

Data Warehouse Manager Job Interview Questions And Answers



Interview Questions Answers

<https://interviewquestionsanswers.org/>

About Interview Questions Answers

Interview Questions Answers . ORG is an interview preparation guide of thousands of Job Interview Questions And Answers, Job Interviews are always stressful even for job seekers who have gone on countless interviews. The best way to reduce the stress is to be prepared for your job interview. Take the time to review the standard interview questions you will most likely be asked. These interview questions and answers on Data Warehouse Manager will help you strengthen your technical skills, prepare for the interviews and quickly revise the concepts.

If you find any **question or answer** is incorrect or incomplete then you can **submit your question or answer** directly with out any registration or login at our website. You just need to visit [Data Warehouse Manager Interview Questions And Answers](#) to add your answer click on the *Submit Your Answer* links on the website; with each question to post your answer, if you want to ask any question then you will have a link *Submit Your Question*; that's will add your question in Data Warehouse Manager category. To ensure quality, each submission is checked by our team, before it becomes live. This [Data Warehouse Manager Interview preparation PDF](#) was generated at **Wednesday 29th November, 2023**

You can follow us on FaceBook for latest Jobs, Updates and other interviews material.
www.facebook.com/InterviewQuestionsAnswers.Org

Follow us on Twitter for latest Jobs and interview preparation guides.
<https://twitter.com/InterviewQA>

If you need any further assistance or have queries regarding this document or its material or any of other inquiry, please do not hesitate to contact us.

Best Of Luck.

Interview Questions Answers.ORG Team
<https://InterviewQuestionsAnswers.ORG/>
Support@InterviewQuestionsAnswers.ORG



Data Warehouse Manager Interview Questions And Answers Guide.

Question - 1:

Explain me what is meant by Data Analytics?

Ans:

Data analytics (DA) is the science of examining raw data with the purpose of drawing conclusions about that information. A data warehouse is often built to enable Data Analytics

[View All Answers](#)

Question - 2:

Explain me what is conformed fact?

Ans:

Conformed fact is a table which can be used across multiple data marts in combined with the multiple fact tables.

[View All Answers](#)

Question - 3:

Do you know what is called data cleaning?

Ans:

Name itself implies that it is a self explanatory term. Cleaning of Orphan records, Data breaching business rules, Inconsistent data and missing information in a database.

[View All Answers](#)

Question - 4:

Explain me what is real-time datawarehousing?

Ans:

Real-time datawarehousing captures the business data whenever it occurs. When there is business activity gets completed, that data will be available in the flow and become available for use instantly.

[View All Answers](#)

Question - 5:

Explain any five applications of data warehouse?

Ans:

Some applications include:

- * financial services
- * banking services
- * customer goods
- * retail sectors
- * controlled manufacturing

[View All Answers](#)

Question - 6:

Explain me what are the key columns in Fact and dimension tables?

Ans:

Foreign keys of dimension tables are primary keys of entity tables. Foreign keys of fact tables are the primary keys of the dimension tables.

[View All Answers](#)

**Question - 7:**

Explain what is dimensional modeling?

Ans:

Dimensional model consists of dimension and fact tables. Fact tables store different transactional measurements and the foreign keys from dimension tables that qualifies the data. The goal of Dimensional model is not to achieve high degree of normalization but to facilitate easy and faster data retrieval.

Ralph Kimball is one of the strongest proponents of this very popular data modeling technique which is often used in many enterprise level data warehouses.

[View All Answers](#)

Question - 8:

Tell me what are Aggregate tables?

Ans:

Aggregate tables are the tables which contain the existing warehouse data which has been grouped to certain level of dimensions. It is easy to retrieve data from the aggregated tables than the original table which has more number of records.

This table reduces the load in the database server and increases the performance of the query.

[View All Answers](#)

Question - 9:

Explain me what needs to be done when the database is shutdown?

Ans:

Following needs to be done when the database is shutdown:

- * Close the database
- * Dismount the database
- * Shutdown the Instance

[View All Answers](#)

Question - 10:

Do you know what is VLDB?

Ans:

VLDB is abbreviated as Very Large Database and its size is set to be more than one terabyte database. These are decision support systems which is used to server large number of users.

[View All Answers](#)

Question - 11:

Tell us what is called Dimensional Modelling?

Ans:

Dimensional Modeling is a concept which can be used by dataware house designers to build their own datawarehouse. This model can be stored in two types of tables - Facts and Dimension table.

Fact table has facts and measurements of the business and dimension table contains the context of measurements.

[View All Answers](#)

Question - 12:

Explain me the functions of a load manager?

Ans:

A load manager extracts data from the source system. Fast load the extracted data into temporary data store. Perform simple transformations into structure similar to the one in the data warehouse.

[View All Answers](#)

Question - 13:

Tell me what are the different types of datawarehousing?

Ans:

Following are the different types of Datawarehousing:

- * Enterprise Datawarehousing
- * Operational Data Store
- * Data Mart

[View All Answers](#)

Question - 14:

Tell me what is Snowflake Schema?

