An Architecture for Ad-hoc and Collaborative Business Intelligence

Henrike Berthold
SAP AG
Dresden, Germany
henrike.berthold@sap.com

Philipp Rösch
SAP AG
Dresden, Germany
philipp.roesch@sap.com

Stefan Zöller
IBIS Prof. Thome
Wuerzburg, Germany
zoeller@ibis-thome.de

Felix Wortmann
University of St. Gallen
St. Gallen, Switzerland
felix.wortmann@unisg.ch

Alessio Carenini
CEFRIEL
Milan, Italy
alessio.carenini@cefriel.it

Stuart Campbell
TIE Nederland B.V.
Hoofddorp, Netherlands
Stuart.Campbell@tieglobal.com

Pascal Bisson
Thales Group
Palaiseau Cedex, France
pascal.bisson@thalesgroup.com

Frank Strohmaier
BASF SE
Ludwigshafen, Germany
frank.strohmaier@basf.com

ABSTRACT

The success of organizations or business networks depends on fast and well-founded decisions taken by the relevant people in their specific area of responsibility. To enable timely and well-founded decisions, it is often necessary to perform ad-hoc analyses in a collaborative manner involving domain experts, line-of-business managers, key suppliers or customers. Current Business Intelligence (BI) solutions fail to meet the challenges of ad-hoc and collaborative decision support, slowing down and hurting organizations.

The main goal of our envisioned system, which will be designed and implemented in a future research project, is to realize a highly scalable and flexible platform for collaborative, ad-hoc BI over large data sets. This will be achieved by developing methodologies, concepts and an infrastructure to enable an information self-service for business users and collaborative decision making over high-volume data sources within and across organizations.

Categories and Subject Descriptors

J.1 [**Administrative data processing**]: Business

General Terms

Business Intelligence

1. INTRODUCTION

Current Business Intelligence (BI) environments suffer from several shortcomings. One problem is the missing focus on the individual needs of the analysts and decision makers, i.e., the business users. These users are forced to rely on standard reporting and predefined analytical content that often rather insufficiently fulfills the individual information needs. They strongly depend on either IT administration or enhanced technical skills to generate the required business reports.

A second problem follows from the lack of business context information, such as definitions, business goals and strategies as well as business rules or best practices for the provided analytical data. Hence, business users have to understand the semantics of data by themselves and they have to take decisions and derive strategies using additional information sources, which often leads to an escalation of efforts and costs.

Another shortcoming concerns the collaboration aspect: Well-founded decisions are often based upon opinions and expertise of several analysts. This implies that decision making is inherently a collaborative task and the combination of social software with business intelligence (collaborative BI) can dramatically improve the quality of decision making. However, current BI solutions only provide (in the best case) rudimentary collaboration capabilities. Collaborative decision making is done outside of the system using other tools or classical communication channels such as emails or phone conversations. Such cumbersome, lengthy, and expensive processes slow down organizations and have negative impact on the adoption and usage of analytical tools.

Furthermore, the setup and configuration of current BI systems requires deep insight in both the data to be analyzed and the intended analytical tasks. Here, content and data models have to be provided in advance by the IT department and tailored to the needs of a specific group of people within the company. This procedure usually ends up with several configuration (implementation or adaption) cycles; according to the world's largest BI study [2] the modal time for new BI implementations is between 3 and 6 months causing im-

EDBT 2010, March 22–26, 2010, Lausanne, Switzerland.

plementation and support costs that often deter companies of a wider BI deployment.

Finally, today's BI solutions have a strong focus on structured, enterprise-internal data but lack the capability of integrating external and/or unstructured information in an easy, (near) real-time and effective way. As a consequence, a lot of useful information is never included in the analysis. Not considering this information could provide a distorted or incomplete view of the actual world and consequently, it could lead to wrong business decisions.

