#### What is Data Scheme?

Data Scheme is a diagrammatic representation that illustrates data structures and data relationships to each other in the relational database within the data warehouse.

The data structures have their names defined with their data types.

Data Schemes are handy guides for database and data warehouse implementation.

The Data Scheme may or may not represent the real lay out of the database but just a structural representation of the physical database.

Data Schemes are useful in troubleshooting databases.

#### What is Bit Mapped Index?

Bitmap indexes make use of bit arrays (bitmaps) to answer queries by performing bitwise logical operations.

They work well with data that has a lower cardinality which means the data that take fewer distinct values.

Bitmap indexes are useful in the data warehousing applications.

Bitmap indexes have a significant space and performance advantage over other structures for such data.

Tables that have less number of insert or update operations can be good candidates.

The advantages of Bitmap indexes are:

* They have a highly compressed structure, making them fast to read.
* Their structure makes it possible for the system to combine multiple indexes together so that they can access the underlying table faster.

The Disadvantage of Bitmap indexes is:

* The overhead on maintaining them is enormous.

#### What is Bi-directional Extract?

In hierarchical, networked or relational databases, the data can be extracted, cleansed and transferred in two directions. The ability of a system to do this is refered to as bidirectional extracts.

This functionality is extremely useful in data warehousing projects.

**Data Extraction**The source systems the data is extracted from vary in various forms right from their structures and file formats to the department and the business segment they belong to. Common source formats include flat files and relational database and other non-relational database structures such as IMS, VSAM or ISAM.

**Data transformation**  
The extracted data may undergo transformation with possible addition of metadata before they are exported to another large storage area.

In transformation phase, various functions related to business needs, requirements, rules and policies are applied on them. During this process some values even get translated and encoded. Care is also taken to avoid redundancy of data.

**Data Cleansing**In data cleansing, scrutinizing of the incorrect or corrupted data is done and those inaccuracies are removed. Thus data consistency is ensured in Data cleansing.

It involves activities like   
- removing typographical errors and inconsistencies   
- comparing and validating data entries against a list of entities

**Data transformation**This is the last process of Bidirectional Extracts. The cleansed, transformed extracted source data is then loaded into the data warehouse.

**Advantages**- Updates and data loading become very fast due to bidirectional extracting.  
- As timely updates are received in a useful pattern companies can make good use of this data to launch new products and formulate market strategies.

**Disadvantage**- More investment on advance and faster IT infrastructure.   
- Not being able to come up with fault tolerance may mean unexpected stoppage of operations when the system breaks.  
- Skilled data administrator needs to be hired to manage the complex process.

#### What is Data Collection Frequency?

Data collection frequency is the rate at which data is collected. However, the data is not just collected and stored. it goes through various stages of processing like extracting from various sources, cleansing, transforming and then storing in useful patterns.

It is important to have a record of the rate at which data is collected because of various reasons:

Companies can use these records to keep a track of the transactions that have occurred. Based on these records the company can know if any invalid transactions ever occurred.

In scenarios where the market changes rapidly, companies need very frequently updated data to enable them make decisions based on the state of the market and then invest appropriately.

A few companies keep launching new products and keep updating their records so that their customers can see them which would in turn increase their business.

When data warehouses face technical problems, the logs as well as the data collection frequency can be used to determine the time and cause of the problem.

Due to real time data collection, database managers and data warehouse specialists can make more room for recording data collection frequency.

#### What is Data Cardinality?

Cardinality is the term used in database relations to denote the occurrences of data on either side of the relation.

