blogs. The data flow is so fast that the company could not rely on the traditional framework that is typically used to solve "big data" problems.

To address this big and fast data problem, WalmartLabs developed its own in-house proprietary solution code named Muppet, which processes streaming fast data in a lightening fashion, over large clusters of machines (Srikanth, 2011). Muppet can manage and track data streams with billions of updates a day. Using this technology, Wal-Mart will be able to track social media mentions on locations, people or products and use this intelligence to better tailor its product offerings.

As Wal-Mart (as some analysts say) is the future of the giant retailer, WalmartLabs is in an interesting position as more and more people are increasingly being influenced by what their friends say on Facebook or what a group of strangers tweet on Twitter. If Wal-Mart can tap into this torrent of information, and can use this to understand customer behavior better, it can redefine shopping experience.

Indeed, a month after the acquisition, WalmartLabs launched a Facebook app named Shopycat, for social gift suggestions (Geron, 2011). The new app is part of WalmartLab's broader move to bring social and mobile technologies into Wal-Mart. But, how does an algorithm determine what is an appropriate gift for an individual unique person? What may be a special gift to one person may be not such a great choice to another. It is not as simple as it sounds. As senior vice president of Wal-Mart

global ecommerce and cofounder of Walmart Labs Venky Harinarayansays, “Capturing that in an algorithm is a challenge.” A new challenge faced with Wal-Mart in the next decade would be not a matter of data volumes or manipulation but how to figure out data and act on it.

IX. Conclusion and Future

Wal-Mart has continued to add data since 1989. By 2000s about 99% of their data elements were maintained in their enterprise data warehouse that they built in the initial year of the project. An eye for technology was definitely a key element in its success. Wal-Mart’s investment in IT has definitely increased its own productivity.

Now, it is difficult to imagine that Wal-Mart would not keep track of daily records. No company better illustrates the advantages of leveraging massive volumes of data for competitive advantage than Wal-Mart, which now operates a data warehouse with 583 terabytes of sales and inventory data built on a massively parallel 1,000-processor system (Babcock, 2009). While some companies might consider having more than half a petabyte of data to be overkill, at Wal-Mart it is the way to do business. Solid practices for managing stored data, properly indexed tables for query processing, and good data extraction, transformation, and loading techniques to ensure that users are able to correctly interpret data have been the keys to keeping this huge database from getting out of control. Of course, its downside to too much data lingers. The computer hardware, software, and storage needed to manage more than tera- or peta-level of data can run into the millions of dollars. With its big and fast stream, data accumulation can be useless without enough insight and analytical skill to interpret, not only in a cor-

rect way but also in a timely manner. However, it is still the very company that makes the best use of data and related technology in the world. What will be the next move of this company? Will it sustain growth as it has always enjoyed? Or will it eventually fail due to too much information? The non-stopping company is still on the way.

Reference

- Babcock, C., “Data, data, everywhere”, *Information Week*, January 2006, <http://www.informationweek.com/news/175801775>, Last access January 2012.
- Basker, E. and P. H. Van, *Putting a smiley face on the dragon: Wal-Mart as catalyst to U.S.-China Trade*, unpublished paper, University of Missouri, 2007.
- Basker, E., *The causes and consequences of Wal-Mart’s growth*, unpublished paper, University of Missouri, 2007.
- Davenport, T. H. and J. H. Snabe, “How fast and flexible do you want your information, really?”, *MIT Sloan Management Review*, Vol.52, No.3, 2011, pp. 57-62.
- Economist, “Data, data everywhere”, *The Economist*, February 2010, <http://www.economist.com/node/15557443>, Last access January 4, 2012.
- Edwards, D. J. A., “Types of case study work: A conceptual framework for case-based research”, *Journal of Humanistic Psychology*, Vol.38, No.3, 1998, pp. 36-70.
- Geron, T., “WalmartLabs launches Shopycat for social gift suggestions”, *Forbes*, December 2011, <http://www.forbes.com/sites/tomiogeron/2011/12/01/walmartlabs-launches-shopycat-for-social-gift-suggestions/>, Last access, January 2012.
- Hoffer, J. A., M. B. Prescott, and F. R. McFadden,

