Universal Data Warehouse Management System Research Proposal

Nuno Soares Domingues[1,3], Joao C Ferreira[2,3,4], Miguel Nuno Gago[2,3]

***Abstract*—**The author proposes research on the creation of a Data Warehouse Management System (involving all its domains: logical and conceptual data model and language, user interface, software and hardware) that is more flexible than current alternatives in adapting to changes in business requirements, while still improving value to the user.

***Keywords*—**Business Intelligence, OLAP, data warehouse, Database, Data Model, Database Management System, SQL, MDX

I. Introduction

All organizations need relevant, correct and timely information to support decision-making in order to survive and thrive [16] [17] [52] [79] [83]. Business Intelligence (BI) addresses this need [31] [68] [83], by regularly loading organization related activity data into a data warehouse (DW) where it remains kept in an organized manner (data model), so it can be processed by Online Analytical Processing (OLAP) software tools that enable its easy exploration, analysis and visualization by final users so they can make decisions [2] [36].

By definition, the DW must replicate the organization’s business structure [30], thus changes to the latter force changes to the DW data model [21] [72]. Furthermore, these changes to the DW data model imply, for compatibility reasons, the need for changes to the OLAP tools [24] [74], additional to those derived from the evolution of exploration, analysis and visualization requirements. All these changes, implying specialized administration, design and development activities [66] concerning both the DW and the OLAP tools, involve time, risk and costs at levels that jeopardize the success or even the adoption of BI [21] [45] [74] [87].

Thus, it is imperative to create an OLAP oriented Database Management System (DBMS) that maximizes the automation of data model and dependent software adaptation to accelerating changes in organizations, business, and requirements for data exploration, analysis and visualization [21] [37] [53] [87], while still improving user value.

[1]Instituto Politécnico de Lisboa/ Instituto Superior de Engenharia de Lisboa, Rua Conselheiro Emidio Navarro, 1, 1959-007 Lisbon, Portugal

[2]Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, 1649-026 Lisbon, Portugal

[3]Inov Inesc Inovação—Instituto de Novas Tecnologias, 1000-029 Lisbon, Portugal

[4]Business Research Unit (BRU-IUL), Instituto Universitário de Lisboa (ISCTE-IUL), 1649-026 Lisbon, Portugal

The creation of such system is the goal of the research project presented in the current proposal, encompassing 7 layers: logical data model, conceptual data model, data language, user interface and experience, software development, hardware optimization and embedded data analysis/mining.

II. Motivation

This research will add a novel theoretical vision on data through new algebraically founded generic multidimensional logical and conceptual models. It will benchmark current BI practices and bring a new more efficient set of data languages (for queries, manipulation, definition, navigation, calculation, data mining and visualization) and a new DBMS optimized for performance and scalability of the new logical model and for intuitive exploration, analysis and visualization tools reflecting the conceptual model.

IT market analyst Gartner reported a ratio between 70% to 80% for BI project failures [43], and Watson and Ariyachandra [85] mentioned an industry percentage of 30% to 50% of data warehousing projects either belated or out of budget. The current research will contribute to counter such disastrous trend, as the intended solution will allow adopting organizations to significantly lower time and costs on system resetting, database and programming skills, business loss due to downtime and risk of data inconsistency or loss derived from task complexity and dispersion. The solution will help avoid BI projects failure and abandon that are mostly due to their unsuitability to business while resorting to a corrective plan turns out to be too costly, too slow or too late.

The solution will permit very short implementation times for full BI systems or just upgrades. It will ease data integration and migration between organizations, as well as between integrated systems [10], laying the path for extending data warehousing concepts beyond BI. Another aim of the presented research is to minimize disk waste [82] by physical data storage optimization.

For all mentioned reasons the solution is candidate to a new standard framework for data warehousing.

III. State of the Art

A. Business Intelligence

Commonly, BI means automatic delivery of information to decision-makers in an easy way so they can quickly decide [64], despite many not always coincident definitions [46] [69] [83], and lack of academic systematization [40] [58]. Nevertheless, there is consensus that BI encompasses 3 areas: data gathering, data storage and information delivery [33] [70] [59]. As these areas require software, they can benefit from agile development methodologies [3] [4] [18].

B. Database Management Systems and Data Models

DBMSs address limitations of independent system files for data storage [73]. Nevertheless, the first DBMSs, based on the hierarchical (HM) and the network models (NM), required

https://doi.org/10.17758/EARES16.EAP1223138

complex development for simple queries and had no solid theoretical grounding [12] [19].

