www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 7 Issue 7 July 2018, Page No. 24130-24151
Index Copernicus Value (2015): 58.10, 76.25 (2016) DOI: 10.18535/ijecs/v7i7.05

Perception and conceptualization of Data Warehouse of a day to day Transaction

Sumita Mukherjee

Associate Director
University Advancement & Internationalization Riara University P.O. Box 49940-00100 Nairobi, Kenya

Objective: The value, which is shown by organization handling Data warehouse, has more significance when the database becomes older. The fully developed database has a positive effect and consequences on data analyst, any business world user or any association related to data analytics. Day to day the preferment is going through new developments on software, hardware, live ware and firmware. This generation has moved through the transition of bytes to yottabytes. Any kinds of prediction on many fields like medicine, biometrics, metrology, business, artificial intelligence etc. are determined more on data warehouse. Thus Data warehouse is the mandatory stage for any kind of prediction and business intelligence. The data Warehouse does the accomplishment of data from the source system, categorizing it, scrubbing it in such a manner so that any users can understand and manipulate the abstraction easily.

In this paper example of a Furniture shop in Nairobi is taken and tried to explain the various aspects of Data Warehouse. This Paper helps to understand the basic concepts, modeling design architecture and general implementation of Data Warehouse online analytical process overall (OLAP).

Abstract: This paper incorporates procedures, protocols, system and constructs a utilitarian tool to generate a factual resource, from one big storehouse by collecting, organizing, storing data from different databases and is available for various kind of quarries which can directly be handled and is utilized for relevant analysis. Various literature is studied and evaluated related to data warehouse concepts, design and application before wiring this paper.

Introduction: The method of dealing with past and present data is based on availability utilizing the same for day to day business transaction and generating new data for future use is known as data mining**.**

Transaction might have a negative effect with data in a database. Database might not access its own data accurately and process properly as the entire data is not used for day-to-day transaction, whereas most of the data is utilized for the mining of data, analysis and reporting issues.

Storing historical data in everyday database will cause a huge increase of its size, which leads to a slower performance. A good practice is to move the old data from different sources and integrate the whole in another repository called data warehouse moving the data from operational databases to a data warehouse involves three steps: 1) Cleaning, 2) Transformation and 3) Integration.

Data Warehouses is used to induce and strengthen data in many various dimensional capacity . The Construction of data warehouse involves data cleaning, data integration, and data transformation, and can be viewed as an important preprocessing step for data mining.

Moreover data warehouses provide online analytical processing (OLAP) tools for the interactive analysis of multidimensional data of varied granularities, which facilitates effective data generalization and data mining.

Many other data mining functions such as association, classification, and prediction and clustering can be integrated with OLAP operations to enhance interactive mining of knowledge at multiple levels of Abstraction. Therefore OLAP and data warehouse together form an essential package in the knowledge discovery process.

Warehousing of Data: Data warehousing is the action of establishing and managing a data warehouse. A data warehouse is constructed by integrating data from multiple heterogeneous sources that reinforce analytical reporting, structured and/ or ad hoc queries, and decision-making. Data warehousing involves data cleaning, data integration, and data consolidations.

Data warehouse is distinctly differentiated from different data repository systems such as relational database systems, transaction processing systems and open file systems.

The Four Keywords ---object aligned, unsegregated, time-variant, and consistent.

1. **Object-Aligned**: This phenomena incorporates around main objects and formulates data ,and information for the creates a warehouse of data and information around main objects such as client , supplier, retailer, buyer, product and sales. These Objects are having data members and functions associated with them. Rather than handling every day's operation on business transaction any company is keen to extract data from data warehouse and emphasizes on the modelling of data and inspection of data for conclusion makers and system analysts. Thus data warehouse briefs a succinct and compact view of a particular object. It solves the problem of handling the data member, having any complications and also defines the data members and their characteristics.

2. **Un-segregated:** This phenomena creates a data warehouse by amalgamating various objects from different data sources. It can involve torrent files, relational databases, indexed files and online transaction records.

Heterogeneous Databases: There are two methods described based on amalgamating databases from different sources of data–i) Enquiry-driven method ii) Revise-driven method

i)Enquiry Driven Method: This method involves the processes which are filtered and assembled together. They are available for heterogeneous systems and heterogonous records retrieved temporarily for consideration. It is not feasible and practically turns costly for periodical enquiries, mostly for those enquiries who need aggregations and clustering.

ii)Revise Driven Method: This method involves various data information derived and retrieved from different sources and then put together for integration. Once integrated, data with information is then reformed and then are placed in a warehouse for further reference. The information is stored, organized, copied, integrated, processed, expounded, summarized and restructured in semantic data store permanently so that the data can be used for future reference. This updated information is suitable for direct enquiry and inspection. The data provided by this method is correct and of good standard.

3**. Data model Described:** An entity relationship(ER) data model is generally adopted by an Online Transaction Process (OLTP) and designs the database in such a manner where data is scrutinized on their application. A star or snowflake model generally considers an online transaction process. It also contains logical and physical design used for data definition language.

4. **Table:** An OLTP system involves and extracts those data which are available immediately in an organization, venture, business house or department, It does not only include old data from one organization but also do not consider old data from various institutes. Different versions of database schema either from one organization or from different organization is handled by online transaction process. OLTP system also extracts date from the transition or movement of data from one department to another in the same organization. OLTP deals with huge volumes of data and is not possible to store the large data in one storage device thus refers to data warehouse. It make sure that the data is to be stored on different devices.

5**. Entry Described**: Here an explanation of an OLTP system process is mainly short unique transactions.

Definition of Metadata:

Data within data is known as metadata. When one data correlates with another data and depicts the main data. For example, date, time, filename, camera settings of a photo image implicates metadata for the photo image clicked. In other words, we can say that when many small information is encapsulated it rises to the complete data and smaller fragments of information represent metadata.

Metadata can be defined as

1. Data warehouse refers metadata for not only drawing out a small entity but also for group of entities and also data members. Metadata also figure outs and designs the data warehouse flowchart and helps for conducting the process.
2. Metadata in data warehouse defines the data entity, data relationship, and also throws back the objects in collaboration with data member.
3. Metadata is a directory in data warehouse. This directory helps to maintain a systemetic order of information.
4. Metadata detects the information of data warehouse and thus also known as the decision support system .

Defination of Metadata Respirotary:

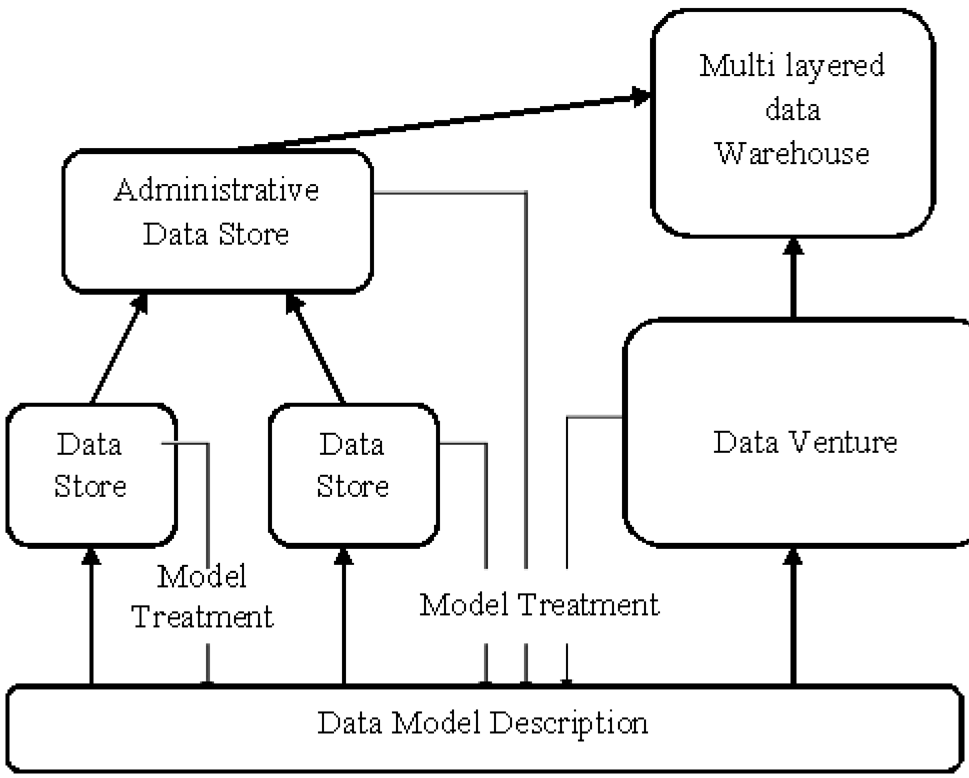
a) Descriptive Metadata: This metadata structure throws explanation on identification,defination,description and discovery.It can include tittle,objective,introduction,abstract,reference,keywords.
b) Operational Metadata:Data lineage and Curreccy of data are two important aspects of Operational metadata. a) Data Linage: The number of steps of transition and migration imposed on past and old data refers to history of data. b)Currency of data: By generating stastical report,error report and aduit trials on active, archived or purged files.

c) Constructed metadata is metadata about containers of data and indicates how compound objects are put together, for example, how pages are ordered to form chapters. It describes the types, versions, relationships and other characteristics of digital materials.