In this paper, we propose a BI platform—that will be implemented in a future research project—allowing business users to derive their own business information in an ad-hoc fashion, discuss the gained knowledge and shape their business strategies in a collaborative manner. This platform will reduce IT dependencies and put information acquisition directly into the business user's hands, be it managers or operative information workers (see Figure 1). To accomplish these goals, we need

A flexible data model to describe and adapt business relevant entities and their relationships within and across organizations. Such a data model allows to react on rapidly changing conditions by adding, removing or modifying new entities and relationships. Further, it allows to model various context information (i.e. information relevant for the decision making process that complements the data to be analyzed).

An efficient and scalable data store for these business data to realize fast response times even for large datasets. This data store has to take the flexible data model into account.

A business configuration methodology to populate the data store in a pay-as-you-go fashion and to empower users to adapt the business data on their own.

An information self-service allowing users to easily retrieve the business information of interest. The information self-service should enable business users to satisfy their information needs in an intuitive, fast, and efficient manner and to define a personalized environment for retrieving and analyzing information.

An integrated collaboration environment to discuss the results, derive conclusions and take decisions in a collaborative fashion. Important properties of this environment should be synchrony, intuitiveness and traceability.

Our goal is to empower business users to be able to configure business data and create analyses based on their specific requirements without technical skills or synchronization efforts using the language and terminology they are familiar with.

In the next section, we will give an overview over the state of the art in the relevant fields. In Section 3, we will address the individual points given above in more detail. Section 4 summarizes the paper.

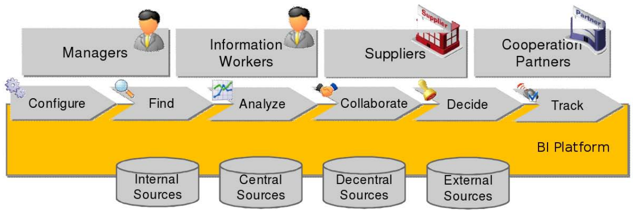

Figure 1: Envisioned platform

2. STATE OF THE ART

In this section, we describe the state of the art in the key areas which are relevant to reach our goal.

2.1 Flexible Data Models for Business Metadata

Data Warehousing and OLAP are the means of choice for the integration and consolidation of multi-dimensional data; their architectures and data models are well known [17]. In addition, OLAP functions allow complex analyses of the data. However, today's analysis tools provide a rather technical view on the analysis content and thus require technical skills of the analysts. Some BI solutions have already incorporated business metadata in a limited manner via a dedicated abstraction level based upon business terminology. Also some applications like dedicated performance management solutions cover enhanced business knowledge, comprising general and domain specific rules and interdependencies.

Current research efforts concentrate on introducing ontology-based approaches to improve decision making, tracking, progress monitoring and usage-information mining. For instance, [18] contributes a first attempt to enable end-user self-service BI with the help of ontology-based infrastructure and reasoning. By providing a target ontology capturing the business view the authors achieve a declarative and flexible decoupling between data sources and business oriented models. Furthermore, [16] proposes an approach for strategy-driven business process analyses by providing an ontology formalizing performance management related aspects like strategies, goals, KPIs, processes, measures, etc. as well as their interrelationships. [6] is an alternative approach to achieve ontology-based integration of BI considering data warehouses, data mining and reporting systems as well as OLAP tools. The authors state that current BI solutions focus on structural integration, and, thus, enable semantic integration mechanisms to facilitate user-friendly and adaptive analyses. Albeit being powerful, ontologies are expensive to setup and maintain.

In the Web community, the RDF (Resource Description Framework) [20] data model is well established; it has been designed as a flexible representation of schema-relaxable or even schema-free information for the Semantic Web. To achieve our goals, we leverage the flexibility and expressiveness of RDF to add business metadata—i.e. a business view—to the technical terms. The resulting relationships between technical and business terms allow to abstract from the technical view.