Ans:

Snowflake schema which has primary dimension table to which one or more dimensions can be joined. The primary dimension table is the only table that can be joined with the fact table.

[View All Answers](#)

Question - 15:



Do you know what is ODS?

Ans:

ODS is abbreviated as Operational Data Store and it is a repository of real time operational data rather than long term trend data.

[View All Answers](#)

Question - 16:

Tell me what is ER Diagram?

Ans:

ER diagram is abbreviated as Entity-Relationship diagram which illustrates the interrelationships between the entities in the database. This diagram shows the structure of each tables and the links between the tables.

[View All Answers](#)

Question - 17:

Tell me what are the benefits of data warehouse?

Ans:

A data warehouse helps to integrate data and store them historically so that we can analyze different aspects of business including, performance analysis, trend, prediction etc. over a given time frame and use the result of our analysis to improve the efficiency of business processes.

[View All Answers](#)

Question - 18:

Tell me what are loops in Datawarehousing?

Ans:

In datawarehousing, loops are existing between the tables. If there is a loop between the tables, then the query generation will take more time and it creates ambiguity. It is advised to avoid loop between the tables.

[View All Answers](#)

Question - 19:

Tell me how can we load the time dimension?

Ans:

Time dimensions are usually loaded through all possible dates in a year and it can be done through a program. Here, 100 years can be represented with one row per day.

[View All Answers](#)

Question - 20:

Tell me what needs to be done while starting the database?

Ans:

Following need to be done to start the database:

- * Start an Instance
- * Mount the database
- * Open the database

[View All Answers](#)

Question - 21:

Explain load manager?

Ans:

A load manager performs the operations required to extract and load the process. The size and complexity of load manager varies between specific solutions from data warehouse to data warehouse.

[View All Answers](#)

Question - 22:

Tell us what are the tools available for ETL?

Ans:

Following are the ETL tools available:

- * Informatica
- * Data Stage
- * Oracle
- * Warehouse Builder
- * Ab Initio
- * Data Junction

[View All Answers](#)

Question - 23:

Tell us what is a core dimension?



Ans:

Core dimension is nothing but a Dimension table which is used as dedicated for single fact table or datamart.

[View All Answers](#)

Question - 24:

Tell me what is the difference between OLTP and OLAP?

Ans:

Following are the differences between OLTP and OLAP:

OLTP:

Data is from original data source
Simple queries by users
Normalized small database
Fundamental business tasks

OLAP:

Data is from various data sources
Complex queries by system
De-normalized Large Database
Multi-dimensional business tasks

[View All Answers](#)

Question - 25:

Tell me why facts table is useful in representing the data?

Ans:

Fact table allows the measurement and the values of the facts of the data to be contained inside the table. This table consists of the foreign keys and primary keys of the dimension tables. It is located in between the star schema or snowflake schema. It provides values that are additive and independent variables through which the dimensional attributes are analyzed. This table consists of the grains, which consist of atomic level of data and through which the facts in the tables are defined. Each record defines the independent facts that provide higher level of data to be given to the user. It is useful in representing the data due to easy storage and less memory to be taken to the facts of the data that are associated with it.

[View All Answers](#)

Question - 26:

Tell me what are the reasons for partitioning?

Ans:

Partitioning is done for various reasons such as easy management, to assist backup recovery, to enhance performance.

[View All Answers](#)

Question - 27:

What is ETL?

Ans:

ETL is abbreviated as Extract, Transform and Load. ETL is a software which is used to reads the data from the specified data source and extracts a desired subset of data. Next, it transform the data using rules and lookup tables and convert it to a desired state.

[View All Answers](#)

Question - 28:

Do you know what is the difference between Datawarehouse and OLAP?

Ans:

Datawarehouse is a place where the whole data is stored for analyzing, but OLAP is used for analyzing the data, managing aggregations, information partitioning into minor level information.

[View All Answers](#)

Question - 29:

Tell me what is the definition of Cube in Datawarehousing?

Ans:

Cubes are logical representation of multidimensional data. The edge of the cube has the dimension members, and the body of the cube contains the data values.

[View All Answers](#)

Question - 30:

Explain me what is the use of dimensional modeling in data warehousing?

Ans:

Dimensional modeling is a set of techniques that is used in designing the overall structure of data warehousing. It doesn't involve relational database but at the logical level uses the physical form of the database. It is used to support user queries and to increase the performance and understanding of a particular database concept. It uses facts and dimensions to support the measures and the context of the database. The facts define the values that can be aggregated and dimensions represent the group of hierarchies and the descriptors that define the facts in return. This type of models is built by business process model and consists inside the process area. This process area consists of the same design and operation details as others.

[View All Answers](#)

**Question - 31:**

Explain me how does a Data Cube help?

Ans:

Data cube helps us to represent the data in multiple dimensions. The data cube is defined by dimensions and facts.