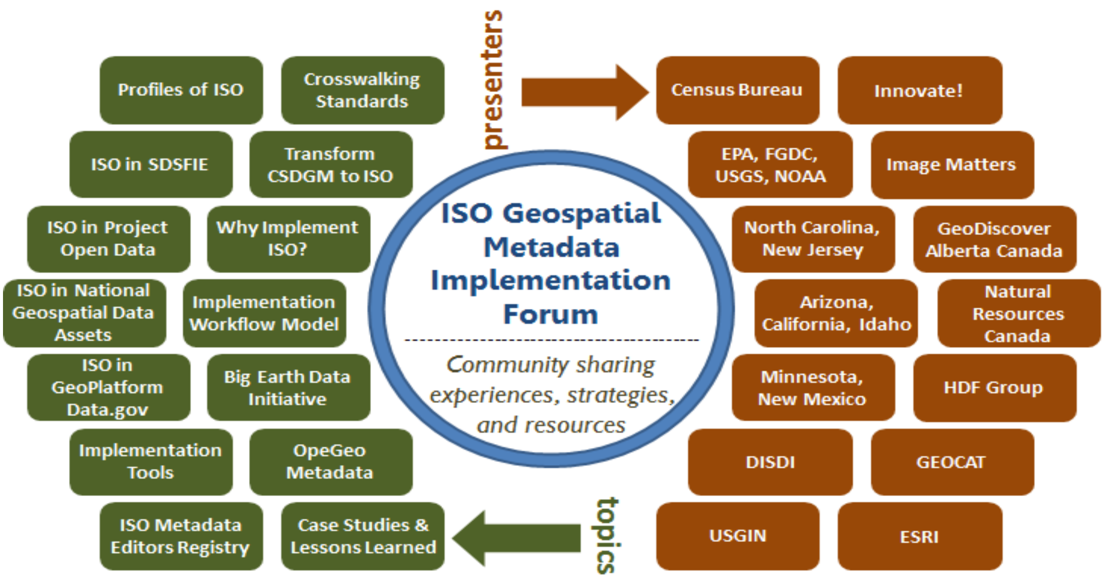
d) Administrative metadata: This metadata provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it.

e) Reference metadata: This describes the contents and quality of mathematical, numerical data.

Example of a Corporate Data Model Fig 1.1



Example of Metadata Respiratory Fig 1.2



Source:https://www.fgdc.gov/metadata/events/iso-geospatial-metadata-implementation-forum/index_html

Definition of Data Store:

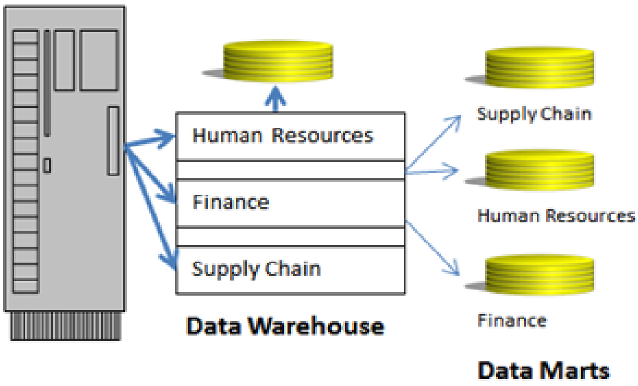
A data store contains a group and subset of organizational data which has wide range of brackets and index from which the value is taken and referred for an individual value or a group of values so that users accept the values only from the storage pools. This pool contains proved and determined values and objects. Let's explain with an example of a furniture data store may confine to its objects to customer, product, retail, supply, location, sales, inventory, procurement etc. The data which is present in data mart tend to be summarized in such a manner, which can be utilized immediately. Data stores are usually referred for departmental servers as less money is spent on them which are generally working on different operating systems. The execution pattern of a data store is more likely to be measured in weeks or fortnight rather than months or years.

Data stores can be categorized in to two groups as dependent and independent on the basis of sources through which data is retrieved. When the data are gathered from internal sources and directly from various sources like department, surveys, questionnaire, individual interviews or data recoding etc. are known as Dependent data stores. Though they are derived directly from venture data warehouses still applicable and useful data.

When data are sourced indirectly from different organization or operational system or from different exterior providers dealing with information support system, or locally originated data in a particular department or demographic area is known as Independent Data Store. Though they are external but relevant and useful for reference.

Definition of Indirect Warehouse: A virtual warehouse is a set of views over operational databases. When the few confirmed and suitable and effective summary views are materialized and accepted it is believed that an effective query processing takes place. A virtual warehouse is built for effective query processing though requires a huge capacity. A virtual warehouse depends on various database servers, which have a regularized and valid operational and analytical system.

Example of Data Mart   Fig 1.3



Source : Martin Gibbs ***https://study.com/academy/lesson/what-is-a-data-mart-design-types-example.html***

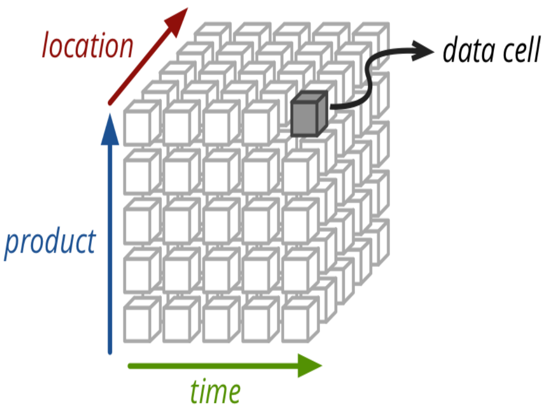
Definition of Data Cube: Dimensions and facts are represented through a data cube. It is designed and viewed in a manner which depicts a clear ideas and images on multiple dimensions. It works on

A multidimensional database develops a concept and frames a functional form which enables OLAP to interact smoothly with database.

Within a relational database a table is called a cube. It can be 2 dimensional, 3 dimensional or higher dimensional.

a) Measure: It represents some numbers facts and figures like cost, units, square foot, etc.

b) Dimension: It represents measures, which describes data item such as date, time, and location.



Example of a Data Cube Figure 1.4

Source: http://okfnlabs.org/blog/2014/01/20/olap-cubes-and-logical-model.html

Data Warehousing: Multitier Architecture Defined

Data warehousing nominates different layer tier architecture.

a) The first layer is represented by a relational database system. Data are picked from operational databases provided by back end tools or other external sources. The second tire is then enriched with this Data. Gateway takes the support of DBMS. Examples of gateway include ODBC and OLEDB by Microsoft and JDBC. This tier also contains a metadata repository which stores the data warehouse and information of its contents.

b) The center tier is an OLAP server which is ideally executed using either 1.) a relational OLAP or ROLAP model (an extended relational DBMS that maps operations on multidimensional data to standard relational operations. 2.) a multidimensional OLAP or MOLAP model a special purpose server that directly handles multidimensional data and operations.

c) The Upper tier is a front end client layer, which contains query and reporting tools, analysis tools, and or data mining tools like trend analysis, prediction etc.

