

# A Design Comparison: Data Warehouse Schema versus Conventional Relational Database Schema

Abdulrahman Yusuf  
Yobe State University Damaturu,  
Yobe State, Nigeria.  
audura33@yahoo.com

## ABSTRACT

Initially, relational database is for both operational and decision support system, as the information society experiences exponential growth in the amount of data/information to be stored in a database, a line has been drawn between transactional database and decision support database. Unlike traditional database, data warehouse aims to come from a number of pre-existing databases (developed from relational schemas). This conceptual paper discusses traditional database schema design and that of data warehouse schema architectural designs strategies that could be a guiding principles for both learners and beginners in database management system. It has explored the stages in development processes of the two. Subject orientation, data integration, non-volatility of data, and time variations are the key issues under consideration that could differentiate between traditional databases and data warehouse schema designs. It has also presented Design Modelling Techniques as well as addressing logical data models for data warehouse schema and traditional relational database.

## CCS Concepts

• Information systems → Database management system engines

## Keywords

Database; Data Warehouse (DW) ; Schema Design; Relational; Dimensional; Modelling; Normalization.

## 1. INTRODUCTION

In the words of [1] formerly, data had been kept in files (databases) which could not be accessed without programming language knowledge. When the idea of relational database has eluded, the uneasy organizing and designing of database in the underlying database structure transformed to very less troublesome. The father of relational model, E. F. Godd brought the basis for solving a lot of database problems, this is because relational database have offered data independencies, data consistency and data integration, thus capable of stockpiling data in rows and column tables, data normalization and tables' relationship.

According to [8] for a long, databases and data theory have been in existence; originally, database was for information processing community (transaction and analysis); but fundamentally centered around transaction processing. The concept of elegant notion of

database has erected – one for handling operational needs and the other for analytical needs. In 1990s, Inmon and Kimbell transformed and extended the issue of data warehouse to become more sophisticated as a result of advent of PCs, 4GL technology and end users encouragement [6]; [8].

Data warehouse development is seemed to be iterative development process as it involves business requirement identification, solution development of identified requirements, as well as implementation of architectural design of the data warehouse. Both relational schema and DW schema can be design using relational database design. Most DW involves the use of pre-compiled tables of summary of materialized views. [11] stated that in designing data warehouse, existing databases are to be taken into consideration; however, this is not necessary in operational database design (unless bottom up design is to be used to incorporate previously created views). We should understand that DW can be designed using either of the entity-relational model or dimensional model. But star (dimensional) schema is more structurally easier to understand. And in term of query, it is less cost, correctness and friendly. The data warehouse schema development could be in accordance with the relational model, based on data normalization, or the multidimensional based on de-normalization.

This research paper is generally intended to clearly demonstrate the road map for successful design as well as implementation of schema of data warehouse and relational database. The paper has been approached theoretically, using the existing literature, explored and presented the technicalities in database design for better comprehension. The study has compared and exposed the **Design Issues** of relational database and that of data warehouse in which the designers could be able to get quick knowledge for the simplification of the schema design of both transactional database (relational database) and decision support system (data warehouse). In this connection therefore, existing literature on relevant rr techniques were found from authentic sources for this work. Furthermore, the study produced both theoretical and pictorial presentations of the design comparison of database schema and data warehouse schema that can stand as a guide to especially both learners and beginners in field of database management system.

Consequently, the study has concretely revealed that either of traditional database or data warehouse schema design requires requirement analysis and specification. Unlike traditional database, data warehouse aims to combine a number of pre-existing databases (developed from relational schema). It has been explained that Subject orientation, data integration, non-volatility of data, and time variations are the key issues under consideration that can give base to differentiate between traditional databases and data warehouse schema design. Star and relational schemas are logical data models for data warehouse. The research has found two great disparities of schema design of the two models. In data warehouse which uses star schema emphasis is on one table

(fact table) with the structural rule describing the organization unlike traditional relational schema no particular table has total consideration. In the other hand, the relationships established between the tables in traditional relational database are based logically while relationship is based on 1-Many in star schema.

Before indulging into design issues (schemas' comparison) of the relational database and data warehouse, let have vivid understand of the meaning of operational database and data warehouse.

