International Journal of Management (IJM)
Volume 11, Issue 7, July 2020, pp. 463-471, Article ID: IJM_11_07_044
Available online at http://iaeme.com/Home/issue/IJM?Volume=11&Issue=7
ISSN Print: 0976-6502 and ISSN Online: 0976-6510
DOI: 10.34218/IJM.11.7.2020.044

A COMPREHENSIVE SURVEY TO DESIGN EFFICIENT DATA WAREHOUSE FOR BETTERMENT OF DECISION SUPPORT SYSTEMS FOR MANAGEMENT AND BUSINESS CORPORATES

Abhishek Gupta
Senior Analyst, Software Development, Accenture Services Pvt. Ltd, Chennai, India
Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies, Chennai, India

Arun Sahayadhas
Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies, Chennai, India

ABSTRACT

Data warehousing (DWH) is an extensively used and crucial practice in business corporations that support the data analytic and decision-making process. But still there are some corporations based on domains like insurance and life-science which has a richness of data from their business process. But still these corporations lack usage of approaches/process that can handle data effectively, which makes them difficult to take right decision at right time. So, we have done research and study in this area and it was observed that DWH is the approach that can enable corporates to receive and integrate information from heterogeneous data sources and to query very large databases efficiently. After extensive research and study, we have identified 52 relevant papers on DWH and related areas. The purpose of the study is to make clear understanding on building of efficient DWH. We have explored and identified the priority areas of research and the research gaps in this area. The research gaps that have been identified can become a source of opportunities to start new lines of research, which can help to make more efficient DWH.

Key words: Decision Support System, Data Warehousing, Dimensional Modeling, Slowly Changing Dimensions

Cite this Article: Abhishek Gupta and Arun Sahayadhas, A Comprehensive Survey to Design Efficient Data Warehouse for Betterment of Decision Support Systems for Management and Business Corporates, *International Journal of Management,* 11(7), 2020, pp. 463-471. http://iaeme.com/Home/issue/IJM?Volume=11&Issue=7

1. INTRODUCTION

Decision Support Systems (DSS) uses the summary information, exceptions, patterns, and trends using the analytical models that are intended to help corporations/business units in decision-making process by accessing current as well historical data gathered from various heterogeneous sources involved in business processes. DSS improves correctness, speed and efficiency of decision-making activities. Mostly used DSS are database oriented where database contains highly structured data.

These databases can be categorized in two types based on nature of data it handles: - 1. On-Line Transactional processing (OLTP): - It is used for handling current data. It uses traditional DBMS. In the OLTP database most of the tables are in normalized form.

2. On-Line Analytical Processing (OLAP): - It is used for analysis of data, which includes current as well as historical data. It uses data warehouse (DWH). In the OLAP database most of the tables are in non-normalized form.

In the decision-making process, there is the requirement of current as well as historical data. So, in this paper we have mainly concentrated on the DWH.

Data Warehouse: - Data warehouse is centralized repository which contains subject-oriented, integrated, non-volatile and time-variant collection of data, for online business analytical processing (OLAP).

There are two approaches to design the data warehouse: -

1. Top-Down approach: - It is data-driven approach as in this the first process is to gather and integrate data as centralized repository and then as per business requirements these data are classified based on subject areas like sales etc., which is called as data-marts. The advantageous part of this approach is that it supports a single integrated data source.

2. Bottom-Up approach: - It is business-driven approach as in this the first process is to categorize the data subject-wise, also referred as data-marts and then the data is integrated as centralized repository. The advantageous part of this approach is that it has quick return of investment (ROI) as developing subject-wise data takes less time than and effort than developing an enterprise-wide data warehouse.

There are mainly seven steps to build a Data Warehouse: -

1. Requirement gathering.

2. Setting the physical environments.

3. Defining schema: - Schema is logical description of the whole database. DWH uses mainly three types of schema: Star, Snowflake, and Fact Constellation schema.

