Integrated IS Final Revision part 2 by Eng Ihap EL-Galaly

List some of Data analysis problems ?

- The same data found in many different systems
- Data is suited for operational systems
- Data are "volatile"
- Heterogeneous sources
- Data quality is bad

Define Data warehouse and explain why it's separated about operational data?

- A data warehouse is a **database**, which is kept separate from the organization's operational database.
- There is **no frequent updating** done in a data warehouse.
- It possesses **consolidated historical data**, which helps the organization to analyze its business.
- A data warehouse helps executives to organize, understand, and use their data to take strategic decisions.

A data warehouses is kept separate from operational databases due to the following reasons:

- **An operational database** is constructed for well-known tasks and workloads such as searching records, indexing, etc. In contract, data warehouse queries are often complex, and they present a general form of data.
- **An operational database** query allows to read and modify operations, while an OLAP query needs only read only access of stored data.
- **An operational database** maintains current data. On the other hand, a data warehouse maintains historical data.

What is a Data Warehouse?

- Data warehouse provides architectures and tools for business executives to systematically organize, understand, and use their data to make strategic decisions.

Data warehouse systems are built by integrating data from multiple heterogeneous sources and, in addition to centralization, performs the task of structuring data, supporting analytical reporting and structuring decision-making.

What are the Data Warehousing Features?

a subject-oriented, integrated, time-variant, and non- volatile collection of data in support of management's decision-making process.

Data Warehousing Features –

Subject-oriented

▪ A data warehouse focuses on the modeling and analysis of data for decision makers.

▪ data warehouses typically provide a simple and concise view around subject issues by excluding data that are not useful in the decision support process.

Integrated

▪ A data warehouse integrates various heterogeneous data sources like RDBMS, flat files, and online transaction records.

▪ It requires performing data cleaning and integration during datawarehousing to ensure consistency in naming conventions, attributes types, among different data sources.

Time-variant: Data are stored as a snapshots or views to provide information from a historical perspective

Non-volatile

▪ A data warehouse is always a physically separate store of data transformed from the application data found in the operational environment.

▪ Due to this separation, a data warehouse does not require transaction processing, recovery, and concurrency control mechanisms.

List some of Data warehouse Name ?

- Decision Support System (DSS)
- Executive Information System
- Management Information System
- Business Intelligence Solution
- Analytic Application
- Data Warehouse

What are the Types of Data Warehouse Applications?

1. Information Processing – A data warehouse allows to process the data stored in it. The data can be processed by means of querying, basic statistical analysis, tables, charts, or graphs.

2. Analytical Processing – A data warehouse supports analytical processing of the information stored in it. The data can be analyzed by means of basic OLAP operations, including slice-anddice, drill down, drill up, and pivoting.

3. Data Mining – Data mining supports knowledge discovery by finding hidden patterns and associations, constructing analytical models, performing classification and prediction. These mining results can be presented using the visualization tools.

What are the Benefits of Data Warehouse?

1.Understand business trends and make better forecasting decisions.
2.Data Warehouses are designed to perform well enormous amounts of data.
3.The structure of data warehouses is more accessible for end-users to navigate, understand, and query.
4.Queries that would be complex in many normalized databases could be easier to build and maintain in data warehouses.
5.Data warehousing is an efficient method to manage demand for lots of information from lots of users.
6.Data warehousing provide the capabilities to analyze a large amount of historical data.

List the advantages and challenges of adapting data warehouse?

▪ **Advantages of adopting data warehouses include:**

– Integrating data from multiple sources;
– Performing new types of analytical analysis;
– Reducing costs to access historical data;
– Improving turnaround time for analysis and reporting;
– Sharing data and allowing others to easily access data;

challenges resulting from the adoption of data warehouses including:

– Time consuming preparation and implementation;
– High maintenance costs;
– Limited use due to confidential information;
– Data ownership and data security;
– Underestimation of ETL processing time;
– Inability to capture the required data;

List some Needs of Data warehouse?

1. Business User: require a data warehouse to view summarized data from the past. Since these people are non-technical, the data may be presented to them in an elementary form.

2. **Store historical data:** Data Warehouse is required to store the time variable data from the past. This input is made to be used for various purposes.

3. Make strategic decisions: Some strategies may be depending upon the data in the data warehouse. So, data warehouse contributes to making strategic decisions.

4. For data consistency and quality: Bringing the data from different sources at a commonplace, the user can effectively undertake to bring the uniformity and consistency in data.

5. High response time: Data warehouse has to be ready for somewhat unexpected loads and types of queries, which demands a significant degree of flexibility and quick response time.

Describe the classes of data ?

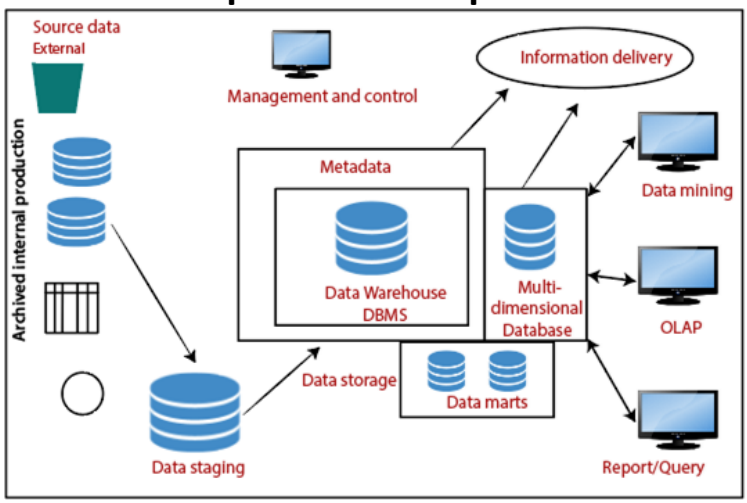
1. Structured Data: Have a predefined model, which organizes data into a form that is easy to store, process, retrieve and manage [e.g., relational data].

