Revolution or Evolution? Reflections on In-Memory Appliances from an Enterprise Information Logistics Perspective

Robert Winter, Stefan Bischoff, Felix Wortmann

Institute of Information Management
Mueller-Friedberg-Strasse 8
CH-9000 St. Gallen
{robert.winter; stefan.bischoff; felix.wortmann}@unisg.ch

Abstract: While the conceptual architecture of enterprise information logistics has been stable since the late 1980ies, IT hardware has been subject to radical change recently. Since in-memory appliances as a new technology might address many challenges of information logistics, we discuss its potentials first. Based on a case study of a global automotive company, we then compare potentials with actual achievements. We conclude that there are situations where in-memory appliances are a useful extension to existing IT support concepts, while other situations do not require such support. As a consequence, we regard in-memory appliances as an evolution, but not revolution of IT support from an enterprise information logistics perspective.

1 Introduction

Starting from a rather technical (tool) perspective in the mid 1990ies, the understanding of Business Intelligence (BI) information systems has widened and now covers "a broad category of technologies, applications, and processes for gathering, storing, accessing, and analyzing data to help its users make better decisions." [WW10] Since BI is focused on decision-making, it usually has a 'local' perspective, i.e. its design is focused on a specific business domain, a specific business problem, a specific business process, or a specific organizational role. In order to cover the overall perspective of a business unit, company or even business network on data gathering, data integration, data storage and data analysis, the concept of Enterprise Information Logistics (EIL) has been proposed [Wi08]. Since the understanding of BI has developed from a technical to an information system (IS) perspective, BI and EIL do not differ in this respect. Their difference is that EIL has an enterprise-wide data provisioning perspective, while BI focuses on the 'local' use of provisioned data.

Rapidly growing amounts of data and more complex analyses are only two of the current EIL challenges. In times of rapid change and volatile economy, decision makers rely on timely information [Sc10]. Different approaches exist for reducing data entry latency, data analysis latency and decision latency. While some improvements focus on the manual parts of data provisioning and decision making, new technology in the form of in-memory appliances promises to reduce time for data analysis significantly [Pl09].

IT vendors claim that in-memory is the end of classical EIL [El11]. They argue on the basis of radical hardware changes and foresee a new era of business analytics. While there is certainly evidence that radical hardware innovation has an impact on EIL, we want to challenge this statement from two lenses. First, hardware innovation has to be put into the context of people and tasks – in order to create impact, changes of one IS component type need to be matched by changes of the others to create improvements. Second, an enterprise wide focus might put specific hardware innovations into perspective, even if they are fundamental for a certain business domain, business problem, or business process. Both lenses are combined in a case study from a large automotive company that recently introduced in-memory appliances for business analytics.

The paper is structured into five sections. Section 2 discusses the state of the art of EIL, EIL challenges, fundamentals of in-memory appliances and the potentials of in-memory appliances in relation to EIL architecture challenges. Section 3 introduces the case study and compares the potentials of in-memory appliances to actual achievements. These findings are discussed in section 4. A summary and outlook concludes the paper in section 5.

2 Technology challenges of EIL and potentials of in-memory appliances

2.1 Conceptual EIL architecture

EIL IS are based on a data warehousing infrastructure that collects, stores, and integrates relevant data that is further modified and enhanced within specific reporting processes. The resulting conceptual (hub and spoke) EIL architecture usually is comprised of five layers [BG09; DM88; PR03]:

- Source systems: These software systems contain operational data and other data that is important for the reporting and analysis process. Usually internal and external information systems are used as sources [In05].
- Data transformation (ETL - extract, transform, load processes): In order to move data from different heterogeneous sources into one central data warehouse, different transformation steps are required in order to normalize, harmonize and integrate data.
- Data warehouse (DWH): The DWH is a logically centralized database, which consolidates all data that could be useful for analytical applications as well as decisions and management tasks [In05; KR02].
- Data Marts (DM): Based on the DWH, data marts are copies of data that relate to a specific domain. For example, a purchasing DM contains only relevant data for purchasing analyses. DMs usually store data in a form that supports interactive data analyses [Ba09; KBM10].
- Reporting systems: These software systems access DMs (or sometimes the DWH directly) in order to present data to users. There are different reporting approaches like standard reporting [GGD08], ad-hoc reporting [GK06] or advanced, interactive reporting like OLAP (online analytical processing) [CC93].

