ISSN 2229-5518

Designing Decision Support Systems Approaches using Entity/Relationship Data Schema: A Survey

Apollinaire Bamana Batoure, Kolyang

Abstract— Data warehouse are used for knowledge extraction and decision support. Their design is based on existing operational systems or needs expressed by users, or both. In order to better understand how approaches use Entity/Relationship data schema to produce multidimensional ones, we studied and presented eight of them. In addition to these approaches, we presented multidimensional canonical partitioning approach, earlier proposed. We compare approaches based on criteria regarding the way they determine multidimensional elements; how they tackle modeling phases; and their automation degree.

Index Terms— Entity/Relationship, Decision support system, Data warehouse design, Transactional system

1 Introduction

THE need for data analysis, for extracting new knowledge and decision-making purposes, has required the restructuring of information systems from their basic schema. Thus, the so-called transactional data schema have been transformed in order to obtain usable data schema for decision support systems. Several scientific studies, largely conducted around the 2000s, have focused on this issue. The issue is the transition from transactional systems (**TS**) to decision support systems (**DSS**), in order to get best benefit from numerous data present in companies.

In this article, we study and present some of the approaches, then present the approach we proposed. Therefore, we make a comparative study of these different approaches. To achieve it, we start by a recall of TS and DSS concepts and different design types for decision support systems or data warehouses. Then, we present eight most referenced approaches, in addition to the approach we proposed. Finally, we compare these approaches, based on some criteria. These criteria allow us to better understand how approaches determine multidimensional elements (facts, measures, dimensions, and hierarchies); how they reach, eventually, the logical or physical modeling levels; and their automation degree.

2 Transactional and Decision Support Systems Concepts

In this section, we review the concepts of transactional and decision support systems before comparing them.

2.1 Transactional systems

A transaction is a grouping, in one set, of changes to be made on a database [1]. On-line transactional processing (OLTP), implemented in a computer application, allows a large number of users to submit transactions via their terminals to a system. The OLTP allows in priority, insertion, modification, fast, efficient and secure interrogation on database. The systems store the current data of the organization and do not constitute archives [2]. Queries are simple, non-aggregative and implemented according to a relational structure, normalized to different degrees (minimum redundancy, data integrity, facility of updating). Transactional operations should check properties of atomicity, consistency, isolation, and durability. We talk of ACID properties [1], [3]. These properties transform a system from a coherent state to another coherent state.

To get to Database Management Systems (DBMS), automatic information management has come a long way, which started with files and systems to manage them. Methods to access information contained in files are then sequential, indexed or hashed [4]. What stands for information systems (IS) was reserved only for banking and industrial management applications. In view of the amount of information produced by companies, which is becoming increasingly important and difficult to manage by files, databases (DB) emerge. First databases types were network and hierarchy [Silberschatz2010]. The development of DB, to the detriment of files, is also motivated by the need of separation between data and programs, and especially the reduction of the redundancy produced in files[23]. Other TS types are logical DBs and deductive DBs.

A DBMS is a set of systems software that supports structuring, storing, updating, and maintaining data [5]. The first generation of DBMS, marked by the separation of data description and manipulation by programs, appeared in the late 1970s [4]. The second generation, driven by the relational approach, was commercialized from the 1980s. The third generation, on the other hand, supports extensible data models. It integrates relational and object, as well as distributed architectures. The fourth generation, current one, focuses on Web, Internet of things, cloud computing, badly or non-structured information, multimedia objects, decision-making and

- *Apollinaire Bamana Batoure is currently completing a PhD program in Computer Science in University of Ngaoundere, Cameroon, PH-01123456789. E-mail: apollinaire.batoure@univ-ndere.cm*
- *Kolyang is Prof.Dr-Ing. Habil. in Computer Science in University Maroua, Cameroon, PH-01123456789. E-mail: kolyang@univ-maroua.cm*

knowledge extraction from big data [6], [7], [8], [9]. These new technologies also bring new types of DBMS with specific features [10].

2.2 Decision-support Systems

Main idea behind introduction of decision support systems in the 1990s was to help companies with a large amount of archived information, but not always well organized, to make the most of in order to help in making decisions in relation to the facts observed previously. We talk of Business Intelligence (BI) [12], [12], [13]. A data warehouse is a collection of historical, time-varying, subject-driven, aggregated data in a single database, managed in a distributed storage environment, and helping in business decision-making [14], [15], [16]. According to I. Comyn-Wattiau and J. Akoka in [17], [18] "The data warehouse is probably, with the Internet, one of the recent trends that companies will increasingly exploit in the years to come. The data warehouse is the heart, the backbone of the decision support system".

Decision support systems architecture makes possible the consolidation of various components [19], [16]. It influences several factors such as the availability of data and the effectiveness of treatments [20]. The standard architecture consists of databases taken as sources, a central data warehouse, and multiple clients who use the data each [21]. For transition from sources to storage, Extraction-Transformation-Loading (ETL) processes are used. From storage to exploitation or representation, multidimensional servers are used [22].