2. Unstructured Data: Opposite of structured data [e.g., Flat binary files containing text, video or audio].

3. Dynamic Data: Data that changes relatively frequently [e.g., office documents and transactional entries in a financial database]

4. Static Data: Opposite of dynamic data [e.g., Medical imaging data from MRI or CT scans]

Draw and explain the component of Data Warehouse ?



 Source Data refer to any electronic repository of information that contains data of interest for management use or analytics.

Data needs to be passed from these systems to the data warehouse either on:

– transaction-by-transaction basis for real-time data warehouses or

– regular cycle (e.g., daily/weekly) for offline data warehouses.

 Staging Area The data staging layer hosts the ETL processes that extract, integrate, and clean data from operational sources to feed the data warehouse layer.

– In a three-layer architecture, ETL processes feed the Reconciled Data Layer - a single, detailed, comprehensive, top-quality data source—that feeds the data warehouse.

 Data Warehouses/Data Storage (in the middle)

– This element not only stores and manages the data; and keeps track of data using the metadata repository.

 Information Delivery (on the right)

– That consists of all the different ways of making the information from the data warehouses available to the users.

– subscribing for data warehouse files and having it transferred to one or more destinations

Metadata

Metadata in a data warehouse is equal to the Data Dictionary or the Data Catalog in a DBMS.

– Data Dictionary keeps the data about the:

o logical data structures,

o records and addresses, the information about the indexes.

Source data coming into the data warehouses may be grouped into four broad categories list them ?

- **Production Data:** comes from the different operating systems of the enterprise. Based on the data requirements in the data warehouse, we choose segments of the data from the various operational modes.
- **Internal Data:** In each organization, the client keeps their "private" spreadsheets, reports, customer profiles, and sometimes even department databases. This is the internal data, part of which could be useful in a data warehouse.
- **Archived Data:** Operational systems are intended to run the current business. In every operational system, we periodically take the old data and store it in achieved files.
- **External Data:** Most executives depend on information from external sources for a large percentage of the information they use. They use statistics associating to their industry produced by the external department.

What are the Data Warehouse Architecture properties are necessary for a data warehouse system?

Separation: analytical and transactional processing should be kept apart as much as possible.
Scalability: Hardware and software architectures should be simple to upgrade the data volume, which must be managed and processed, and the number of user's requirements, which must be met, progressively increase.
Extensibility: The architecture should be able to perform new operations and technologies without redesigning the whole system.
Security: Monitoring accesses are necessary because of the strategic data stored in the data warehouses.
Administrability: Data Warehouse management should not be complicated.

Draw the Goals of Metadata ?

Figure 18. Goals of Metadata

What in details the phases of ETL?

1. Extraction

 Relevant data is obtained from sources in the extraction phase using either:

– Static extraction when a data warehouse needs populating for the first time,

– Incremental extraction to update data warehouses regularly and capture the changes applied to source data since the latest extraction,

2. Cleansing

This phase is crucial in DW system because it improves data quality. The following list includes the most frequent mistakes and inconsistencies that make data "dirty ":

Duplicate data, Inconsistent values, Missing data, Unexpected use of fields Impossible or wrong values

– Specific Dictionaries to rectify typing mistakes and to recognize synonyms

– Rule-based Cleansing to enforce domain-specific rules and define appropriate associations between values.

3. Transforming

 Transformation is the core of the Reconciliation Phase (as shown in Figure

 It converts data from operational source format into a specific data warehouse format.

– If you implement a three-layer architecture, this phase outputs the Reconciled Data Layer.

4. Loading

– When we complete the structure and construction of the data warehouse and go-live for the first time -> we do the initial loading of the information into the data warehouse storage.

♣ **With all these uses of metadata, it also has several challenges list them?**

– Metadata in a big organization is scattered across the organization. This metadata is spread in spreadsheets, database, and applications;

– Metadata could be presented in text files or multimedia files. To use this data for information management solutions, it must be correctly defined;

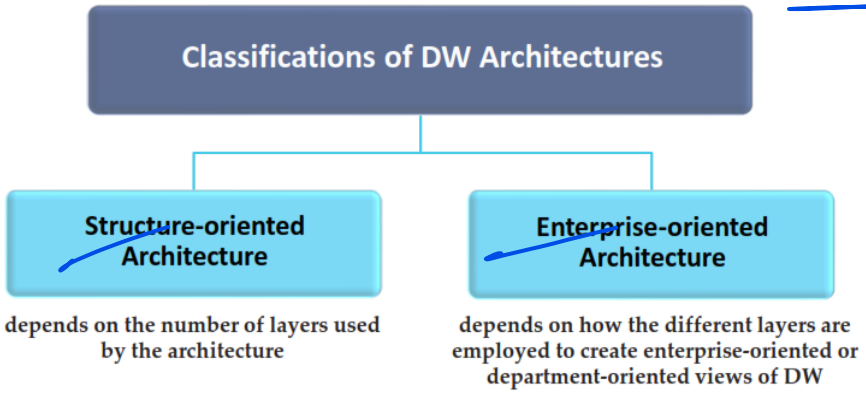
– There are no industry-wide accepted standards. Data management solution vendors have a narrow focus;

– There are no easy and accepted methods of passing metadata.