Tools and Utilities required by Data Warehouse:

- **Extraction of Data** – Different heterogonous databases are responsible for collecting and sourcing data.
- **Cleaning of Data** – Once data is collected next step is of finding the errors present in collected data and removing the errors so that the data which is collected is useable. **Transformation of Data**– Once clean and useable data is available next step is the conversion i.e. from provision format is converted to warehouse format.
- **Loading of Data** – With warehouse format of data next step is of indexing, reporting, compression, examining integrity, and constructing indices and partitions.
- **Refreshing of Data** – Methodically after loading the data, the data must be modified and then the modified data from various sources are sent to warehouse.

Difference between OLAP and OLTP

Points	**OLAP(online application process)**	**OLTP(online transaction process)**
1.	Applicable in business analysis.	Useful in running the business.
2.	Provides summarized and multidimensional view of data. Contains ancient information.	Imports/allocates complete and flat relational view of data. Contains current data.
3	Number or users is in hundreds.	Number of users is in thousands.
4.	Well established employees such as executives, managers and analysts preferred to use OLAP system.	OLTP systems are used by clerks, DBAs, or database professionals.
5	Highly flexible and is established on the principle of Star Schema, Snowflake, Schema and Fact Constellation Schema.	High performance oriented and is established on Entity Relationship Model.

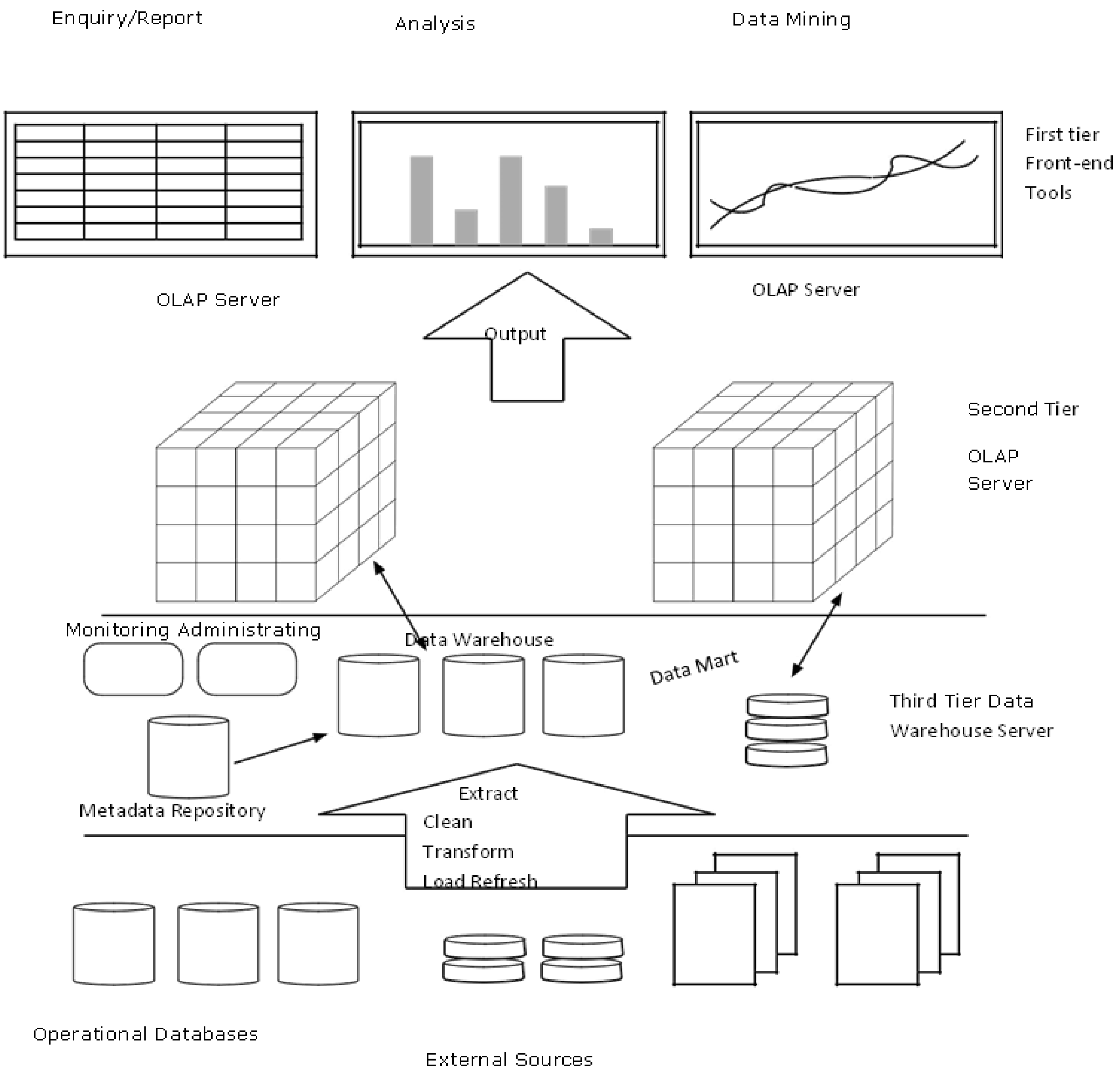


Fig 1.5 The Following Figure shows the three tier Warehouse Architecture

Explanation of Data Cube

Location "Nairobi"				
Time (Quarter)	Product (type)			
	Dining Table (DT)	Study Table (ST)	Bed	Sofa Set 9 (SS)
Quarter 1	20	8	11	9
Quarter 2	17	8	6	4
Quarter 3	19	6	8	2
Quarter 4	11	3	7	3

Example: A furniture enterprise wishes to be keeping track of inventory records with the help of data warehouse which reflects products in store referring to time frame, product, subsidiary, and location. These dimensions allow to keep track of monthly inventory and at which branch the products are available to sell. A diagram is represented for every feature. This diagram is known as dimension table. For example the above diagram "product" dimension table may have attributes such as time, Product_type etc. The above table represents the 2-D view of Inventory Data for a furniture company with respect to time, product, and location dimensions.

But here in this 2-D table, we have records with respect to time and product only. The inventory for Nairobi are shown with respect to time, and product dimensions according to type of product available.

Time	Location ”Nairobi”				Location ”Mombasa”				Location ”Kisumu”			
	Product				Product				Product			
	DT	ST	Bed	SS	DT	ST	Bed	SS	DT	ST	Bed	SS
Quarter1	10	10	10	8	17	7	10	9	12	17	7	9
Quarter2	17	7	9	11	14	12	14	7	9	3	11	8
Quarter3	19	9	10	12	9	10	11	8	7	19	6	9
Quarter4	11	13	11	7	9	8	10	12	10	12	10	8

For getting detailed inventory information containing with an extra aspect example, the location dimension, then the 3-D view would be useful. The 3-D view of the inventory data with respect to time, product and location is shown in the table below