The main data models in decision support systems are star and snowflake [2], [19]. These models divide the data warehouse (DW) into stores or marts. Data marts are the smallest piece of business intelligence [24].

On-Line Analytical Processing (OLAP) is a category of software technology that enables analysts [25], [26]. There are several models for data analysis [27]. The main ones are Multidimensional OLAP (MOLAP) and Relational OLAP (ROLAP). Other models include Hybrid OLAP (HOLAP), Dynamic OLAP (DOLAP) and Spatial OLAP (SOLAP). More recently, Dehne et al. [28] presented Velocity OLAP (VOLAP), a real-time OLAP for high velocity data; and Zeng et al. [29] proposed Incremental OLAP (IOLAP) to improve performance of query execution.

OLAP operations are performed on cubes or hypercubes [30]. Basic operators are roll-up, drill-down, slice and dice [31]. We can combine them and write complex expressions on cubes. However, there are other operators which include screening, scoping or pivoting. An OLAP algebra, with graphical representation is defined in [32]. OLAP query language in multidimensional databases is Multi Dimensional eXpression (MDX) [33]. Unlike the SQL language that returns a record-set in tabular form, the MDX language returns a multidimensional data flow (scalar, dimension and hierarchy, level, member, tuple and set). Dimensions, hierarchies and levels are for MDX what tables and columns are for SQL. XML standard is also used to represent and query multidimensional data, using *XQuery* and *XML for Analysis* (**XMLA**) [33], [34]. Table 1 summarizes the main differences between these two systems.

TABLE 1
DIFFERENCES BETWEEN OLTP AND OLAP

Characteristics	**OLTP (Transactional system)**	**OLAP (decision support system)**
Application	Ordinary management, production	Analysis / Decision-support
Users	Information system experts	Decision-makers
Data schema	Entity / Relationship	Star / Snowflake / Constellation
Normalization	Frequent	Scare
Data	Up to date / Raw	Archived / Aggregated
Up dating	Immediate / Real time	Delay or postpone
Queries	Simple / Regular / Predefine / Predictable	Complex / Irregular / Non-Predictable / Ad-hoc
Query language	SQL, QBE, QUEL	MDX, XQuery, XMLA
Analysis axis	Uni- or bi-directional	Multidimensional or multi-axes
Operations	Modification / Up to date / Cancelling / Insertion	Lecture / Cross analysis / Refreshment
Data size	Mega or Gigabytes	Tera, Peta or Zetabytes

In order to have full advantage of decision support technology, bridges have been created. Decision support systems can be obtained from transactional systems or from user needs.

3 DECISION SUPPORT DESIGN APPROACHES

Decision support systems approaches are classified in three main categories, defined and adopted by researchers and industrialists [13], [27], [35]. They are top-down, bottom-up or mixed approaches.

The requirement-driven approach, also called user-driven or top-down approach, defines the conceptual schema of the data warehouse from user needs. Difficulties with these approaches is instability and constant evolution of user needs [36]. Among these approaches, we can mention the one of Kimball [16]. The approach is not formalized but can be seen as a detailed guide for identifying the multidimensional concepts that give rise to the data warehouse schema. In addition to this approach, we have approaches defined by Gam, Mazon, Prat and Tsois [37], [38], [39], [40].

The supply-driven approach, also called date-driven or bottom-up approach, defines multidimensional schema from operational data sources. These data sources can be on Entity/Relationship (ER) [41], [42], [43], XML [34], [44], [45] or UML data schema [39], [77]. These approaches directly take into account the data and operational systems used in the company. But if these sources are very numerous, it will require more human, temporal and financial resources for their efficient exploitation [46], [47]. We have several proposed approaches that take into account the ER data schema. We focus on these approaches in the following sections.

Finally, hybrid, combined or mixed approaches consider user needs and existing data schema to produce multidimensional model [35], [36].

Analysis of sources produced candidate schema and analysis of needs produced ideal schema. Confrontation follows, for production of definitive schema. Among these approaches, we can cite that of Annoni et al. [48] which specifies the requirements in tabular form. From this formalization, an automatic process guides the choice of decision support system architecture. The author proposes a catalog of patterns, which capitalizes the development process for the purpose of systematic reuse. We can also mention in this category, approaches proposed by Carneiro, Khouri and Phipps [49], [50], [51].

Works carry out in [27], [52], [53], [54], [55], show that most design approaches were developed between 1998 and 2010 (Fig. 2, adapted from [52]). According to [56] and [57], data warehouse design are still considered as research trend. The following works aim at improving the first one or to study other types of sources to consider (i * framework, ontologies, Web, etc.) [38], [50], [58], [59], [60], [62], [63], [64].