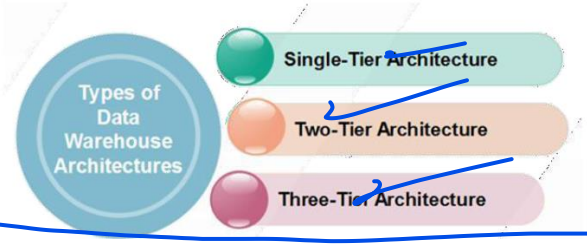
In the context of a DW project the metadata should keep set of information list them?

– The structure of the data according to the view of the programmer;
– The structure of the data according to view of DSS analysts;
– The data sources of the DW;
– The transformation undergone by the data at the moment of its migration to DW;
– The data model;
– The relationship between the data model and the DW;
– The history of data extraction.

Explain in details the Classifications of Data Warehouse Architectures?



▪ There are mainly Three types of Structured-oriented DW Architecture.



▪ **Single-tier Architecture is not periodically used in practice.**
– Its purpose is to minimize the amount of data stored.
– To reach this goal; it removes Data Redundancies.
– In this method, Data Warehouses are VIRTUAL.
– This means that the DW is implemented as a Multidimensional View of operational data created by specific Middleware, or an Intermediate Processing Layer.

Two-tier Architecture to highlight a separation between physically available sources and data warehouses.

▪ **Consists of FOUR subsequent data flow stages:**

1. Source Layer -a DW system uses a Heterogeneous Source of data.
– That data is stored initially to corporate relational databases or legacy databases
2. Data Staging The data stored to the source should be:

– extracted, cleansed to remove inconsistencies and fill gaps, and
– integrated to merge heterogeneous sources into one standard schema.
3. Data Warehouse Layer
– Information is saved to one logically centralized individual repository: a Data Warehouse.
– The data warehouses can be directly accessed, but it can also be used as a source for creating Data Marts.
4. Analysis
– In this layer, integrated data is efficiently, and flexible accessed to:
o issue Reports, dynamically analyze Information, and simulate Hypothetical Business Scenarios.

▪ **The Three-tier Architecture** consists of the Source Layer, that is containing multiple source system, including:
1. The Reconciled Layer which sits between the source data and data warehouse, and
2. The Data Warehouse Layer containing both data warehouses and data marts.

What are is Advantage and Disadvantage Three-Tier Architecture?

▪ **The main Advantage of the Reconciled Layer is that:**
1. It creates a Standard Reference Data Model for a whole enterprise.
2. At the same time, it Separates The Problems of Source Data Extraction and Integration from those of data warehouse population.
3. This layer may also be used to Accomplish some operational tasks

▪ **A Disadvantage of this structure is:**
1. the Extra File Storage Space used through The Extra Redundant Reconciled Layer
2. It also makes the Analytical Tools a little further away from being Real-time.

DW usually have a Three-level (Tier) Architecture explain in details ?

Bottom-Tier

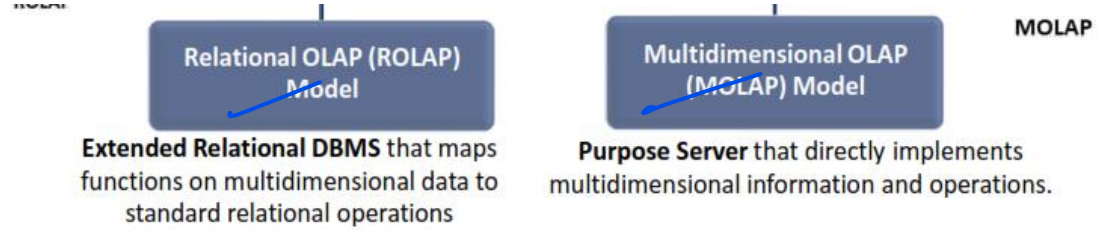
▪ A Bottom-tier that consists of the DW Server, which is always an RDBMS.

– It includes Data Marts and a Metadata Repository.

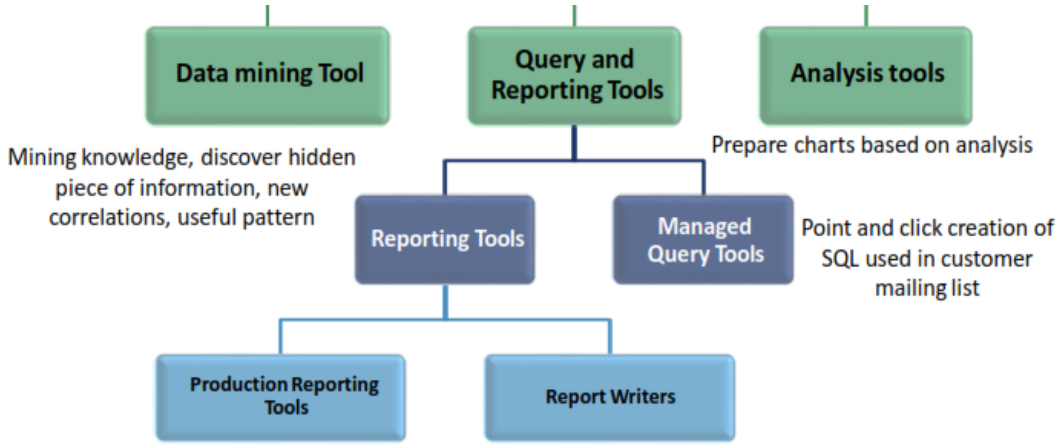
▪ Data from operational DB and external sources are extracted using Application Program Interfaces called a Gateway.

– A Gateway is provided by the DBMS and allows customer programs to generate SQL code to be executed at a server.

A Middle-tier which consists of an OLAP Server for fast querying of the data warehouse.



