Sindh Univ. Res. Jour. (Sci. Ser.) Vol. 50 (3D) 110-115 (2018)

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

A Metadata Architecture over cloud for enabling knowledge support for SMEs

Z. A. MAHER, A. P. MALIK*, G. A. RAHU, S. SHAIKH, P. K. BUTT, I. A. MAHAR, H. SHAIKH

Information Technology Centre, Sindh Agriculture University Tandojam, Pakistan

Received 10th June 2018 and Revised 15th September 2018

Abstract: Recent surges in the global economy led to organizations adopting decision support systems as an integral part of their daily operations. These decision support systems were initially based on operational data store, but as the time and requirements progressed, a need was felt to embed knowledge into the data warehouses that resulted in the birth of Knowledge Warehouses. The resultant was a bonded application that facilitated decision makers to not only process on knowledge but also share it on organizational level. However, the resultant bonded sharing evolved an architecture that although was good but as it required large organizational knowledge base and extensive infrastructure hence was suitable for large organizations. As a result, SMEs were left behind as they could not afford expensive hardware and neither did they have large organizational base to develop large organizational knowledge base. What they required was a low-cost infrastructure that would enable them to have knowledge support and have the ability to gather knowledge base from their competitors. Until recently, no suitable architecture was present to provide such a facility to SME, but the advent of hybrid main memory database has now made it possible. The aim of this study is to provide an architectural add-on to the already available hybrid main memory database (HYRISE) that will solely be used to store knowledge and will have the ability to communicate with SMEs sister concerns over cloud. T study focuses on building a new meta-data architecture that will work in-conjunction with the original meta-data architecture of HYRISE to store knowledge information and have the ability to communicate securely over cloud.

Keywords: Decision Support Systems, Main Memory Database, (SMEs), Meta-data Architecture, HYRISE Hybrid Main Memory Database, Cloud, Security

1. INTRODUCTION

The advent of new modern age has to change the dynamics on which things go about in everyday life of a person. Similar is the case for the enterprises and companies that they need to be more adaptive and more dynamic in nature. For the companies to cope to this dynamic nature they need personals with good decision making capabilities so that they can steer themselves through thick and thin towards a bigger success. The decision makers face a daunting task on everyday basis, where they are constantly looking to gather knowledge and ways of how to improve their efficiency so that to get maximum output be given to the company. For decision making they need help in the form of analytical tool that can help them analyze and support their daily decisions (Nemati, *et al.,* 2002). To make it possible Data Warehouses came into existence that helped decision makers with data analysis. The data warehouse though had the ability to extract, transform and load data from different sources but it lacked knowledge and as said by (Nemati *et al.,* 2002) it lacks knowledge that is available in the firm's intellectual mind because it constitutes only a fraction of information. Over recent years, multiple advancements have been made in this context and people have now realized that knowledge organizing is important in acquiring, creating and sharing of knowledge for formulating strategy and making strategic decisions (López-Nicolás and Meroño-Cerdán, 2011).

Until recently, it was thought that same knowledge warehouses or systems that could cater for large organizations could be used for enterprises (Massa and Testa, 2011). But the mere fact that a knowledge warehouse can be scaled down so that it can be a replica of large company's experience (Massa and Testa, 2011) is not correct. As a result there is a general consensus that unlike large companies SMEs have fallen behind and benefits of KM have not been fully exploited by them (Evangelista, Esposito, Lauro, and Raffa, 2010). The mere fact that SMEs do not require this kind of knowledge is also not true because they need knowledge to enable them to acknowledge and understand the requirements so that they can survive, grow and have a sustainable competitive advantage over their competitors (Evangelista *et al.,* 2010). As a result multiple efforts have been made in identifying the gaps as to why knowledge is not managed properly by small enterprises. Small enterprises are known to have limited resources and work within very confined limits of the unlike large organizations. We need to realize the fact that small enterprises work on different scale

*Faculty of Information & Communication Technology, International Islamic University Malaysia

and have different requirements with respect to infrastructure, user complexity, fast changing dynamic environments and difference in data access necessity (Grabova, *et al.,* 2010). They need a low cost infrastructure (Lee and Lan, 2011) Hence require a cheap, light weight, flexible, simple and efficient solution (Grabova *et al.,* 2010).