The relational model (RM), supported on relational algebra [12], is the basis of the relational second generation DBMS [19]. Its fundament is the independence between data representation and machine implementation, which, coupled with a high-level data access language (SQL) [6] constitutes an important advancement over its original competitor, Bachman's NM. Further research on relational databases normalization [12] [23] reinforced the RM simplicity and robustness.

Notwithstanding, Kimball [44] confronts and bashes the RM against the dimensional model (DM) he defends, though the DM is in fact just a denormalized RM. When extending a DM, new tables, columns and relationships must as well be created, requiring downtime, specialized work and reprogramming [44].

In fact, promises are not entirely fulfilled, neither by Codd's RM "to free users from the frustration of having to deal with the clutter of storage representation details" [15], neither by Kimball's DM that "a (...) strength of the dimensional model is that it is gracefully extensible to accommodate new data elements and new design decisions" [44]. Though the OLAP concept was already defined by Codd since 1993 [16], there is still work to be done to fullfill the 12 rules he established for it: multidimensional conceptual view; transparency; accessibility; consistent reporting performance; client-server architecture; generic dimensionality; dynamic sparse matrix handling; multi-user support; unrestricted cross-dimensional operations; intuitive data manipulation; flexible reporting; and unlimited dimensions and aggregation levels.

A third generation of DBMS features object-based logic: Object-Relational (ORDBMS) and Object-Oriented (OODBMS). They combine data structure and dynamic processes, promising flexibility though at the expense of complexity and cost [19] [87].

Traditional relational databases have some maximal limit to the number of columns in a table, and it's awkward to add constantly new attributes as columns [57] [87]. Most Electronic Patient Record Systems (EPRS) have this problem solved by using the Entity-Attribute-Value (EAV) representation model, or Row modeling, where attributes are treated as data [26]. Such generic idea, aiming at recording all kinds of entities, attributes, values and relationships with maximum flexibility, has led to several model concepts: Universal Data Model (UDM), Generic Data model (GDM), EAV with classes and relationships (EAV/CR) and vertical design (VD) [28]. The potential advantages of such approaches are simplicity and easier maintenance, despite lack of consensus and discussion [28].

Anchor modeling, an agile data modeling technique [66], allows extending the DW while keeping all previous versions, that existing software can keep accessing. Despite the authors' claim of simplicity, the Anchor Model (AM) is built upon a large number of concepts, and its classification is complex and more controverse than in traditional models [29]. Like EAV, a reduced and fixed number of columns avoids data access inefficiency and the need for a database administrator (DBA) or architect [9] [19]. However, the fact that AM versioning is achieved through dynamic SQL that adds fields and tables, indicates that this task could as well be done with any other model, namely a third normal form (3NF) RM [67]. Furthermore, AM tables become quickly very numerous and are not fully generic, as it is always necessary to create more than one new table for each new entity or event [29].

Jovanovic et al. [42] criticize AM for the excessive number of concepts, and for its lack of real-world experience compared to Data Vault (DV) modeling, though they recognize both are adequate to intermediate storage as defined in DW 2.0 philosophy [38]. DV patented model, initially presented in industry by Linstedt [50] [51], and conceptually reviewed and systemized by Jovanovic and Bojicic [41] as Conceptual Data Vault (C-DV), only uses three concepts, Hubs (entities and events), Links (relationships) and Satellites (attributes), and functions as AM, though in a simpler way.

Gago [29], after reviewing the above-mentioned models, conceived the Zero Effort Entity Network (ZeEN) logical model, where all dimensions, relationships and facts are recorded in only two tables, in a generic implementation that fits any business case, and compares it favorably to traditional alternatives (relational and dimensional) and AM, demonstrating its ability to support complex multidimensional analysis and scalability needs, while dramatically reducing maintenance costs. The author suggests further research: formal algebraic proof of ZeEN validity and conceptual equivalence to known models; automatic data source detection; new data definition, manipulation, analysis, calculation and navigation languages; support to various data domains and additional metadata; fully generic query algorithms; and full development of a ZeEN DBMS.