Recent BI innovations combine technical and organizational improvements regarding overall responsiveness of BI (e.g. real-time BI [Di09] or active BI [Di07]). Although new trends have emerged, looking back at the first data warehouse publications (e.g. [DM88]) one has to conclude that the fundamental conceptual EIL architecture has been stable for more than 20 years.

2.2 EIL challenges

While the stability of conceptual EIL architecture indicates a high level of maturity, the depicted conceptual EIL architecture poses three key challenges [Wi11]:

- ***Speed***: Today, data processing and enrichment (extraction, transformation, preparation, integration, provisioning and analysis) cannot be performed in real-time in larger organizations, even when using powerful organizational concepts and powerful technology. Underlying inhibitors are rising data volumes, increasing needs for data integration and ever more diverse data analysis possibilities. For efficiency reasons, analyses have to be optimized carefully and data updates are often organized as batches. Iterative, interactive analytics of live data, as required by many decision makers, are therefore only inadequately supported [Fi10].

- ***Integration***: The many integration and processing steps that are caused by specialized tools and infrastructures, lead to high complexity as well as high operating and development costs [Ec09]. In order to achieve substantial complexity encapsulation and cost reductions, data processing and enrichment functions have to be brought together in integrated IT systems. However, with existing technologies this seems to be possible only to a very limited extent, as numerous processing steps are indispensably simple because of performance reasons (e.g. pre-computed aggregates in DMs).

- ***Flexibility***: Due to the need for optimizations in today's IT landscapes, only certain specific analysis functions and paths can be supported efficiently [Fi10]. However, in increasingly dynamic decision situations, decision makers ask more and more for solutions that enable them to analyze any desired business object from any perspective, e.g. on the basis of individual and spontaneous merge of data.

2.3 In-memory appliances – fundamentals and potentials

While conceptual EIL architecture has been stable for a long time, hardware is driven by radical change. During the last decades Moore's law [Mo65], which predicts the expansion rate of available processing speed, has been valid at all times. The processing performance has increased by more than the factor 4.000 [NP11] over the last 20 years. Even with the processor's clock speed having reached its peak, new concepts like massive parallel processing enable new improvements that help to increase processing speed even further [Pl09]. Additionally, the exponential price drop of processing power and main memory fosters breakthrough innovation. In-memory databases leverage these developments in order to optimize performance of analytical systems. Therefore, they build upon massive parallel processing (MPP), in-memory technology and a new type of data storage [Pl09].

In the environment of data processing, multi-core processors and MPP offer huge acceleration possibilities. In order to exploit this potential, software can use multiple processors in parallel to speed up calculations and data manipulation. In order to take full advantage of such capabilities, special programming techniques as well as processor specific coding need to be used. Thus programs need to be adapted to the corresponding processing platforms [Pl09]. Compared to hard drives, main memory – that is used as the storage media in in-memory databases – has much faster access times (20-50 ns vs. 8-15 ms) [BU10]. Moreover, columnar databases have huge advantages in analytical environments because usual reports are rather set oriented than record oriented [Pl09]. In order to calculate the overall number of sold products in a classical, record / row store database, every sales order has to be processed and the overall amount of accessed data is calculated as D1 = number of records * size of all columns. In contrast, in a column store database only the column with the corresponding quantity has to be read from the storage device. Thus the amount of accessed data is only D2 = number of records * size of column "quantity".