[View All Answers](#)

Question - 32:

Tell me what are the stages that are required in Data warehousing?

Ans:

There are four different kinds of stages that are required in data warehousing and they are:

Offline Operational Databases: This is the top most and initial stage that allows the database to be viewed offline without going to online. This copy the database to the operational system and an offline server that processes the load of the online and offline and allow the performance to be balanced.

Offline Data Warehouse: This is the second stage where the updation of the time cycle that is regular takes place. The settings are given through which the data can be set like daily, weekly, monthly and yearly. This data is taken from the operational system. The data is stored in the report oriented data structure.

Real Time Data Warehouse: This allows the transaction update on the event basis. It means on an event an updation occurs. The transaction is performed in the operational system as well.

Integrated Data Warehouse: This is the final stage and it is used to generate activity or transactions. After generation they are again put back to the operational system to be used by the user on the daily basis

[View All Answers](#)

Question - 33:

Tell me what does subject-oriented data warehouse signify?

Ans:

Subject oriented signifies that the data warehouse stores the information around a particular subject such as product, customer, sales, etc.

[View All Answers](#)

Question - 34:

Tell me why is dimensional normalization not required?

Ans:

Dimensional normalization allows the database related problems to be solved. It is used to remove the redundant attributes that are used as de-normalized dimensions. Dimensions consist of sub-dimensions that are joined together in one. The dimensional normalization is not used due to the fact that it makes:

* Data structure more complex and due to which the performance can be degraded as it requires lots of joining of tables and keep the relations intact.

* The space is not utilized properly and use of more space is required.

* The query performance suffers when aggregating or retrieving many dimensional values. This requires proper analysis and making of operational reports are necessary.

[View All Answers](#)

Question - 35:

Tell me what is dimension?

Ans:

A dimension is something that qualifies a quantity (measure).

For an example, consider this: If I just say... "20kg", it does not mean anything. But if I say, "20kg of Rice (Product) is sold to Ramesh (customer) on 5th April (date)", then that gives a meaningful sense. These product, customer and dates are some dimension that qualified the measure - 20kg.

Dimensions are mutually independent. Technically speaking, a dimension is a data element that categorizes each item in a data set into non-overlapping regions.

[View All Answers](#)

Question - 36:

Define me a warehouse manager?

Ans:

Warehouse manager is responsible for the warehouse management process. The warehouse manager consist of third party system software, C programs and shell scripts. The size and complexity of warehouse manager varies between specific solutions.

[View All Answers](#)

Question - 37:

Tell me what are the steps involved in designing a fact table?

Ans:

The fact table allows the data to be represented in the detailed manner. It has an association with the dimensional table.

The steps that are required to design the fact table includes:

Identify the business process for analysis so that all the process can be defined and used with the complete details provided by the business process.

Identify the measures, constraints, facts and keys before designing the fact table. The questions have to be asked during the business process regarding the table that has to be created and the purpose of it.

Identify the dimensions for facts like product dimensions, location, time, etc. This phase also include the analysis of the components that are required to be involved.

List of the columns takes place that describe the dimension and the lowest level of detail is to be found out to be used in the fact table.

[View All Answers](#)

Question - 38:



Tell me why is Business Intelligence (BI) important in data warehousing?

Ans:

Business intelligence is required everywhere and it is most importantly required in data warehousing due to the technologies that are getting used in it. It allows the applications and the practices to be performed according to the business rules and regulations. Business intelligence consists of collection, integration, analysis and the presentation of the information that is used to make business related decisions. This is also used to build business related system like decision support system that provide historical, current and predictive data that can be used in the favor of setting up the environment of the database system. It uses computer based techniques to extract the information from many resources and analyze it according to the industry. It is used to provide nearly accurate decision making system.

[View All Answers](#)

Question - 39:

Explain what is the very basic difference between data warehouse and operational databases?

Ans:

A data warehouse contains historical information that is made available for analysis of the business whereas an operational database contains current information that is required to run the business.

[View All Answers](#)

Question - 40:

Explain what is the difference between OLTP and OLAP?

Ans:

OLTP is known as online transaction processing and it refers to the class of systems that manages the transaction oriented applications used for the data entry and processing of the transaction whereas, OLAP stands for online analytical processing that defines an approach to give the reply to multi-dimensional analytical queries. It is a part of business intelligence. It allows the modification to be done of the traditional database terms.

OLTP uses the original data source to be taken in the database. It doesn't create any copy or uses any virtual data. OLAP uses many data sources that are taken from many places and gets stored in the database.

OLTP uses the business process snapshots that will handle the recovery process if any that has to be done in future. Whereas, OLAP uses the snapshots of multi-dimensional views of business activities of planning and decision making that is required to be used in the database.

OLTP uses the normalization database that will slow down the system due the size of the database and the normalization also degrades the performance. Whereas, OLAP uses the de-normalized process and uses the large database so the speed also increases and it improves the overall performance of the system as well

[View All Answers](#)

Question - 41:

Tell me how many fact tables are there in a star schema?