2.2 Efficient and Scalable Data Stores

Representing business terms and their relations to technical terms and other business terms in a large organization or company may lead to a high data volume. In such a scenario, an efficient and scalable data store such as a distributed or a column-oriented main-memory data management system has to be used.

A distributed data management system consists of a collection of subsystems whose distribution is transparent to the user. The subsystems process the data in parallel. Therefore high data volumes can be handled efficiently. The distributed system appears as one local system. The user may be unaware of the fact that there are several machines. Their location, storage replication, load balancing and functionality is transparent [15].

Main-memory databases like MonetDB [5] store and thus efficiently process the data in main memory. Column-store databases like MonetDB, Sybase IQ [19] or SAP NetWeaver BI Accelerator (TREX) [14] store the data in a column-oriented way which allows high compression rates and a more effective exploitation of main memory and modern CPU features, like SIMD or pipelining [21]. However, as the tuples are fragmented some query types require expensive joins.

In our system, we apply main-memory processing to improve the overall performance of our RDF-based business data store. To effectively utilize the main memory we will apply column-store techniques; to be efficient we will take special care on joins, e.g. with enhanced index structures.

2.3 Business Configuration Methodologies

In the past, the usage of customized standard software (packaged applications) has been established in particular in the area of transaction-oriented information systems. Configuration and customizing approaches allow an enterprise-specific adaptation of the respective packaged application thereby addressing the individual needs of the enterprise [1, 3]. Recently, basic customization and configuration functionalities have also been incorporated into analytical information systems in order to reduce development and implementation efforts of these systems and to meet user requirements in a better way [10]. The reference content which is provided within packaged applications is called business content and forms the foundation for any customization and configuration task. For instance, SAP defines analytical business content as role- and task-specific information models. These information models may comprise amongst others queries, OLAP cubes, measures, dimensions etc. for a specific purpose like sales planning. In an implementation project business content has to be adapted to the needs of the enterprise. This is often a time-consuming and costly process [1].

To facilitate an effective and efficient configuration process a couple of concepts for configuring business content for transaction-oriented systems evolved over the years such as checklists, requirements navigation, reference models, and business configuration sets [7]. All these approaches build upon one key idea: Enterprise applications have to be adapted to business needs and business processes of respective enterprises. Therefore the configuration of business software is first of all driven by business questions. Technical software settings are just a result of answering business questions and can thus be mostly derived automatically if all relevant business questions are answered properly. Requirements navigation for example is a checklist mechanism based on a configuration wizard and a knowledge-based system. While using the wizard, an end user with business expertise will be asked about business-related topics in a comprehensive way. During this configuration process potential dependencies of functional configurations will be checked in the background and automatically resolved if necessary. This dependency check and resolving mechanism is based on a dedicated rule system.

Until today, these *business driven* configuration approaches have not been applied to the domain of analytical information systems. Comparable to the aforementioned approaches for configuring transaction-oriented systems, we will first provide business content for analytical information systems and second, develop a configuration approach which is adequate for business users.

2.4 Information Self-Service

In today's BI environments, there is often a clear separation of roles between (technical) BI experts—who configure the BI system and specify the underlying queries such that the reports and analyses are created according to the information needs of the (non-technical) business users—and the business users themselves who consume the standard reports and predefined analytical content. While the entrance barrier for defining information queries has been lowered in recent years by query generation tools such as MySQL Query Browser or Microsoft Access 2007 Query Wizard or Query Designer, these tools nevertheless require the user to have a detailed knowledge of technical data schemas as well as query syntax and semantics, which is still too technical and complex for the business users we target with our approach. At the same time, there is a clear need and demand in companies to enable information self-service for business users.