In-memory appliances are understood as IT systems which combine software (i.e. in-memory databases) and hardware (i.e. in-memory technology). On the basis of the depicted innovations, they promise to address the above mentioned EIL challenges [Pl09]:

- ***Speed***: Data updates are propagated incrementally and in real-time into the analytical environment. The asynchronous actualization of persistent aggregates (pre-computed interim results) can be omitted. Thus, iterative, interactive analyses can be supported much better: Actual data is available at any point in time.

- ***Integration***: Transactional and decision-related data is managed in an integrative manner. Data redundancy, which has to be actively managed, is eliminated and possible inconsistencies are thereby avoided. Processing stages can be reduced and the architectural complexity can be cut down.

- ***Flexibility***: Analysis paths are not limited by pre-fabricated aggregates, so that data can be integrated and analyzed from any perspective. Thus an individual and spontaneous fusion of data is possible.

In order to assess whether these promises can be met, a case study is presented in the next section.

3 An in-memory appliance at an international automotive company

3.1 Company characteristics and current situation

Our case study is based on the analysis of a large Germany-based, globally present automotive company. The company can be classified as an early adaptor of in-memory appliances. On both, the organizational and the technical level the corporation prepared itself for a "first class BI environment." On an organizational level they established an internal SAP competence center that is responsible for any operational and analytical SAP-based system. In addition an BI department[1] has been created with company-wide responsibility that not only supports the operations of analytical systems, but also starts initiatives to optimize and improve the BI system. From a technical perspective, the company has clearly an innovate attitude: The BI department is always willing to implement new technology and acts regularly as a showcase or ramp-up partner in case a technology promises significant improvements.

The current EIL landscape of our case study is visualized in figure 1. At the moment the EIL landscape's structure conforms to the traditional architecture introduced in section 2. There are various source systems on the basis of SAP and non-SAP applications that feed data into the analytical components. The company runs two data warehouses. While one is based on the SAP Business Information Warehouse (SAP BW) that is used for SAP-based source systems, another DWH (based on Oracle technology) integrates all other source systems. For different analytical purposes, various data marts (SAP BW-based) exist that contain business scenario specific extractions from the two main DHWs. For reporting and analysis purposes, different applications and tools are used in the various functional areas. Analytical tools like SAP BW Query, Xcelsius Dashboards, Crystal Reports access not only the DMs, but also other data sources.

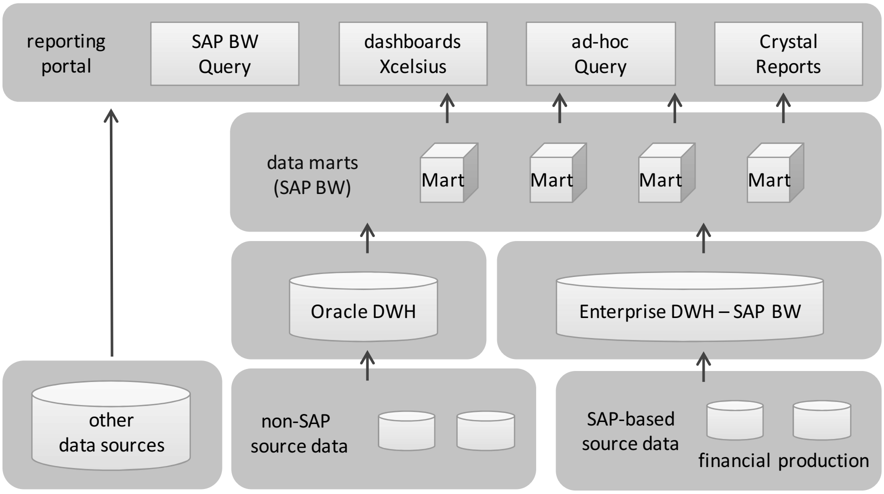

Figure 1: Basic structure of the current EIL landscape

In order to improve EIL and offer improved services for the business users, the company is in the process of implementing an in-memory appliance. In the following, we focus on the in-memory initiative as well as its business improvements.