NoSQL represents a class of databases that don't follow the established RDBMS principles [81], like ACID (Atomicity, Consistency, Independence and Durability) constrained transactions [20], as they relax transaction consistency in favor of higher availability and scalability [11]. Main reasons for the NoSQL movement are avoidance of unneeded complexity and achievement of high throughput and horizontal scalability over cheap hardware [77], enabling distributed parallel computing and massively scalable internet applications [81]. Though RDBMS have been efficient tackling with geospatial data, geographic information systems (GIS) applications are also turning to NoSQL [65]. In clinical data management, NoSQL has potential to become a key database technology [48], facilitating high query speed, superior to Extensible Markup Language (XML) technology. More than 150 different NoSQL engines exist, with their own data models and interfaces [5] [27] [76], though they share common features [7] [11]. NoSQL data stores can be classified according to their data model [11]. Some authors use a broader definition of NoSQL, including graph database systems, object-oriented database systems, distributed object-oriented stores [11], array databases, and column oriented datastores [65]. In face of such diversity, standard frameworks are emerging, namely a common benchmark [20] and a common interface [7]. Cattell [11] defends that any variant of SQL or NoSQL can be best suited to a specific application. Notwithstanding, NoSQL still attracts some skepticism both on the business side [77] and in academy [75].

C. *Data definition, manipulation and navigation languages*

SQL was an important breakthrough in the 1970's as it facilitated querying relational data, with a syntax near to natural language, while operating on entire data sets [12] [13] [14] [22], so it became the most popular language for databases, albeit with limitations in navigating multidimensional data. Later, Multidimensional Expressions (MDX) language [86] tried to address such limitations by providing a potentially easier way to specify dimensions, hierarquies and metrics. However, there's enough consensus that it didn’t fully succeed [32] [54] [79] [84], though MDX became the standard de-facto due to market reasons [79].

As an alternative to MDX logic, Thomsen [79] [80] proposes a multidimensional model including a dedicated language for navigation, Located Contents (LC), that is fully grounded in mathematics and logic. The author claims that LC has the capacity to generalize any OLAP model, and that its features are being gradually implemented by OLAP vendors, though LC remains confined to academia. Alternative multidimensional models with language logic have been proposed, like AMD [87] and SQLM OLAP [60]. There have also been some attempts at developing tools for maintaining NoSQL databases without having to deal directly with APIs, as an unified calculus [8], a generic schema evolution interface and language, and a generic database programming language [71].

D. *Exploration tools*

In general, information power users use SQL to access operational databases and relational data warehouses, while the remaining information workers use special purpose exploration and visualization tools. Power users also normally use multidimensional navigation tools for OLAP rather than the complex MDX. These tools commonly make use of own technology for dealing with proprietary information cubes.

Software such as Excel allows standard multidimensional data manipulation (albeit with some limitations [61]), thanks to Open Database Connectivity (ODBC), a protocol that enables access to proprietary cubes and thus makes OLAP transparent to the user, one of the 12 Codd’s requirements for evaluating OLAP products [16].

There is additional research on transparent front ends, tools, languages and interfaces to OLAP over various data models [56] [39] [78] [35] [47] [70] [71].

E. *Conclusion*

There is still need for multidimensional DW conceptual and logical models and corresponding DBMS that respect all fundamental requirements of OLAP [41] [79] [1] [61] [63]. User interface for OLAP is an important aspect of the DBMS where many areas are still open [47] [70] [78] [35], and where the possibility of directly obtaining state [87], not just attributes, from data, i.e. data mining, represents a promising research avenue.

IV. OBJECTIVES

The research goal is to deliver an OLAP oriented DBMS for data warehousing – the Universal Data Warehouse Management System (UDWMS) – that maximizes automation of data model and dependent software evolution, is based on novel theory and improves OLAP user value, encompassing 7 layers (logical data model, conceptual data model, data language, user interface and experience, software development, hardware optimization and embedded data analysis/mining), while answering these research questions:

A. *Generic*

How can we conceive an OLAP DBMS in which implementing business rules changes has lower maintenance impact in the data model and dependent tools than in traditional systems, while still improving the value of OLAP?

B. *Specific*

- How can we formally define a new generic logical data model that theoretically fully abstracts any conceptual multidimensional model?
- How will such model test?
- What Extract, Transform and Load (ETL) operations will be eliminated or simplified by using the new logical model (namely pre-aggregation of data cubes)?
- Will it require additional operations?
- How will it impact data storage devices utilization?
- How will it impact BI projects’ risk?
- How will it impact query performance?
- What will be the most efficient way to manage metadata in this model?
- How can we formally define a new conceptual multidimensional data model that fully supports OLAP requirements, namely for hierarchies and navigation?
- How can we design a new unambiguous, and simpler than existing ones, multidimensional data management language with sublanguages for definition, manipulation, navigation, querying and calculation, based on the new conceptual model?
- How can we prototype new OLAP interfaces and tools (visual and other) that interact with the new conceptual model and languages while improving OLAP value?
- How can we develop a full usable OLAP DBMS based on the data model and prototyped interfaces and languages?
- Will ETL operations be easier than current practice when using the new DBMS?
- How will it benchmark in query performance?
- How will it impact data storage devices utilization?
- How will it impact BI projects’ risk?
- How will user experience improve?
- Will the user obtain higher information value?
- Are there specific hardware technologies and configurations that can optimize the implementation of the new DBMS?
- How will it scale to “Big Data”?
- How can we seamlessly integrate data mining algorithms and processed calculations in the DBMS, through modeling, software and hardware techniques?