Ans:

There is only one fact table in a star Schema.

[View All Answers](#)

Question - 42:

Explain me what is Normalization?

Ans:

Normalization splits up the data into additional tables.

[View All Answers](#)

Question - 43:

Tell us which one is faster, Multidimensional OLAP or Relational OLAP?

Ans:

Multidimensional OLAP is faster than Relational OLAP.

[View All Answers](#)

Question - 44:

Tell me which language is used for defining Schema Definition?

Ans:

Data Mining Query Language (DMQL) is used for Schema Definition.

[View All Answers](#)

Question - 45:

Please explain what is Summary Information?

Ans:

Summary Information is the area in data warehouse where the predefined aggregations are kept.

[View All Answers](#)

Question - 46:

Do you know what is Virtual Warehouse?

Ans:



The view over an operational data warehouse is known as virtual warehouse.

[View All Answers](#)

Question - 47:

Define data warehouse in your own words?

Ans:

Data warehouse is a subject oriented, integrated, time-variant, and nonvolatile collection of data that supports management's decision-making process.

[View All Answers](#)

Question - 48:

Do you know what is the difference between metadata and data dictionary?

Ans:

Metadata is defined as data about the data. But, Data dictionary contain the information about the project information, graphs, abinito commands and server information.

[View All Answers](#)

Question - 49:

Explain me what is surrogate key?

Ans:

Surrogate key is nothing but a substitute for the natural primary key. It is set to be a unique identifier for each row that can be used for the primary key to a table.

[View All Answers](#)

Question - 50:

Tell me what is Star Schema?

Ans:

Star schema is nothing but a type of organizing the tables in such a way that result can be retrieved from the database quickly in the data warehouse environment.

[View All Answers](#)

Question - 51:

Explain me what is Data Mining?

Ans:

Data Mining is set to be a process of analyzing the data in different dimensions or perspectives and summarizing into a useful information. Can be queried and retrieved the data from database in their own format.

[View All Answers](#)

Question - 52:

Explain what are the different types of SCD's used in data warehousing?

Ans:

SCD (Slowly changing dimensions) that provides different attributes that are used for the record that varies over time and doesn't remain stable.

There are three types of SCDs are used in data warehousing. These are defined as:

SCD1: This is the record that is used to replace the original record even there is only one record exists in the database. The current data will be replaced and the new data will take its place.

SCD2: This is the new record file that is added to the dimension table. This record exists in the database with the current data and previous data that is stored in the history.

SCD3: This uses the original data that is modified to the new data. This consists of two record one record that exist in the database and another record that will replace the old database record with the new information.

[View All Answers](#)

Question - 53:

Explain what are the different models used in cluster analysis?

Ans:

There are many algorithms that can be used to analyze the database to check the maintenance of all the data sets that are already present.

The different types of cluster models include as follows:

* Connectivity models: These are the models that connect one cluster to another cluster. This includes the example of hierarchical clustering that is based on the distance connectivity of one model to another model.

* Centroid models: These are the models that are used to find the clusters using the single mean vector. It includes the example of k-means algorithm.

* Distribution models: It includes the specification of the models that are statistically distributed for example multivariate normal distribution model.

* Density models: Deals with the clusters that are densely connected with one another in the regions having the data space.

* Group models: Specifies the model that doesn't provide the refined model for the output and just gives the grouping information.

[View All Answers](#)

Question - 54:

Tell me what language is the base of DMQL?

Ans:



DMQL is based on Structured Query Language (SQL).

[View All Answers](#)

Question - 55:

Explain the types of OLAP server?

Ans:

There are four types of OLAP servers, namely Relational OLAP, Multidimensional OLAP, Hybrid OLAP, and Specialized SQL Servers.

[View All Answers](#)

Question - 56:

Explain the process that are involved in Data Warehousing?

Ans:

Data Warehousing involves data cleaning, data integration and data consolidations.

[View All Answers](#)

Question - 57:

Tell me can we take backup when the database is opened?

Ans:

No, We cannot take full backup when the database is opened.

[View All Answers](#)

Question - 58:

Explain the Schema that a data warehouse system can implements?

Ans:

A data Warehouse can implement star schema, snowflake schema, and fact constellation schema.

[View All Answers](#)

Question - 59:

Tell me what is SCD?

Ans:

SCD is defined as slowly changing dimensions, and it applies to the cases where record changes over time.

[View All Answers](#)

Question - 60:

Explain me what is Dimension Table?

Ans:

Dimension table is a table which contain attributes of measurements stored in fact tables. This table consists of hierarchies, categories and logic that can be used to traverse in nodes.

[View All Answers](#)

Question - 61:

Explain me what is the goal of Optimizer?

Ans:

The goal to Optimizer is to find the most efficient way to execute the SQL statements.

[View All Answers](#)

Question - 62:

Explain what is Fact?

Ans:

A fact is something that is quantifiable (Or measurable). Facts are typically (but not always) numerical values that can be aggregated.