[1] The BI department has an enterprise perspective. In the terminology of this paper, we would designate it as EIL department.

Within our analysis we were faced with two major usage scenarios that are currently about to be implemented. On the one hand, the in-memory appliance will be used for improving operational analytics. On the other hand, it will be used to increase the overall performance of the DWH landscape. According to our IS lens, we differentiate between the business perspective (task) and the technological solution for these problems (technology) for each of the usage scenarios.

3.2 Operational analytics

The first usage scenario for the in-memory appliance is the support of different business functions in the production, purchasing and financials areas.

Today, transactional ERP systems have significant performance problems with specific tasks like *quality and fault analysis* or *calculation of procurement commissions*. In order to be effective, both analyses are rather complex and require a full bill of material explosion. Current ERP systems are not capable of providing this information within an appropriate time period. Furthermore, due to long calculation times only static decision making is possible. Iterative recalculation of reports to explore different scenarios and alternatives is restricted by query run times. Existing BI systems could be used to accelerate query times. However, in order to leverage these BI systems, data has to be extracted from the underlying ERP systems and stored redundantly in the BI environment. The resulting solution would have such a complexity and incur so much higher costs, that business is reluctant to realize this scenario.

In the *financial environment* the monthly closing of accounts is a complex task that is not supported efficiently by today's EIL landscape. In this case relevant account data needs to be analyzed and the results of the analysis need to be posted back (after performing manual checks and adjustments) into the source systems (closed-loop approach). Today, the complete process is performed in the ERP system. Long calculation times prevent a fast close of books and limit the quality of the overall process (static process without iterations). An easy integration of existing BI systems into this scenario is not possible, as the data has to be queried, manipulated and finally posted back into the ERP system. In order to accomplish such a fully integrated solution that perfectly supports the accountants' tasks, a new technological solution is needed. This solution needs to increase reporting speed and thereby support iterative decision making.

In order to solve the previously described problems and to enhance the operational decision processes, the in-memory appliance can be used for reducing the access time to operational data.

Figure 2 illustrates the technical solution for the first task (quality and fault analysis, calculation of procurement commissions). To improve decision making, the operational ERP database is replicated into an in-memory appliance, so that a new record is instantly available in both databases. Now, reports can be retrieved from the high performance in-memory appliance via BI reporting frontend tools. As a consequence, query times can be reduced significantly.

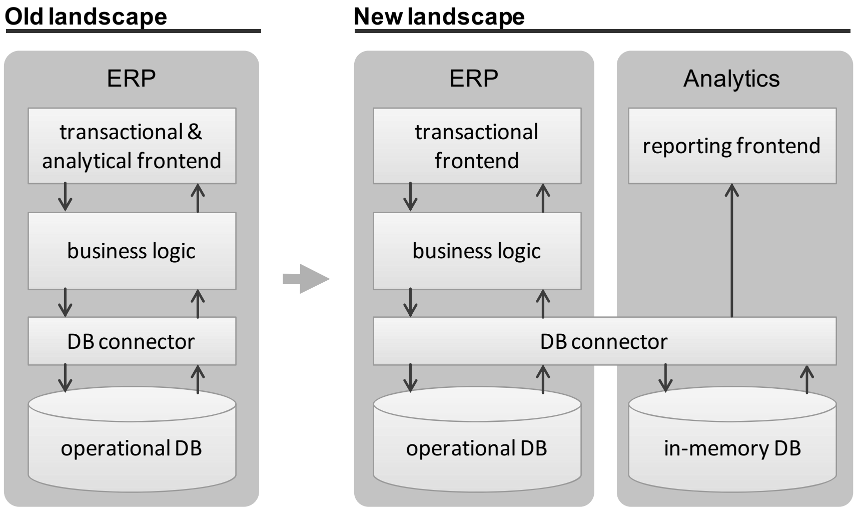

Figure 2: Enabling operational analytics on the basis of an in-memory appliance [on the basis of SA11]

Figure 3 illustrates the technical solution for the second task (financial environment). Again, the operational ERP database is replicated into an in-memory appliance. To enable closed-loop analytics, query results have to be available in the ERP system where they can be checked, modified and processed in a next business transaction. Thus the analyzed data needs to be transferred from the in-memory appliance back to the operational ERP system.