V. RESEARCH PROJECT PROPOSAL

A. *Basis for work development*

For the stated research goal, as it aims at answering questions relevant to a human problem (organizing and exploring information) via the creation of innovative artifacts, the appropriate research paradigm is Design Science Research (DSR) as proposed by Hevner & Chatterjee [34]. These authors acknowledge that Design Science Research Process Methodology (DSRM) [62] provides a commonly accepted framework and a mental model for the presentation of research under such approach, and we found it to be an applicable systematic guideline to our design situation and problem context, as shown in the following description.

B. Work description

Our work will be pursued along the six activities of DSRM, synthesized by Peffers et al. [62] from previous efforts in Design Science Research:

- Activity 1 - Problem identification and motivation. This is covered in the previous sections Introduction, Motivation, State of the Art and Objectives, which render what is the currently existing problem in data warehousing that hinders BI processes, and why it is utterly important to overcome it.
- Activity 2 - Define objectives of a solution. Section Objectives enumerates the questions the problem raises and declares the intended artifactual solution that the respective answers will lead to.
- Activity 3 - Design and Development. Deductive and inductive reasoning will lead iteratively to full conceptualization of the new data architecture. Data models traditionally used for data warehousing, the Relational Model and the Dimensional Model [41] [61] and respective DBMSs will be reviewed, as well as alternative approaches proposed in literature and practice, like Anchor Modeling [66], Data Vault [41] [50] [51], Entity-Attribute-Value [55], Associative Model of Data [87], NoSQL and NewSQL (ACID enforced NoSQL), in order to serve as a basis for the definition of both logical and conceptual new data models. Existing high-level data languages will be reviewed, along with multidimensional conceptual modeling research, in order to lay the foundation for a new conceptual model and language comprehending sublanguages for navigating, querying and visualizing multidimensional databases, defining schema, manipulating data, calculating expressions and aggregations, and real time data mining. State of the art data mining techniques will be reviewed and adapted to enable embedding in the DBMS prototype. Existing user interface paradigms and metaphors will be analyzed and compared in order to merge best features and eventually trigger a new concept that permits the smoothest, most intuitive and richest human / machine interaction in the DBMS. Low-level programming languages and physical data storage and access optimization techniques will also be explored, as to enable the concretion of a fully functional autonomous DBMS prototype made to fit the new multidimensional logical model and the new interface concept.
- Activity 4 - Demonstration. Algebra literature will be reviewed and compared, especially related to data models, such as relational algebra [22], graph theory [25], Located Contents logic [79], Multidimensional SQL (SQLM OLAP) [60], and relational algebra with groupings and aggregates (ALGaggr) [49], in order to allow theoretical understanding and validation of the researched models, and a foundation for the new data artifacts: the logical model will be formally defined and proven to satisfy logical independence, the conceptual model will be proven to be theoretically grounded on first-order logic, and the new data languages will be proven to map to a self-contained algebra that adheres to the conceptual model. Proofs of concept of these artifacts will be further validated experimentally using simulated data representing example cases of real-world requirements. The DBMS interface will be demonstrated in a DBMS mockup first, then together with the DBMS itself when the software is ready, then finally on top of the new hardware artifacts. These demonstrations will contemplate not only OLAP manipulation, but also the new OLAM functionalities.
- Activity 5 - Evaluation. An experimental approach will be used to benchmark the new DBMS artifacts against reviewed alternatives on a specific business scenario, in a first step using simulated data. Then, after optimization of the data implementation, language and user interfaces, prototypes will be tested with real-world business cases from organizations that adhere to the project, in parallel with their current systems, allowing to assess impact in ETL, storage, risk and performance. The criteria for reaching the research goal, that will be used for testing and benchmarking the solution, will be: cost minimization, simplicity and easiness of required administration tasks, effort and time needed for implementing and maintaining the system, risk of system being abandoned or failing to meet requirements and deadlines, and maximization of simplicity and elegance of the solution. Complementarily, other important aspects will be evaluated in order to apprehend the effective value of the solution: query performance, scalability, regular ETL time and effort, and volume of physical storage required.
- Activity 6 – Communication. The design process results for each artifact (logical model, conceptual model, data language, user interface, DBMS software, dedicated hardware, and data mining embedded feature) comprehending the previous activities will be, at successive moments along the four years of the research, published in appropriate journals and presented in conferences of a relevant area, as well as being showcased in the companies that collaborated in the research. A PhD thesis will be delivered and defended at the end of the third year of research, covering the themes that are closed by then.