In star schema, fact table (it is a table which contains business measures and foreign keys to join dimension tables) is surrounded by dimension tables (it is a table which contains business prospective, which helps to measure a fact), but there is no connectivity between dimension to dimension tables. So, querying to the star schema is easy and takes less time as it contains less joins. But this design also faces a consequence of more data redundancy is as tables are not hierarchically divided.

Snow-flake schema is extension to star schema, with a modification that dimension tables are hierarchically divided. Means it has dimension to dimension connectivity. Because of this most of the tables are in normalized form and so has less data redundancy issue. But disadvantage of this design is that it contains a greater number of tables connected hierarchically and so query requires more joins, which make query more complex.

Fact-Constellation/ Galaxy schema has more than one fact tables, that are connected based on a common dimension, having same level of grain (data summarized at the same level).

4. Choosing Extract, Transform, Load (ETL) Solution: - From a long time ETL is being used as traditional approach for data warehousing and analytics. There is a new approach also has been introduced that is ELT (extract, load, transform) approach.

The technological difference between these two approaches are their stages of Transformation. The ETL first Extract data from various heterogeneous sources then transform the data and after transformation loads the data to DWH. Whereas the ELT first Extract data from various heterogeneous sources then loads the data to DWH and then transform the data after which it passes this data to business intelligence (BI) tools for analysis purpose.

ELT is more efficient than ETL in many ways. With ELT, users can work with new transformation rules, test and enhance queries as per the requirements, as it has to be applied to raw data that has been loaded/ integrated in one place, at the same time, without the time and complexity that is required in case of ETL. That strengthen usage of ELT architecture due to its Ad Hoc, Agile, Flexible nature.

5. Creating the Front End, to gather the data in more meaningful way for analysis purpose.

6. Optimization of Queries.

7. Establishment of a Rollout.

When organizing data warehouse, an issue can occur that information of the dimension can vary over time. To manage this issue Slowly Changing Dimensions (SCD) are used. These can be further categorized as per its nature/way to handle the data that varies with time, are as below:

SCD Type 0 – Does not changes over time.

SCD Type 1 - Over-write the data.

SCD Type 2 – Maintain all history with current data, this data can be recognized based on version number/ effective date/ flags.

SCD Type 3 – Maintain limited history (only just previous data) with current data, in this SCD two columns are maintained one has current data and other has previous data.

SCD Type 4 – Maintain all historical data in a separate mini table and current data in dimension table.

SCD Type 5 – It is the hybrid dimension, combination of SCD type 1 and SCD type 4.

SCD Type 6 – It is the hybrid dimension, combination of SCD type 1, SCD type 2 and SCD type 3.

But out of these SCD type 2 and 3 are most widely used. But both have its own drawback, as SCD type 2 maintains all historical data with time it grows, and it becomes difficult to handle this huge amount of data. Whereas SCD type 3 does not maintain all historical data. So, to trace back of whole (history and current) data is not possible. So, there are so many opportunities to work on this area. One of the substitutes of this SCDs has been introduced in a research paper we have studied is usage of temporal dimension.

Similarly, while working with data warehouse there is requirement to handle information security and data integrity check (Data integrity check is a feature of the hash functions, where hash functions are used to generate the checksums on data files. By this it provides assurance to the user about correctness of the data). The information security (like password storage feature) and data integrity check features can be handled with the use of hash functions. Hash function is a mathematical function that encrypt the data at source level and decrypt the data at destination level. Input of this hash function (say h) is of varying length (say x) but output is always of this hash function is always a fixed length (say h(x)). There are so many popular hash functions which are: 1. Message Digest (MD), in this group mostly

used function is MD5 which generates 128-bit output. 2. Secure Hash Function (SHA), there are so many versions are available in this group the basic version was SHA0 with 160-bit output. 3. RACE Integrity Primitives Evaluation Message Digest (RIPEMD), original RIPEMD (128 bit) was based on the design principle used in MD4 (whose upgraded version is MD5), improved version of RIPEMD (128 bit) was RIPEMD-160, which is most widely used version in this group. 4. Whirlpool, with 512-bit output.