Another very important issue that SMEs face is with respect to organizational memory as they are limited in resources in terms of manpower. Due to this restriction they tend to outsource requirements i.e. turning specific function over to another party to implement and support for a fee (Dragoi, *et al.,* 2011). To manage their work outsourcing of services like that of recruiting, job-hunting, knowledge exchange and development depending on the objective of the enterprises is done. Although outsourced, but the SMEs can only function successfully if they perform as a single integrated unit and also keep their independence so that they can exchange information and inter-operate with their partners in real-time. To have this ability they need an infrastructure that has the ability to communicate over Cloud infrastructure so that virtual environments can be made for communication with typical requirements of Quality of Service, High-availability and Security (Dragoi *et al.,* 2011).

So far, however, there has been little discussion about the fact as to what will best suitable architecture for the SMEs so that they have the ability to store and utilize knowledge to their advantage. To provide decision makers of SMEs with a decision making tool that will be light-weight and cheap we need to shift from conventional database architecture to a more modern architecture namely Main-Memory or In-Memory databases (Loos *et al.,* 2011). This main stream shift is mainly owned to the cheap and extraordinary growth of main memory after the processor advancements stopped way back in 2002, when the processor clock hit 3 GHz (Plattner, 2009) that resulted in the introduction of blade servers. This led to extraordinary growth of main memory and parallel computing, which in-turn led the researchers to focus back on once dormant area Main Memory Databases and the igniter came when HANA an In-Memory Database was proposed (Loos *et al.,* 2011).

2. BACKGROUND

2.1 KM in SMEs and Data Warehouse OLAP

Like knowledge, it is also difficult to explain the managing knowledge in terms of definition and conceive. However, (Davenport andPrusak, 1998) mention it as a fluid mix of framed experiences, values, contextual information, and expert inside that provides a framework for evaluating and incorporating new experiences and information. On an organizational level it is the information that is available in documents or repositories, organizational routines, processes, practices and norms that are being followed (Evangelista *et al.,* 2010). Usage of this knowledge within the current dynamic environment is very essential and its importance and utilization by decision makers in daily routine is increasing day by day (Nemati *et al.,* 2002). As a result this ingredient has hence become a managerial literature for the organizations success (Massa and Testa, 2011). The problem that however has been faced is that literature has traditionally been focused on the domain of larger organizations (Massa andTesta, 2011), (Stonebraker, 2011), (Evangelista *et al.,* 2010) and most of the researchers have not paid enough or little relevance in small and medium enterprises. The reason for this is that KM in enterprises was thought to be a small scale model of bigger organizations, which in fact is not true (Massa and Testa, 2011).

As identified by (Dragoi *et al.,* 2011), one of the major reason why warehouse constructed or built for organization cannot be used by enterprises is the availability of less manpower within an organization. Fewer resources transform into a problem when deploying a large-scale warehouse on enterprises because few resources mean less intellectual knowledge available. As mentioned by (Evangelista *et al.,* 2010) the research on KM can be structured down into (i) SME Entrepreneur, (ii) the knowledge systems and routines embedded within the context of the firm and their embedded networks These enterprises have distinct working on their own but also need to communicate within themselves on a single grid in one single unit (Dragoi *et al.,* 2011), and as proved cloud support needs to made available for enterprises to work effectively The under shown diagram **(Fig.1)** shows the exact location as where the knowledge information is identified.

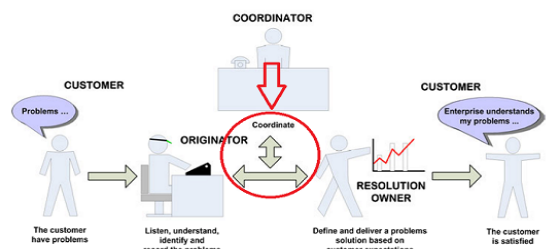


Fig. 1. Knowledge information identification process

Another factor that is also plays an important in distracting small and medium enterprises from decision support systems as highlighted by (Grabova *et al.,* 2010) are (i) high price, (ii) high requirements for a hardware infrastructure, (iii) complexity for most users, (iv) irrelevant functionality, (v) low flexibility to deal with fast changing dynamic environment and (vi) low

.attention to difference in data access necessity in SMEs and large-scaled enterprises. To address this problem the research performed suggests that and In-Memory solution that will in-turn not only be light-weight in terms of architecture requirements but also cheap.