REFERENCES

[1] Abelló, A., Samos, J., & Saltor, F. (2006). "YAM2: a multidimensional conceptual model extending UML". Information Systems, 31(6), 541–567.

https://doi.org/10.1016/j.is.2004.12.002

[2] Agrawal, V., Sundararaghavan, P. S., Ahmed, M. U., & Nandkeolyar, U. (2009). "View Materialization in a Data Cube: Optimization Models and Heuristics". K. Siau & J. Erickson (Eds.), Advanced Principles for Improving Database Design, Systems Modeling, and Software Development. Hershey, PA: IGI Global. doi:10.4018/978-1-60566-172-8

https://doi.org/10.4018/978-1-60566-172-8

[3] Ambler, S. (2002). "Agile modeling: effective practices for extreme programming and the unified process". New York, NY: John Wiley & Sons, Inc.

[4] Ambler, S. (2003). "Agile database techniques: effective strategies for the agile software developer". Indianapolis, IN: Wiley Publishing.

[5] Anand, S., & Ramble, P. (2012). "The State of NoSQL in 2012". Retrieved January 24, 2012, from http://practicalcloudcomputing.com/post/16109041412/the-state-of-nosql-in-2012

https://doi.org/10.1145/320455.320457

[6] Astrahan, M. M., Blasgen, M. W., Chamberlin, D. D., K.P.Eswaran, J.N.Gray, Griffiths, P. P., King, W. F., et al. (1976). "System R: relational approach to database management". ACM Transactions on Database Systems (TODS), 1(2), 97–137.

[7] Atzeni, P., Bugiotti, F., & Rossi, L. (2013). "Uniform access to NoSQL systems". Information Systems, 1–17. doi:10.1016/j.is.2013.05.002

https://doi.org/10.1016/j.is.2013.05.002

[8] Benzaken, V., & Castagna, G. (2013). "Static and dynamic semantics of NoSQL languages". ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL) (pp. 101–113).

https://doi.org/10.1145/2429069.2429083

[9] Bouman, R., & Dongen, J. Van. (2009). "Pentaho solutions: business intelligence and data warehousing with pentaho and mysql". Indianapolis, IN: Wiley.

[10] Caetano, T. V, & Costa, C. J. (2012). "Data Warehousing num contexto de Sistemas Integrados". CAPSI 2012.

[11] Cattell, R. (2010). "Scalable SQL and NoSQL data stores". ACM SIGMOD Record, 39(4), 12–27.

https://doi.org/10.1145/1978915.1978919

[12] Codd, E. F. (1970). "A relational model of data for large shared data banks". Communications of the ACM, 13(6), 377–387.

https://doi.org/10.1145/362384.362685

[13] Codd, E. (1971). "A data base sublanguage founded on the relational calculus". SIGMOD Workshop on Data Description, Access and Control.

https://doi.org/10.1145/1734714.1734718

[14] Codd, E. (1972). "Relational completeness of data base sublanguages". Computer Sciences.

[15] Codd, E. F. (1979). "Extending the database relational model to capture more meaning". ACM Transactions on Database Systems, 4(4), 397–434.

https://doi.org/10.1145/320107.320109

[16] Codd, E., Codd, S., & Salley, C. (1993). "Providing OLAP to User-Analysts: An IT Mandate".

[17] Cody, W. F., Kreulen, J. T., Krishna, V., & Spangler, W. S. (2002). "The integration of business intelligence and knowledge management". IBM Systems Journal, 41(4), 697–713. doi: 10.1147/sj.414.0697

https://doi.org/10.1147/sj.414.0697

[18] Collier, K. (2011). "Agile Analytics: A Value-Driven Approach to Business Intelligence and Data Warehousing". Crawfordsville, IN: Addison-Wesley.

[19] Connolly, T., & Begg, C. (2005). "Database systems: a practical approach to design, implementation, and management (4th ed.)". Essex, England: Pearson Education.

[20] Cooper, B. F., Silberstein, A., Tam, E., Ramakrishnan, R., & Sears, R. (2010). "Benchmarking cloud serving systems with YCSB". Proceedings of the 1st ACM symposium on Cloud computing - SoCC '10 (p. 143). New York, New York, USA: ACM Press.

https://doi.org/10.1145/1807128.1807152

[21] Curino, C. A., Tanca, L., Moon, H. J. M., & Zaniolo, C. (2008). "Schema evolution in wikipedia: toward a web information system benchmark". International Conference on Enterprise Information Systems.