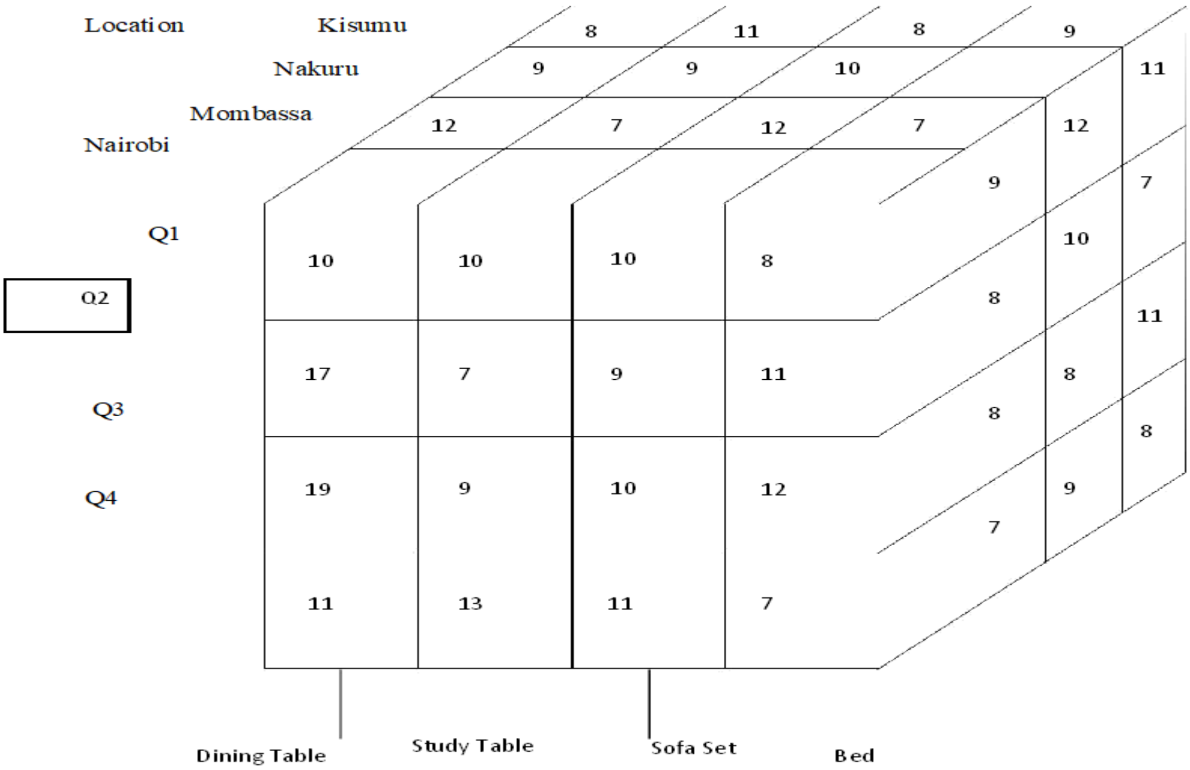


Fig 1.6 an example of a 3 D Data Cube

Fig1.7 Pictorial presentation 4 D Data Cube

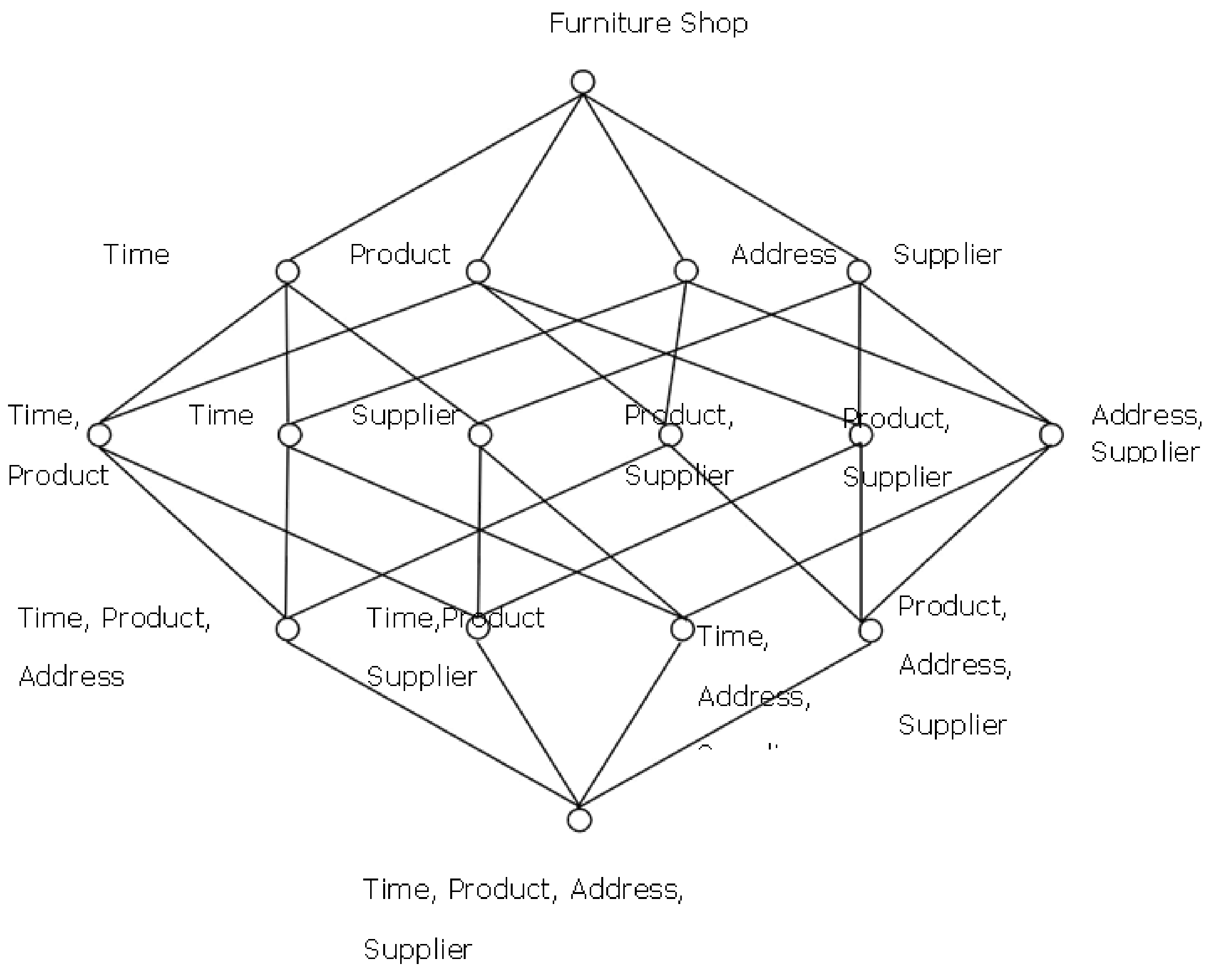
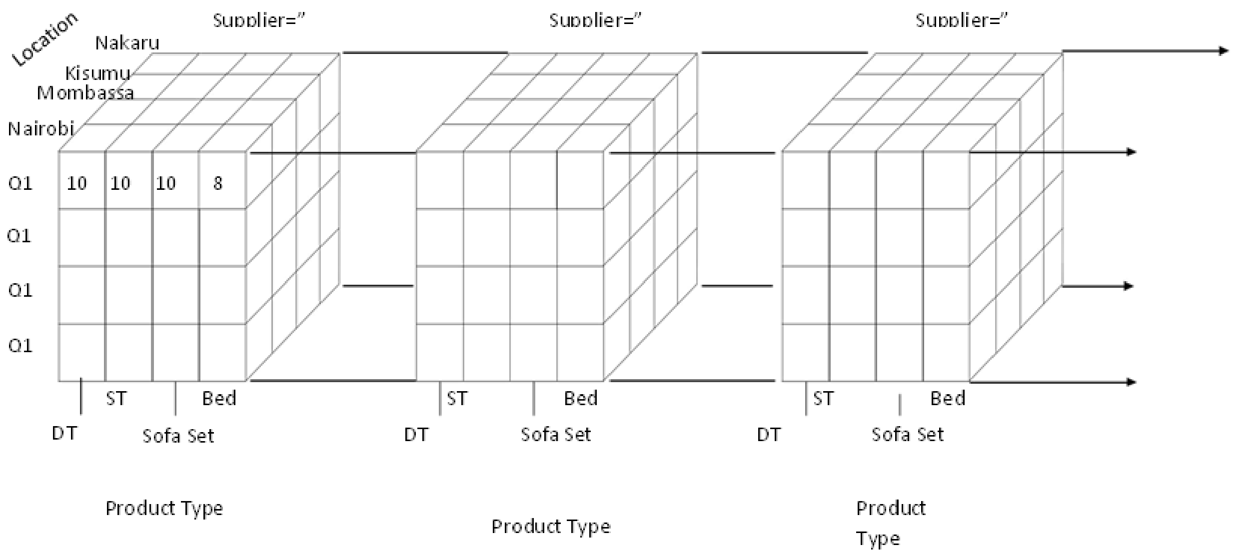


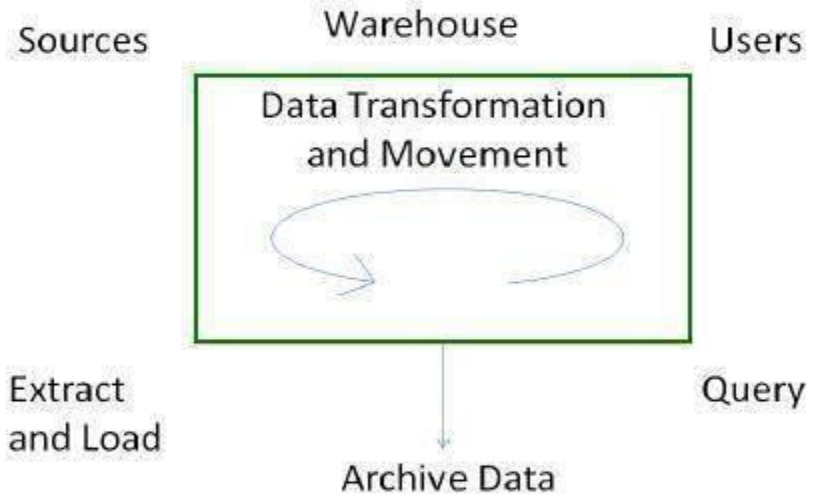
Fig 1.8 Data Cube of Inventory Data



Data Warehouse and its Associated Process

A data warehouse is comprised of 4 major processes. 1. Data extraction
2. Data loading 3.Data cleaning and transformation 4.Data backing formulates Archived data.