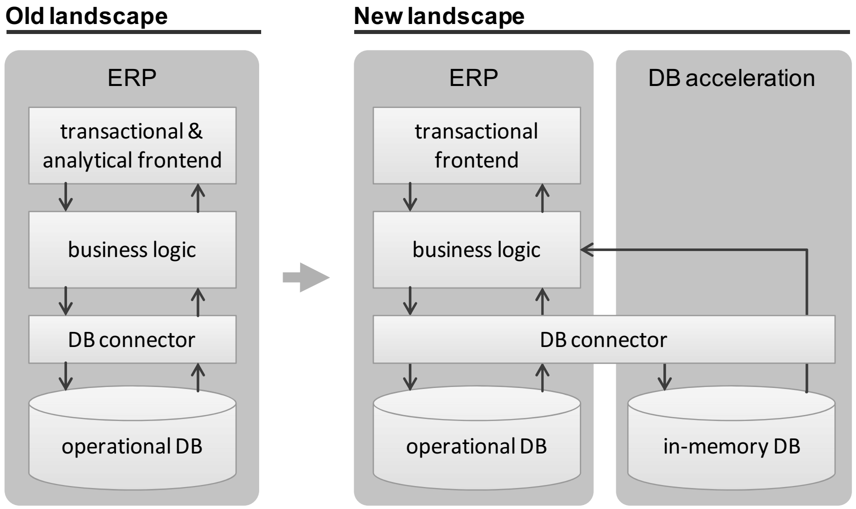

Figure 3: Enabling closed loop analytics on the basis of an in-memory appliance [on the basis of SA11]

3.3 Data warehousing

In addition to the mentioned weaknesses of current operational systems, there are also disadvantages that classical EIL architectures reveal. In analytical tasks the analysts and business users of the studied company demand fast and flexible analyses without restrictions of analyzable key performance indicators (KPIs) or pre-defined drill-down paths. Due to performance problems and slow response times for complex reports, the international automotive company used DMs and materialized cubes for analyses. These concepts extract a previously defined amount of data and KPIs from the data warehouse and pre-aggregate it into DMs or cubes that can be used for analyses. Thus the possible dimensions of reports are fixed, based on the ones that are copied into the DMs, only these are accessible at an acceptable speed. This configuration is not satisfying for business users because they are limited in the scope of their analysis. They demand more flexibility and better response times when it comes to analyzing data. In order to transfer data from the DWH into DMs and cubes, specific ETL processes are necessary. These processes have to be executed after the load process of the DWH has been completed successfully. In the complex environment of our case study partner, the runtime of both ETL processes (source systems to DWH, DWH to DMs) almost exceeds the amount of time that is available every night. Thus their aim is also to reduce the runtime of ETL processes.