There are 3 basic types of cardinality:

High data cardinality:  
Values of a data column are very uncommon.   
e.g.: email ids and the user names

Normal data cardinality:  
Values of a data column are somewhat uncommon but never unique.   
e.g.: A data column containing LAST\_NAME (there may be several entries of the same last name)

Low data cardinality:  
Values of a data column are very usual.   
e.g.: flag statuses: 0/1

Determining data cardinality is a substantial aspect used in data modeling. This is used to determine the relationships

Types of cardinalities:

The Link Cardinality - 0:0 relationships   
The Sub-type Cardinality - 1:0 relationships  
The Physical Segment Cardinality - 1:1 relationship   
The Possession Cardinality - 0: M relation   
The Child Cardinality - 1: M mandatory relationship  
The Characteristic Cardinality - 0: M relationship  
The Paradox Cardinality - 1: M relationship.

#### What is Chained Data Replication?

In Chain Data Replication, the non-official data set distributed among many disks provides for load balancing among the servers within the data warehouse.

Blocks of data are spread across clusters and each cluster can contain a complete set of replicated data. Every data block in every cluster is a unique permutation of the data in other clusters.

When a disk fails then all the calls made to the data in that disk are redirected to the other disks when the data has been replicated.

At times replicas and disks are added online without having to move around the data in the existing copy or affect the arm movement of the disk.

In load balancing, Chain Data Replication has multiple servers within the data warehouse share data request processing since data already have replicas in each server disk.

#### What are Critical Success Factors?

Key areas of activity in which favorable results are necessary for a company to reach its goal.

There are four basic types of CSFs which are:

Industry CSFs  
Strategy CSFs  
Environmental CSFs  
Temporal CSFs

A few CSFs are:  
Money  
Your future  
Customer satisfaction  
Quality  
Product or service development  
Intellectual capital  
Strategic relationships  
Employee attraction and retention  
Sustainability

The advantages of identifying CSFs are:  
they are simple to understand;   
they help focus attention on major concerns;   
they are easy to communicate to coworkers;   
they are easy to monitor;   
and they can be used in concert with strategic planning methodologies.

***Data warehousing - What is Data Warehousing? - Jan 09, 2009 at 18:10 pm by Rajmeet Ghai***

#### What is Data Warehousing?

A data warehouse can be considered as a storage area where interest specific or relevant data is stored irrespective of the source. What actually is required to create a data warehouse can be considered as Data Warehousing. Data warehousing merges data from multiple sources into an easy and complete form.

***Data warehousing - What is Data Warehousing? - Jan 12, 2010 at 14:44 pm by Vidya Sagar***

#### What is Data Warehousing?

Large enterprises organize and analyze larger data sets. Small enterprises wishes to analyze one subject, for example, accessing data of marts. Data warehouse is a repository of an enterprise’s data storage that is stored electronically. Data warehouses are designed / used for facilitating report analysis. Data warehouse also focuses on data storage, data dictionary, data analysis, transformation of data. In this context, data warehousing is also known as business intelligence tool.

***Data warehousing - What is Virtual Data Warehousing? - Jan 09, 2009 at 18:10 pm by Rajmeet Ghai***

#### What is Virtual Data Warehousing?

A virtual data warehouse provides a collective view of the completed data. A virtual data warehouse has no historic data. It can be considered as a logical data model of the containing metadata.

***Data warehousing - What is Virtual Data Warehousing? - Jan 12, 2010 at 14:44 pm by Vidya Sagar***

#### What is Virtual Data Warehousing?

Virtual data warehousing is a ‘de facto’ information system strategy for supporting analytical decision making. It is one of the best ways for translating raw data and presenting it in the form which decision makers can use. It provides semantic map – which allows the end user for viewing as virtualized.

***Data warehousing - fundamental stages of Data Warehousing - Jan 09, 2009 at 18:10 pm by Rajmeet Ghai***

#### Explain in brief various fundamental stages of Data Warehousing.

Stages of a data warehouse helps to find and understand how the data in the warehouse changes.

At an initial stage of data warehousing data of the transactions is merely copied to another server. Here, even if the copied data is processed for reporting, the source data’s performance won’t be affected.

In the next evolving stage, the data in the warehouse is updated regularly using the source data.