A Top-tier that contains front-end tools for displaying results provided by OLAP



There are Five types of Technical Enterprise/Department-oriented view of DW Architecture describe them in details ?

1. Independent Data Marts Architecture

different data marts are separately designed and built in a non-integrated approach.

▪ Limitations of this architecture include:

– Department-based, that it lacks enterprise focus.

– No Consistency or data integration

– Do not share dimensions

– Data is sourced independently.

2. Bus Architecture

like the preceding architecture, with one important difference.

– Data is organized in Dimension Tables, Facts Tables, Flat Files etc. that are stored in Data Marts

– A basic set of conformed dimensions: or analysis dimensions that preserve same meaning during all the facts they belong to

– Dimensions are derived by a careful analysis of the main enterprise processes

– Adopted and shared as a common design guideline.

3. Hub-and-Spoke Architecture

one of the most used in medium to large contexts.

▪ Atomic, normalized data is stored in a Reconciled Layer that feeds a set of Data Marts containing summarized data in Multidimensional Form.

4. Centralized Architecture

can be seen as a particular implementation of the hub-and-spoke architecture.

– where the Reconciled Layer and the Data Marts are collapsed into a Single Physical Repository.

5. Federated Architecture

Is distributing information by organizational areas And adopted in Dynamic Contexts where pre-existing data warehouses/data marts are to be integrated to provide a single, cross organization decision support environment.

What is the meaning of Data Warehouse Modelling and what are the reason for modeling ?

Data Warehouse Modeling is the process of designing the schemas of the detailed and summarized information of the Data Warehouse.

The GOAL of DW modeling is to develop a schema describing the reality, or at least a part of the fact, which the data warehouse is needed to support.

▪ **Data warehouse modeling is an essential stage of building a data warehouse for TWO main reasons:**

1. Firstly, through the Schema, DW clients can visualize the relationships among the warehouse data, to use them with greater ease.
2. Secondly, a Well-designed Schema allows an effective data warehouse structure to emerge, to help decrease the cost of implementing the warehouse and improve the efficiency of using it.

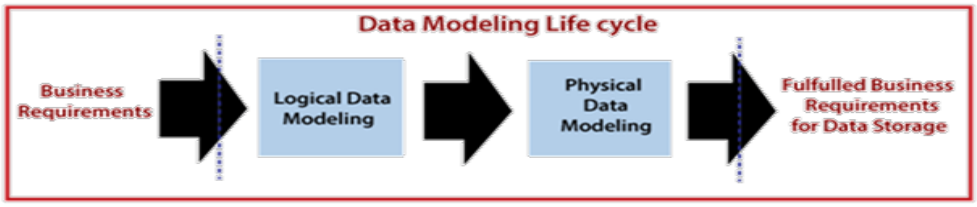
Describe the various levels of summarization of Data Warehouse Modelling?

- **Older detail data** is stored in some form of mass storage, and it is infrequently accessed and kept at a level detail consistent with current detailed data.
- **Lightly summarized data** is data extract from the low level of detail found at the current, detailed level and usually is stored on disk storage. When building the data warehouse have to remember what unit of time is summarization done over and also the components or what attributes the summarized data will contain.
- **Highly summarized data** is compact and directly available and can even be found outside the warehouse.
- **Metadata** is the final element of the data warehouses and is really of various dimensions in which it is not the same as file drawn from the operational data

explain the data modeling life cycle?

It is a straightforward process of transforming the business requirements to fulfill the goals for storing, maintaining, and accessing the data within IT systems.

– The result is a Logical and Physical Data Model for an Enterprise Data Warehouse.



- **A Logical Data Model** defines the information in as much structure as possible, Without Observing how they will be physically achieved in the database.
- **The primary Objective of logical data modeling** is to document the business data structures, processes, rules, and relationships by A Single View - the Logical Data Model.
- **Physical Data Model** describes how the model will be presented in the database.

– **A physical database model** demonstrates all table structures, column names, data types, constraints, primary key, foreign key, and relationships between tables.

– **The Purpose of Physical Data Modeling** is the mapping of the logical data model to the physical structures of the RDBMS system hosting the data warehouse.

List the Characteristics of the conceptual data mode?

- It contains the essential entities and the relationships among them.
- No attribute is specified.
- No primary key is specified.

♣ The Only Data Shown via the Conceptual Data Model is the ENTITIES that define the data and the relationships between those entities.

♣ No Other Data, as shown through the conceptual data model.

List the Features of the conceptual data mode?

- It involves all entities and relationships among them.
- All attributes for each entity are specified.
- The primary key for each entity is stated.
- Referential Integrity is specified (FK Relation).
- The phase for designing the logical data model which are as follows:
- Specify primary keys for all entities.
- List the relationships between different entities.
- List all attributes for each entity.
- Normalization.
- No data types are listed

List the Characteristics and steps of the physical data mode?

Characteristics of a Physical Data Model:

- Specification all tables and columns.
- Foreign keys are used to recognize relationships between tables.

The steps for physical data model:

- Convert entities to tables.
- Convert relationships to foreign keys.
- Convert attributes to columns.

What are the Nine Steps DW Design Methodology?

Step 1: Choosing Process (function) – Refers to the subject matter of a particular data mart, for example: a Bill Payment Process

Step 2: Choosing The Grain – Decide what a record of the fact table is to represent, i.e.. the grain. For example, the grain is a single Payment

Step 3: Identifying and conforming the dimensions – Dimensions set the context for asking questions about the facts in the fact table. For example, Who made the Bill Payment

Step 4: Choosing the Facts – Facts should be numeric and additive.