(Levy, *et al.,* 2011) advocate the fact that data warehouse and as organizations move operations to the internet and establish partnerships with external stakeholders data warehouse roles need to extended i.e. to a KM perspective so that the knowledge created by knowledge workers within and outside the enterprises can be incorporated. They propose a conceptual model and meta-data architecture that extends the DW with a knowledge layer without any major changes in the architecture. The KW4DS@CBP architecture relies on the organizational memory, seeking to improve knowledge transfer and overcome KM barriers by harnessing knowledge required to support decisions. As per (Levy *et al.,* 2011), KM facts are the OLAP products and KM dimension are the meta-data of the OLAP products implemented with organizational ontology whose concepts are installed in KM. Based on these facts the devise elicitation tool and present and architecture that can work as an add-on on the data warehouse and contains information for KM fact. The basic fact schema of KM shown in figure.2 below. Additional information that the induction of KM includes in this case are the service parameters, process categories and their relation with specific departments.

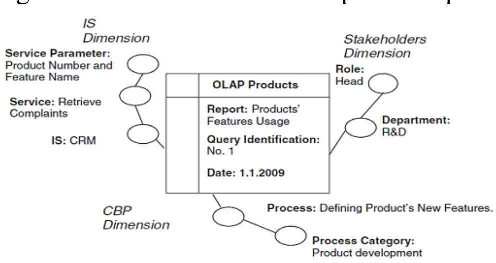


Fig. 2 . Basic fact schema of KM

2.2 In-Memory or Main Memory Databases

Organization of data in applications for enterprises was hardly changed whether it was for OLTP or OLAP. However, the processing and optimization was all stuck when the 3.0 GHz level clock speed was hit (Plattner, 2009). The introduction of Moore's law some 20 years back led to the identification a new area of virtualization namely that of main-memory and massive parallelism. This new trend of vertical partitioning (Krueger *et al.,* 2011) resulted in the emergence of the dormant in-memory databases as the DRAM storage density increased (Loos *et al.,* 2011). As a consequence a new line of databases emerged and the ignitor of in this case was HANA's in-memory database (namely SAPHANA). This resulted in a shift of thoughts as to whether the developed database be operational, analytical or both in nature. Another major difference in the databases developed was whether the database should support column-storage or row-storage.

Row-oriented stores the contents as the rows of a table, whereas column-oriented database stores its content by column rather than by row. The same column-oriented structured is already in used by database dealing with Online analytical processing (OLAP) data. According to (Plattner, 2009) column-oriented database stores are more suited to modern processors. The reason why it is more suited is due to high availability of parallel processing by multiple servers especially RAID server. As seen from the figure underneath that highlights the difference between column-store and row-store, a row-store is much more compatible if distributed and utilized for processing over multiple processors. The reason for this is that while storing and processing the data is physically stored inversely that is rows are converted to columns. (Stonebraker, 2011) also mentions the fact that column-stores will dominate warehouse market over time and will replace row-stores.

As stated by (Krueger *et al.,* 2011), column-store are best suited for both OLTP and OLAP applications. The reason behind this is that SELECT queries is mostly used in both cases and mentioned as 85% for OLTP and 94% for OLAP. Whereas rest of the operations within the database are equivalent to 16% in OLTP and 4% in OLAP. (Plattner, 2009) proved his findings TPC-C benchmark tool, with findings based on distinct values returned/stored within each column for a randomized data set of transport system. It is suggested that the data in the column-store be based on distinct values (compressed form) and the resultant bonded column structure will result such that approximately 75% of all columns have a relatively low value distribution. The research also highlights the fact that column-based architectures of in-memory databases work more effectively with increasing number of aggregates and there is very little time difference if the aggregates increase or decrease.

As identified by (Loos *et al.,* 2011) vast amount of research activity is currently been going on in in-Memory database solutions (also known as Main Memory databases). Times Ten, solid DB and HANA all already available commercially by IBM, Oracle and SAP respectively and is based on the original design proposed by (Agrawal *et al.,* 2008). The oldest in this league is MonetDB that was developed in 2009 at CWI in Amsterdam and is based on column-based approach. The storage model is based on vertical fragmentation and is supported with CPU-turned query execution engine and modular software architecture. One of the most crucial point that is highlighted by (Stonebraker,

2011) is the fact that in-memory databases systems typically work with snapshots for recording database states, with an add-on measures of replication to cope with volatile RAM and make it in non-volatile.