[View All Answers](#)

Question - 63:

Explain what do you mean by Data Extraction?

Ans:

Data extraction means gathering data from multiple heterogeneous sources.

[View All Answers](#)

Question - 64:

Tell me how many dimensions are selected in dice operation?

**Ans:**

For dice operation two or more dimensions are selected for a given cube.

[View All Answers](#)

Question - 65:

Do you know why is chameleon method used in data warehouse?

Ans:

Chameleon is a hierarchical clustering algorithm that overcomes the limitations of the existing models and the methods present in the data warehousing. This method operates on the sparse graph having nodes that represent the data items and edges represent the weights of the data items. The representation of it allows large data set to be created and operated on successfully. The method finds the clusters that are used in the data set using the two phase algorithm. The first phase consists of the graph partitioning that allows the clustering of the data items into large number of sub-clusters. Second phases use an agglomerative hierarchical clustering algorithm to search for the clusters that are genuine and can be combined together with the sub-clusters that are produced.

[View All Answers](#)

Question - 66:

Explain me what are the different types of fact tables used in data warehousing?

Ans:

There are three types of fact tables that are used and the characterization is given below:

Transactional Table: It is the most basic table of data warehousing. It consists of grain that has the association with the transactional fact table and it is defines to use one row per line in a transactional. It consists of the most detailed data that have many number of dimensions associated with it.

Periodic snapshots: It is used to take the snapshot of the currently working data set. It depends on the period of time when the snapshots are taken. It increases the performance. This table is dependent on the transactional table as it has to be synchronized to keep all the detailed data that is mentioned in transactional table. The snapshot can be looked if any problem take place in the database and it helps in recovering the state that is lost.

Accumulating snapshots: This type of table is used to show the activity of the process that is defined from the beginning till end. For example, order process. The snapshots consist of multiple date columns that represent a timeline of the process. It keeps and entry of the associate date dimension representing the unknown date.

[View All Answers](#)

Question - 67:

Explain me what are the different types of schemas used in Data warehousing?

Ans:

Schema defines the type of structure that is used by the database to hold on some data that can be related or can be different.

There are three types of schema that exists in the database as follows:

BUS schema: This is the schema that is composed of the master file that is totally confirmed with the dimension and with standardized definition including the facts.

Star schema: This is the schema that defines the organization type of the tables and it is used to retrieve the result from the database quickly in a controlled environment.

Snow flake schema: This schema is used to show the primary dimension table, that includes one or more dimensions that can be joined. Primary dimension table only allows the joining of the fact table.

[View All Answers](#)

Question - 68:

Explain what are additive, semi-additive and non-additive measures?

Ans:

* Non-additive Measures:

Non-additive measures are those which can not be used inside any numeric aggregation function (e.g. SUM(), AVG() etc.). One example of non-additive fact is any kind of ratio or percentage. Example, 5% profit margin, revenue to asset ratio etc. A non-numerical data can also be a non-additive measure when that data is stored in fact tables, e.g. some kind of varchar flags in the fact table.

* Semi Additive Measures:

Semi-additive measures are those where only a subset of aggregation function can be applied. Let's say account balance. A sum() function on balance does not give a useful result but max() or min() balance might be useful. Consider price rate or currency rate. Sum is meaningless on rate; however, average function might be useful.

* Additive Measures:

Additive measures can be used with any aggregation function like Sum(), Avg() etc. Example is Sales Quantity etc.

[View All Answers](#)

Question - 69:

Explain the functions performed by OLAP?

Ans:

OLAP performs functions such as

- * roll-up
- * drill-down
- * slice
- * dice
- * and pivot

[View All Answers](#)

Question - 70:

Do you know what is the purpose of cluster analysis in Data Warehousing?

Ans:

Cluster analysis is used to define the object without giving the class label. It analyzes all the data that is present in the data warehouse and compare the cluster with



the cluster that is already running. It performs the task of assigning some set of objects into the groups are also known as clusters. It is used to perform the data mining job using the technique like statistical data analysis. It includes all the information and knowledge around many fields like machine learning, pattern recognition, image analysis and bio-informatics. Cluster analysis performs the iterative process of knowledge discovery and includes trials and failures. It is used with the pre-processing and other parameters as a result to achieve the properties that are desired to be used.

[View All Answers](#)

Question - 71:

Explain me what is the function of ETL when used in data warehousing?

Ans:

ETL is used for extract, transform, and load and it is used to enable the business models to consolidate their data that is moving from place to place and it allows the saving of the data in different forms and different formats. The data can be used from any source and can be included by any source that is defined. It is a powerful method to handle the data disparities and other problems that are related to the data.

There are few steps that are involved in using the function of ETL and it is as follows:

The extract function reads the data from the source that is given and stores the extracted information in the database of the path that is given.

The transformation of the data occurs that is done through some rules keeping the lookup tables to see the combination of the data that is getting formed and the transformed data are arranged and kept in order of the choice that is made while transforming it.