Administering queries and directing them to the appropriate data sources. Figure 1.9



Data Warehouse and Data Store Comparison

Points	Data Warehouse	Data Store
1.	Data Warehouse collects information about objects which includes data members and data functions that require the entire organization to operate. Thus it is Enterprise centric.	Data Store a partial subdivision of a data Warehouse that focuses on selected subject. Thus it is Department centric.
2.	Examples are Customers, Products, Inventory, Sales, Purchase, Quality control etc.	Examples are Sales Department, Inventory Department, Shipping Department etc.
3.	Data warehouse can model multiple inter related objects thus commonly used in Fact Constellation Schema.	Data store is suitable model for single object thus commonly used in Star or Snowflake Schema.

Dimensions: Concept Hierarchies Explanation:

A hierarchy abstraction configures the steps of calculating result from a range of smaller-level concepts to bigger-level, and explains a more of general concepts. Consider a concept hierarchy for the dimensions location. City values for location include Nairobi, Mombasa, Naibasha, Nayanza. One town however can be mapped to the provinces, ward, county to which it belongs. For example Mombasa can be mapped to Kilifi and Naibasha to Nakuru. The Provinces and ward can in turn be mapped to County to which they belong. The mappings for the dimension locations depicts hierarchies a standard of smaller –rank concepts (e.g. cities) to bigger-rank, more general concepts (i.e. counties) .This concept ranking is illustrated in the figure 2.0

Many concept hierarchies are implicit means handled internally within the database schema. For example suppose that the dimension location is described by the attributes, number, street, city, ward, province, county etc. The number of columns are reflected by a total order forming a concept hierarchy such as <city<ward<province<county. This ranking is shown below. Number of columns of dimension may be organized in a partial order, forming a lattice. An

example of partial order for the time dimension based on the attributes day, week, month, quarter, and year is day<week<month<quarter<year. This structure is shown in the figure. A concept hierarchy that is a total or partial order among attributes in a database schema is called a schema hierarchy. Concept hierarchies are common to many applications may be predefined in the data mining system. By discretizing or grouping hierarchy traverses concept ranking. Concept hierarchies that are common to many applications may be pre -defined in the data mining system. A given dimension and a column can be read through its values reflecting in a set grouping hierarchy. A total or partial order can be defined among group of values. An example of set grouping hierarchy is shown in the figure. 2.0 For the dimension Price.

Hierarchies are conceptualized and represented physically by system users, domain expert, or knowledge engineers ,system analyst or may be automatically generated based on statistical analysis of the data distribution.

Fig 2.0 Concept Hierarchy For Location

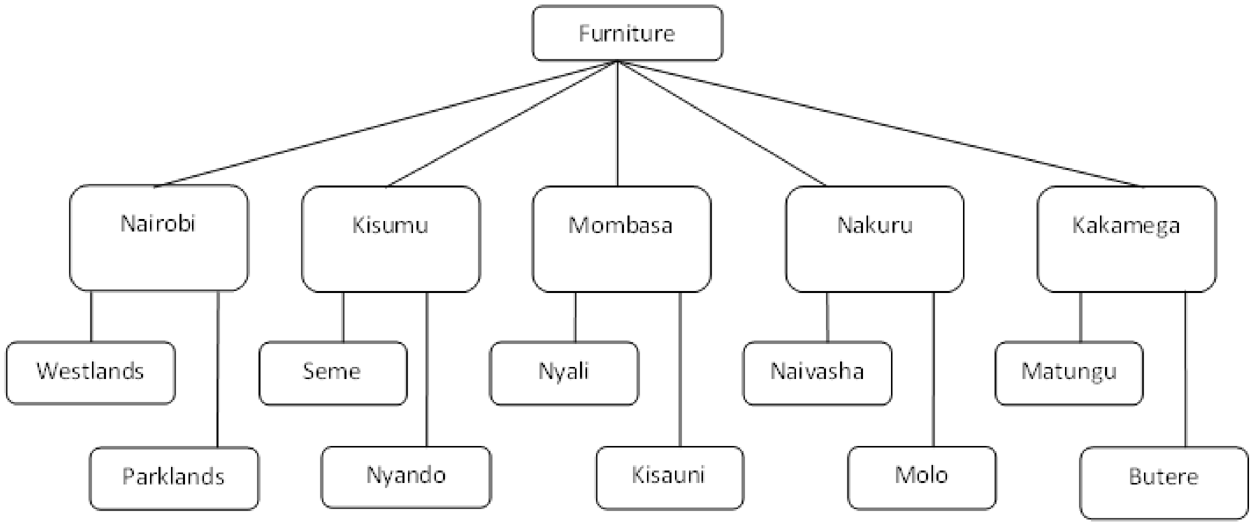
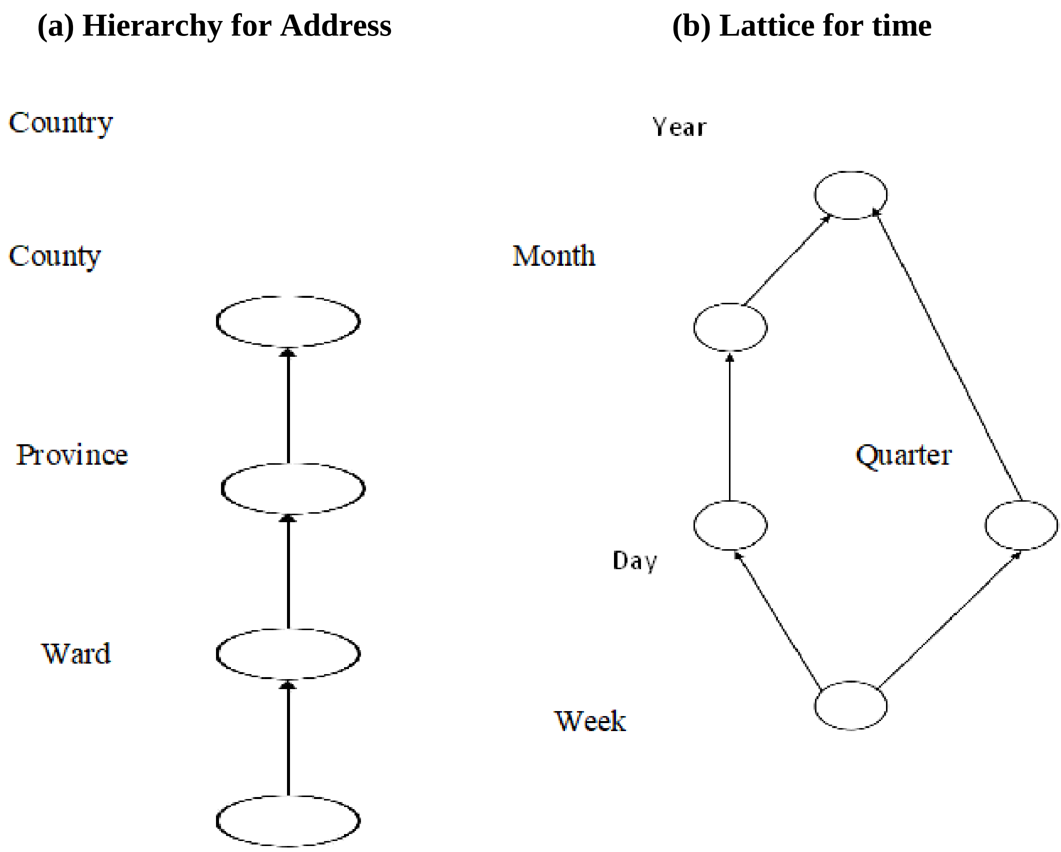