## 2. MEANING OF DATABASE AND DATA WAREHOUSE

Database is the data storage that consists of set of tables for operational data storage and is usually related to a single application. In the word of [15] end users can retrieve and store data in tables from relational database and is built using relational model for business processing applications. Traditional database sometime called OLTP - Online Transaction Processes. DW is a central location (repository) of subjectively collected data from existing operational data storage relevant to qualitative and quantitative data spanning from several sources and time periods. This can allow accomplishment of answering ad hoc, statistical, and complex analytical queries [15]; [9]. DW is the root of Decision Support System (DSS) which is subject oriented, integrated, non-volatile and time variant [6]. It allows for creation of ease use environment for management to conduct decision analysis, identify new opportunities, or make tactical and strategic choices based on both internal and external information [9]. DW also known with OLAP – Online Analytical Processes. **Table 1** gives clear dissimilarities of traditional databases and data warehouse.

## 3. TRADITIONAL DATABASES AND DATA WAREHOUSE DESIGN

The stages used in traditional databases design are also used to handle the design goals of data warehouse [10]. [17] mentioned the following stages as complete development processes of a data warehouse: development of a feasibility study, business line analysis, data warehouse architecture design, selection of the technological solution, planning the project iterations, detail designing, data warehouse testing and implementation, deployment and roll-out.

In another view [10] assumed data warehouse begins with conceptual design which is divided into requirement analysis and design, and then the individual stages of logical and physical phases are performed.

### 3.0.1 Requirement Analysis

Any design begins with obtaining necessary information/data from a business environment for analysis and specification. [10] said in data warehouse analysis and specification of data requirements are performed to agree with approaching data warehouse. He further explained requirement analysis for traditional database design is to analyze and specify the activities of the overall business operations. Unlike in traditional database, requirement analysis for data warehouse aims to bring together a number of already exist operational sources of data (pre-existing databases), so that relevant data warehouse attributes are chosen and outlined initial OLAP queries on the information discovered in operational database schemata. It has been seen that reporting and analyzing of users' needs have to be considered in creating data warehouse else it will be "data jail".

### 3.0.2 Conceptual Data Model

Individual entities are formed logically from requirement analysis, an entity appear to be very significant and is stored data [13]. Rows uniqueness in the entities are identified (unique identifier or candidate key). According to [10] in data warehouse design semi-formal requirement specification from requirement analysis is converted into conceptual schema using a formalized multidimensional model. The result is in multidimensional diagram (fact tables and dimensions). For data warehouse this phase is just for functional dependencies analysis among measures as well as analysis of chosen dimension level. Due to the dynamic nature of data warehouse requirements object oriented multidimensional is best to deal with the criteria [7].

### 3.0.3 Logical Data Model

Technical issues are not taken into consideration; the model tries to depict all identified entities with their associated attributes and the relationship between the entities that stands for business requirement. More attention is on business problems and come up with a dynamic design [13]. In [10], data warehouse schema design, conceptual design is converted to logical data warehouse schema. For each terminal dimension level, its dimension level is developed and come up with update-independent of a set of view definitions.

### 3.0.4 Physical Data Model or Database Schema design

This is a graphical model that can be used for implementation of database. And it is product of fully normalized logical design or model [13]. Demormalization is usually carried out at this phase. This is adding redundancy tables for query performance improvement. And it is done by expert data architect especially in data warehouse where huge amount of data are sorted and summarized. The data warehouse schema development could be in accordance with the relational model, based on data normalization, or the multidimensional based on de-normalization. Semantic meaning has been conveyed by a drawing in graphical models as past research hypothesized and it is more advantageous in term of easy comprehension.

## 3.1 Schemas Design Issues of Relational Databases and Data Warehouses

The environment within which the relational database or data warehouse to be used needs to be explored for designing. This is done to find necessary details requirements for actual design. The following are issues to consider:

### 3.1.1 Subject Oriented Data

In data warehouse the subject area are considered in the design which could be students, employees, learning materials; in contrast, transaction operational systems are planned around organizational activities like health, payroll processing, students enrollment [6]; [16].

### 3.1.2 Integrated Data

Integrated data means "de-duplication information and merging it from many sources into consistent locations" [16]. In the words of [6] this happened to be of primary important out of all other aspects in data warehouse design because the data from different applications with various design decisions are to be considered for inconsistencies coordination. He further explains given specific example, an application may encode M for male and F for female others may use 1 and 0 respectively; and. However, the issue is

different from relational database perspective where only one particular design is taken out of thousands available.