(Loos *et al.,* 2011) also identifies Hyper database system that is currently under development at Technical University of Munich that puts an emphasis on transaction processing. As suggested by (Kemper andNeumann, 2011) can make transactional gurantees and executes multiple OLAP queries on the same, arbitrarily current and consistent snapshots. Hyper engine relies on hardware support page shadowing that is in-turn controlled by memory unit of the underlying processor and it processes transactions sequentially on individual database partitions so that locks are no longer needed (Loos *et al.,* 2011). Hyper has an advantage over previous databases such that it focuses on both the categories of databases systems namely OLTP and OLAP unlike previous main-memory databases that focus on OLAP only.

(Kemper andNeumann, 2011) advocate the fact that with the help of snapshot mechanism an effective database system is in-place that allows both OLTP and OLAP simultaneously and with guarantee of ACID properties for OLTP transactions with facility to execute multiple query sessions. Using combined TPC-C and TPC-H benchmarking namely TPC-CH it has been identified that unprecedentedly high transactions rates can be achieved in conjunction with less response time for OLAP query. However (Krueger, Grund, Boissier, Zeier, andPlattner, 2010) have a different opinion. They advocate that due to the fact that HyPer uses a virtual memory snapshot mechanism for fast data modification and duplication it has a severe workload overhead. The result is that a redundant data storage is utilized and a two-phase commit protocol has to be in place to make both storages consistent and that in-turn generates additional costs and complexity on writes and might result in data loss.

HYRISE (Grund *et al.,* 2010), on the other hand is also a hybrid database system like HyPer. However HYRISE automatically partitions tables into vertical partitions of varying widths depending on how the columns of the table are accessed. They advocate the fact that analytical queries perform better in narrow partitions which in conjunction to (Stonebraker, 2011) is true because better and quick results are retrieved if values are store in a heap or close. Whereas for transactional queries HYRISE uses highly accurate model of cache misses, to predict performance of different partitions. This is for the fact that transactional queries are more effective if handled in wider partitions. Using the realistic workload that has been derived, HYRISE has been tested and performance improvement from 20% to 400% over all-column or all-row designs has been proved. This results in a more scalable and produces a better design than previous vertical partitioning approaches for main memory systems.

The aim of this research is to combine the already proposed KW4DS@CBP architecture (Levy *et al.,* 2011) with the scalable Main Memory hybrid database HYRISE (Grund *et al.,* 2010), so that we can provide SMEs with a database that will be light-weight and affordable. The resultant database will have knowledge support with no additional costs.

3 PROBLEM STATEMENT

Over the years, a general acknowledgement between researchers and organizations have developed that KM is very important in the current dynamic environment to gain competitive advantage and improve performance. This acknowledgement resulted in the evolution of Data Warehouse to a Knowledge Warehouse that had the capability to not only include analytical data but also apply and leverage the knowledge that has been captured, organized and stored, distributed and shared. The same is true for small and medium enterprises as they play an important role in any national economy and also require an effective Knowledge Warehouse that can help them cope with fast changing dynamic environment and the decision need a tool that will help them analyse and support their decision makers. However, the resultant knowledge warehouses that have been proposed are expensive in terms of price, hardware requirements and require large organizational memory. Latest research have identified and proved (Grabova *et al.,* 2010), that SMEs although do require knowledge warehouse but have different parameters not the ones that are suitable for organizations. While there are numerous factors that have been identified the most critical are high price, high requirements for hardware infrastructure and less human resource resulting in low organizational intellectual knowledge. Therefore the purpose of this study is to propose a database architecture for the knowledge workers at SMEs that will cheap and lightweight. It will also be able to cope up with problem of deficient intellectual knowledge by encapsulating knowledge from its sister concerns to form single organizational knowledge.

4 PROPOSED ARCHITECTURE AND METHODOLOGY

Before focusing on the research methodology it is imperative to mention as to what the resultant architecture will look like. The figure underneath highlights as to what will be the resultant architecture.