Out of these hash functions SHA and other functions are more robust than MD function as MD5 has faced intruder attack and so its security feature cannot be trusted. But if we check for the second feature offered by hash functions is data integrity check, then MD5 is faster than any other hash functions like SHA, as it generates 128-bit as output, which is lesser than any other group introduced above. So MD5 can still be a better option for the data integrity check, at internal levels where information security is not a major issue.

2. LITERATURE REVIEW

Decision-making is the most essential part of any organization to achieve their goals. So, it's imperative to know the past, present and future of Decision Support System, which has been articulated in [50]. While making any decision then it does not depend only on the current data but also depends on the historical data too, through which datamining (analysis of data, pattern findings etc.) can be done and the concrete decision can be taken. For current data OLTP (Online Transaction Processes) are used. Where the data is resided in the form of entity/table and columns. For designing of each table first proper requirements are gathered and then each entity/table goes through a rigorous process of design can be defined in steps as conceptual, logical and finally physical modelling, these things (requirement gathering and modeling techniques) have been defined in the [28,38]. After design of the tables, the relationship between these tables must be established which is called as entity-relationship-modeling, it has been defined in [37]. Similarly, for historical data OLAP (Online Analytical Processing) is used, which is used to analyze current as well as historical data. The data-model which contains these kinds of tables which contains historical data is called as Data Warehouse (DWH). In the DWH tables are defined as Facts or Dimensions. The facts and dimensions can be further classified based on their behavior. To handle historical as well as current data slowly changing dimensions (SCD) are used, because of which the current and historical data can be tracked. By the nature of data handling SCDs can be further classified like SCD type 1/2/3 etc., have been elaborated in [12,21,46]. But as SCDs must handle current as well as historical data, sometimes its performance get degraded, so for this a solution has been presented by [32], which is to use the temporal dimensions. Similarly, cryptography is very crucial requirement while working with sensitive data, which is handled with the help of hash algorithms, it has been elaborated in [30,52]. The relationship between Fact and Dimension tables are defined as schema, there are three kinds of schema exists which are star schema, snow-flake schema and combination of both is called as fact-constellation/ galaxy schema. [22] has discussed the Integration of Star and Snowflake Schemas in Data Warehouses. [31,48] has discussed about the importance of snowflake schema. The modeling which is used in DWH is called as Dimensional modeling, it is also defined as multi-dimensional data (MDD) modeling. The Multidimensional Modeling Methodologies has been discussed in [35]. There are so many articles and papers have been published till now for the designing of data warehouse in various domains like health care, dairy farming etc. some of them are [1,6,7,8,14,15,17,18,20,25,39,51]. Similarly, [2] has introduced a data model with a hierarchical structure for clinical research. After defining the relationship, the data needs to be extracted from various sources, and data-cleansing and transformations needs to be performed before loading it. This whole process of Extraction-

Transformation and Load is called as ETL. In some places use of ETL is not so much fruitful/ advantageous rather than that it's better to use the ELT, in which the data is first extracted from sources then loaded in target and then data-cleansing or transformation takes place, [34] has showcased the comparative analysis on ETL vs ELT techniques. [36] has summarized the present the research works performed in the field of ETL technology. Sometimes it's highly required to extract data at the real-time, which is very difficult to implement due to various reasons likewise because of technical/architectural issues etc., some of these points have been discussed in [9,33]. To perform ETL or ELT operations there are so many tools are available in the market and out of which Informatica is the ETL tool preferred by most of the IT companies, through which these operations can be done easily. But still if design is poor or data is huge then it takes time to Extract-Transform and Load the data. So, to tune the performance of Informatica there are so many papers have been published are [3,11]. After designing the DWH it's very crucial to test it, [24,29] have introduced multiple DWH testing processes. There are multiple ways to access DWH data for analysis, the very first is to use the Structured-Query-Language (SQL). But during accessing the data using SQL sometimes a small mistake can show the wrong data and it would become cause of wrong analysis, and consequently become cause of wrong decision making, that can impact whole organization. [4] has showcased an example that how the data can be taken more accurately. At present there are so many reporting tools are also available in the market for better visualization, crystal reports can be generated based on some key performance indicators (KPIs), some of the papers have discussed about the crystal reports, are [10,16].