### 3.1.3 Non-Volatile Data

The nature of data in the warehouse are only loaded and accessed but not updated [6]. This non-volatility characteristic makes it different from operational database in which data are constantly accessed, loaded as well as updated. [16] indicated that users can get the actual history of organization and reach at particular business decision due to this unique feature.

### 3.1.4 Time Reference Data

One crucial issue of consideration in schema design is time variant. [6] have identified the following issue of time variance in comparison of operational database system and data warehouse:

- The time horizon in data warehouse is of interval of at least 5 -10 years. However, for operational is shorter than data warehouse of 60 – 90 days.
- Data warehouse contains just sequence of refined snapshots of data at certain interval while operational databases do carry “current value” and its correctness is at the time of access hence updatable.
- The essential structure of data warehouse is the present of some feature of time (example days, month, and year). However, the fundamental structure of operational data is may not have those time element.

### 3.1.5 Granularity

This refers to the “level of details or summarization held in the unit of data in the data warehouse” [6]. This reflects the facts that details increase with low level of granularity and vice versa. Therefore, granularity is a design issue but less considered in traditional relational database design. For efficiency the amount of data to be kept need to be considered at low level of details in DW.

After exploring the design issues here are some tasks for DW. As DW data are from a number of OLTP sources, Extract, Transfer, and Load (ETL) tools are used for extraction and loading. The subsets of data warehouses (Data marts) are adopted through a data warehouse bus which is a standard set of attribute declarations. The integrate all of data warehouse components are from a metadata repository. The metadata is where definitions of the source data, data models for target databases, and transformation rules that transfigure source data into target data are stored [15]. Further stated a data warehouse need to be well-equipped with metadata management tools due to the size of metadata.

## 3.2 Design Modelling Techniques

There are number of modeling techniques that pave the way for actual schema design of either relational database or data warehouse. Prior chosen any technique, the designer should have clear understanding of the applicability and incapability of the techniques. As [6] pointed out two prominent models that can be used for schema design of either relational database or data warehouse are **process model** and **data model**.

### 3.2.1 Process model

Process model in data warehouse design appears to be problematic due to the fact that it is dealing with the requirement assumptions and such cannot be worked in data warehouse environment rather could be hold in operational environment.

From software engineering view point, a typical process model encompasses all or some of the following:

- A functional decomposition
- A context level diagram
- A data flow diagram
- A structural chart
- A state transition diagram
- An HIPO chart and
- Pseudocode [6].

### 3.2.2 Data model

The most prominent data modeling techniques are two for data warehouse, Entity-Relational and Dimensional modeling [15]. OLTP database schema design strategies give base for the philosophy of data warehouse architectural design. [3] found that data models have three (3) characteristics that make them easily understandable: (i) uncomplicatedness to structure types (tables and tuples) (ii) very negligible structure types in the model are described by styles (iii) pictorial representation used in the model. They also mentioned that there are two competitive prominent models for data warehouse design: **dimensional model or star schema** and **relational model**. **Snowflake schema** appears to be extension of star schema. **Multidimensional schema** is a generic term used to denote both star schema and snowflake schema and they are shown in **Figure 2**. Hence, both traditional relational schema and the star schema are logical data models.

#### 3.2.2.1 Dimensional modeling (star schema)

This is dealt with the fact table having a kind of relation with multi-attribute keys. An attribute primary key (mostly surrogate) in dimensional table corresponds exactly to one attribute in the fact table with dimensional key. This feature appears to be star-like structure; therefore, called dimensional model or star join schema or in short **star schema** [15]. In the word of [3] entity relationship constrained version are used for developing dimensional data warehouses. **The Figure 1** has central fact table depicting the business event (flight reservation) on which data warehouse is constructed. This pronounces a reservation as a specific flight, with a specific passenger, on a specific airline, at a specific time. This diagrammatic representation is looked like star and is called star schema. A business event is always being captured within the central entities from which DW is built. In another exemplification of purchase transaction where Purchase is the fact table and Task, Part plus Supplier are dimensions tables with their associated attributes as PURCHASE (**S#**, **P#**, **T#**, **PRICE**); TASK (**T#**, **TNAME**); PART (**P#**, **PNAME**, **PWT**) and SUPPLIER (**S#**, **SNAME**, **SZIP**). This is represented pictorially in **Figure 2A**.