In Real time Data warehouse stage data in the warehouse is updated for every transaction performed on the source data (E.g. booking a ticket)

When the warehouse is at integrated stage, It not only updates data as and when a transaction is performed but also generates transactions which are passed back to the source online data.

***Data warehousing - fundamental stages of Data Warehousing - Jan 12, 2010 at 14:44 pm by Vidya Sagar***

#### Explain in brief various fundamental stages of Data Warehousing.

The following are the stages of data warehousing:

**Offline Operational Database:** It is the stage where copying the data off an operational system to another server where the report processing load against the copied data takes place and OS performance does not impact.

**Offline Data Warehouse:** In this stage, data warehouses are updated from data in the OS and the data of data warehouse is stored in a data structure that is designed for facilitating reports.

**Real time Data Warehouse:** The data warehouses are updated often when an OS performs a transaction.

**Integrated Data Warehouse:** The data warehouses are updated by OS, at the time of performing a transaction. Then transactions are generated which are passed back into the operational systems.

***Data warehousing - What is active data warehousing? - Jan 09, 2009 at 18:10 pm by Rajmeet Ghai***

#### What is active data warehousing?

An active data warehouse represents a single state of the business. Active data warehousing considers the analytic perspectives of customers and suppliers. It helps to deliver the updated data through reports.

***Data warehousing - What is active data warehousing? - Jan 12, 2010 at 14:44 pm by Vidya Sagar***

#### What is active data warehousing?

A form of repository of captured transactional data is known as ‘active data warehousing’. Using this concept, trends and patterns are found to be used for future decision making. Active data warehouse has a feature which can integrate the changes of data while scheduled cycles refreshes. Enterprises utilize an active data warehouse in drawing the company’s image in statistical manner.

***Data warehousing - What is data modeling and data mining? - Jan 09, 2009 at 19:30 pm by Venkatesh Raman***

#### What is data modeling and data mining? What is this used for?

Data Modeling is a technique used to define and analyze the requirements of data that supports organization’s business process. In simple terms, it is used for the analysis of data objects in order to identify the relationships among these data objects in any business.

Data Mining is a technique used to analyze datasets to derive useful insights/information. It is mainly used in retail, consumer goods, telecommunication and financial organizations that have a strong consumer orientation in order to determine the impact on sales, customer satisfaction and profitability. Data Mining is very helpful in determining the relationships among different business attributes.

***Data warehousing - What is data modeling and data mining? - Jan 12, 2010 at 14:44 pm by Vidya Sagar***

#### What is data modeling and data mining? What is this used for?

Data modeling is the analysis of data objects and their relationships with other objects. It is the primary step for database design and using OOP, the conceptual design model is created to depict how the data items are related to each other. It is the involvement of progression from conceptual model to logical model to that of physical metadata.

Analyzing data from various perspectives and summarizing it into the required and useful information is known as data mining / knowledge discovery.

Data mining allows the users to analyze data from many different dimensions or angles, categorizing data and summarize the identified relationships. The correlations or patterns among different fields in large RDBMS are the technical aspect of data mining.

***Data warehousing - ER Modeling and Dimensional Modeling - Jan 09, 2009 at 19:30 pm by Venkatesh Raman***

#### Difference between ER Modeling and Dimensional Modeling

The entity-relationship model is a method used to represent the logical flow of entities/objects graphically that in turn create a database. It has both logical and physical model. And it is good for reporting and point queries.

Dimensional model is a method in which the data is stored in two types of tables namely facts table and dimension table. It has only physical model. It is good for ad hoc query analysis.

***Data warehousing - ER Modeling and Dimensional Modeling - Jan 12, 2010 at 14:44 pm by Vidya Sagar***

#### Difference between ER Modeling and Dimensional Modeling

Dimensional modeling is a form of modeling of data that is more flexible for the perspective of user. The ER modeling is for databases that are OLTP databases which uses normalized data using 1st or 2nd or 3rd normal forms.