3. ANALYSIS AND DISCUSSION

After a rigorous study we become to know that there is a vast scope for the Decision-Support-Systems, which enables management to take right decision at the right time to achieve their organization goals. Data Warehouse plays a key role for maintaining the historical as well as current data and so many papers have been published for the same. The same logics for designing of DWH can be applied to any domain like life-sciences etc. and a DWH can be designed. To design the ETL/ELT model multiple ETL tools like Informatica/data-stage etc. ETL tools are available. We are focusing on the Informatica ETL tool due to its vast usage. So, while designing the DWH, the designs build-up by Informatica can be further optimized (performance can be further tuned), while will make DWH more efficient. Similarly, we also have observed that not only newly build designs, but also existing designs can be tuned performance-wise. After studying review papers, it has been observed that there are two architectures are available first is ETL and second is ELT, and both of the methods have their own advantages and disadvantages, if we will do further research then we can find a better solution, where capability of ETL and ELT can be used combinedly to have better outcome. Further-more we have studied about the Slowly Changing Dimensions, that handles the historical as well as current data, but its performance degraded when huge amount of the data required to handle. To compensate this problem temporal dimension has been proposed. But if we further analyzed the Slowly Changing Dimensions then SCD Type 2/3 are the dimensions which holds the current data. Where SCD type 2 holds all historical data, where the data can be traced for all history and so it becomes huge day after day, Whereas SCD type 3 holds only current and last one historical data, in SCD type 3 it's not possible to trace data before one history so it holds less amount of data than SCD type 2. So, ultimately, we can use SCD type 2/3 combinedly, and can make two data storage houses like DWH1 and DWH2 in serial manner which can be like (Source -> DWH1 -> DWH2). Where DWH1 will have limited history (it will be based on SCD type 3) and DWH2 will have all history (it will be based on SCD type 2). So as per the requirements (how much history is required) queries can be fired, which will upgrade the performance. Additionally, we have studied about the hash

algorithms which are used for the cryptography. After analysis we found that Passwords and all security features shouldn't be implemented using MD5 function, as this cryptography algorithm can be easily decoded by attackers/intruders, but MD5 algorithm also has a benefit that it's very fast comparative to other algorithms so rather than using it at external level, if it is used at internal level than we can get the advantage of this algorithm and make our DWH operations more efficient. Further-more [4] has explained queries to take accurate data, but its significant to one part of area like it efficiently handles the blank (‘’) values, but the same query can't be used while working with NULL values.

4. FUTURE SCOPE

Before choosing any research work, it's essential to know about its usability in the future, [50] has discussed about the future of the Decision-Support-System. [1,6,7,8,14,15,17,18,20,25,39,51] have introduced DWH design for various domains, but none of them have discussed about performance tuning of DWH through which the DWH can become more efficient. The design can be build-up using any ETL tools available in the market. But as Informatica is used by most of the organizations, we have also discussed about Informatica ETL tool to design the DWH, and its design can be more performance tuned by using the tips available in [11], the same points can be applied on other ETL tools too. Similarly, sometimes it is required to update the DWH on real-time basis, which has so many challenges, that have been introduced in [9,33], on which the research can be performed.

5. CONCLUSION

In this paper we have showcased the importance of Decision-Support-Systems. After which we have discussed about various terminologies used in data-science, likewise OLTP, OLAP and DWH that are used handle current as well as historical data. After which we have discussed about Slowly Changing Dimension types, Schemas, ETL and ELT architectures, MD5 function. Then we have summarized the papers and articles have published/presented in the same area of research. Furthermore, we have shown that even so much work has been done in this area, but still there are so many open future aspects of research in the same area.