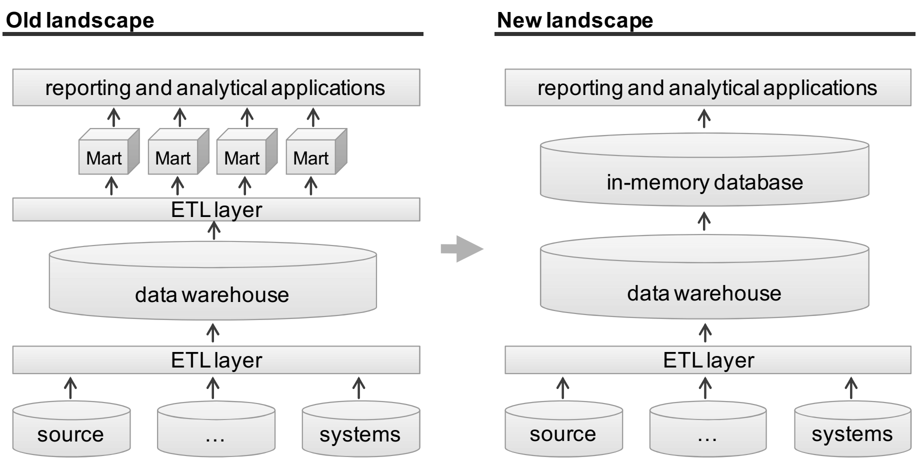

Figure 4: Enabling in-memory data warehousing [on the basis of SA11]

In order to fulfill these requirements, the case study company is about to implement an in-memory-based application that speeds up the DWH queries. Based on the fact that this solution offers much faster data access, there is no need for pre-aggregation and data extraction into DMs or other materialized components that limit reporting flexibility. In addition the elimination of DMs obviates the need for the second ETL level and reduced the overall ETL runtime. Figure 4 visualizes the architecture of that solution where an in-memory appliance replaces traditional DM technology. In this architecture the classical EIL stack (see section 1) is slightly modified, i.e. the DM layer is replaced by an in-memory appliance. The whole DWH that is stored in a classical, hard drive-based database, is copied into an in-memory database that now represents the basis for analysis and reporting. The response time is now significantly reduced because data can be retrieved from the in-memory database much faster than it could be from a traditional DM.

In order to assess the impact of in-memory appliances from an EIL perspective, the following section discusses our findings.

4 Discussion

In our case study we discovered that in-memory appliances indeed can have a significant impact on the way people use IT systems to support their tasks. Static decision making can become dynamic, i.e. iterative analysis and lightweight simulation are facilitated by these technologies. The case study reveals two scenarios which should be differentiated:

- ***Operational analytics*** is about enhancing the performance of a single transactional system. Operational analytics is often performed in ERP systems, which are not able to process more sophisticated queries in a short period of time. Established BI solutions can be used to solve this problem. However, huge investments are necessary to bring that operational data into the existing BI/DWH environments. Therefore, in-memory appliances are perceived as a lightweight and easy to implement/maintain technology to increase operational query performance.

- ***Data warehousing*** is about integrating data from multiple sources in order to support decision making. While in-memory appliances can replace the performance function of DMs, in-memory appliances do not address the fundamental issue of integration. Therefore, in-memory databases complement the DWH and, by replacing DMs, slightly change the traditional hub-and-spoke architecture.

So how do in-memory appliances change the EIL landscape? On the basis of the case study two essential characteristics can be identified: To a large extent, the underlying IS landscape is determined by the need for integration (*required degree of integration*). If the performance of a single IS has to be improved, other approaches should be implemented than in the case of optimizing the performance in a domain where several IS are tightly integrated. The *data volume*, which has to be analyzed, is a major driver for the use of in-memory appliances. However, there are scenarios where data volumes are moderate so that no advanced, high performance technology is necessary. On the basis of these characteristics four stereotype patterns can be identified (see Fig. 5):

- ***Low/moderate data volume, no need for integration***: This pattern is about operational analytics. The transactional system at hand has no performance issues. There is no significant need for an in-memory appliance.

- ***Hugh data volume, no need for integration:*** This pattern is also about operational analytics. High amounts of data can be handled by introducing an in-memory appliance (see section 3.2).

- ***Low/moderate data volume, need for integration:*** The need for integration drives the application of BI/DWH. There is no significant value added by an in-memory appliance.

- ***Hugh data volume, no need for integration:*** This pattern is also about BI/DWH. High amounts of data can be handled by introducing an in-memory appliance (see section 3.3).