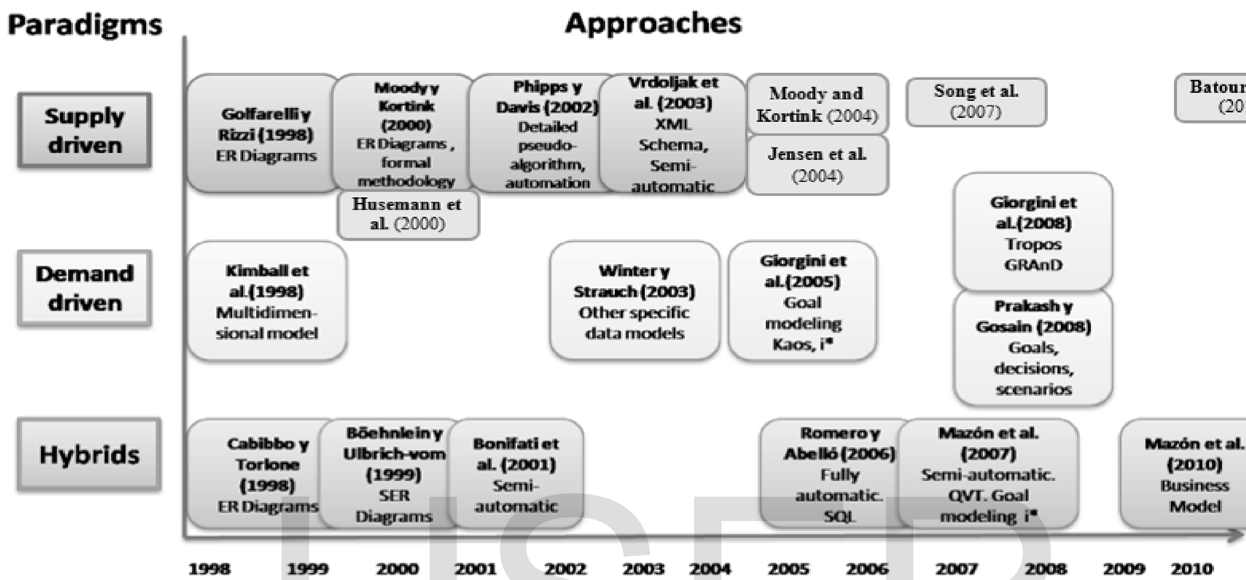

Fig. 1. Chronological diagram of data warehouse design approaches

4 Approaches using Entity/Relationship data schema

These approaches are mostly supply-driven. However, some are mixed. Sources taken into account are ER. Eight of these approaches are presented, arranged in ascending chronological order.

4.1 Cabibbo et al., 1998

The Cabibbo and Torlone approach ([65]) builds the multidimensional model from an ER schema. This schema presents an integrated view of transactional databases. The two logical schema (relational and multidimensional) are generated. The method consists of four steps:

- definition of facts and dimensions following a global and manual analysis of the source diagrams;
- restructuring of the ER schema to reveal facts, dimensions and hierarchical levels;
- derivation of the dimensional graph from an extended entity-association schema;
- transformation of the dimensional graph into a multidimensional diagram.

4.2 Golfarelli et al., 1998

Golfarelli et al. [66], [67] proposes the approach called Dimensional Fact Model (DFM). It uses direct, acyclic, and weakly connected graphs to represent the multidimensional data schema. Nodes are attributes and entities. They are related to fact which is the graph root. Dimensions and their hierarchies are defined from 1-N cardinality associations. The method consists of six steps:

- analysis of data sources to generate the conceptual diagram describing them;
- collection and analysis of user needs;
- construction of the multidimensional conceptual diagram, represented by a fact diagram (FD) model;
- defining the logic model of the data warehouse, by translating each identified entity into a relational table;
- generation at the physical level by a ROLAP software model and the definition of optimal structures using indexes;
- the last step estimates the query workload in order to validate the multidimensional conceptual schema generated.

This approach is one of the few that defines data warehouse modeling up to the physical level, and estimates queries workload.

4.3 Boehnlein, 1999

In [58], Boehnlein begins by restructuring the input ER model. This restructuring, called the Structured Entity Relationship Model (SERM), is used to derive the multidimensional model more easily. Its purpose is to design model extensions, visualize dependencies order between object types, and remove inconsistencies and superfluous relationships. The approach is summarized in the following steps:

- transformation of the ER model into a structured one;
- identification and definition of facts and measures;
- identification and production of dimensions and hierarchies through the use of direct or transitive functional dependencies.

ISSN 2229-5518

From the multidimensional elements obtained, the data schema is generated.

4.4 Romero et Abello, 2000

Romero and Abello [68] propose an automatic hybrid method called Multidimensional Design By Examples (MDBE). To generate multidimensional schema, this method takes as input, on one hand, the needs of the decision-makers expressed as SQL queries, and on the other hand, the relational data source. Source interrogation is provided by SQL queries and requires good knowledge of the relational schema. Therefore, construction of the multidimensional schema involves a computer expert to formulate the SQL queries and interrogate data sources.