REFERENCES

[1] Ozaydin, Bunyamin & Zengul, Ferhat & Oner, Nurettin & Feldman, Sue. (2020). Healthcare Research and Analytics Data Infrastructure Solution: A Data Warehouse for Health Services Research (Preprint). 10.2196/preprints.18579.

[2] Danese, Mark & Halperin, Marc & Duryea, Jennifer & Duryea, Ryan. (2019). The Generalized Data Model for clinical research. BMC Medical Informatics and Decision Making. 19. 10.1186/s12911-019-0837-5.

[3] Gupta, Abhishek. (2019). A Complete Reference for Informatica Power Center ETL Tool. International Journal of Trend in Scientific Research and Development. Volume-3. 1063-1070. 10.31142/ijtsrd19045.

[4] Pasimeni, Francesco. (2019). SQL query to increase data accuracy and completeness in PATSTAT. World Patent Information. 57. 1-7. 10.1016/j.wpi.2019.02.001.

[5] Baumer, Ben. (2017). A Grammar for Reproducible and Painless Extract-Transform-Load Operations on Medium Data. Journal of Computational and Graphical Statistics. 10.1080/10618600.2018.1512867.

[6] Schütz, Christoph & Schausberger, Simon & Schrefl, Michael. (2018). Building an active semantic data warehouse for precision dairy farming. Journal of Organizational Computing and Electronic Commerce. 28. 122-141. 10.1080/10919392.2018.1444344.

[7] Cheowsuwan, T. & Rojanavasu, Pornthep & Srisungsittisunti, B. & Yeewiyom, S. (2017). Development of data warehouses and decision support systems for executives of educational facilities in northern Thailand to increase educational facility management capacity. International Journal of Geoinformatics. 13. 35-43.

[8] Arifin, S. M. Niaz & Madey, Gregory & Vyushkov, Alexander & Raybaud, Benoit & Burkot, Thomas & Collins, Frank. (2017). An online analytical processing multi-dimensional data warehouse for malaria data. Database: The Journal of Biological Databases and Curation. 2017. 10.1093/database/bax073.

[9] Sabtu, A. & Azmi, N.F.M. & Sjarif, N.N.A. & Ismail, S.A. & Mohd Yusop, Othman & Sarkan, Haslina & Chuprat, Suriayati. (2017). The challenges of extract transform and load (ETL) for data integration in near real-time environment. Journal of Theoretical and Applied Information Technology. 95. 6314-6322.

[10] Crystal Reports: Formatting Multidimensional Reporting Against OLAP Data. http://www.informit.com/articles/article.aspx?p=1249227

[11] Syed. (2016). Informatica Performance Optimization Techniques. Clearpeaks Bi Lab.

[12] Bhide et al. (2016). Slowly Changing Dmension Attributes in Extract, Transform, Load Processes. United States Patent, (10) Patent No.: US 9,311,368 B2, Date of Patent: Apr. 12, 2016

[13] Schütz, Christoph & Neumayr, Bernd & Schrefl, Michael & Neuböck, Thomas. (2016). Reference Modeling for Data Analysis: The BIRD Approach. International Journal of Cooperative Information Systems. 25. 10.1142/S0218843016500064.

[14] Aljawarneh, Isam. (2015). Design of a data warehouse model for decision support at higher education: A case study. Information Development. 32. 10.1177/0266666915621105.

[15] Gao, Liang & Chen, Yun. (2015). Application Research of University Decision Support System Based on Data Warehouse. 10.1007/978-3-662-45402-2_89.

[16] Miranda, Eka & Suryani, Eli & Rudy, Rudy. (2014). Implementation of datawarehouse, datamining and dashboard for higher education. Journal of Theoretical and Applied Information Technology. 30.

[17] Vaisman, Alejandro & Zimanyi, Esteban. (2014). Data Warehouse Systems: Design and Implementation. 10.1007/978-3-642-54655-6.