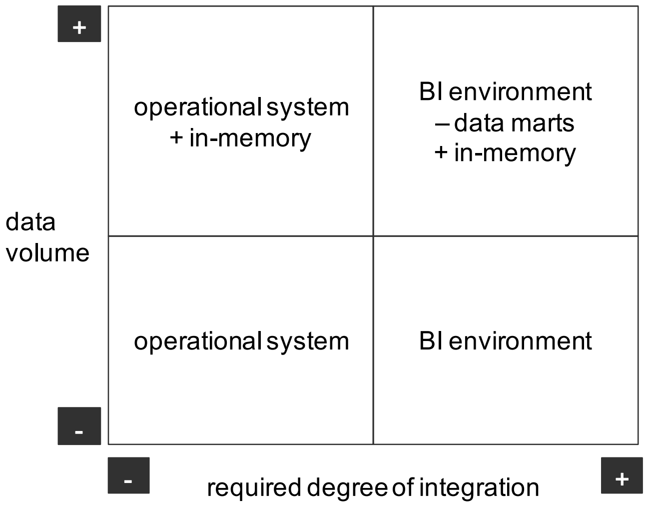

Figure 5: Four system architectures stereotypes

Summing up, in-memory appliances complement traditional operational software systems as well as BI/DWH systems. Therefore, in-memory appliances can be regarded as an incremental evolution of EIL architecture rather than a revolutionary change.

5 Summary and outlook

In this article we conducted a case study that helped us to assess the impact of in-memory appliances on the EIL landscape of a global automotive company. We identified two usage scenarios where in-memory appliances enhance the functionality and speed of EIL. Reflecting the case study results, we proposed to differentiate four architecture patterns. While in-memory appliances promise significant improvements in two of these architecture stereotypes, their impact is very limited in the other two. Therefore, in-memory should be regarded not to be disruptive or as “the end of” existing EIL architectures and approaches, but rather as an evolution that creates significant progress under certain circumstances.

The current discussion of the potentials and challenges of in-memory appliances fits to the increasingly important role of analytics, that is being reflected especially in the U.S.

under the label “Big Data.” However, this technology driven analysis has its risks: IT-driven transformations often fail exactly because of their focus on technology [LWU11]. Therefore, further research should investigate in scenarios where decision making is significantly improved by exploiting the full potential of all available data assets. On the basis of these scenarios, reference models and methods can be build, which finally can be leveraged by software companies in order to develop innovative business applications. Moreover, there is a need for (1) examining the effects of user groups with distinct preferences on how systems are constructed and (2) moving from pure requirements orientation to more differentiated constructional considerations.

References

[Ba09] Bauer, A. et al.: Referenzarchitektur. Aspekte einer Referenzarchitektur. In: (Bauer, A.; Günzel, H. eds.): Data-Warehouse-Systeme. Architektur, Entwicklung, Anwendung. 3. Ed., dpunkt.verlag, Heidelberg, 2009, pp. 33-38.

[BG09] Bauer, A.; Günzel, H. (eds.): Data-Warehouse-Systeme. Architektur, Entwicklung, Anwendung. dpunkt.verlag, Heidelberg, 2009.

[BU10] Brinkschulte, U.; Ungerer, T.: Mikrocontroller und Mikroprozessoren. 3. Ed., Springer, Heidelberg, 2010.

[CC93] Codd, E. F.; Codd, S. B.; Salley, C. T.: Providing OLAP to User-Analysts: An IT Mandate. E.F. Codd & Associates, 1993.

[Di07] Dittmar, C.: Integration von Real Time Business Intelligence in eine SOA – Auf dem Weg zum Active Business Intelligence. In: (Gluchowski, P. et al. eds.): Schlaglichter der Wirtschaftsinformatik. 1. Ed., guc, Chemnitz, 2007, pp. 129-142.

[Di09] Dittmar, C.: Phasen des Data Warehousing. Exkurs: Realtime Data-Warehouse-Systeme. In: (Bauer, A.; Günzel, H. eds.): Data-Warehouse-Systeme. Architektur, Entwicklung, Anwendung. 3. Ed., dpunkt.verlag, Heidelberg, 2009, pp. 101-104.