A key element of such an information self-service environment is the ability to seamlessly integrate (or mash-up) different information sources (that may come from different information providers). By combining several independent queries, business users will be able to develop their own analytical applications and dashboards. Existing mash-up technologies enable the combination of different types of information in order to provide a new integrated information or view [12]. For instance, geographic information can be combined with real estate services and yellow page services for a person looking for an apartment to rent. Yahoo! Pipes provides a graphical user interface for building applications that aggregate Web feeds, Web pages, and other services, creating Web-based applications from various sources, and publishing those applications. Pipes can be seen as communication channels between data sources and gadgets. Users can create mash-ups by dragging pipes from a toolbox and dropping them in work area, specifying data input, interconnecting resources through pipes, and specifying data output formats. However, current mash-up technologies and in particular mash-up applications created using technologies like Yahoo! Pipes usually do not have an enterprise focus and do not support more complex data queries.

2.5 Integrated Collaboration Environments

It is widely agreed that the decision making process of a company has major impact on the success of a company in the market. Companies will only survive if they take the right decisions and if they can rely on a sound decision making process. Within the next decade the meaning of having a good and sophisticated decision making process in an enterprise will become even more important as companies have to face the increasing globalization and the impact of strongest market competitions.

Business Intelligence, data warehousing, and related technologies and approaches provide the technical mechanisms for decision makers to examine historical trends, to measure results, and to look for patterns in the data that might be missed otherwise [9]. The kinds of decisions which need to be made in organizations and which are supported by these mechanisms are manifold. Decisions are made on different hierarchy levels of a company such as on the strategic level, the management level ,and the operative level. Furthermore, the data which form the base of decisions differ regarding the degree of structuring (unstructured, semi-structured, and structured) [11, 13].

A typical collaborative decision making process including different decision makers and business analysts is rather unstructured and dominated by efforts for coordination and inquiries. Today, email is the preferred communication means in such a process, so that the overall decision process is distributed in the form of multiple email threads over several mailboxes. Intransparent decision processes and the loss of information are the ultimate result of these practices. Collaborative decision environments thrive to overcome these challenges by bringing together all relevant information and people in one place for effective and efficient decision making.

In the portfolios of analytical software vendors a couple of collaboration techniques can be found that are partly integrated into their products. In addition to that, in the field of social software, "much of the technological foundation is [already] in place", which could be used for further enhancements of collaborative functionalities of analytic software. However, "no commercial offerings [exist] that comprehensively incorporate the vision of collaborative decision making" [4]. Collaborative decision making is much more than the integration of an instant messaging component into analytical applications as realized in e.g. Microsoft SharePoint Server 2007.

3. CONCEPTS AND ARCHITECTURE

In this section, we first discuss the relevant points for our envisioned BI platform. Afterwards, we introduce the architecture of our proposed BI platform.

3.1 Flexible Data Model

The global data model of our proposed platform consists of two concepts: *DataSpaces* that manage the base data and *InfoSpaces* that model the relations and provide context information for DataSpaces. This is motivated by our goal to have a clear separation of the IT point of view (DataSpaces) from the business point of view (InfoSpaces).

DataSpaces

Our platform builds upon data feeds from central and decentral data sources. Central data sources are managed by IT and cover internal transactional data sources (e.g. ERP, CRM) as well as data warehouses and data marts. Decentral data sources may be added by business users and may cover locally managed data (e.g. data managed by business departments or business users) as well as external data (e.g. external RSS feeds). Both the central and the decentral data sources are encapsulated as DataSpaces. Beside the data itself DataSpaces comprise meta data like ETL definitions or basic post-processing rules like anonymization and data cleaning.

InfoSpaces

InfoSpaces are business representations of data abstracting from technical concepts such as data formats or schemas. They integrate data from different sources for the needs of a specific user group. In general, InfoSpaces may contain business objects (e.g. orders, customers), measures (e.g. quantities, amounts, KPIs) and strategic goals. The aim of InfoSpaces is to provide access to data on an appropriate end-user-oriented data-abstraction level.