Dimensional Modeling is used in data warehouses that uses 3rd normal form. It contains denormalized data.

***Data warehousing - difference between data warehousing and business intelligence - Jan 09, 2009 at 19:30 pm by Venkatesh Raman***

#### What is the difference between data warehousing and business intelligence?

Data warehousing relates to all aspects of data management starting from the development, implementation and operation of the data sets. It is a back up of all data relevant to business context i.e. a way of storing data

Business Intelligence is used to analyze the data from the point of business to measure any organization’s success. The factors like sales, profitability, marketing campaign effectiveness, market share and operational efficiency etc are analyzed using Business Intelligence tools like Cognos, Informatica, SAS etc.

***Data warehousing - difference between data warehousing and business intelligence - Jan 12, 2010 at 14:44 pm by Vidya Sagar***

#### What is the difference between data warehousing and business intelligence?

**Business Intelligence:** BI is a blend of technologies and applications for gathering, integrating, storing, analyzing and providing the access to the data to enterprise uses for better decision making in an enterprise. BI is a combination of decision support system, online analytical processing, query and reporting, statistical analysis, forecasting and data mining. BI is a group of collection for knowledge and data mart and data warehousing.

**Data warehousing:** Data warehousing is all about dimensional modeling by extraction, conform delivering for building data warehouse that are subject oriented, non volatile and time variant.

Data warehouse is a RDBMS. It designs analysis and transformation processing.

***Data warehousing - dimensional Modeling - Jan 09, 2009 at 19:30 pm by Venkatesh Raman***

#### Describe dimensional Modeling.

Dimensional model is a method in which the data is stored in two types of tables namely facts table and dimension table. Fact table comprises of information to measure business successes and the dimension table comprises of information on which the business success is calculated. It is mainly used by data warehouse designers to build data warehouses. It represents the data in a standard and sequential manner that triggers for high performance access.

***Data warehousing - dimensional Modeling - Jan 12, 2010 at 14:44 pm by Vidya Sagar***

#### Describe dimensional Modeling.

Dimensional modeling is a design concept which is used by designers of building data warehouses. The data is stored in two types of tables - fact table and dimension table. Facts or measurements of the business are persisted in fact table and the context of measurements such as the dimensions on which the calculation of facts takes place.

***Data warehousing - What is snapshot with reference to data warehouse? - Jan 09, 2009 at 19:30 pm by Venkatesh Raman***

#### What is snapshot with reference to data warehouse?

Snapshot refers to a complete visualization of data at the time of extraction. It occupies less space and can be used to back up and restore data quickly.

***Data warehousing - What is snapshot with reference to data warehouse? - Jan 12, 2010 at 14:44 pm by Vidya Sagar***

#### What is snapshot with reference to data warehouse?

A snap shot is a process of knowing about the activities performed. Snap shot is stored in a report format from a specific catalog. The report is generated soon after the catalog is disconnection.

***What is SQL Server 2005 Analysis Services (SSAS)? - May 27, 2009 at 11:20 AM by Rajmeet Ghai***

#### What is SQL Server 2005 Analysis Services (SSAS)?

SSAS gives the business data an integrated view. This integrated view is provided by combining online analytical processing (OLAP) and data mining functionality. SSAS supports OLAP and allows data collected from various sources to be managed in an efficient way. Analysis services, specifically for data mining, allow use of a wide array of data mining algorithms that allows creation, designing of data mining models.

***What is SQL Server 2005 Analysis Services (SSAS)? - Jan 10, 2010 at 18:50 PM by Vidya Sagar***

#### What is SQL Server 2005 Analysis Services (SSAS)?

SQL Server 2005 Analysis Services is a service that provides to view business data by unifying and integrating. It is the foundation for traditional reporting, OLAP analysis, data mining and Key Performance Indicator scorecards.