4.5 Husemann et al., 2000

The Husemann et al. [69] approach uses the definition of functional dependencies to determine multidimensional concepts. It addresses the transformation into two main stages:

- specification and needs analysis: from the relevant data sources retained by consultation between users and designers, a semi-formal schema of multidimensional concepts and a dictionary, listing attributes characteristics are produced. The semi-formal schema is obtained from functional dependencies between measures defined by decision-makers, and attributes defined in the dictionary. A minimal subset of attributes is linked to each measure. These subsets are used to determine the finer levels of hierarchies.
- conceptual modeling: the semi-formal schema is transformed into a multidimensional conceptual schema. This diagram presents each fact with its measures and related dimensions. This approach uses functional dependencies to determine facts and dimensions. The conceptual schema is then refined into a logical schema that can be relational or multidimensional. Thereby, the physical implementation of the system is done.

4.6 Jensen et al., 2004

Jensen et al. [70] begin their approach by consulting the catalog of supply database. The structure of the database is enriched by the explanation of functional and inclusive dependencies. These dependencies are identified using data mining techniques and are necessary for the identification of dimensions. The data retained are then classified into three categories: measures, keys and descriptive data. An algorithm is proposed to generate a snowflake schema by analyzing the meta-data of the DB. Facts tables are identified by a semi-automatic process, according to the cardinalities of the relationships and the number of measures identified. Inclusive dependencies identified are represented by different related graphs. A graph is considered as a dimension if a dependency exists between facts table and a graph node. This node will be considered as the first hierarchy level of the dimension. Dimension hierarchies are analyzed to verify the aggregation of data across each hierarchy.

4.7 Moody et Kortink, 2004

Moody and Kortink [42], [71], [72] use a generic business management ER model to illustrate their approach. The construction of the data marts is done in four steps:

- the classification of entities into three categories: transactional entities that represent the facts in the final schema; component entities that represent the dimensions; and classification entities that represent the dimension hierarchies;
- the identification of hierarchies through association types between entities;
- the definition of the dimensional schema by the use of merge and aggregation operations;

the evaluation and refinement of the dimensional schema is used to improve the model in order to propose, as a last resort, the desired multidimensional schema. Five types of schema can be produced. These are star, snowflake, constellation, cluster or terrace schema.

4.8 Song et al., 2007

The approach of Song et al. [73] is called Semi-automated lexical method for generating star schema from an entity-relationship diagram (SAMSTAR). The design process follows these steps:

- ER schema are redefined to transform ternary associations into binary associations;
- entities with the number of associations 1-N greater than a threshold value, are considered as facts;
- entities related to facts by 1-N relations are considered as fact dimensions. Wordnet ontology is used to identify dimension hierarchies.

The authors subsequently suggested an improvement entitled Connection Topology Value (CTV). It identifies candidate facts automatically, by analyzing the number of links from each entity.

5 Multidimensional Canonical Partitioning Approach

In [74], Batouré et al. propose a supply-driven approach, called Multidimensional Canonical Partitioning (MCP). It takes into account, without distinction, ER data schema. From universal relations assumption, the schema is derived into a universal relation (UR). This step provides a flat schema, where all features are grouped into a single entity. This entity is then partitioned vertically, according to characteristics or attributes semantics. To achieve it, we use a heuristic greedy type algorithm. Resulting partitions are candidates for being dimensions in the future data schema. To do this, we use an algorithm that matches the attributes present in the partitions and those that must actually be in the dimensions. Obtained dimensions are, if necessary, snowflaked (normalized), using the third normal form algorithm (3NF). Because all the multidimensional elements are obtained, the data schema is generated using model transformations from QVT (Query View Transformation) and a multidimensional and spatio-temporal design pattern.

The approach is recaped into six steps:

1. Verification of provided ER schema. If it is on universal relation form, we go straight to step 2, otherwise we restructure it according to universal relation assumption;
2. Vertical partitioning by fragmentation and distribution of attributes in obtained partitions;
3. Transformation of partitions into dimensions;
4. Normalization of dimensions;
5. Construction of facts table;
6. Generation of multidimensional schema.

To coordinate the complete modeling phases (conceptual, logical and physical) of the decision support system, model-driven engineering (MDE) is used. By successive model transformations, one goes from the multidimensional annotation to obtain implementation and ETL codes of data warehouse, according to a chosen platform.

6 Comparative synthesis of approaches

The approaches studied adopt different techniques to define multidimensional elements from input schema. Before applying the approach, a restructuration of ER schema is performed. It is a matter of eliminating ternary or more, recursive, many-to-many relationships or judging the relevance of certain entities and associations. This is the case of the Boehnlein, Cabibbo and Torlone, Jensen et al. and Song et al. ([58], [65], [70], [73]) approaches. The Batoure et al. [74] approach begins by transforming input data schema into a universal relation.

In Golfarelli et al., Jensen et al., Moody and Kortink, Romero and Abello and Song et al. ([67], [68], [70], [72], [73]) approaches, cardinality of relationships between entities is taken into account to determine facts, dimensions and their hierarchies.
Except cardinality, some approaches use functional dependencies. Its the case of the Boehnlein, Golfarelli et al., Husemann et al., Jensen et al. and Batoure et al. ([58], [67], [69], [70], [74]) approaches. Only Jensen et al. ([70]) approach uses both relationship cardinality and functional dependencies to derive elements for multidimensional schema.