[22] Darwen, H. (2012). "SQL: A Comparative Survey". bookboon.com.

[23] Date, C. J. (2012). "Database Design and Relational Theory: Normal Forms and All That Jazz". Sebastopol, CA: O'Reilly.

[24] De Vries, D., & Roddick, J. F. (2007). "The case for mesodata: An empirical investigation of an evolving database system". Information and Software Technology, 49(9-10), 1061–1072. doi:10.1016/j.infsof.2006.11.001

https://doi.org/10.1016/j.infsof.2006.11.001

[25] Diestel, R. (2010). "Graph Theory (4th ed.)". New York, NY: Springer. https://doi.org/10.1007/978-3-642-14279-6

[26] Dinu, V., Nadkarni, P., & Brandt, C. (2006). "Pivoting approaches for bulk extraction of Entity-Attribute-Value data". Computer methods and programs in biomedicine, 82(1), 38–43.

https://doi.org/10.1016/j.cmpb.2006.02.001

[27] Edlich, S. (2014). "NoSQL - Your Ultimate Guide to the Non-Relational Universe". Retrieved June 19, 2014, from http://nosql-database.org/

[28] Eessaar, E., & Soobik, M. (2008). "On Universal Database Design". Baltic DB & IS.

[29] Gago, M. R. (2013). "Modelo ZeEN: uma abordagem minimalista para o desenho de data warehouses". Universidade Nova de Lisboa.

[30] Golfarelli, M. (2008). "DFM as a Conceptual Model for Data Warehouse". Encyclopedia of Data Warehousing and Mining (3), 638–640.

https://doi.org/10.4018/978-1-60566-010-3.ch100

[31] Golfarelli, M., Mandreoli, F., Penzo, W., Rizzi, S., & Turricchia, E. (2012). "OLAP query reformulation in peer-to-peer data warehousing". Information Systems, 37(5), 393–411.

https://doi.org/10.1016/j.is.2011.06.003

[32] Halpin, T., & Morgan, T. (2008). "Information modeling and relational databases (2nd ed.)". San Francisco, CA: Morgan Kaufmann Publishers.

[33] Hannula, M., & Pirttimäki, V. (2003). "Business intelligence empirical study on the top 50 Finnish companies". Journal of American Academy of Business, (March).

[34] Hevner, A., & Chatterjee, S. (2012). "Design Research in Information Systems". New York, NY: Springer. doi:10.1007/978-1-4419-6108-2

https://doi.org/10.1007/978-1-4419-6108-2

[35] Hirte, S., Seifert, A., Baumann, S., Klan, D., & Sattler, K.-U. (2012). "Data3 -- A Kinect Interface for OLAP Using Complex Event Processing". 2012 IEEE 28th International Conference on Data Engineering, 1297–1300. doi:10.1109/ICDE.2012.131

[36] Inmon, W. H. (1999). "Data Marts and the Data Warehouse: Information Architecture for the Millenium". California: Informix.

[37] Inmon, W. H. (2005). "Building the data warehouse (4th ed.)". Indianapolis, IN: Wiley Publishing.

[38] Inmon, W. H., Strauss, D., & Neushloss, G. (2008). "DW 2.0: The Architecture for the Next Generation of Data Warehousing". Burlington, MA: Morgan Kaufman.

[39] Johnston, T., & Weis, R. (2010). "Managing time in relational databases: how to design, update and query temporal data". Burlington, MA: Morgan Kaufmann Publishers.

[40] Jourdan, Z., Rainer, R. K., & Marshall, T. E. (2008). "Business Intelligence: An Analysis of the Literature". Information Systems Management, 25(2), 121–131.

https://doi.org/10.1080/10580530801941512

[41] Jovanovic, V., & Bojicic, I. (2012). "Conceptual Data Vault Model". SAIS Conference, 131–136.

[42] Jovanovic, V., Bojicic, I., Knowles, C., & Pavlic, M. (2012). "Persistent Staging Area Models for Data Warehouses". Issues in Information Systems, 13(1), 121–132.

[43] Kernochan, W. (2011). "Why Most Business Intelligence Projects Fail". Enterprise Apps Today. Retrieved November 15, 2012, from http://www.enterpriseappstoday.com/business-intelligence/why-most-business-intelligence-projects-fail-1.html

[44] Kimball, R. (1997). "A Dimensional Modeling Manifesto: Drawing the Line Between Dimensional Modeling and ER Modeling Techniques". DBMS and Internet Systems.