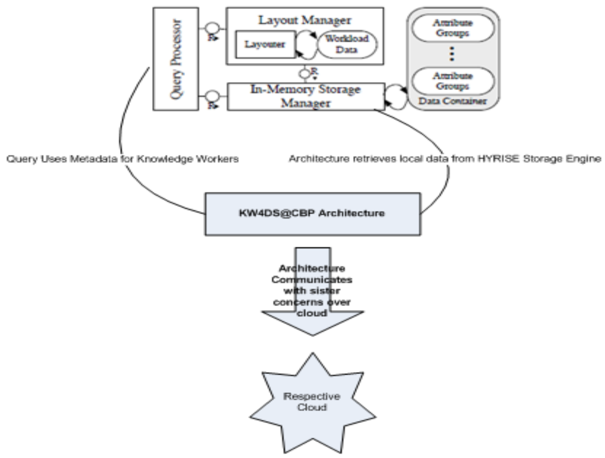


Fig. 3. Proposed Metadata Architecture over cloud

As shown in the Fig.3 above the proposed architecture will be add-on to the already available HYRISE, hybrid Main Memory Database. The same storage engine that is being used by HYRISE will be used to store meta-data information about knowledge and the query processor will also work as it is working for HYRISE. The only difference in this case will be the added add-on architecture that will determine the knowledge required by the knowledge worker and store patterns required.

5 EXPERIMENTAL DESIGN AND BASELINE

As the aim of this study is to effectively, establish a knowledge support database that is suitable for SMEs means light-weight and cheap in terms of financial resources, hence we will be focusing on a specific case. The resultant database will be tested on a cluster machine that will be built from combining two or three multiple rack servers so that we can have sufficient main memory to test our database with the benchmark tool. These servers will be connected to another set of rack servers so that we can form a cloud. It is important to note here that normal enterprises over the world do already have or can afford rack servers at least 2-3 depending upon the requirement and company size.

As mentioned earlier we are dealing with a specific case so in this case we will be using the same data set that has been used in HYRISE benchmarking. The same benchmarking tool that has been specifically built for HYRISE will be used in this case and will determine if there is any difference in the performance of the database with respect to load i.e. CPU cycles and Number of Partitions used. An additional overhead in the database in terms of memory usage is already there due to the induction of new architecture and its related data. Hence, the design that will be followed in this case will be within-subject true experimental design (Design 7).

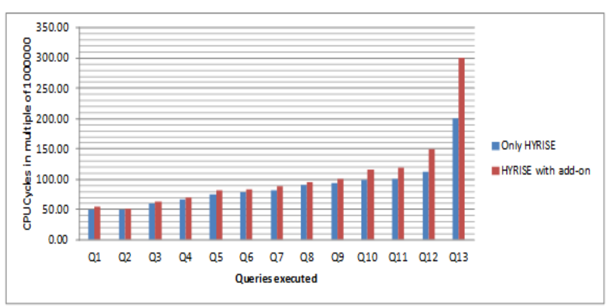
6 DATA SET AND ITS ORGANIZATION

The schema for the work load is based on a CRM application, and the tables represent with 3.6 M entries, 144 M entries, materials text with 600,000 materials and its related text, addresses with 180,000 entries, partners with 144,000 entries and finally materials hierarchy of 1M elements. This will constitute to a memory size of 28GB. Another approximately 15GB based on the number of entries will also be added to the database in terms of knowledge and will contain information of complaints.

Based on the above set of data a tweak in the already developed bench-marking tool will be performed and new queries added that will enable us to retrieve the additional 15GB information available in the database as part of our add-on. The already 13-queries that have been initially constructed to test HYRISE will be used to test and analyze the effects of resultant add-on first with add-on and second without add-on. It is supposed that the resultant add-on will give approximately same results in terms of CPU Cycles and Data Morphing. Secondly out of the 13 queries the ones with SELECT will be tweaked to retrieve additional 15GB information and the results will be compared in terms of CPU Cycles and number of partitions used. A results will be presented in tabular form and bar format for explanation as it will be clearly visible.

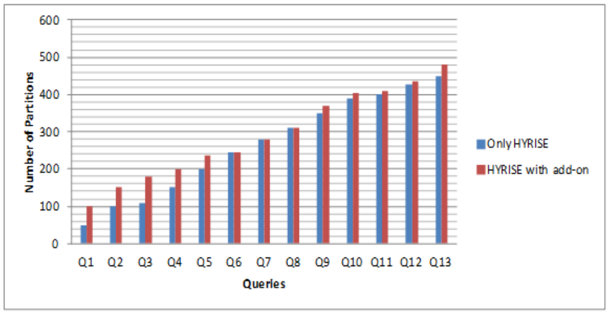
7 RESULTS

In the representation of results it is important the firstly we identify the effect of change in the queries that has had on the number of CPU cycles. It is important to note that there is very little difference in the execution of Q6, Q7 and Q8 owing to the fact that these are INSERT queries and no change in this case has been done but the overall addition of new add-on effects on the total number of cycles needed by the query to execute on the database show in **(Graph.1).**