Step 5: Storing pre-calculations in the fact table – Once the facts have been selected each should be re-examined to determine whether there are opportunities to use pre-calculations. (denormalization)

Step 6: Rounding out the dimension tables – What properties to include in dimension table to best describe it. Should be intuitive and understandable

Step 7: Choosing the duration of the database – How long to keep the data for

▪ Step 8: Tracking slowly changing dimensions –

– Type 1: where a changed dimension attribute is overwritten

– Type 2: where a changed dimension attribute causes a new dimension record to be created

Step 9: Deciding the query priorities and the query modes –

– Consider physical decision issues

– Indexing for performance, Indexed Views, partitioning, physical sort order,

– Storage, backup, security

What is the meaning of OLAP ?

▪Online Analytical Processing (OLAP) is a classification of software technology which authorizes analysts, managers, and executives to gain insight into information

▪OLAP implement the Multidimensional Analysis of business information and support the capability for complex estimations, trend analysis, and sophisticated data modeling.

▪ OLAP enables end-clients to perform ad hoc analysis of record in multiple dimensions, providing the insight and understanding they require for better decision making.

Why Uses OLAP?

1. The first is to provide business users with a data model more instinctive to them than a tabular model. This model is called a Dimensional Model.
2. The second purpose is to enable fast query response that is usually difficult to achieve using tabular models.

Compare between OLAP and OLTP?

OLAP Online Analytical Processing	OLTP Online Transaction Processing
▪Describes processing at warehouse ▪ Mostly reads ▪ Queries long, complex ▪ Gb-Tb of data ▪ Summarized, consolidated data ▪ Decision-makers , analysts as users	▪Describes processing at operational sites ▪ Mostly updates ▪ Many small transactions ▪ Mb-Tb of data ▪ Raw data ▪ Clerical users ▪ Up-to-date data ▪ Consistency, recoverability critical

Define Dimensional Modeling, it's element and state it's purposes ?

▪ Every dimensional model (DM) is composed of ONE table with a composite primary key, called the FACT table, and a set of smaller tables called DIMENSION tables.

▪ A Data Cube, such as Sales, allows data to be modeled and viewed in multiple dimensions.

– Dimensional table, such as item (item_name, brand, type), or time (day, week, month, quarter, year)

– Fact table contains measures (such as dollar_sold) and keys to each of the related dimension tables.

Elements of Dimensional Modeling

▪ Fact: It is a collection of associated data items, consisting of measures and context data. It typically represents business items or business transactions.

▪ Dimensions: It is a collection of data which describe one business dimension. Dimensions decide the contextual background for the facts, and they are the framework over which OLAP is performed.

▪ Measure: It is a numeric attribute of a fact, representing the performance or behavior of the business relative to the dimensions.

▪ **The purposes of dimensional modeling are:**

1. To produce Database Architecture that is easy for end-clients to understand and write queries.
2. To maximize the efficiency of queries. It achieves these goals by minimizing the number of tables and relationships between them.

What is Multidimensional Data Model?

♣ Each dimension has a table related to it, called a Dimensional Table, which describes the dimension further.

♣ A Multidimensional Data Model is organized around a central theme

What are the Advantages and disadvantages of Dimensional Modeling?

Advantages of Dimensional Modeling

▪ **Dimensional modeling is simple:** Dimensional modeling methods make it possible for warehouse designers to create database schemas that business customers can easily hold and comprehend.

▪ **Dimensional modeling promotes data quality:** The star schema enable warehouse administrators to enforce referential integrity checks on the data warehouse.

▪ **Performance optimization is possible through aggregates:** As the size of the data warehouse increases, performance optimization develops into a pressing concern.

Disadvantages of Dimensional Modeling

▪ To maintain the integrity of fact and dimensions, loading the data warehouses with a record from various operational systems is complicated.

▪ It is severe to modify the data warehouse operation if the organization adopting the dimensional technique changes the method in which it does business.

Define fact table and list it's Characteristics?

Fact tables are used to data facts or measures in the business. Facts are the numeric data elements that are of interest to the company.

Characteristics of the Fact table

– The fact table includes numerical values of what we measure.

– Each fact table includes the keys to associated dimension tables. These are known as foreign keys in the fact table.

– Fact tables typically include a small number of columns.

– When it is compared to dimension tables, fact tables have many rows.

Define Dimension table and list it's Characteristics?

Dimension Tables establish the context of the facts. Dimensional tables store fields that describe the facts.

Characteristics of the Dimension table

– Dimension tables contain the details about the facts. That, as an example, enables the business analysts to understand the data and their reports better.

– The dimension tables include descriptive data about the numerical values in the fact table. That is, they contain the attributes of the facts. analysis function might include attributes such as time, marketing region, and product type.

– Since the record in a dimension table is denormalized, it usually has many columns. The dimension tables include significantly fewer rows of information than the fact table.

– The attributes in a dimension table are used as row and column headings in a document or query results display.

Define data cube, OLAP Cube?

A Data Cube enables data to be modeled and viewed in Multiple Dimensions.

– A Multidimensional Data Model is organized around a central theme, like sales and transactions.

An OLAP Cube is a data structure that combines data from disparate data sources and allows fast analysis of data.

 The arrangement of data into cubes overcomes a limitation of relational databases.

 It consists of numeric Facts called measures which are categorized by Dimensions.

Define Dimensional Cuboids

♣ Let us suppose that we would like to view our sales data with an additional fourth dimension, such as a supplier.

– For example, the 4-D cuboid in the Figure is the base cuboid for the given time, item, location, and supplier Dimensions.

♣ In data warehousing, the data cubes are n-dimensional.

♣ The n-D Base Cube which holds the lowest level of summarization is called a Base Cuboid.