[45] Kimball, R., & Ross, M. (2010). "The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence". Indianaplois, IN: Wiley Publishing.

[46] Kobielus, J. (2010). "What's Not BI? Oh, Don't Get Me Started....Oops Too Late...Here Goes...." Retrieved from http://blogs.forrester.com/james_kobielus/10-04-30-what%E2%80%99s_not_bi_oh_don%E2%80%99t_get_me_startedoops_too_latehere_goes, April 30, 2010.

[47] Lafon, S., Bouali, F., Guinot, C., & Venturini, G. (2013). "3D and Immersive Interfaces for Business Intelligence: The Case of OLAP". 2013 17th International Conference on Information Visualisation, 272–277. doi:10.1109/IV.2013.36

https://doi.org/10.1109/IV.2013.36

[48] Lee, K. K.-Y., Tang, W.-C., & Choi, K.-S. (2013). "Alternatives to relational database: comparison of NoSQL and XML approaches for clinical data storage". Computer methods and programs in biomedicine, 110(1), 99–109.

[49] Libkin, L. (2003). "Expressive power of SQL". Theoretical Computer Science, 296, 379–404.

https://doi.org/10.1016/S0304-3975(02)00736-3

[50] Linstedt, D. E. (2001). "Patent-Method And System of Data Warehousing". Pat: US 2002/0161778 A1

[51] Linstedt, D. E. (2002). "Data Vault Series 1 - Data Vault Overview". The Data Administration Newsletter. Retrieved from http://www.tdan.com/view-articles/5054

[52] Lönnqvist, A., & Pirttimäki, V. (2006). "The Measurement of Business Intelligence". Information Systems Management, 23(1), 32–40.

https://doi.org/10.1201/1078.10580530/45769.23.1.20061201/91770.4

[53] Moody, D. L., & Kortink, M. A. R. (2000). "From enterprise models to dimensional models: a methodology for data warehouse and data mart design". Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'2000) (Vol. 2000, pp. 1–12).

[54] Mundy, J., & Thornthwaite, W. (2008). "The Microsoft Data Warehouse Toolkit: With SQL Server2005 and the Microsoft Business Intelligence Toolset" (Vol. 2006, pp. 1–692).

[55] Nadkarni, P. (2011). "Metadata-driven Software Systems in Biomedicine: Designing Systems that Can Adapt to Changing Knowledge". (K. J. Hannah & M. J. Ball, Eds.). London: Springer-Verlag London Ltd.

https://doi.org/10.1007/978-0-85729-510-1

[56] Nadkarni, P. , & Brandt, C. (1998). "Data extraction and ad hoc query of an entity-attribute-value database". Journal of the American Medical Informatics Association : JAMIA, 5(6), 511–27.

https://doi.org/10.1136/jamia.1998.0050511

[57] Nadkarni, P., Marenco, L., Chen, R., Skoufos, E., Shepherd, G., & Miller, P. (1999). "Organization of heterogeneous scientific data using the EAV/CR representation". Journal of the American Medical Informatics Association : JAMIA, 6(6), 478–93.

https://doi.org/10.1136/jamia.1999.0060478

[58] Negash, S. (2004). "Business Intelligence". Communications of the Association for Information Systems, 13, 177–195.

https://doi.org/10.17705/1CAIS.01315

[59] Negash, S., & Gray, P. (2008). "Business Intelligence". F. Burstein & C. W. Holsapple (Eds.), Handbook on Decision Support Systems 2: Variations. Berlin: Springer-Verlag.

[60] Pedersen, D., Riis, K., & Pedersen, T. (2001). "A powerful and SQL-compatible data model and query language for OLAP". X. Zhou (Ed.), Thirteenth Australasian Database Conference (ADC2002). Melbourne: Australian Computer Society, Inc.

[61] Pedersen, T. B., & Jensen, C. S. (2001). "Multidimensional database technology". Computer, (December)

https://doi.org/10.1109/2.970558

[62] Peffers, K., Tuunanen, T., Rothenberger, M. a., & Chatterjee, S. (2008). "A Design Science Research Methodology for Information Systems Research". Journal of Management Information Systems, 24(3), 45–77. doi:10.2753/MIS0742-1222240302

[63] Pendse, N. (2008). "What is OLAP ? The BI Verdict". Retrieved November 15, 2012, from http://www.bi-verdict.com/fileadmin/dl_temp/0493635e66cd5c7af3f2626b88033c8a/fasmi.htm?user_id=

[64] Power, D.J. (2007). "A Brief History of Decision Support Systems". DSSResources.COM, World Wide Web, http://DSSResources.COM/history/dsshistory.html, version 4.0, March 10, 2007.