[DM88] Devlin, B. A.; Murphy, P. T.: An Architecture for a Business and Information System. In: IBM Systems Journal 27 (1988) 1, pp. 60-80.

[Ec09] Eckerson, W. W.: 2009 TDWI BI Benchmark Report. TDWI, Chatsworth, 2009.

[El11] Elliott, T.: Preparte to be blown away: The 4.0 releases of BI & EIM solutions have landed!, 2011.

[Fi10] Finucane, B. et al.: The BI-Survey 9. Business Application Research Center - BARC GmbH, Würzburg, 2010.

[GGD08] Gluchowski, P.; Gabriel, R.; Dittmar, C.: Management Support Systeme und Business Intelligence. 2. Ed., Springer, Berlin, Heidelberg, 2008.

[GK06] Gluchowski, P.; Kemper, H.-G.: Quo Vadis Business Intelligence? In: BI-Spektrum 1 (2006) 1, pp. 12-19.

[In05] Inmon, W. H.: Building the Data Warehouse. 4. Ed., Wiley Computer Publishing, Indianapolis, IN, 2005.

[KBM10] Kemper, H.-G.; Baars, H.; Mehanna, W.: Business Intelligence - Grundlagen und praktische Anwendungen. Eine Einführung in die IT-basierte Management-unterstützung. 3. Ed., Vieweg + Teubner, Wiesbaden, 2010.

[KR02] Kimball, R.; Ross, M.: The Data Warehouse Toolkit. 2. Ed., John Wiley & Sons Inc., New York et al., 2002.

[LWU11] Lahrmann, G.; Winter, R.; Uhl, A.: Transformationsmanagement – Aktueller Stand der Praxis und Potenziale. In: Wirtschaftinformatik Und Management 3 (2011) 3, pp. 62-67.

[Mo65] Moore, G. E.: Cramming more components onto integrated circuits. In: Electronics 38 (1965) 8.

[NP11] Nambiar, R.; Poess, M.: Transaction Performance vs. Moore's Law: A Trend Analysis. In: (Nambiar, R.; Poess, M. eds.): Performance Evaluation, Measurement and Characterization of Complex Systems. Springer, Berlin, 2011, pp. 110-120.

[Pl09] Plattner, H.: A common database approach for OLTP and OLAP using an in-memory column database. In: (Binnig, C.; Dageville, B. eds.): Proceedings, SIGMOD/PODS '09 International Conference on Management of Data, Providence, RI, (USA) 2009.

[PR03] Propach, J.; Reuse, S.: Data Warehouse: ein 5-Schichten-Modell. In: WISU – Das Wirtschaftsstudium 32 (2003) 1, pp. 98-106.

[SA11] SAP AG (ed.): SAP In-Memory Computing and SAP High-Performance Analytical Appliance (HANA). http://www.slideshare.net/SAPanalytics/sap-hana-overview-nd-update-dec-1-2010, Access on 23.10.2011.

[Sc10] Schelp, J.: Near Real-Time Warehousing. In: (Chamoni, P.; Gluchowski, P. eds.): Analytische Informationssysteme. Business Intelligence-Technologien und -Anwendungen. 4. Ed., Springer, Berlin, 2010, pp. 463-480.

[Wi08] Winter, R. et al.: Das St. Galler Konzept der Informationslogistik. In: (Dinter, B.; Winter, R. eds.): Integrierte Informationslogistik. Springer, Berlin, 2008, pp. 1-16.

[Wi11] Winter, R.: In-Memory-Appliances – ein Fortschritt in der Informationslogistik? wird erscheinen in: Wirtschaftinformatik 53 (2011) 6.

[WW10] Wixom, B. H.; Watson, H. J.: The BI-Based Organization. In: International Journal of Business Intelligence Research 1 (2010) 1, pp. 13-28.