♣ The topmost 0-D cuboid, which holds the highest level of summarization, is known as the Apex Cuboid.

Choose the correct answer

A ……..is a database, which is kept separate from the organization's operational database.

data warehouse

Data mart

Database

There is no frequent updating done in a …….

data warehouse

Data mart

Database

……..It possesses consolidated historical data, which helps the organization to analyze its business.

data warehouse

Data mart

Database

A …… helps executives to organize, understand, and use their data to take strategic decisions.

data warehouse

Data mart

Database

A …… does not require transaction processing, recovery, and concurrency controls.

data warehouse

Data mart

Database

…….. systems are built by integrating data from multiple heterogeneous sources and, in addition to centralization

data warehouse

Data mart

Database

Data Warehousing Features - ……..Data are stored as a snapshots or views to provide information from a historical perspective
Time-variant
Non-volatile
Integrated
Subject-oriented

Data Warehousing Features - ……..Every key structure in the data warehouse contains, either implicitly or explicitly, an element of time
Time-variant
Non-volatile
Integrated
Subject-oriented

Data Warehousing Features - ……..A data warehouse is always a physically separate store of data transformed from the application data found in the operational environment.
Time-variant
Non-volatile
Integrated
Subject-oriented

Data Warehousing Features –……..A data warehouse integrates various heterogeneous data sources like RDBMS, flat files, and online transaction records.
Time-variant
Non-volatile
Integrated
Subject-oriented

Data Warehousing Features –……..It requires performing data cleaning and integration during data warehousing to ensure consistency in naming conventions, attributes types
Time-variant
Non-volatile
Integrated
Subject-oriented

Data Warehousing Features - ……A data warehouse focuses on the modeling and analysis of data for decision makers.
Time-variant
Non-volatile
Integrated
Subject-oriented

Data Warehousing Features - ……data warehouses typically provide a simple and concise view around subject issues by excluding data that are not useful in the decision support process.
Time-variant
Non-volatile
Integrated
Subject-oriented

……– A data warehouse allows to process the data stored in it. The data can be processed by means of querying, basic statistical analysis, tables, charts, or graphs.
Information Processing
Analytical Processing
Data Mining

A data warehouse supports ……..of the information stored in it. The data can be analyzed by means of basic OLAP operations, including slice-and dice, drill down, drill up, and pivoting.
Information Processing
Analytical Processing
Data Mining

……– supports knowledge discovery by finding hidden patterns and associations, constructing analytical models, performing classification and prediction. These mining results can be presented using the visualization tools.
Information Processing
Analytical Processing
Data Mining

……: require a data warehouse to view summarized data from the past. Since these people are non-technical, the data may be presented to them in an elementary form.

Business User
Store historical data
Make strategic decisions
data consistency and quality

……: Data Warehouse is required to store the time variable data from the past. This input is made to be used for various purposes.

Store historical data
Make strategic decisions
data consistency and quality
High response time

……: Some strategies may be depending upon the data in the data warehouse. So, data warehouse contributes to making strategic decisions.

Store historical data
Make strategic decisions
data consistency and quality
High response time

……: Bringing the data from different sources at a commonplace, the user can effectively undertake to bring the uniformity and consistency in data.

Store historical data
Make strategic decisions
data consistency and quality
High response time

……: Data warehouse has to be ready for somewhat unexpected loads and types of queries, which demands a significant degree of flexibility and quick response time.

Store historical data
Make strategic decisions
data consistency and quality
High response time

……..is a collection of data that is used primarily in organizational decision making
data warehouse
Decision Support Database
all of the above
non of the above

……..: Have a predefined model, which organizes data into a form that is easy to store, process, retrieve and manage
Structured Data
Unstructured Data
Dynamic Data
Static Data:

……..: Opposite of structured data [e.g., Flat binary files containing text, video or audio].
Structured Data
Unstructured Data
Dynamic Data
Static Data:

……..: Data that changes relatively frequently [e.g., office documents and transactional entries in a financial database]
Structured Data
Unstructured Data
Dynamic Data
Static Data:

……..Opposite of dynamic data [e.g., Medical imaging data from MRI or CT scans]
Structured Data
Unstructured Data
Dynamic Data
Static Data:

……: comes from the different operating systems of the enterprise. Based on the data requirements in the data warehouse, we choose segments of the data from the various operational modes.

Production Data
Internal Data
Archived Data
External Data

……: In each organization, the client keeps their "private" spreadsheets, reports, customer profiles, and sometimes even department databases.

Production Data
Internal Data
Archived Data
External Data

……: Operational systems are intended to run the current business. In every operational system, we periodically take the old data and store it in achieved files.

Production Data
Internal Data
Archived Data
External Data

……: Most executives depend on information from external sources for a large percentage of the information they use.

Production Data
Internal Data
Archived Data
External Data

Relevant data is obtained from sources in the ……

extraction phase
Cleansing phase
Transformation phase
Loading phase

……..is crucial in DW system because it improves data quality.
extraction phase
Cleansing phase
Transformation phase
Loading phase

……The following list includes the most frequent mistakes and inconsistencies that make data "dirty"
extraction phase
Cleansing phase
Transformation phase
Loading phase

……..is the core of the Reconciliation Phase.
extraction phase
Cleansing phase
Transformation phase
Loading phase

……It converts data from operational source format into a specific data warehouse format.
extraction phase
Cleansing phase
Transformation phase
Loading phase

When we complete the structure and construction of the data warehouse and go-live for the first time
extraction phase
Cleansing phase
Transformation phase
Loading phase

The Traditional Database implements ……approach, which requires complex information filtering and integration processes, and competes for resources with processing at local sources.
query-driven
update-driven