Fig 2.1 Ranked and Lattice structure of Attributes in Warehouse Dimensions



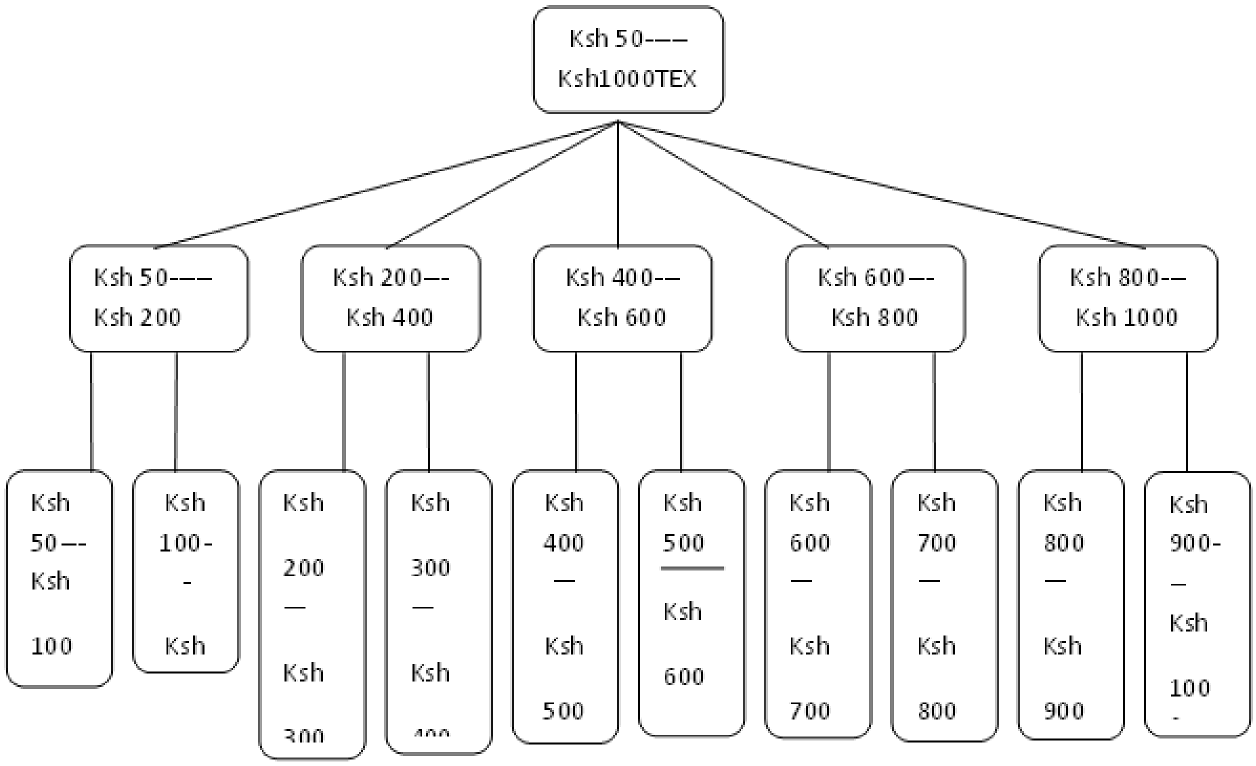


Fig 2.2 Example of a Concept Hirerarchy for Price

Estimates:

A data cube is a measurement unit formulates a function based on numbers, which is calculated and processed at each point and at center of a data cube space. A measure value is computed for a given point by aggregating and accumulating the data corresponding to the respective dimension taking the value of pairs defining the specified point. Measures is bifurcated into three groups.

a) Distributive: An aggregate function or accumulation is distributive if it can be computed in a distributive manner, suppose the data are partitioned into N sets. We apply the function to partition, resulting in n aggregates values. When applying the function to the aggregate attains the result. If the result derived by applying the function to the N aggregate values is the same as that deprived by applying the function to the entire data set the function can be computed in a distributive manner. For example

Sum () can be computed for a data cube by first partitioning the cube into a set of sub-cubes, computing sum () for each sub-cubes and then Summing the counts obtained for each sub-cube.
Hence sum () is a distributive aggregate function. For the same count (), min () and max () are distributive aggregate functions.
Algebraic: An aggregate function is algebraic if it can be computed by an algebraic function with N arguments where N is an enclosed positive integer each of which is obtained by applying a distributive aggregate function. For example avg (), is average can be computed by sum ()/count () where both sum () and count () are distributive aggregate functions. A measure is algebraic if it is obtained by algebraic aggregate derivation.
Holistic: A total function on summation is holistic if there is no constant bound on the storage size needed to describe a sub aggregate. It means there does not exist an algebraic function with N arguments which can be used for the calculation. The examples of holistic functions are median (), mode ( ) and rank ()

Multidimensional Data Models with different Schemas:

a)**Stars Constellations:** The simple, easy and common modeling in a large pattern represents the star schema in which the data warehouse contains (1) there will be a large central table (fact table) containing the huge data and there will be no repetition of the data. (2) It will have a range of smaller attendant tables (dimension tables) one for each dimension. The Schema graph resembles a starburst with the dimension tables displayed in a radial pattern around the central fact table.

Example of Star Schema:

A star schema for all furniture inventory shown in the figure .Inventory are considered along four dimensions instant, product, branch, and venue. The schema contains a central fact table, for inventories that contain keys to each of the four dimensions, along with two measures Shillings spent for storing the product and numbers of units stored.

Each dimension in Star Schema is represented by only one table and each table contains a set of attributes. For example the location dimension table contains attribute set (location key, city, province, ward, county, country).The constrains may introduce some redundancy. It is represented in the following figure 2.3