According to [12] for two reasons the star schema is considered to be more efficient design than relational design: firstly, it uses small join operations for the fact that has de-normalized table design. Secondly, it has efficient access plan generated for the operation because majority of the optimizers perfectly work with the star schema. According to [18] the most commonly relational representation uses for multi-dimensional database is star schema, which connect fact table directly to the dimensional table (attributes descriptive table).

[5] opined that star and relational schema are logical data models. These schemas have two things that make them different the **Figures 1 and 3** are used to highlight the disparity diagrammatically. In star schema emphasis is on one table (fact

table) with the structural rule describing the organization unlike traditional relational schema no particular table has total consideration. The relationships established between the tables in traditional relational database are based logically. Conversely, the relationship is based on 1-Many in star schema.

However, having a more complex data warehouse with a lot of dimensions could weaken the advantages of a star schema's presentation by the total number of elements. A complexity in data warehouse means having many several smaller star schemas that share a common dimension [16].

**Snowflake schema** – A perfectly star schema may encounter problem due to the increased number of de-normalized tables; therefore, performance problem may result, snowflake could be the solution [12]. Snowflake is an extension of star schema [5]; [15]. Low level cardinality attributes are moved from dimension table and placed in other tables that can be linked against dimensional table and shown in **Figure 2B** which is the extension of instance given previously. **Dept**, **Proj** and **Region** are the outer-level extensions that make snow-flake with their associated attributes as PURCHASE (S#, P#, T#, PRICE); **DEPT** (D#, DNAME, DBUDGET); **PROJ** (PJ#, PJNAME) TASK (T#, TNAME, D#, PJ#); PART (P#, PNAME, PWT); **REGION** (R#, RNAME); and SUPPLIER (S#, SNAME, SZIP, R#). **Figure 1** and **2A** show the star is placed at the center. The presence of outer-level dimension table is the significant extension. Hierarchy is placed in which 1-to-many relationship from every outmost table to the fundamental fact table. The de-normalized table might be reduced due the introduction of outer-level dimension table(s). Although, introduction of outer-level dimension table mostly reduces if not essentially eliminating the number of de-normalized tables.

### 3.2.2.2 Entity-Relational model

This model follows same procedure with OLTP database design process, conceptual entity relationship is translated into relational schema before normalization as well as de-normalization possible [15]. [3] explained that entity relationship schema is not like star schema, has no focus on central table as shown in the **Figure 3** (flight reservation). One important aspect of relational model is that normalization is used to preserve data integrity. However, this is different in data warehouse because data update is not an issue in DW. Clear differences can be seen when compared **Figure 1 & 3** as feature the same information. From an information-content perspective, the two schemas of different models could spotlight the same objective. In [18] Path Relational Schema which is transformation of relational database schema exhibited promising properties in term of query answering time compare with star and snowflake schemas. This is because; it is suitable for handling inconsistencies in the multi-dimensional databases.

Design difficulties, inefficient use of I/O and disk space, challenges in maintenance as well as designers' effort to find trade-off between optimization of query performance and maximization of query flexibility [1]

## 3.3 Schema Architectural Design

Architectural design happens to be blueprint that makes communication, planning, maintenance, learning and reuse possible. Data design, technical design and hardware and software infrastructure are some several aspects that involved in architecture. [15] mentioned top-down, bottom-up, inside out and mixed are schema design strategies available.

The architecture is logical and physical formation. The existing components as well as missing ones of the organisation in question are identified for data warehouse architecture completion. Data warehouse architecture need to be designed with a minimal impact on the existing model, so that it can allow further development.

The two have different optimization techniques because of different data access patterns. Different accesses perform require various optimization techniques greatly. In the OLAP bundle, the queries are usually involve significant aggregation and joining for decision support. Therefore, for performance improvement, de-normalization is usually encouraged in a data warehouse environment.

The use of combination top-down and bottom-up of data design strategy facilitate the creation of data mart individually in a bottom-up manner, which is in conformity with a skeleton schema called **data warehouse bus**. The amalgamation of adapted data marts of whole organization is DW. A hub-and-spoke architecture, enterprise warehouse with operational data store (real-time access support), and distributed enterprise data warehouse architecture are various types of basic architectural design presented by [15] as shown **Figure 4**.