The load function is used load the data to the database or the user location where it needs to be located and then the resulting data is to be loaded to the target location.

[View All Answers](#)

Question - 72:

Explain me what is factless fact tables?

Ans:

A factless fact tables are the fact table which doesn't contain numeric fact column in the fact table.

[View All Answers](#)

Question - 73:

Do you know what is OLTP?

Ans:

OLTP is abbreviated as On-Line Transaction Processing, and it is an application that modifies the data whenever it received and has large number of simultaneous users.

[View All Answers](#)

Question - 74:

Tell me what is the function of surrogate key in data warehousing?

Ans:

Surrogate key is the key that can be substituted with the primary key. It is a unique key that uniquely identifies the each row that is used. This also allows the unique search of the primary key to the table. It is used in the database to uniquely identify an entity in a model or an object that is present in the database. This key is not derived from any of the data entity and not even from the application data. These are represented either in the form of numbers or integers. This kind of keys is generated from the system and it is not at all visible to the outside world. This key can be used as a primary key for a given database and there are few differences between primary key and surrogate key. There remains many to one relationship between primary and surrogate keys.

[View All Answers](#)

Question - 75:

Tell me what is Fact Table?

Ans:

Fact table contains the measurement of business processes, and it contains foreign keys for the dimension tables.

Example - If the business process is manufacturing of bricks

Average number of bricks produced by one person/machine - measure of the business process

[View All Answers](#)

Question - 76:

Tell me what are the different ways to represent the loading of Dimension tables?

Ans:

The loading of the dimension table can be represented in two different ways:

Conventional (Slow): In this way, the constraints and keys goes for validations that are done against the data. These validations are performed before the loading of the data takes place in the data warehousing. This ensures the data integrity that is maintained throughout the loading phase. As the name suggests it is slower method that has to be implemented. This way deals in too much complexity in creating the database.

Direct (Fast): In this, all the constraints and the keys that are used with the data are disabled. The validations are performed after the data is loaded in data warehousing. These validations include everything that are applied on constraints and keys. Data are checked for invalid bit or dirty bit, if anything like this is found then the data is not included in the index and all the processes that require the data are skipped.

[View All Answers](#)

Question - 77:

Explain me what is the difference between agglomerative and divisive Hierarchical Clustering?

Ans:

Agglomerative Hierarchical clustering method allows the clusters to be read from bottom to top and it follows this approach so that the program always reads from



the sub-component first then moves to the parent whereas, divisive uses top-bottom approach in which the parent is visited first then the child. Agglomerative hierarchical method consists of objects in which each object creates its own clusters and these clusters are grouped together to create a large cluster. It defines a process of merging that carries on till all the single clusters are merged together into a complete big cluster that will consist of all the objects of child clusters whereas, in divisive the parent cluster is divided into smaller cluster and it keeps on dividing till each cluster has a single object to represent.

[View All Answers](#)

Question - 78:

Explain what kind of costs are involved in Data Marting?

Ans:

Data Marting involves hardware & software cost, network access cost, and time cost.

[View All Answers](#)

Question - 79:

Explain how many dimensions are selected in Slice operation?

Ans:

Only one dimension is selected for the slice operation.

[View All Answers](#)

Question - 80:

Explain the phases involved in the data warehouse delivery process?

Ans:

The stages are:

- * IT strategy
- * Education
- * Business Case Analysis
- * technical Blueprint
- * Build the version
- * History Load
- * Ad hoc query
- * Requirement Evolution
- * Automation
- * Extending Scope

[View All Answers](#)

Question - 81:

What is additive Measures?

Ans:

Additive measures can be used with any aggregation function like Sum(), Avg() etc. Example is Sales Quantity etc.

[View All Answers](#)

Question - 82:

Do you know why Data Warehouse is used?

Ans:

For a long time in the past and also even today, Data warehouses are built to facilitate reporting on different key business processes of an organization, known as KPI. Today we often call this whole process of reporting data from data warehouses as "Data Analytics". Data warehouses also help to integrate data from different sources and show a single-point-of-truth values about the business measures (e.g. enabling Master Data Management).

[View All Answers](#)

Question - 83:

Explain me what is defined as Partial Backup?

Ans:

A Partial backup in an operating system is a backup short of full backup and it can be done while the database is opened or shutdown.

[View All Answers](#)

Question - 84:

Tell me what are the types of Dimensional Modeling?

Ans:

There are three types of Dimensional Modeling and they are as follows:

- * Conceptual Modeling
- * Logical Modeling
- * Physical Modeling

[View All Answers](#)

Question - 85:

Tell me what is Active Datawarehousing?



Ans:

An active datawarehouse is a datawarehouse that enables decision makers within a company or organization to manage customer relationships effectively and efficiently.

[View All Answers](#)

Question - 86:

Tell us what are the stages of Datawarehousing?

Ans:

There are four stages of Datawarehousing:

- Offline Operational Database
- Offline Data Warehouse
- Real Time Datawarehouse
- Integrated Datawarehouse

[View All Answers](#)

Question - 87:

Tell us what is the difference between View and Materialized View?