Data Warehousing employs an ……..approach in which information from multiple, heterogeneous sources is integrated in advance and stored in a warehouse for direct querying and analysis
query-driven
update-driven

……..: analytical and transactional processing should be kept apart as much as possible.
Separation
Scalability
Extensibility
Security

……..: Hardware and software architectures should be simple to upgrade the data volume, which must be managed and processed, and the number of user's requirements, which must be met, progressively increase.
Scalability
Extensibility
Security
Administrability

……..: The architecture should be able to perform new operations and technologies without redesigning the whole system.
Scalability
Extensibility
Security
Administrability

……..: Monitoring accesses are necessary because of the strategic data stored in the data warehouses.
Scalability
Extensibility
Security
Administrability

……: Data Warehouse management should not be complicated.
Scalability
Extensibility
Security
Administrability

In Three -Tier Architecture The …….which sits between the source data and data warehouse
Reconciled Layer
Data Warehouse Layer

The main Advantage of ……is that: It creates a Standard Reference Data Model for a whole enterprise.
Reconciled Layer
Data Warehouse Layer

This layer may also be used to Accomplish some operational tasks
Reconciled Layer
Data Warehouse Layer

A Disadvantage of this structure is: the Extra File Storage Space used through The Extra Redundant Reconciled Layer
Reconciled Layer
Data Warehouse Layer

In Three -Tier Architecture The …….containing both data warehouses and data marts.
Reconciled Layer
Data Warehouse Layer

…….depends on the number of layers used by the architecture
Structure-oriented Architecture
Enterprise-oriented Architecture

…….depends on how the different layers are employed to create enterprise oriented or department-oriented views of DW
Structure-oriented Architecture
Enterprise-oriented Architecture

لا يستخدم بكثرة

…… is not periodically used in practice.
Single-tier Architecture
Two-Tier Architecture
The Three-tier Architecture

…… Its purpose is to minimize the amount of data stored.
Single-tier Architecture
Two-Tier Architecture
The Three-tier Architecture

…… In this method, Data Warehouses are VIRTUAL.
Single-tier Architecture
Two-Tier Architecture
The Three-tier Architecture

……its failure to meet the requirement for Separation between analytical and transactional processing.
Single-tier Architecture
Two-Tier Architecture
The Three-tier Architecture

……highlight a separation between physically available sources and data warehouses.
Single-tier Architecture
Two-Tier Architecture
The Three-tier Architecture

……consists of the Source Layer that is containing multiple source system
Single-tier Architecture
Two-Tier Architecture
The Three-tier Architecture

– Information is saved to one logically centralized individual repository:……
a Data Warehouse.
Data Marts.
Meta-data Repositories
Gateway

The data warehouses can be directly accessed, but it can also be used as a source for creating ……
a Data Warehouse.
Data Marts.
Meta-data Repositories
Gateway

local DW

…… are as small, local Data Warehouses replicating the part of a primary DW required for a specific application domain.
a Data Warehouse.
Data Marts.
Meta-data Repositories
Gateway

……store information on sources, access procedures, data staging, users, data mart schema, etc.
a Data Warehouse.
Data Marts.
Meta-data Repositories
Gateway

Data from operational DB and external sources are extracted using Application Program Interfaces called a……..
a Data Warehouse.
Data Marts.
Meta-data Repositories
Gateway

…….is provided by the DBMS and allows customer programs to generate SQL code to be executed at a server.
a Data Warehouse.
Data Marts.
Meta-data Repositories
Gateway

A ……..in The Three-tier Architecture that consists of the DW Server, which is always an RDBMS.
Bottom-tier
Middle-tier
Top-tier

A ……..in The Three-tier Architecture consists of an OLAP Server for fast querying of the data warehouse.
Bottom-tier
Middle-tier
Top-tier

A ……..in The Three-tier Architecture contains front-end tools for displaying results provided by OLAP
Bottom-tier
Middle-tier
Top-tier

Include the warehouse schema, dimension, hierarchies, data mart locations, and contents, etc.
DW Structure
Operational Metadata
System Performance Data
Mapping Information

……..Describes the currency level of the stored data, i.e., active or archived, and warehouse monitoring information, i.e., usage statistics, error reports, audit, etc.
DW Structure
Operational Metadata
System Performance Data
Mapping Information

……..Includes indices, used to improve data access and retrieval performance.
DW Structure
Operational Metadata
System Performance Data
Mapping Information

……From operational databases (source of RDBMSs) and their contents, cleaning and transformation rules, etc.
DW Structure
Operational Metadata
System Performance Data
Mapping Information

…….It's tends to be replaced by other architectures that better achieve Data Integration and Cross-reporting.
Independent Data Marts Architecture
Bus Architecture
Hub-and-spoke Architecture
Centralized Architecture

……Data is organized in Dimension Tables, Facts Tables, Flat Files etc. that are stored in Data Marts
Independent Data Marts Architecture
Bus Architecture
Hub-and-spoke Architecture
Centralized Architecture

In the……, one of the most used in medium to large contexts.
Independent Data Marts Architecture
Bus Architecture
Hub-and-spoke Architecture
Centralized Architecture

…….Atomic, normalized data is stored in a Reconciled Layer that feeds a set of Data Marts containing summarized data in Multidimensional Form.
Independent Data Marts Architecture
Bus Architecture
Hub-and-spoke Architecture
Centralized Architecture

……where the Reconciled Layer and the Data Marts are collapsed into a Single Physical Repository.
Independent Data Marts Architecture
Bus Architecture
Hub-and-spoke Architecture
Centralized Architecture