From a node, usually facts table, some approaches determine the multidimensional schema, by constructing a graph. This is the case of the Cabibbo and Torlone, Golfarelli et al. and Jensen et al. approaches ([65], [67], [70]).

The approaches of Cabibbo and Torlone, Jensen et al., Moody and Kortink and Song et al. ([65], [70], [72], [73]) use some operations to define the multidimensional data schema. These operations include filters, fusions, aggregations, generalizations, specifications and pruning.
Approaches use algorithms and/or guidelines in their constituent steps. Some of them are automatic or semi-automatic. Sometimes, a design tool is provided. This is the case of Golfarelli et al., Jensen et al., Batoure et al., Romero and Abello ([67], [68], [70], [74]).

Some of the approaches are illustrated from a concrete example. This is the case of Moody and Kortink approach, in [72] who illustrate their approach, in all proposed articles, by a management system of companies, taking into account sale, localization, production, customers, subsidiaries and periods (temporal aspect).

Among all these approaches, only that of Golfarelli et al., described in 6 steps [67], goes as far as the definition of the physical level from the ROLAP software model; index optimization; and estimation of the queries workload. Step 4 of the approach defines the logic model and steps 5 and 6 are devoted to definition of physical model. The Batoure et al. [74] approach uses a design pattern and the QVT language to generate data schema from multidimensional annotation. It uses Model-Driven Engineering (MDE) for the development and implementation of DSS, in addition to ETL processes.

The approaches of Cabibbo and Torlone, Golfarelli et al. and Boehnlein ([58], [65], [67]) propose an extension of the known ER model, to adapt it to multidimensional concepts. Among these approaches, those of Golfarelli et al., Romero and Abello and Boehnlein ([58], [67], [68]) are hybrid. Table 2 summarizes the characteristics of the approaches. Some comparison criteria are defined in [61], and completed by those of [53], [76], [77].

The following notation is used in table 2: [A] = (Cabibbo et al., 1998) ; [B] = (Golfarelli et al., 1998) ; [C] = (Boehnlein, 1999) ; [D] = (Romero et Abelló, 2000) ; [E] = (Husemann et al., 2000) ; [F] = (Jensen et al., 2004) ; [G] = (Moody et Kortink, 2004) ; [H] = (Song et al., 2007) ; [I] = (Batouré et al., 2017).

Table 2: Comparison of Design Works

Criteria ↓ / **Approach →**	[A]	[B]	[C]	[D]	[E]	[F]	[G]	[H]	[I]
Approach type (Supply-driven: SD; Hybrid: HY)	HY	HY	HY	HY	SD	SD	SD	SD	SD
ER schema Restructuration (No: N; Yes: Y)	Y	N	Y	N	N	Y	N	Y	Y
Facts (Guidelines: GL; Heuristic: HE)	GL	HE	HE	HE	GL	HE	GL	HE	GL
Dimensions (Guidelines: GL; 1-N relationship: 1-N; Functional dependencies: FD; Algorithm: AL)	GL	FD	FD	FD	GL	FD	1-N	1-N	AL
Hierarchies (Heuristic: HE; 1-N relationship: 1-N; Functional dependencies: FD)	HE	FD	1-N	1-N	FD	FD	1-N	HE	FD
Logical formalism (Relationnl: R; Multidimensionnal: M)	R,M	R	R	R	R	R	R	R	R
Physical representation (No: N; Yes: Y)	N	Y	N	N	N	N	N	N	Y
Implementation (Mono-layer : MO ; Multi-layer : MU)	-	MO	-	-	-	-	-	-	MO
Special notation of data shema (No: N; Yes: Y)	Y	Y	N	Y	N	N	N	Y	N
Automation degree (Manuel: M; Semi-automatic: SA; Automatic: A)	M	SA	SA	SA	M	SA	M	A	SA
Proposed tool (No: N; Yes: Y)	N	Y	Y	N	N	Y	N	Y	N

All the above compared approaches have a conceptual design level abstraction. In other words, they all produce a conceptual data schema when been applied. This schema is produced from an ER data schema. Also, all these approaches range from conceptual representation to logical representation. Only the approaches of Golfarelli et al. ([B]) and Batoure et al. ([I]) go up to the physical representation, through a mono-layer implementation. This implementation uses a relational DBMS while the multi-layered one uses a multidimensional DBMS. The physical representation of the MCP approach is made through use of model-driven engineering.

7 Conclusion

In this paper, we made a state of the art of decision support design approaches using ER data schema. This comparative study allowed us to understand how they determine the multidimensional elements, how they realize different modeling phases and their automation level. Through this work, we understand that the current trend is to propose mixed approaches. It is the goal assigned to the MCP approach.

References

[1] B. Claybrook, *OLTP (online transaction processing)*. Wiley Publishing, 2014.