3.2 Efficient and Scalable Data Store

The data store faces the challenge to easily and efficiently manage the large amount and high diversity of InfoSpaces. The latter aspect makes RDF—due to its flexibility—a promising solution. Hence, for the management of business context information we propose a technological shift from heavy-weight ontologies and reasoning to a light-weight approach based upon a combination of RDF data and database technology in order to support performing BI analyses also over large scale data sets. We realize efficiency by heavy use of main memory and distributed query processing. As main memory is rare we apply column-store and compression techniques to effectively utilize it.

3.3 Business Configuration Methodology

While in traditional BI environments data integration is performed by IT departments, our proposed business configuration methodology enables business users to perform such configuration tasks by themselves. As data integration can be a complex task, a specific business-user-oriented design will be necessary. Our business configuration process is based on interaction paradigms that allow business users to express their information needs in a simple and intuitive way. In order to facilitate a quick and easy configuration process, the business configuration uses a three step approach. In the first step the scope of an InfoSpace is defined (scoping) while in the second step the detailed configuration is specified (fine-tuning). With the help of a wizard, the configuration process can be guided by questions like "Do you want analytical support for the sales planning?" (scoping) and "Do you perform a monthly or quarterly sales planning?" (fine-tuning). The third and final step is to select the DataSpaces to be used in the analysis, out of a set of candidates that have been suggested by the system as able to deliver the needed data. Beside the integration of new data sources, the business configuration also manages the creation and adaption of business metadata (business strategies, KPIs) and InfoSpaces.

BI Life Cycle

Our envisioned platform supports the life cycle of a BI analytical task as shown in Figure 2. The life cycle visualizes the recurrent adaptation of BI environment to business needs. The lifting of the BI life cycle on a business level results in an environment that can be primarily used by business users. This entails that also configuration activities, which so far have been assigned to the software engineers with a direct effect on the increase of the business/IT coordination costs, is manageable by business users.

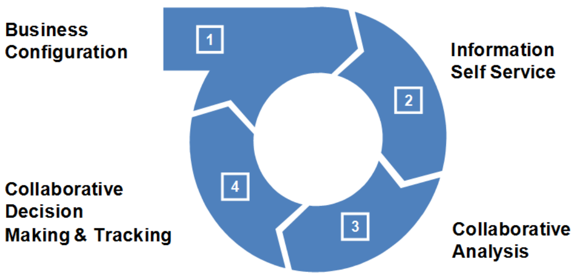

Figure 2: BI life cycle

3.4 Information Self-Service

The envisioned information self-service (ISS) environment is the business user environment for self-service query design, query execution and information mash-ups; it enables business users to find required information using enhanced navigation and search techniques and to build lightweight analytical applications according to their individual needs based upon InfoSpaces. Intuitive portals and wizards are going to help the users to formulate their information needs on the basis of InfoSpaces and to browse the data along different dimensions. An intuitive query consumption environment allows the user to consume information in a table format or as graphical chart. Additionally to the requested data, the ISS environment automatically provides context information, such as related reports, business goals or general/company specific business knowledge (KPI definitions, interdependencies), in order to improve and accelerate the business decision processes. The built-in evaluation of report usage information allows the system to pro-actively make recommendations to business users, e.g. on related reports. Furthermore, different information elements like tables or charts could be composed into dashboards, based on simple patterns.

Regarding ad-hoc analysis, the combination of several data sources and linkages between single queries resembles an advanced development process by itself that may discourage business users. A viable alternative that we aim to investigate is a pattern-based approach, which provides templates of pre-defined and pre-linked queries that fulfill the needs of an explicit scenario or use case. In a Web 2.0-based approach, these could be collected from other users via a shared repository. Moreover, semi-automatic linking of information queries can be a vital approach to business user mash-ups. Usage analyses ("How have other users linked information?") could serve as a basis to automatically generate information mash-ups and propose them to a user. Furthermore, the analysis of data models could also serve as a basis for automatic query linking and mash-up generation. One of the key objectives of our approach is therefore a business-oriented composition approach of analytical mash-ups comprising information queries while technical aspects are handled "behind the scenes" as much as possible.