……..where the Reconciled Layer and the Data Marts are collapsed into a Single Physical Repository.
Independent Data Marts Architecture
Bus Architecture
Hub-and-spoke Architecture
Centralized Architecture

Each Data Warehouse/Data Mart is either virtually or physically integrated with the others, leaning on a variety of advanced techniques such as distributed querying, ontologies, and meta-data interoperability.
Independent Data Marts Architecture
Bus Architecture
Hub-and-spoke Architecture
Federated Architecture

……are used to data or measures in the business.
Fact Table
Facts
Hierarchy

……are the numeric data elements that are of interest to the company.
Fact Table
Facts
Hierarchy

A…… is a directed TREE whose nodes are dimensional attributes and whose arcs model Many To One Association between Dimensional Attributes team.
Fact Table
Facts
Hierarchy

……It contains a dimension, positioned at the Tree's Root, and all the dimensional attributes that define it.
Fact Table
Facts
Hierarchy

……..enables data to be modeled and viewed in Multiple Dimensions.
A Data Cube
An OLAP Cube
Dimensional Cuboids

……A Multidimensional Data Model is organized around a central theme, like sales and transactions.
A Data Cube
An OLAP Cube
Dimensional Cuboids

……..is a data structure that combines data from disparate data sources and allows fast analysis of data.
A Data Cube
An OLAP Cube
Dimensional Cuboids

……The arrangement of data into cubes overcomes a limitation of relational databases.
A Data Cube
An OLAP Cube
Dimensional Cuboids

……Let us suppose that we would like to view our sales data with an additional fourth dimension, such as a supplier.
A Data Cube
An OLAP Cube
Dimensional Cuboids

True and false

Question	T&F
A Decision Support System (DSS) is a set of expandable, interactive IT techniques and tools designed for processing and analyzing data and for supporting managers in decision making.	
A Decision Support Database Technique for assembling and managing data from various sources for the purpose of answering business questions.	
Data Warehousing maintained separately from the organization's operational database.	
making decisions that were not previous possible in Data Warehousing	
Relational databases are usually used for structured data	
File systems or No. SQL databases can be used for (static), unstructured data	
Information Delivery This element not only stores and manages the data; and keeps track of data using the metadata repository.	
Data Warehouses/Data Storage That consists of all the different ways of making the information from the data warehouses available to the users.	
Incremental extraction when a data warehouse needs populating for the first time, this looks like a snapshot of operational data; or	
Static extraction to update data warehouses regularly and capture the changes applied to source data since the latest extraction, based on the log maintained by the operational DBMS.	
A timestamp is associated with operational data to record exactly when the data is changed or added.	
Cleansing and Transformation processes are closely connected in ETL Tools.	
update is normally used in combination with static extraction to initially populate a data warehouse.	
his technique is used in combination with incremental extraction to Refresh data warehouses regularly.	
Data storage for the data warehousing is a Metadata.	
split repository in a data warehouse is equal to the Data Dictionary or the Data Catalog in a DBMS.	
Standardization of data components forms a large part of data transformation from different sources. We combine data from single source record or related data parts from many source records.	
Data Warehouse applications are designed to support the user ad-hoc data requirements, an activity recently dubbed online analytical processing (OLAP).	

Enterprise-oriented Architecture depends on the number of layers used by the architecture	
Structure-oriented Architecture depends on how the different layers are employed to create enterprise-oriented or department-oriented views of DW	
Data Marts are as small, local Data Warehouses replicating the part of a primary DW required for a specific application domain.	
The Traditional Database implements query-driven approach	
Data Warehousing employs an update-driven approach	
query processing in data warehouses interfere with the processing at local sources.	
data base can store and integrate historical information and support complex multidimensional queries.	
In Independent Data Marts Architecture, different data marts are separately designed and built in a non-integrated approach.	
he Bus Architecture is like the preceding architecture, with one important difference. Data is organized in Dimension Tables, Facts Tables, Flat Files etc. that are stored in Data Marts	
Data base Modeling is the process of designing the schemas of the detailed and summarized information of the Data Warehouse.	
he GOAL of DW modeling is to develop a schema describing the reality, or at least a part of the fact, which the data warehouse is needed to support.	
Data Modeling in DW is different from Data Modeling in Operational Database Systems.	
DW is designed for the customer with general information knowledge about the enterprise	
Operational Database Systems are more oriented toward use by software specialists for creating distinct applications.	
Highly summarized data is data extract from the low level of detail found at the current, detailed level and usually is stored on disk storage.	
Lightly summarized data is compact and directly available and can even be found outside the warehouse.	
A Logical Data Model recognizes the highest-level relationships between the different entities.	
A Conceptual Data Model defines the information in as much structure as possible, Without Observing how they will be physically achieved in the database.	
Physical Data Model describes how the model will be presented in the database.	
A physical database model demonstrates all table structures, column names,	

data types, constraints, primary key, foreign key, and relationships between tables.	
Online Analytical Processing (OLAP) is a classification of software technology which authorizes analysts, managers, and executives to gain insight into information	✓
OLAP don't enable end-clients to perform ad hoc analysis of record in multiple dimensions, providing the insight and understanding they require for better decision making.	✗
OLAP implement the single dimension Analysis of business information and support the capability for complex estimations, trend analysis, and sophisticated data modeling.	✗
Dimensional modeling is complicated	✗
The n-D Base Cube which holds the lowest level of summarization is called a Apex Cuboid.	✗
The topmost 0-D cuboid, which holds the highest level of summarization, is known as the Base Cuboid.	✗
A data warehouse does not require transaction processing, recovery, and concurrency controls.	✓
A data warehouse it not is physically stored and separate from the operational database.	✗