Graph . 1. Effect of change in the queries on the number of CPU cycles

Another important comparison that needs to be done is the effect of new add-on on the number of partitions used. HYRISE generates partitions based on the number of tables and its data and it is assumed that there will be a slight increase in the number of partitions used against the queries but the final complex queries will have no effect on the number of partitions that will be used as shown in **(Graph.2).**



Graph . 1. Effect of new add-on on the number of partitions used

8 CONCLUSION

We extended Metadata Architecture of HYRISE (a Hybrid Main Memory database) to encapsulate knowledge information and include network searching to provide a database to SMEs that will be affordable, light-weight and will be able to communicate over cloud. The focus of this research is to provide an add-on architecture to the already HYRISE, Main Memory database. No changes to the database architecture will be made and neither change in the storage engine will be made. Same structure as that being used by HYRISE will be used to store Knowledge data.

REFERENCES:

Agrawal, R., A. Ailamaki, P. A. Bernstein, E. A., Brewer, M. J.Carey, S., Chaudhuri, . H.. Garcia-Molina, (2008). The Claremont report on database research. *ACM Sigmod Record, 37*(3), 9-19.

Davenport, T. H., and L. Prusak, (1998). *Working knowledge: How organizations manage what they know*: Harvard Press.

Dragoi, G., A. Draghici, S. M.,Rosu, C. E.,Cotet, and L. Ivascu, (2011). *Knowledge Applications Development for SMEs* Management *System Improvement.* Paper presented at the International Conference on ENTERprise Information Systems.

Evangelista, P., E. Esposito, V. Lauro, and M. Raffa, (2010). The adoption of knowledge management systems in small firms. *Electronic Journal of Knowledge* Management, *8*(1).

Grabova, O., Darmont, and I. Zolotaryova, (2010). intelligence for small and middle-sized entreprises. *ACM Sigmod Record, 39*(2), 39-50.

Grund, M., J. Krüger, H. Plattner, A. Zeier, P., Cudre-Mauroux, and S. Madden, (2010). HYRISE: a main memory hybrid storage engine. *Proceedings of the VLDB Endowment, 4*(2), 105-116.

Kemper, A., and T. Neumann, (2011). *HyPer: A hybrid OLTP&OLAP main memory database system based on virtual memory snapshots.* Paper presented at the Data Engineering (ICDE), IEEE 27th International Conference .

Krueger, J., M. Grund, M. Boissier, A. Zeier, and H. Plattner, (2010). *Data structures for mixed workloads in in-memory databases.* Paper presented at the Computer Sciences and Convergence Information Technology (ICCIT), 5th International Conference on.

Krueger, J., F. Huebner, A. Zeier, and H. Plattner, (2011). *Main memory databases for enterprise applications.* Paper presented at the Industrial Engineering and Engineering Management (IE&EM), IEEE 18Th International Conference on.

Lee, M. R., and Y.C. Lan, (2011). Toward a unified knowledge management model for SMEs. *Expert systems with applications, 38*(1), 729-735.

Levy, M., N. Pliskin, and G. Ravid, (2011). *Knowledge warehouse for decision support in critical processes: Conceptual modeling and requirements elicitation.* Paper presented at the Decision Support.

Loos, P., J. Lechtenbörger, G. Vossen, A. Zeier, J. Krüger, J. Müller, O. Günther, (2011). In-memory databases in information systems. and *Information Systems Engineering, 3*(6), 389-395.

López-Nicolás, C. and Á. L. Meroño-Cerdán, (2011). Strategic knowledge management, innovation and performance. *International journal of information 31*(6), 502-509.

Massa, S., and S. Testa, (2011). Knowledge domain and innovation behaviour: A framework to conceptualize KMSs in small and medium enterprises. *Vine, 41*(4), 483-504.

Nemati, H. R., D. M. Steiger, L. S. Iyer, and R. T. Herschel, (2002). Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and data warehousing. *Decision Support Systems, 33*(2), 143-161.

Plattner, H. (2009). *A common database approach for OLTP and OLAP using an in-memory column database.* Paper presented at the Proceedings of the 2009 ACM SIGMOD International Conference .

Stonebraker, M. (2011). Stonebraker on data warehouses. *Commun. ACM, 54*(5), 10-11.