- Modern Database Management*, 6th ed., Prentice Hall, 2002.
- Holmes, T., “The diffusion of Wal-Mart and economies of density”, *unpublished paper*, University of Minnesota, 2006.
- Inmon, B., *Building the Data Warehouse*, Wiley, 1992.
- Jia, P., “What happens when Wal-Mart comes to town: An empirical analysis of the discount industry”, *unpublished paper*, Yale University, 2005.
- Jick, T. D., “Mixing qualitative and quantitative methods: Triangulation in action”, *Administrative Science Quarterly*, Vol.24, No.4, December, 1979, pp. 602-611.
- Kimball, R., *The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*, John Wiley and Sons, 1996.
- Lee, S., J. Park, and M. Choi, “Technician or leader?: A review of the evolving role of the CIO”, *Seoul Journal of Industrial Relations*, Vol.20, 2010, pp. 157-179.
- Luftman, J., “Assessing IT/business alignment”, *Information Systems Management*, Vol.20, No.4, 2003, pp. 9-15.
- Luftman, J., “Key issues for IT executives 2004”, *MIS Quarterly*, Vol.4, No.2, 2005, pp. 269-285.
- McKendrick, J., “Retailer or a data company? Wal-Mart is now both”, *IBM Smart Planet*, December 2011, <http://www.smartplanet.com/blog/business-brains/retailer-or-a-data-company-wal-mart-is-now-both/20850>, Last access, January 2012.
- Ohlinger, P., “Wal-Mart’s data warehouse”, in *Proceedings of the 2nd Student Conference on Data Warehousing (SCODAWA 2006)*, June 2006.
- Reich, B. H. and I. Benbasat, “Factors that influence the social dimension of alignment between business and information technology objectives”, *MIS Quarterly*, Vol.24, No.1, March 2000, pp. 81-113.
- Roush, W., “Inside WalmartLabs: How the former Kosmix team plans to help the world’s largest retailer get social and mobile”, *Economy*, January 2011, <http://www.xconomy.com/san-francisco/2011/08/01/inside-walmartlabs-how-the-former-kosmix-team-plans-to-help-the-worlds-largest-retailer-get-social-and-mobile/>, Last access, January 2012.
- Schuman, E., “At Wal-Mart, world’s largest retail data warehouse gets even larger”, *Enterprise Applications News*, October 2004, <http://www.eweek.com/c/a/Enterprise-Applications/At-Wal-Mart-Worlds-Largest-Retail-Data-Warehouse-Gets-Even-Larger/>, Last access January 3, 2012.
- Schuman, E., “Wal-Mart plans its 4 PB data warehouse”, *Enterprise Applications News*, August 2007, <http://www.eweek.com/c/a/Data-Storage/WalMart-Plans-for-Its-4PB-Data-Warehouse/>, Last access January 3, 2012.
- Srikanth, R. P., “How Walmart plans to use big data”, *Information Week*, December 19, 2011, http://www.informationweek.in/Storage/11-12-19/How_Walmart_plans_to_use_Big_Data.aspx, Last access January 3, 2012.
- Teorey, T., *Database Modeling and Design*, 3rd ed., Morgan Kaufmann Publishers, 1998.
- Walton, S. and J. Huey, *Sam Walton: Made in America*, Doubleday Publishing, 1992.
- Westerman, P., *Data Warehousing: Using the Wal-Mart Model*, Morgan Kaufmann Publishers, 2001.
- Yin, R. K., “The case study crisis: Some answers”, *Administrative Science Quarterly*, Vol.26, No.1, March 1981, pp. 58-65.

Data Warehousing Revisited: A Wal-Mart Case

Jinsoo Park* · Minjung Choi**

Abstract

The concept of data warehousing was born in the 1970-80s at the same time when Wal-Mart's growing database gave birth to its need to view records in a more effective way. As the world's largest retail warehouse, Wal-Mart represents the world's largest data warehouse. A large portion of the discussion around Wal-Mart's database management heretofore has only been process-centric, however, not its data per se. The study largely involves technical concepts around database and takes a deep look at Wal-Mart's chronicle of pre-implementation (preparation), implementation, and post-implementation (evaluation) of the data warehousing project. The highlight underlies how Wal-Mart acted against solving endless problems and was able to gear their data management to business contribution. By revisiting their data warehousing project case from a current point of view, we expect their lessons to be still available and applicable. With the renewed interest in (big) data and business intelligence, Wal-Mart's high respect for data and technology holds a lot more meaning than a couple of decades ago.

Keywords: *Data Warehousing, Data Management, Big Data*

* Graduate School of Business/College of Business Administration. Seoul National University

** College of Business Administration. Seoul National University

◎ 저 자 소 개 ◎



박진수 (snumj04@snu.ac.kr)

The University of Arizona에서 경영정보시스템을 전공하여 경영학 박사를 취득했으며, University of Minnesota의 Carlson School of Management에서 조교수, 고려대학교 경영대학에서 조교수를 역임했다. 현재 서울대학교 경영전문대학원/경영대학에 부교수로 재직 중이다. 현재 국제저널인 Journal of Database Management의 편집위원으로 활동하고 있으며 MIS Quarterly, IEEE Transactions on Knowledge and Data Engineering(TKDE), IEEE Computer, ACM Transactions on Information Systems(TOIS), Data and Knowledge Engineering, Journal of Database Management 등 국내외 우수 전문학술지에 다수의 논문을 게재하였다. 주요 관심분야는 온톨로지, 빅 데이터 관리, 정보 시스템 통합, 지식 공유, 데이터 모델링 등이다.



최민정 (minjungchoi.is@gmail.com)

서울대학교 경영학과를 졸업하고 동 대학원에서 석사(MIS)를 마쳤다. 출간일 현재 Temple University Fox School of Business에서 박사과정을 밟고 있다. 관심분야는 digital innovation, 지식관리시스템 등이다.

논문접수일 : 2012년 04월 04일

게재확정일 : 2012년 08월 11일

1차 수정일 : 2012년 06월 28일