## 3.4 Schema Evolution

According to [12] over time the way things being carried out changes, the idea behind data model design of associated application may experience drastic change. Dimension changes, instant changes, fact changes, level changes, attribute changes, constraint changes and quality changes are possible arising changes in schema design especially for data warehouse [2]. The schema evolution challenges is everlasting and applicable to every database practically possible. Data warehouse is developed iteratively so that strategic decision in the future could be possible [12].

## 4. CONCLUSION

Initially data had been kept in files (databases) with a lot of maintenance difficulties. The idea of relational database have come up and simplified the underlying database structures to be less troublesome where the developed database is for both transaction and analytic (information processing community) principally on transaction processing. OLTP database schema design strategies give base for the philosophy of data warehouse architectural design.

The exponential growth of data storage plus the advancement of information and communication technology (ICT) pave the way, a line has been drawn between operational system (transactional database) and analytical system (decision support database). Inmon and Kimbel transformed and extended the issue of data warehouse and become more sophisticated following the advent of PCs and 4GL technology [6], [8]. Before traditional database or data warehouse schema design requirement analysis and specification should be carried out. Unlike traditional database, data warehouse aims to combine a number of pre-existing database.

This paper was approached theoretically, the technicalities in database and data warehouse designs were presented from explored existing literature for easy and better comprehension. The study compared and exposed the **Design Issues** of relational database and that of data warehouse in which the designers can be able to get quick knowledge for the simplification of the schema

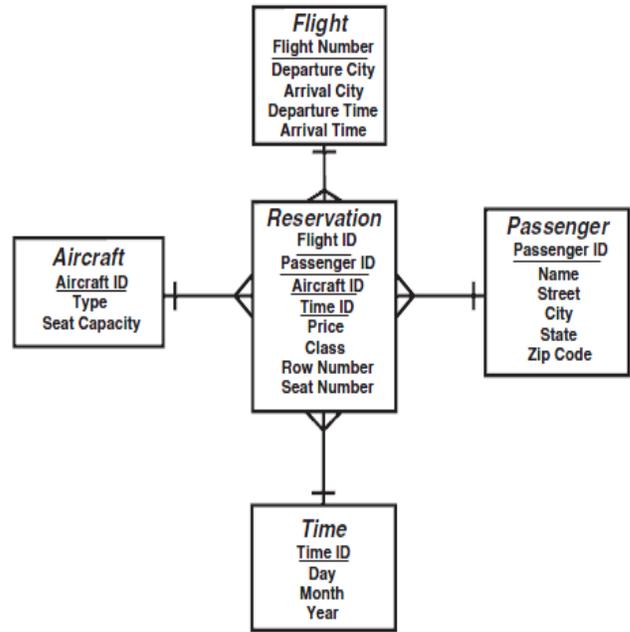
design of both transactional database (relational database) and decision support system (data warehouse).

The study produced both theoretical and diagrammatical presentations of the design comparison of database schema and data warehouse schema that can stand as a guide, especially to both learners and beginners in field of database management system. Consequently, the paper was concretely revealed that either of traditional database or data warehouse schema design requires requirement analysis and specification. Unlike traditional database, data warehouse aims to combine a number of pre-existing databases (developed from relational schema). It explained Subject orientation, data integration, non-volatility of data, and time variations as the key issues under consideration that can give base for the differentiation between traditional databases and data warehouse schema design. Star and relational schemas are logical data models for data warehouse. The research found two great disparities of schema design of the two models. In data warehouse which uses star schema emphasis is on one table (fact table) with the structural rule describing the organization unlike traditional relational schema no particular table has total consideration. In the other hand, the relationships established between the tables in traditional relational database are based logically while relationship is based on 1-Many in star schema. **Figure 1 & 3** highlight the disparity.