Due to the multitude of queries that could be generated by business users, a management regarding their semantic content is essential for e.g. realising synergies by reusing predefined queries. For this purpose, metadata can be used for describing the data in the integration layer as well as the queries. [8] differentiates between the following categories of metadata, that is relevant for our purposes: definitional (what does this data mean from a business perspective?), data quality (does this data possess sufficient quality for me to use it for a specific purpose?), navigational (where can I find the data I need?), and lineage (where did this data originate and what's been done to it?).

3.5 Integrated Collaboration Environment

The proposed platform aims to provide collaboration rooms where business users are able to share, comment and take joint decisions. We aim to allow users to work on business reports simultaneously and to allow them to collaborate in all phases of the decision making process including the formulation of queries and the interpretation of results. Live interaction facilities allow users of our system to exchange ideas on-the-fly. Users may bring in information from various public and private sources and may use them as an information source when taking decisions in a team. As such, we want to add a new level to the collaborative decision making process lifting it to the same level as if people would work together in the same room. Furthermore, collaboration rooms are the means to capture user feedback, community knowledge and best practices to improve decision making. Exploiting information usage data to proactively distribute information is a key feature of collaboration rooms.

3.6 Architecture

The overall architecture of our envisioned platform consists of two major layers—the *Ad-hoc and Collaborative Analysis* and the *Integration and Enrichment* layer—and one key repository—the *Global Business Data Model*, see Figure 3.

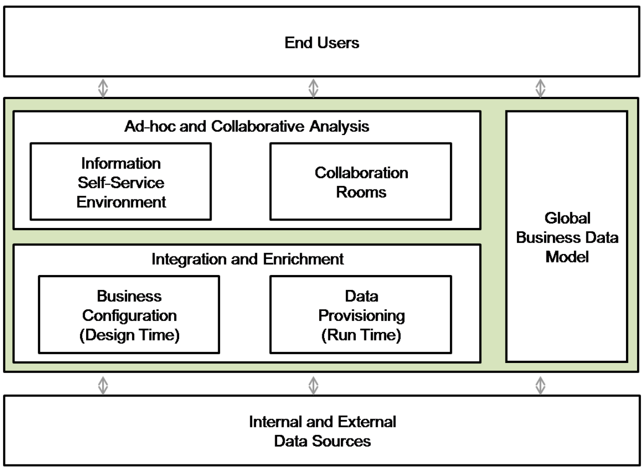

Figure 3: Proposed architectural components

Ad-hoc and Collaborative Analysis

The Ad-hoc and Collaborative Analysis layer facilitates information self-service and collaboration of business users. It consists of two main components: the *Information Self-Service Environment* for intuitive ad-hoc analyses and the *Collaboration Rooms* as seamlessly integrated means for collaborative decision making.

Integration and Enrichment

The Integration and Enrichment layer integrates and semantically enriches heterogeneous data sources within and across organizations. It consists of two main components, namely the *Business Configuration*—the environment for business users with moderate data management skills to setup and configure InfoSpaces and DataSpaces—and the *Data Provisioning* that ensures high flexibility and efficiency.

Global Business Data Model

The Global Business Data Model is the central repository of the platform. On one hand, it serves to "translate" the IT terminology (e.g. data models, fields, etc.) into business terminology and thus to support business users in configuring new information models (InfoSpaces) and create new reports. On the other hand, it provides extended general, company- and user-specific context information for faster and more accurate business decisions. It contains InfoSpace definitions, navigational and lineage metadata (descriptions of ETL processes), definitional metadata (business object, KPIs, dimensions), integrated organizational data (contact persons; business functions, access rights), advanced business data (decisions, legal data strategies), and usage and evaluation data (logs on reports and KPI usages).

4. SUMMARY

In this paper, we have outlined the importance of providing business users with the means for ad-hoc and collaborative decision making. We presented the coarse-grained architecture for the envisioned system and identified the main building blocks that need to be developed. We further described ideas of how to realize the envisioned goals.