[18] Azwa, Abdul & Aziz, Azwa & Jusoh, Julaily & Hassan, Hasni & Rizhan, Wan & wan idris, wan mohd rizhan & Md Zulkifli, Addy & Zulkifli, M & My, shahrul anuwar & Yusof, Mohamed. (2014). A Framework For Educational Data Warehouse (EDW) Architecture Using Business Intelligence (BI) Technologies. Journal of Theoretical and Applied Information Technology. 1. 50.

[19] Rahman, Nayem & Marz, Jessica & Akhter, Shameem. (2012). An ETL Metadata Model for Data Warehousing. Journal of Computing and Information Technology. 20. 10.2498/cit.1002046.

[20] Kochar, Barjesh & Chhillar, Rajender. (2012). An Effective Data Warehousing System for RFID Using Novel Data Cleaning, Data Transformation and Loading Techniques. International Arab Journal of Information Technology. 9.

[21] Braden et al. (2012). Systems and Methods for Storing and Querying Slowly Changing Dimiensions. United States Patent, Patent No.: US 8,260,822 B1, Date of Patent: Sep. 4, 2012

[22] Garani, Georgia & Helmer, Sven. (2012). Integrating Star and Snowflake Schemas in Data Warehouses. International Journal of Data Warehousing and Mining. 8. 22-40. 10.4018/jdwm.2012100102.

[23] Bateni, Mohammad Hossein & Golab, Lukasz & Hajiaghayi, Mohammad Taghi & Karloff, Howard. (2011). Scheduling to Minimize Staleness and Stretch in Real-Time Data Warehouses. Theory Comput. Sysems. 49. 757-780. 10.1007/s00224-011-9347-2.

[24] Golfarelli, Matteo & Rizzi, Stefano. (2011). Data warehouse testing: A prototype-based methodology. Information & Software Technology. 53. 1183-1198. 10.1016/j.infsof.2011.04.002.

[25] El-Sappagh, Shaker & Hendawi, Abdeltawab & El-Bastawissy, Ali. (2011). A proposed model for data warehouse ETL processes. Journal of King Saud University - Computer and Information Sciences. 23. 91–104. 10.1016/j.jksuci.2011.05.005.

[26] Liu, Liang & Andris, Clio & Ratti, Carlo. (2010). Uncovering cabdrivers' behavior patterns from their digital traces. Computers, Environment and Urban Systems. 34. 541-548. 10.1016/j.compenvurbsys.2010.07.004.

[27] Rönnbäck, Lars & Regardt, Olle & Bergholtz, M. & Johannesson, Paul & Wohed, Petia. (2010). Anchor modeling — Agile information modeling in evolving data environments. Data & Knowledge Engineering. 69. 1229-1253. 10.1016/j.datak.2010.10.002.

[28] Golfarelli, Matteo. (2009). From User Requirements to Conceptual Design in Data Warehouse Design–a Survey. Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction. 10.4018/978-1-60566-756-0.ch001.

[29] Manoj Philip Mathen. (2010). Data warehouse testing. Infosys DeveloperIQ Magazine, pages 1–8.

[30] Coles, Michael & Landrum, Rodney. (2009). Expert SQL Server 2008 Encryption. 10.1007/978-1-4302-3365-7.

[31] R, Suneetha & R., Krishnamoorthi. (2009). Data Preprocessing and Easy Access Retrieval of Data through Data Warehouse. Lecture Notes in Engineering and Computer Science. 2178.

[32] Golfarelli, Matteo & Rizzi, Stefano. (2009). A Survey on Temporal Data Warehousing. IJDWM. 5. 1-17. 10.4018/jdwm.2009010101.

[33] Vassiliadis, Panos & Simitsis, Alkis. (2008). Near Real Time ETL. 10.1007/978-0-387-87431-9_2.

[34] Vikas Ranjan. (2009). A Comparative Study between ETL (Extract-Transform-Load) and E-LT (Extract Load-Transform) approach for loading data into a Data Warehouse.MS Candidate in Computer Science at California State University, Chico, CA 95929.