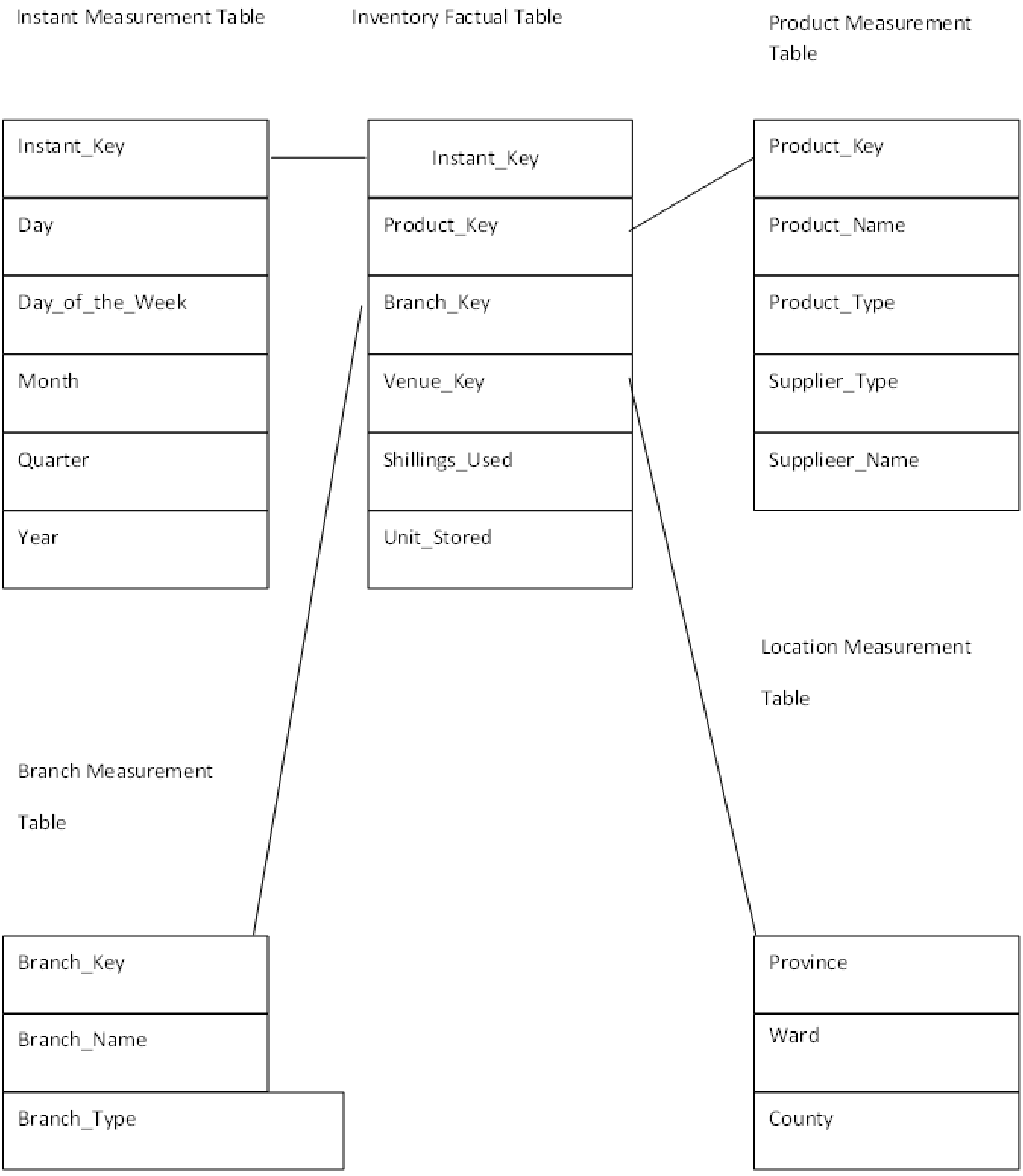


Fig 2.3 **Example of Explanation of Star Schema**

Snowflake Schema: The snowflake schema is different from the star schema model, where few of the dimension tables are normalized where splitting takes place giving rise to an extra table.

A resulting schema graph forms a shape similar to a snowflake.

Example: 2.4 Snowflake Schema a snow flake schema for a furniture inventories are shown in the figure. The distinction of the two schemas is the definition of Dimension Tables. The single dimension table for product in the star schema is normalized in the snowflake schema resulting in new product and supplier tables

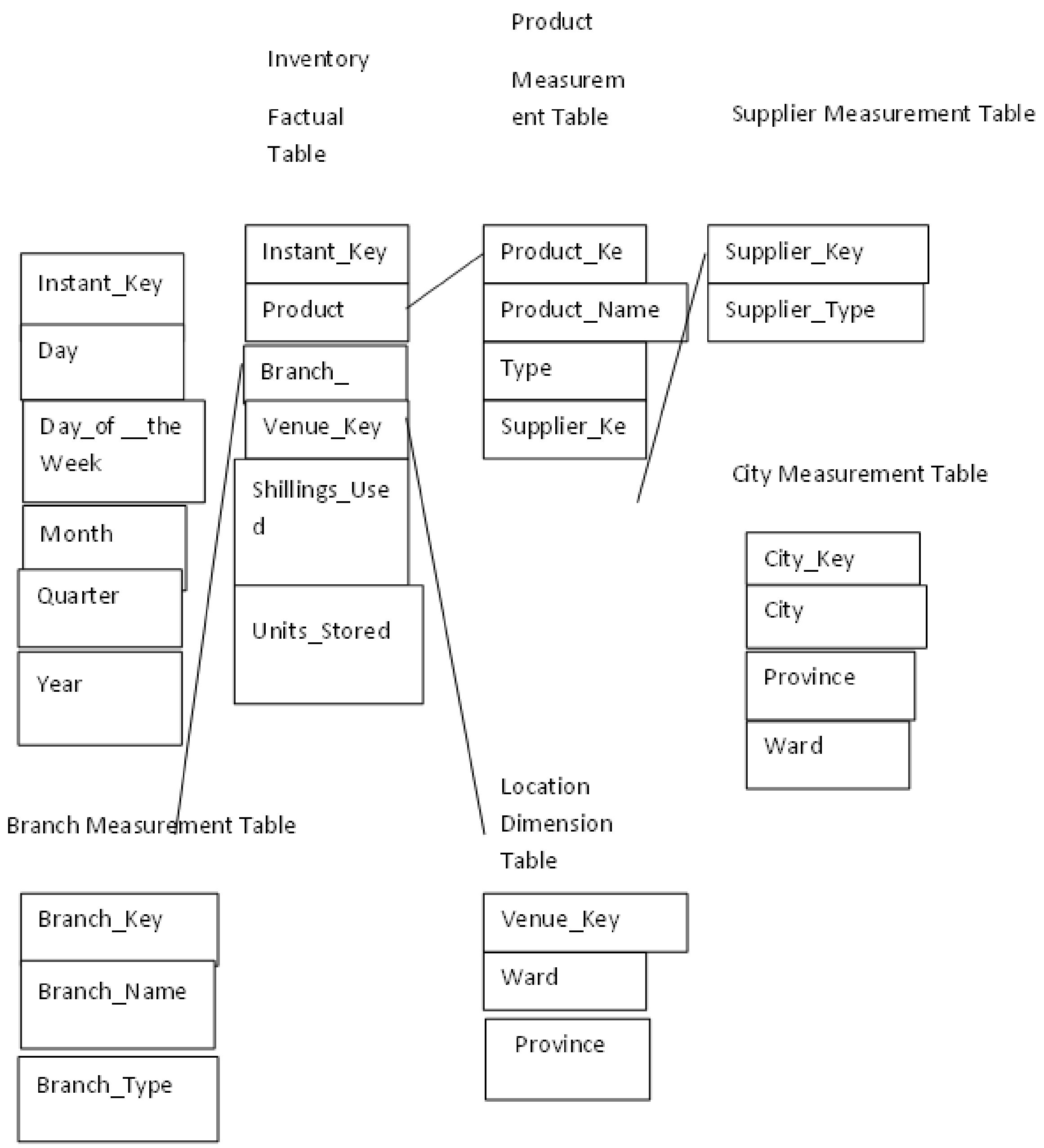


Fig 2.4 Explanation and Example of Snow Flake

The product espect diagram now contains the attributes product_ key , product_name, supplier _key, where supplier _key is linked to the supplier dimension table.

Snowflake Schema of an Inventory Data Warehouse

Fact constellations: There are some advanced applications which include many factual diagram to share dimension tables. This Kind of schema can be considered as a collection of stars, and hence is called a galaxy schema or a fact constellation.

Example: 3 A fact constellation schema is shown in Figure. 2.5

This schema contains two fact tables, sales and shipping. The sales table definition is identical to that of the star schema. The shipping table has five dimensions or key product key, time_key, shipper key to location, and shipper key from location and two measures Shillings Cost and Units _of _products shipped .A fact constellation schema allows dimension tables to be shared between fact tables. For example the dimension tables for time, product, and location are shared between the inventory and shipping of fact tables.