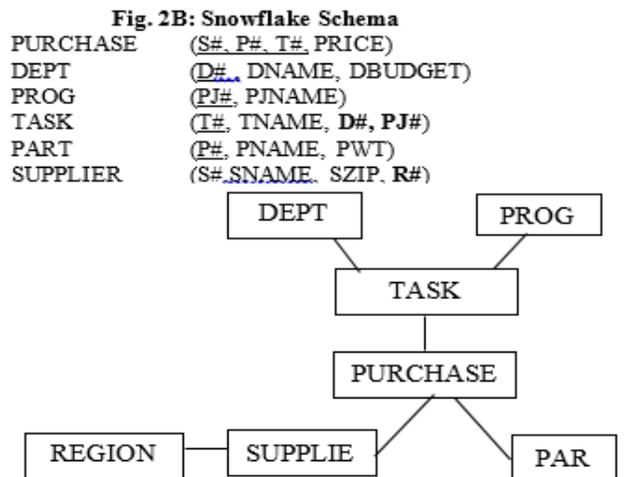
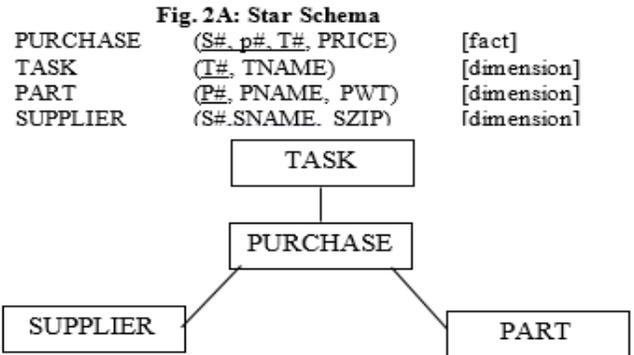
### 5. TABLE AND FIGURES

**Table 1: Distinction between OLTP and OLAP systems**

	<b>OLTP System (Operational Data)</b>	<b>OLAP System (Data Warehouse)</b>
<b>Data source</b>	From Traditional or original source	From diverse databases
<b>Aim of data</b>	Organizational basic activities running	For planning and problem solving known Business support system
<b>Data type</b>	images of ongoing business transaction	Business activities from various sections
<b>Insertion and updating</b>	Insert and update are by end users which is fast	Periodic refreshing inform of Batch jobs
<b>Queries</b>	Simple queries that return small records	Complex aggregate queries
<b>Efficiency</b>	Very high speed to answer queries	Takes some time depending on the requirement but can be improved using index
<b>Space hold</b>	holds very small history	Comprehensive history due to aggregation
<b>Databa se design</b>	Greatly normalized due to many tables involvement	Highly denormalised with few tables using Star and snowflake schema.



**Figure 1. Simple star schema**



**Figure 2. Star and snowflake schema**

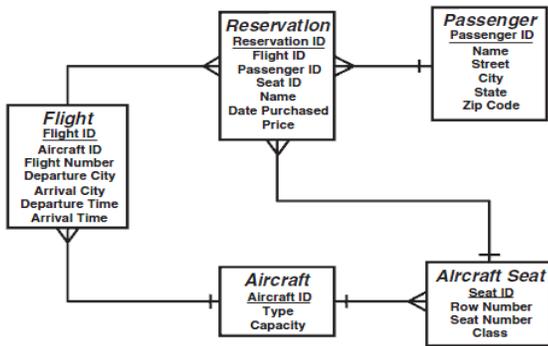


Figure 3. Simple traditional relational schema Captions

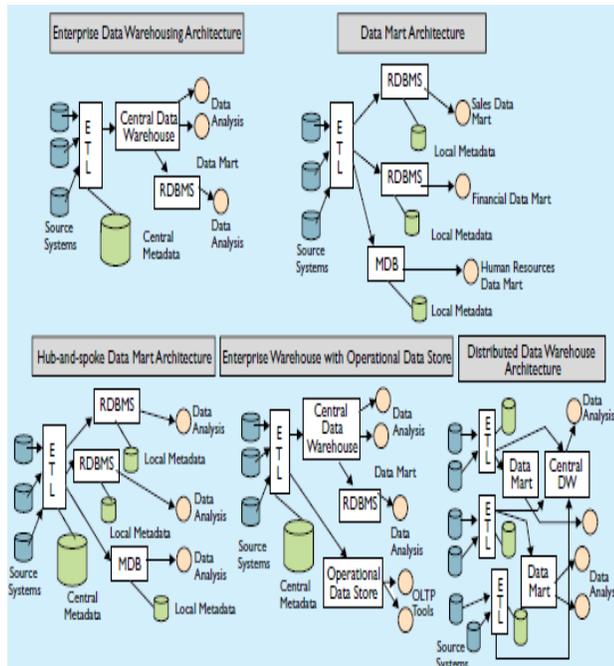


Figure 4: Various architectural design of DW

## 6. ACKNOWLEDGMENTS

My special thanks to Prof. Carl Dedley and Dr. Mary G. of Wolverhampton University, UK who have accessed part of this paper as their module assessment.2013.