[2] R. Sharda, D. Delen, E. Turban, *Business Intelligence and Analytics Systems for Decision Support*, Pearson Education Limited 2014, 4th Ed.

[3] D. Martin, *La Performance transactionnelle : Problèmes, Techniques, État de L'art*. Masson, 1997.

[4] A. Silberschatz, H.F. Korth, and S. Sudarshan, *Database System Concepts* – 6th. ed., volume 4. McGraw-Hill Education, 6th edition, 2011.

[5] A. Meier, *Introduction pratique aux bases de données relationnelles*. Springer Science & Business Media, 2006.

[6] D. Agrawal, S. Das, and A. El Abbadi, "Big data and cloud computing: current state and future opportunities". In *Proceedings of the 14th International Conference on Extending Database Technology*, pages 530–533, 2011.

[7] A. Cuzzocrea, L. Bellatreche, and Il Y. Song, "Data Warehousing and OLAP over Big Data: Current challenges and future research directions". In *DOLAP'13*, pages 67–70, San Francisco, CA, USA, 2013.

[8] A. Cuzzocrea, I.Y. Song, and K.C. Davis, "Analytics over large-scale multidimensional data: The big data revolution!" *14th ACM International Workshop on Data Warehousing and Online Analytical Processing, DOLAP'11*, pages 101–103, 2011.

[9] R. Kimball. *The Evolving Role of the Enterprise Data Warehouse in the Era of Big Data Analytics*. Ralph Kimball Group, 2011.

[10] R. Valencia-García, M.A. Paredes-Valverde, M.P. Salas-Zárate, G. Alor-Hernández, *Exploring Intelligent Decision Support Systems. Current State and New Trends*, Springer 2018.

[11] C. Carlsson and E. Turban, "DSS : Directions for the next decade". *Decision Support Systems*, 33(2) :105–110, 2002.

[12] W.H. Inmon. *Building the Data Warehouse*. Wiley Publishing, fourth edition, 2002.

[13] W. Oueslati, J. Akaichi, "A survey on data warehouse evolution". *Int J Database Manag Syst* (IJDMS) 2(4):11–24, 2010.

[14] J.-M. Franco and S. de Lignerolles, *Piloter l'entreprise grâce au data warehouse*. Eyrolles, 2000.

[15] S. Chaudhary, D.P. Mural, V.K. "Srivastav VK, A critical review of data warehouse". *Glob J Bus Manag Inf Technol* 1(2):95–103, 2011.

[16] R. Kimball and J. Caserta, *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning Conforming, and Delivering Data*. Wiley Publishing, 2004.

[17] I. Comyn-Wattiau and J. Akoka, *Entrepôts de données et business intelligence*. Hermès Lavoisier, 2014.

[18] J. Akoka and I. Comyn-Wattiau, *Business Intelligence. Un état de l'art*. Ingénierie des systèmes d'information, 19(5) :9–43, 2014.

[19] C. Adamson, Mastering *Data Warehouse Aggregates: Solutions for Star Schema Performance*. Wiley, 1st edition, 2012.

[20] R. Alshboul, "Data Warehouse Explorative Study". *Applied Mathematical Sciences*, 6(61) :3015–3024, 2012.

[21] M. Humphries, M. Hawkins, and M. Dy, *Data Warehousing : Architecture and Implementation*. Prentice Hall, 1999

[22] W.H. Inmon. *Building the Data Warehouse*. Wiley Publishing, 4th Ed., 2002.

[23] Kolyang. *Introduction aux bases de données*. Kolyang et Ka'arang : Editions et Média, 2005.

[24] M. Arora, A. Gosain, "Schema evolution for data warehouse: a survey". *Int J Comput Appl* 22(6):6–14, 2011.

[25] E.F. Codd, S.B. Cood, C.T. Salley, "Providing OLAP (On-Line Analytical Processing) to user-analysts: an IT mandate", 1993 (white paper).

[26] B. Kuijpers, L. Gomez, A. Vaisman, "Performing OLAP over graph data: query language, implementation, and a case study". In: *BIRTE '17 proceedings of the international workshop on realtimebusiness intelligence and analytics*, no 6. ACM, New York, 2017.

[27] N. Arfaoui and J. Akaichi, "Data Warehouse: Conceptual And Logical Schema -Survey-". *International Journal of Enterprise Computing and Business Systems* (IJECBS), 2(1) :31, 2012.

[28] F. Dehne, D. Robillard, A. Rau-Chaplin, N. Burke, "VOLAP: a scalable distributed system for real-time OLAP with high velocity data". In: *2016 IEEE international conference on cluster computing (CLUSTER)*. IEEE, pp 354–363, 2016.

[29] K. Zeng, S. Agarwal, I. Stoica, "IOLAP: managing uncertainty for efficient incremental OLAP". In: *Proceedings of the 2016 international conference on management of data*. ACM, pp 1347–1361, 2016.

[30] N. Geary, B. Jarvis, C. Mew, H. Gore, "Method and apparatus for automatically creating a data warehouse and OLAP cube". *Precision point Software Limited*. US Patent 9,684,703, 2017.