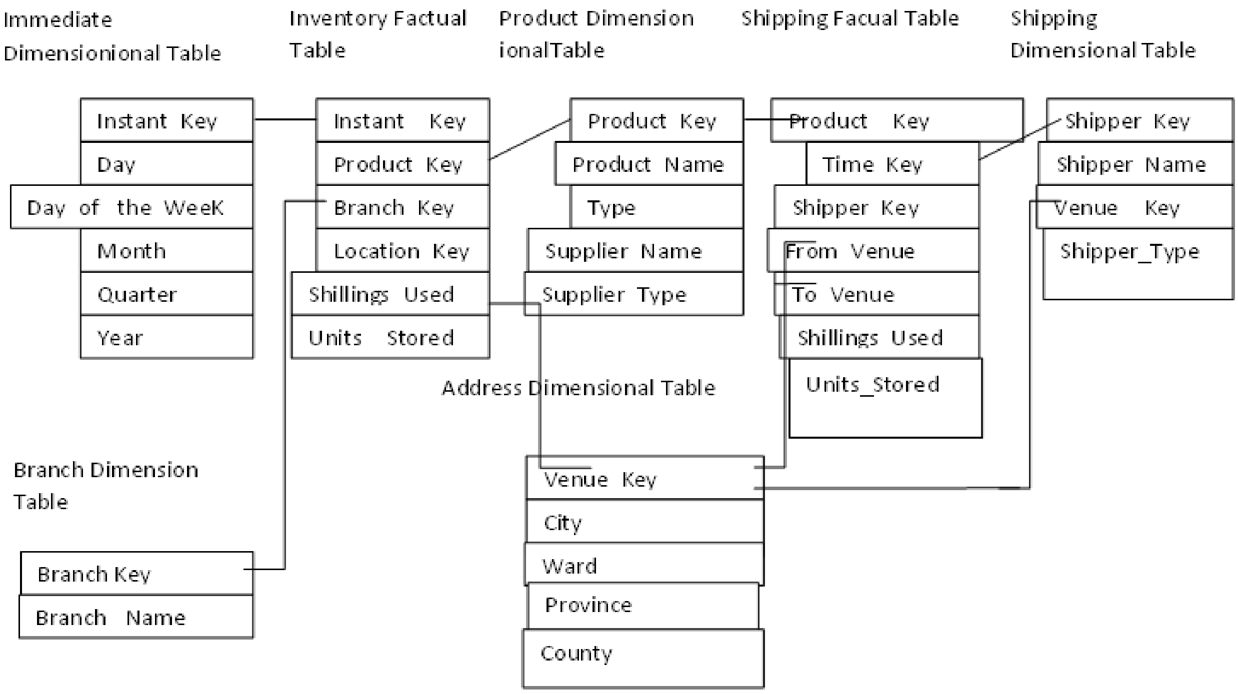


Fig 2.5 Fact Constellation Schema

OLAP Architecture

Few Examples of OLAP

1. Study of sales (fact) or inventory (fact) of a product (dimension) over years (dimension) either in the same demographic area of the same region or different regions of same location or different regions of different demographic area.
2. How many members (facts)have opened either a savings account or credit card debit card accounts in Kenya branch (measure)over a period(measure).
3. What is the total amount spent (fact) for keeping an inventory of a specified product (dimension) in a particular grocery shop (dimension) for a particular period of time (dimension)

OLAP Servers

- Relational OLAP (ROLAP)
- Multidimensional OLAP (MOLAP)
- Hybrid OLAP (HOLAP)
- Specialized SQL Servers

Relational OLAP

ROLAP servers are put and placed between relational back-end server and client front-end tools. To store and manage warehouse data, ROLAP uses relational or extended-relational DBMS. ROLAP includes the following characteristics.

Accumulation/ aggregation navigation logic is implemented.

Each DBMS back end is optimized.

Tools and services are provided as an additional attachment.

Multidimensional OLAP

As the name signifies MOLAP uses multidimensional storage which can be two or more array-based configuration for many dimensional views of data. With multidimensional data stores, there is a possibility of having low storage utilization. Utilizing storage space will be lowest utilizing storage having a sparse data set. Thus most MOLAP server use two levels of data storage representation to handle dense and sparse data sets.

Hybrid OLAP

Hybrid OLAP is a combination of both ROLAP and MOLAP. This kind of presentation is used when we require faster computation and higher scalability of ROLAP. Hybrid OLAP servers allows to store the high voluminous of detailed information. The aggregations and assimilations of highbred data are conducted individually in MOLAP store.

Specialized SQL Servers

Specialized SQL servers provide advanced query language and query processing support for SQL queries over star and snowflake schemas in a read-only environment.

Differences between MOLAP and ROLAP

Points	**ROLAP**	**MOLAP**
1	Retrieving information is comparatively a slow process.	Retrieving Information is a fast process.
2	Uses relational table. It is best for experienced users.	To store data sets sparse array is used. It is best for in experienced users.
3.	DBMS facility is strong	DBMS facility is weak.

Warehouse Designing:

Warehouse designing may be built using a top down approach a bottom up approach

Or a combination of both.

a) Flows top to bottom: - It starts with overall design and planning. It is useful in cases where the technology is mature and well known and where the business problems that must be solved are clear and well understood.
b) Flows Bottom to up:-It starts with prototypes and experiments. This is useful in the early stage of business modelling and technology development. It allows an organization to move forward with less amount of spending and also helps to evaluate the technological benefits before the final commitments.
c) Combined Approach:-An organization can exploit the planned and strategic nature of the top down approach while retaining the rapid implementation and opportunities application of the bottom up approach.

From the view point of Engineering the structure, construction and design follows:

1) Planning 2) Feasibility Study 3) Problem Definition 4) Warehouse Design 5)Data Integration 6) Testing 7)Development of the Warehouse.

Two methods are used to develop large software systems.

a) Water Fall Method: It performs a structured and systematic analysis at each step before proceeding to the next, which is like a water fall falling from one step to the next.

b) Spiral Method: It involves the rapid generation of increasingly functional systems with short intervals between successive releases. This is the best choice for developing data marts as the turnaround time is short and modifications can be done easily and quickly.

Conclusion:

A data Warehouse is the best supporter of decision making as it supports object oriented entity, integrated data. It covers and collects time vibrant, and nonvolatile data. Assorted factors distinguish data warehouse from performing operational on a database. Warehouse containing data often adopts a three tier architecture.

It also specifies back end tools and utilities for populating and refreshing the warehouse. Metadata are data defining the warehouse objects. A modeling on data items is used for the design of corporate data warehouse and departmental data marts must have many dimensions. Any cuboids and lattice represents a data cube referring to various degree of summarization of the given dimensional data.

Concept hierarchies organize the values of attributes or dimensions into gradual abstraction levels. They are useful in mining at multiple abstraction levels.

Online analytical processing can be performed in data warehouse/marts using the multidimensional data model. A relational OLAP, a multidimensional OLAP or a Hybrid OLAP implementation can be adopted by many OLAP servers.

References:

[1] T. Dasu and T Johnson. Explanatory Data mining and Data Cleaning John Wiley @ Sons. 2003.

[2] T.Dasu, T. Jhonson, S Muthukrishnan and V.shkapenyuk Mining Database Structure or how to build quality browser.Madison,WI, June 2002

[3] M.Dash,H.Liu and J.Yao Dimensionality reduction of unsupervised data. Newport Beach, CA, IEEE, Computer Society,1997

[4] D.Dasgupta and N.S. Majumdar. Anomaly detection in multidimensional data Using negative selection algorithm. Washington DC.,2002

[5] Data Warehousing in the Real world Sam Anahory & Dennis Murray,Pearson,Educstion

[6] Data Warehousing, Data Mining and OLTP Alex Berson,1997,MGH

[7] Data Mining Concepts and Technique Jiawei Han,Micheline Kamber Third Edition

[8] International Journal of Computer Applications (0975 – 8887) Volume 141 – No.8, May 2016 22 0 Data Mining: Analysis of student database using Classification Techniques K. Sumathi, PhD Assistant Professor, Department of CS and IT, Kalasalingam University, Kirishnan Kovil. S. Kannan, PhD Associate Professor, Department of Computer Applications, MKU, Madurai. K. Nagarajan Chief Architect of Business Intelligence, Tata Consultancy Services, Chennai

[9] International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.2, No.5, September 2012 DOI: 10.5121/ijdkp.2012.2502 13 Data Mining and Its Applications for Knowledge Management : A Literature Review from 2007 to 2012 Tipawan Silwattananusarn1 and Assoc.Prof. Dr. KulthidaTuamsuk2 1 Ph.D. Student in Information Studies Program, Khon Kaen University, Thailand 2Head, Information & Communication Management Program, Khon Kaen University, Thailand

[10] ASEE 2014 Zone I Conference, April 3-5, 2014, University of Bridgeport, Bridgpeort, CT, USA. Data Mining and Warehousing Ali Radhi Al Essa School of Engineering University of Bridgeport Bridgeport, CT, United States aalessa@my.bridgeport.edu Bach, Christian School of Engineering University of Bridgeport Bridgeport, CT, United States cbach@bridgeport.edu

[11] Review Article Data Mining for the Internet of Things: Literature Review and Challenges Feng Chen,1,2 Pan Deng,1 Jiafu Wan,3 Daqiang Zhang,4 Athanasios V. Vasilakos,5 and Xiaohui Rong6 sources and clean the noise from data; extract some parts of data into data mining system; preprocess the data to facilitate the data mining. Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2015, Article ID 431047, 14 pages http://dx.doi.org/10.1155/2015/431047

[12] M.Ankerst ,M.Breuning,H-P. Kerigel Visual classification an interactive approach to decision tree construction. In proc.1999 int. conf. Knowledge Discovery and Data Mining