[65] Queiroz, G., Monteiro, A., & Câmara, G. (2013). "Bancos de dados geográficos e sistemas NoSQL: onde estamos e para onde vamos". Revista Brasileira de Cartografia, 65(3), 479–492.

https://doi.org/10.14393/rbcv65n3-44800

[66] Rönnbäck, L., Regardt, O., Bergholtz, M., Johannesson, P., & Wohed, P. (2010a). "Anchor modeling — Agile information modeling in evolving data environments". Data & Knowledge Engineering, 69(12), 1229–1253.

https://doi.org/10.1016/j.datak.2010.10.002

[67] Rönnbäck, L., Regardt, O., Bergholtz, M., Johannesson, P., & Wohed, P. (2010b). "From Anchor Model to Relational Database". Retrieved from http://www.anchormodeling.com/wp-content/uploads/2010/09/AM-RDB.pdf

[68] Rud, O.P. (2009). "Business intelligence success factors: tools for aligning your business in the global economy". Hoboken, NJ: Wiley & Sons.

[69] Sabanovic, A. (2008). "Business Intelligence Software: Customers' Understanding, Expectations and Needs". University of Kristianstad.

[70] Saias, J., Quaresma, P., Salgueiro, P., & Tiago Santos. (2012). "Binli: An ontology-based natural language interface for multidimensional data analysis". Intelligent Information Management, 4(September), 225–230.

https://doi.org/10.4236/iim.2012.45033

[71] Scherzinger, S., Klettke, M., & Störl, U. (2013). "Managing Schema Evolution in NoSQL Data Stores". arXiv Preprint arXiv:1308.0514.

[72] Sen, A., & Sinha, A. (2005). "A comparison of data warehousing methodologies". Communications of the ACM, 48(3), 79–84.

https://doi.org/10.1145/1047671.1047673

[73] Silberschatz, A., Korth, H. F., & Sudarshan, S. (2011). "Database system concepts (6th ed.)". New York, NY: McGraw-Hill.

[74] Simsion, G. C., & Witt, G. C. (2005). "Data modeling essentials (3rd ed.)". San Francisco, CA: Morgan Kaufmann Publishers.

[75] Stonebraker, M. (2010). "SQL databases v. NoSQL databases". Communications of the ACM, 53(4), 10.

https://doi.org/10.1145/1721654.1721659

[76] Stonebraker, M. (2011). "Stonebraker on NoSQL and enterprises". Communications of the ACM, 54(8), 10. doi:10.1145/1978542.1978546

https://doi.org/10.1145/1978542.1978546

[77] Strauch, C., & Kriha, W. (2012). "NoSQL databases". Stuttgart.

[78] Terwilliger, J. F., Delcambre, L. M. L., & Logan, J. (2007). "Querying through a user interface". Data & Knowledge Engineering, 63(3), 774–794. doi:10.1016/j.datak.2007.04.007

https://doi.org/10.1016/j.datak.2007.04.007

[79] Thomsen, E. (2002). "OLAP Solutions: Building Multidimensional Information Systems (2nd ed.)". New York, NY: John Wiley & Sons, Inc.

[80] Thomsen, E. (2011). "Knowing the World Is Language: a theory of logic based on a new paradigm of cognitive processing that also provides a foundation for language and mathematics". dsslab.com. DSSlab.

[81] Tiwari, S. (2011). "Professional NoSQL". Indianapolis, IN: Wiley & Sons

[82] Todman, C. (2001). "Designing a data warehouse: supporting customer relationship management". Upper Saddle River, NJ: Prentice-Hall.

[83] Turban, E., Sharda, R., Aronson J.E., & King, D. (2007). "Business Intelligence: a Managerial Approach". New Jersey: Pearson Education

[84] Vassiliadis, P., & Sellis, T. (1999). "A survey of logical models for OLAP databases". ACM SIGMOD Record, 28(4), 64–69.

https://doi.org/10.1145/344816.344869

[85] Watson, H. J., & Ariyachandra, T. (2005). "Data Warehouse Architectures : Factors in the Selection Decision and the Success of the Architectures". Athens, GA.

[86] Whitehorn, M., Zare, R., & Pasumansky, M. (2002). "Fast track to MDX (2nd ed.)". Nottingham, UK: Springer.

[87] Williams, S. (2002). "The associative model of data (2nd ed.)". Lazy Software.