[31] Z. Wang, Y. Chu, K.L. Tan, D. Agrawal, A.E. Abbadi, "HaCube: extendingMapReduce for efficient OLAP cube materialization and view maintenance". In: *International conference on database systems for advanced applications*. Springer, Cham, pp 113–129, 2016.

[32] F. Ravat, O. Teste, and G. Zurfluh, "Algèbre OLAP et langage graphique". In *CoRR*, volume abs/1005.0, pages 1039–1054, 2010.

[33] B. Espinasse, « Entrepôts de données : Introduction au langage MDX (MultiDimensional eXtensions) pour l'OLAP ». Technical report, Ecole Polytechnique Universitaire de Marseille, Marseille, 2015.

[34] S. Boucher, B. Verhaegen, and E. Zimanyi. "Modélisation et interrogation de données XML multidimensionnelles ». In *Egc* 2010.

[35] Dishek Mankad, Preyash Dholakia, "The Study on Data Warehouse Design and Usage", *International Journal of Scientific and Research Publications*, Volume 3, Issue 3, March 2013.

[36] N. Prakash, D. Prakash, *Data Warehouse Requirements Engineering. A Decision Based Approach*, Springer Nature Singapore Pte Ltd., 2018.

[37] I. Gam and C. Salinesi. "Analyse des Exigences pour la Conception d'Entrepôts de Données". In *INFORSID*, pages 1023–1038, 2006.

[38] M. Piattini J.-N. Mazón, J. Trujillo, M. Serrano, "Designing data warehouses: from business requirement analysis to multidimensional modeling". *Rebnita*, 5 :44 – 53, 2005.

[39] N. Prat and J. Akoka. "From UML to ROLAP multidimensional databases using a pivot model". In *18èmes Journées Bases de Données Avancées*, BDA '02, 2002.

[40] A. Tsois, N. Karayannidis, and T.K. Sellis, "MAC : Conceptual data modeling for OLAP". In *Proceedings of the 3rd Intl. Workshop on Design and Management*, volume 39, page 5, 2001.

[41] Y.-T. Chen and P.-Y. Hsu. "An Efficient and Grain Preservation Mapping Algorithm: From ER Diagram to Multidimensional Model". *Issads* 2005, Lncs 3563, pages 331–346, 2005.

[42] D.L. Moody and M.A.R. Kortink, "From ER Models to Dimensional Models Part II : Advanced Design Issues". *Journal On The Theory Of Ordered Sets And Its Applications*, pages 1–12, 2004.

[43] G. Saxena, "Data Warehouse Designing: Dimensional Modelling and E-R Modelling". *International Journal of Engineering Inventions*, 3(9):28–34, 2014.

[44] G. Chandwani, Designing of Star Schemas from XML Document. International Journal of Multidisciplinary and Current Research, 2 :10, 2014.

[45] M. Golfarelli, S. Rizzi, and B. Vrdoljak, "Data warehouse design from XML sources". Proceedings of *the 4th ACM international workshop on Data warehousing and OLAP* - DOLAP '01, pages 40–47, 2001.

[46] E. Annoni, "Eléments méthodologiques pour le développement des systèmes décisionnels dans un contexte de réutilisation". PhD thesis, Université Paul Sabatier – Toulouse III, 2007.

[47] K. Shahzad and A. Sohail, "A Systematic Approach for Transformation of ER Schema to Dimensional Schema". In *Proceedings of the 7th International Conference on Frontiers of Information Technology*, pages 1–6, New York, NY, USA, 2009. ACM.

[48] E. Annoni, F. Ravat, O. Teste, and G. Zurfluh, "Automating the choice of decision support system architecture". *Lecture Notes in Computer Science*, 4080 :244–253, 2006.

[49] L. Carneiro and A. Brayner, "X-META: A Methodology for Data Warehouse Design with Metadata Management. 4th International Workshop Design and Management of Data Warehouses" (*DMDW'02*), 58 :13–22, 2002.

[50] S. Khouri, L. Bellatreche, S. Jean, and Y. Ait-Ameur, "Requirements driven data warehouse design : We can go further". In *Lecture Notes in Computer Science*, volume 8803, pages 588–603, 2014.

[51] C. Phipps and K.C. Davis, "Automating data warehouse conceptual schema design and evaluation". *Dmdw*, pages 1–10, 2002.

[52] A. Cravero and S. Sepulveda, "Methodologies, techniques and tools for OLAP design : A Systematic Mapping Study". *IEEE Latin America Transactions*, 14(2) :943–950, 2016.

[53] S. Saroop and M. Kumar, "Comparison of Data Warehouse Design Approaches from User Requirement to Conceptual Model: A Survey". In *Communication Systems and Network Technologies, International Conference* on, volume 0, pages 308–312, 2011.

[54] W. Tebourski, W.B.A. Karâa, Ghezala, and H. Ben, "Semi-automatic Data Warehouse Design methodologies: a survey". *IJCSI International Journal of Computer Science Issues*, 10(5) :48–54, 2013.