[35] Romero, Oscar & Abelló, Alberto. (2009). A Survey of Multidimensional Modeling Methodologies. IJDWM. 5. 1-23. 10.4018/jdwm.2009040101.

[36] Vassiliadis, Panos. (2011). A Survey of Extract–Transform–Load Technology. International Journal of Data Warehousing and Mining. 5. 1-27. 10.4018/jdwm.2009070101.

[37] Li, Qing & Chen, Yu-Liu. (2009). Entity-Relationship Diagram. 10.1007/978-3-540-89556-5_6.

[38] Simitsis, Alkis & Vassiliadis, Panos. (2008). A method for the mapping of conceptual designs to logical blueprints for ETL processes. Decision Support Systems. 22-40. 10.1016/j.dss.2006.12.002.

[39] March, Salvatore & Hevner, Alan. (2007). Integrated decision support systems: A data warehousing perspective. Decision Support Systems. 43. 1031-1043. 10.1016/j.dss.2005.05.029.

[40] Rahman, Nayem. (2007). Refreshing data warehouses with near real-time updates. Journal of Computer Information Systems. 47. 71-80.

[41] Giorgini, Paolo & Rizzi, Stefano & Garzetti, Maddalena. (2008). GRAnD: A goal-oriented approach to requirement analysis in data warehouses. Decision Support Systems. 45. 4-21. 10.1016/j.dss.2006.12.001.

[42] Carreira, Paulo & Galhardas, Helena & Lopes, Antónia & Pereira, João. (2007). One-to-many data transformations through data mappers. Data Knowl. Eng.. 62. 483-503. 10.1016/j.datak.2006.08.011.

[43] Abelló, Alberto & Samos, José & Saltor, Fèlix. (2006). YAM2: a multidimensional conceptual model extending UML. Information Systems. 31. 541-567. 10.1016/j.is.2004.12.002.

[44] Fernández-Medina, Eduardo & Trujillo, Juan & Villarroel, Rodolfo & Piattini, Mario. (2006). Access control and audit model for the multidimensional modeling of data warehouses. Decision Support Systems. 42. 1270-1289. 10.1016/j.dss.2005.10.008.

[45] Gupta, Himanshu & Mumick, Inderpal. (1999). Incremental Maintenance of Aggregate and Outerjoin Expressions. Information Systems. 31. 10.1016/j.is.2004.11.011.

[46] Griffin et al. (2005). Method of Managing Slowly Changing Dimensions. United States Patent, Patent No.: US 6,847,973 B2, Date of Patent: Jan. 25, 2005.

[47] Vassiliadis, Panos & Simitsis, Alkis & Georgantas, Panos & Terrovitis, Manolis & Skiadopoulos, Spiros. (2005). A generic and customizable framework for the design of ETL scenarios. Information Systems. 30. 492-525. 10.1016/j.is.2004.11.002.

[48] Levene, Mark & Lozou, George. (2003). Why is the snowflake schema a good data warehouse design?. Information Systems. 28. 225-240. 10.1016/S0306-4379(02)00021-2.

[49] Lechtenbörger, Jens & Vossen, Gottfried. (2003). Multidimensional normal forms for data warehouse design. Information Systems. 28. 415-434. 10.1016/S0306-4379(02)00024-8.

[50] Shim, Jung & Warkentin, Merrill & Courtney, James & Power, Daniel & Sharda, Ramesh & Carlsson, Christer. (2002). Past, Present, and Future of Decision Support Technology. Decision Support Systems. 33. 111-126. 10.1016/S0167-9236(01)00139-7.

[51] Vassiliadis, Panos & Quix, Christoph & Vassiliou, Yannis & Jarke, Matthias. (2001). Data warehouse process management. Information Systems. 205-236. 10.1016/S0306-4379(01)00018-7.

[52] https://exploreinformatica.com/md5-message-digest-algorithm-5-in-informatica/