5. REFERENCES

[1] B. Arinze and M. Anandarajan. A framework for using OO mapping methods to rapidly configure ERP systems. 2003.

[2] BARC. The BI Survey 8. http://www.bi-survey.com/, 2009.

[3] D. Barstow and G. Arango. Designing software for customization and evolution. In *Proceedings of the 6th international workshop on Software specification and design*, pages 250–255. IEEE Computer Society Press Los Alamitos, CA, USA, 1991.

[4] A. Bitterer, K. Schlegel, B. Hostmann, B. Gassman, M. Beyer, G. Herschel, J. Radcliffe, A. White, T. Payne, W. Andrews, et al. Hype Cycle for Business Intelligence and Performance Management, 2007. *Gartner Research, Stamford, CT*, 2007.

[5] P. A. Boncz, M. L. Kersten, and S. Manegold. Breaking the memory wall in MonetDB. *Commununications of the ACM*, 51(12), 2008.

[6] L. Cao, C. Zhang, and J. Liu. Ontology-based integration of business intelligence. *Web Intelligence and Agent Systems*, 4(3):313–325, 2006.

[7] J. Dittrich, M. Hau, P. Mertens, and A. Hufgard. *Dispositionsparameter in der Produktionsplanung mit SAP: Einstellhinweise, Wirkungen, Nebenwirkungen*. Vieweg+ Teubner Verlag, 2006.

[8] N. Foshay, A. Mukherjee, and A. Taylor. Does data warehouse end-user metadata add value? *Communications of the ACM*, 50(11):77, 2007.

[9] M. Golfarelli, S. Rizzi, and I. Cella. Beyond data warehousing: what's next in business intelligence? In *Proceedings of the 7th ACM international workshop on Data warehousing and OLAP*, pages 1–6. ACM New York, NY, USA, 2004.

[10] J. Gómez, C. Rautenstrauch, P. Cissek, and B. Grahlher. *Einfuhrung in SAP Business Information Warehouse*. Springer, 2006.

[11] G. Gorry and M. Morton. A framework for management information systems. *Sloan Management Review*, 13(1):55–70, 1971.

[12] S. Ikeda, T. Nagamine, and T. Kamada. Application framework with demand-driven mashup for selective browsing. *Journal of Universal Computer Science*, 15(10):2109–2137, 2009.

[13] K. Laudon, J. Laudon, and F. Filip. Management information systems: managing the digital firm. *New Jersey*, 8, 2004.

[14] T. Legler, W. Lehner, and A. Ross. Data Mining with the SAP Netweaver BI Accelerator. In *VLDB 2006*, pages 1059–1068, 2006.

[15] M. T. Özsu and P. Valduriez. *Principles of Distributed Database Systems*. Prentice Hall, 1997.

[16] C. Pedrinaci, I. Markovic, F. Hasibether, and J. Domingue. Strategy-Driven Business Process Analysis. 2009.

[17] R. R. Kimball, M. Ross, W. Thornthwaite, and J. Mundy. *The Data Warehouse Lifecycle Toolkit*. Wiley & Sons, 2008.

[18] M. Spahn, J. Kleb, S. Grimm, and S. Scheidl. Supporting business intelligence by providing ontology-based end-user information self-service. In *Proceedings of the first international workshop on Ontology-supported business intelligence*. ACM New York, NY, USA, 2008.

[19] Sybase Inc. Sybase iq. http://www.sybase.com/products/datawarehousing/sybaseiq.

[20] W3C. Resource description framework (rdf). Technical report.

[21] T. Willhalm, N. Popovici, Y. Boshmaf, H. Plattner, A. Zeier, and J. Schaffner. Ultra Fast in-Memory Table Scan using on-Chip Vector Processing Units. In *VLDB 2009*, pages 385–394, 2009.