Ans:

A view is nothing but a virtual table which takes the output of the query and it can be used in place of tables.

A materialized view is nothing but an indirect access to the table data by storing the results of a query in a separate schema.

[View All Answers](#)

Question - 88:

Please explain what is Datamart?

Ans:

A Datamart is a specialized version of Datawarehousing and it contains a snapshot of operational data that helps the business people to decide with the analysis of past trends and experiences. A data mart helps to emphasizes on easy access to relevant information.

[View All Answers](#)

Question - 89:

Do you know what is BUS Schema?

Ans:

BUS schema consists of suite of confirmed dimension and standardized definition if there is a fact tables.

[View All Answers](#)

Question - 90:

Explain me whether Dimension table can have numeric value?

Ans:

Yes, dimension table can have numeric value as they are the descriptive elements of our business.

[View All Answers](#)

Question - 91:

Explain what are the steps to build the datawarehouse?

Ans:

Following are the steps to be followed to build the datawarehouse:

- * Gathering business requirements
- * Identifying the necessary sources
- * Identifying the facts
- * Defining the dimensions
- * Defining the attributes
- * Redefine the dimensions and attributes if required
- * Organize the Attribute hierarchy
- * Define Relationships
- * Assign unique Identifiers

[View All Answers](#)

Question - 92:

Tell me what are the approaches used by Optimizer during execution plan?

Ans:

There are two approaches:

- * Rule Based
- * Cost Based

[View All Answers](#)

Question - 93:



Tell me what is data mart?

Ans:

Data marts are generally designed for a single subject area. An organization may have data pertaining to different departments like Finance, HR, Marketing etc. stored in data warehouse and each department may have separate data marts. These data marts can be built on top of the data warehouse.

[View All Answers](#)

Question - 94:

What is semi Additive Measures?

Ans:

Semi-additive measures are those where only a subset of aggregation function can be applied. Let's say account balance. A sum() function on balance does not give a useful result but max() or min() balance might be useful. Consider price rate or currency rate. Sum is meaningless on rate; however, average function might be useful.

[View All Answers](#)

Question - 95:

Tell me what does Metadata Respiratory contain?

Ans:

Metadata respiratory contains definition of data warehouse, business metadata, operational metadata, data for mapping from operational environment to data warehouse, and the algorithms for summarization.

[View All Answers](#)

Question - 96:

Tell me what does the Query Manager responsible for?

Ans:

Query Manager is responsible for directing the queries to the suitable tables.

[View All Answers](#)

Question - 97:

Tell me what is the benefit of normalization?

Ans:

Normalization helps in reducing data redundancy.

[View All Answers](#)

Question - 98:

Tell me what are the key features of chameleon that separates it from other algorithms?

Ans:

The key features that are in the chameleon are:

The chameleon method determines the pair of similar sub-clusters that can be connected with other clusters. It also finds the closeness of the clusters from one another.

The chameleon with the above property overcomes the limitation that is present in agglomerative hierarchical model.

It uses different methods to take the internal characteristics of the clusters and matches with those which are already present.

It doesn't depend on static model that is supplied by the user and uses automated functions to perform the merging of the clusters that are already associated in the cluster.

[View All Answers](#)

Question - 99:

Tell me what is the difference between data warehouse and operational systems?

Ans:

Operational systems are optimized to preserve the data integrity of the system, whereas data warehouse are optimized to speed up the process of data analysis.

Operational system increases the speed of the business transactions through the use of normalization of the database and using the entity relationship models, whereas data warehouse uses de-normalization and dimension based model to speed the data retrieval.

Operational system uses relational databases to maintain the relationship between the tables. It also consists of insert and update process that takes very less time hence increment in the performance of the system to create the transaction. Whereas, data warehouse store the same data multiple times to keep the aggregation of the data and gather the data from the operational systems.

[View All Answers](#)

Question - 100:

Do you know what is Business Intelligence?

Ans:

Business Intelligence is also known as DSS - Decision support system which refers to the technologies, application and practices for the collection, integration and analysis of the business related information or data. Even, it helps to see the data on the information itself.

[View All Answers](#)

Question - 101:

Explain what is OLAP?



Ans:

OLAP is abbreviated as Online Analytical Processing, and it is set to be a system which collects, manages, processes multi-dimensional data for analysis and management purposes.

[View All Answers](#)

Question - 102:

Tell me what are Non-additive facts?

Ans:

Non-Additive facts are said to be facts that cannot be summed up for any of the dimensions present in the fact table. If there are changes in the dimensions, same facts can be useful.

[View All Answers](#)

Question - 103:

Tell me what are the types of SCD?

Ans:

There are three types of SCD and they are as follows:

SCD 1 - The new record replaces the original record

SCD 2 - A new record is added to the existing customer dimension table

SCD 3 - A original data is modified to include new data

[View All Answers](#)

Question - 104:

Explain me what is Metadata?