## 7. REFERENCES.

[1] Anonymous 2013. Comparison of Data Warehousing DBMS Platforms - ODBMS.org DOI= [www.odbms.org/wp-content/uploads/2013/11/illuminate-Comparison.pdf](http://www.odbms.org/wp-content/uploads/2013/11/illuminate-Comparison.pdf).

[2] Arora, M. and Gosain, A. 2011. Schema Evolution for Data Warehouse: A Survey. International Journal of Computer Applications. [online] 22, 6, p6-14. DOI= <http://wlv.summon.serialssolutions.com/link/>.

[3] Corral, K., Schuff, D. and St. Louis, R.D. 2006. The impact of alternative diagrams on the accuracy of recall: A comparison of star-schema diagrams and entity-relationship diagrams. Decision Support Systems [online], 42(1), pp. 450-468. DOI= <http://wlv.summon.serialssolutions.com/link/>.

[4] Datawarehouse4u 2009. OLTP VS OLAP. DOI= <http://www.datawarehouse4u.co.uk/OLTP-vs-OLAP.html>.

[5] Dedley, C. 2009. Database Technologies. Lecture: Data Warehouse [online]. [Accessed 28th November, 2012]. DOI= <http://wolf.wlv.ac.uk/stech/72043/datawarehouse2009.ppt?menu=833910/>

[6] Inmon, W. H. 1996. Building the Data Warehouse. 2nd Ed. New York: John Wiley & Sons.

[7] Jindal, R. and Taneja, S. 2012. Comparative Study of Data Warehouse Design Approaches : A Survey. International Journal of Database Management Systems [online], 4(1), pp. 33-45. DOI= <http://wlv.summon.serialssolutions.com/link/>.

[8] Johnston T. and Weis R. 2010. Managing Time in Relational Databases: How to Design, Update and Query Temporal Data. 1st Ed. Published: Morgan Kaufmann. DOI= <http://proquestcombo.safaribooksonline.com/book/databases/database-design/9780123750419/>

[9] Kazparaz, R. 2011. What is the differences between database and data warehouse? DOI= <http://techforum4u.com/member.php/>

[10] Lechtenborger, J. (2003) Data Warehouse Schema Design, Dept. of Information Systems, University of Munster Germany. DOI= [www.pcw202.uni-muenster.de/dbms/media/.../2003/diss-jl-btw2003.pdf](http://www.pcw202.uni-muenster.de/dbms/media/.../2003/diss-jl-btw2003.pdf)

[11] Lechtenbörger, J. and Vossen, G. 2003. Multidimensional normal forms for data warehouse design. Information Systems [online], 28(5), pp. 415-434 DOI= <http://wlv.summon.serialssolutions.com/link/>

[12] Martyn, T. 2004. Reconsidering Multi-Dimensional schemas. ACM SIGMOD. 33(1), pp. 83-88 DOI= <http://wlv.summon.serialssolutions.com/link/>

[13] Richert A. 2009. Oracle SQL by Example. 4th Edition. [online] Publisher: Prentice hall pp 960-965. DOI= <http://proquestcombo.safaribooksonline.com/>

[14] Schuff, D., Corral, K. and Turetken, O. 2011. Comparing the understandability of alternative data warehouse schemas: An empirical study. Decision Support Systems. 52(1), pp. 9 DOI= <http://wlv.summon.serialssolutions.com/link/>.

[15] Sen, A. and Sinha, A. 2005. A comparison of data warehousing methodologies. Communications of the ACM [online], 48(3), pp. 79-84. DOI= <http://wlv.summon.serialssolutions.com/link/>

[16] Suri, P. and Sharma, M. (2011) A Comparative Study Between the Performance of Relational & Object Oriented Database in Data Warehousing. International Journal of Database Management Systems [online], 3(2), pp. 116-127 DOI= <http://wlv.summon.serialssolutions.com/link/>

[17] Velicanu, M. 2007. Building a Data Warehouse step by step Informatica Economică, nr. Vol. 2 (42).

[18] Yaghmaie, M., Bertossi, L. and Ariyan, S. 2012. Repair-oriented relational schemas for multidimensional databases [online]. ACM, pp.408-419. DOI= <http://wlv.summon.serialssolutions.com/link/>.