[55] A. Vaisman and E. Zimányi. "A Method for Data Warehouse Design". In *Data Warehouse Systems*, chapter 1. Springer-Verlag Berlin Heidelberg, Berlin, 2014.

[56] P. Chandra, M. K. Gupta, "Comprehensive survey on data warehousing research", *Int. j. inf. tecnol*., Springer 2017, 8p.

[57] M. Golfarelli, S. Rizzi, "From star schemas to big data: 20+ years of data warehouse research". In: *A comprehensive guide through the Italian database research over the last 25 years*. Springer International Publishing, pp 93–107, 2018.

[58] M. Boehnlein and A. Ulbrich-vom Ende, "Deriving initial data warehouse structures from the conceptual data models of the underlying operational information systems". In *Proceedings of the 2nd ACM international workshop on Data warehousing and OLAP* - DOLAP '99, number 0, pages 15–21, 1999.

[59] P. Giorgini, S. Rizzi, and M. Garzetti, "Goal-Oriented Requirement Analysis for Data Warehouse Design". In *Proc. of the 8th ACM International Workshop on Data Warehousing and OLAP*, pages 47–56, Bremen, 2005.

[60] S. Khouri, "Modélisation conceptuelle à base ontologique d'un entrepôt de données". Mémoire de master, Ecole nationale Supérieure d'Informatique, 2009.

[61] S. Khouri, "Cycle de vie sémantique de conception de systèmes de stockage et manipulation de données". PhD thesis, École Nationale Supérieure de Mécanique et d'Aérotechique - Poitiers, 2013.

[62] S. Khouri, I. Boukhari, L. Bellatreche, E. Sardet, S. Jean, and M. Baron, "Ontology based structured web data warehouses for sustainable interoperability : Requirement modeling, design methodology and tool". *Computers in Industry*, 63(8) :799–812, 2012.

[63] J. Pardillo and J.-N. Mazón, "Using ontologies for the design of data warehouses". *International Journal of Database Management Systems* (IJDMS), 3(2), 2011.

[64] D. Skoutas and A. Simitsis, "Ontology-Based Conceptual Design of ETL Processes for Both Structured and Semi-Structured Data". Technical Report 4, Dept. of Electrical and Computer Engineering, 2007.

[65] L. Cabibbo and R. Torlone, "A logical approach to multidimensional databases". In *Advances in Database Technology*—EDBT'98, pages 183–187. Springer Berlin Heidelberg, 1998.

[66] M. Golfarelli, D. Maio, and S. Rizzi, "Conceptual design of data warehouses from E/R schemes". In *Proceedings of the Thirty-First Hawaii International Conference on System Sciences*, volume 7, pages 334–343, 1998.

[67] M. Golfarelli, D. Maio, and S. Rizzi, "The Dimensional Fact Model: A Conceptual Model for Data Warehouses". *International Journal of Cooperative Information Systems* (IJCIS), 7(2) :215–247, 1998.

[68] O. Romero and A. Abelló, "Automatic validation of requirements to support multidimensional design". *Data and Knowledge Engineering*, 69(9) :917–942, 2010.

[69] B. Hüsemann, J. Lechtenbörger, and G. Vossen, "Conceptual data warehouse design". In *Design and Management of Data Warehouses* (DMDW 2000), pages 3–9, 2000.

[70] M.R. Jensen, T. Holmgren, and T.B. Pedersen, "Discovering Multidimensional Structure in Relational Data". *Data Warehousing and Knowledge Discovery*, 3181 :138–148, 2004.

[71] D. Moody and M.A.R. Kortink, "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), 2000 :5–16, 2000.

[72] D. L. Moody and M.A.R. Kortink, "From ER Models to Dimensional Models: Bridging the Gap between OLTP and OLAP Design, Part I". *Business Intelligence Journal*, pages 7–24, 2003.

[73] Il Y. Song, R. Khare, and B. Dai, "SAMSTAR: a semi-automated lexical method for generating star schemas from an entity-relationship diagram". In *Proceedings of the ACM tenth international workshop on Data warehousing and OLAP*, pages 9–16, Lisbon, Portugal, 2007. ACM New York.

[74] A.B. Batoure, Kolyang, and M. Tchotsoua, "Using Multidimensional Canonical Partitioning (MCP) as a Supply-Driven Approach for Data Warehouses Design". *International Journal of Computer* (IJC), 25(1):52–62, 2017.

[75] P. Pahwa, S. Taneja, and S. Jain, "Design of a Multidimensional Model Using Object Oriented Features in UML". *International Research Journal*, 1(1) :14, 2011.

[76] R. Jindal and S. Taneja, "Comparative Study of Data Warehouse Design Approaches: a Survey". *International Journal of Database Management Systems*, 4(1):33–45, 2012.

[77] C. Khnaisser, L. Lavoie, H. Diab, and J.-F. Éthier. "RR0012 DWDMR-Comparison table". Technical report, 2015.