Ans:

Metadata is defined as data about the data. The metadata contains information like number of columns used, fix width and limited width, ordering of fields and data types of the fields.

[View All Answers](#)

Question - 105:

Tell me what is the difference between ER Modeling and Dimensional Modeling?

Ans:

ER modeling will have logical and physical model but Dimensional modeling will have only Physical model.

ER Modeling is used for normalizing the OLTP database design whereas Dimensional Modeling is used for de-normalizing the ROLAP and MOLAP design.

[View All Answers](#)

Question - 106:

Tell me what is Execution Plan?

Ans:

Execution Plan is a plan which is used to the optimizer to select the combination of the steps.

[View All Answers](#)

Question - 107:

What is the difference between OLTP and OLAP?

Ans:

OLTP is the transaction system that collects business data. Whereas OLAP is the reporting and analysis system on that data.

OLTP systems are optimized for INSERT, UPDATE operations and therefore highly normalized. On the other hand, OLAP systems are deliberately denormalized for fast data retrieval through SELECT operations.

In a departmental shop, when we pay the prices at the check-out counter, the sales person at the counter keys-in all the data into a "Point-Of-Sales" machine. That data is transaction data and the related system is a OLTP system.

On the other hand, the manager of the store might want to view a report on out-of-stock materials, so that he can place purchase order for them. Such report will come out from OLAP system.

[View All Answers](#)

Question - 108:

What is non-additive Measures?

Ans:

Non-additive measures are those which can not be used inside any numeric aggregation function (e.g. SUM(), AVG() etc.). One example of non-additive fact is any kind of ratio or percentage. Example, 5% profit margin, revenue to asset ratio etc. A non-numerical data can also be a non-additive measure when that data is stored in fact tables, e.g. some kind of varchar flags in the fact table.

[View All Answers](#)

Question - 109:

Tell me what do OLAP and OLTP stand for?

Ans:



OLAP is an acronym for Online Analytical Processing and OLTP is an acronym of Online Transactional Processing.

[View All Answers](#)

Question - 110:

Explain the functions of a warehouse manager?

Ans:

The warehouse manager performs consistency and referential integrity checks, creates the indexes, business views, partition views against the base data, transforms and merge the source data into the temporary store into the published data warehouse, backs up the data in the data warehouse, and archives the data that has reached the end of its captured life.

[View All Answers](#)

Question - 111:

Explain the functions of data warehouse tools and utilities?

Ans:

The functions performed by Data warehouse tool and utilities are

- * Data Extraction,
- * Data Cleaning,
- * Data Transformation,
- * Data Loading and Refreshing.

[View All Answers](#)

Question - 112:

Explain what are the steps involved in creating dimensional modeling process?

Ans:

The business process of the dimensional modeling includes:

Choose The Business Process:In this, 4-step design method is followed that helps to provide the usability of the dimensional model. This allows the business process to be more systematic in representation and more helpful in explaining it as well. It includes the use of Business Process Modelling Notation (BPMN) or Unified Modelling Language (UML).

Declaring The Grain:After choosing the business process, the declaration of the model comes that consists of grains. The grain of the model provides the accurate description of the dimensional model and allows the focus should be shifted there.

Identify The Dimensions:In this phase, the identification of the dimension takes place in the dimensional model. The dimensions are defined in the grain that is defined in the declaration part above. Dimensions acts as a foundation of the fact table where the data gets collected that comes under the fact.

Identify The Facts:Defining the dimensions provides a way to create a table in which the fact data can be stored. These facts are populated on the basis of the numerical figures and facts.

[View All Answers](#)

Question - 113:

What is Datawarehousing?

Ans:

A Datawarehouse is the repository of a data and it is used for Management decision support system. Datawarehouse consists of wide variety of data that has high level of business conditions at a single point in time.

In single sentence, it is repository of integrated information which can be available for queries and analysis.

[View All Answers](#)

Data Warehouse Most Popular & Related Interview Guides

- 1 : [Warehouse Assistant Interview Questions and Answers.](#)
- 2 : [Ab Initio Interview Questions and Answers.](#)
- 3 : [Data Stage Interview Questions and Answers.](#)
- 4 : [ColdFusion Interview Questions and Answers.](#)
- 5 : [Data Warehouse Supervisor Interview Questions and Answers.](#)
- 6 : [Informatica Data Warehousing Interview Questions and Answers.](#)
- 7 : [Data Warehousing Interview Questions and Answers.](#)
- 8 : [Data Warehouse BI Interview Questions and Answers.](#)
- 9 : [ETL \(Extract, transform, load\) Interview Questions and Answers.](#)
- 10 : [Micro Strategy Interview Questions and Answers.](#)

Follow us on FaceBook

www.facebook.com/InterviewQuestionsAnswers.Org

Follow us on Twitter

<https://twitter.com/InterviewQA>

For any inquiry please do not hesitate to contact us.

Interview Questions Answers.ORG Team

[https://InterviewQuestionsAnswers.ORG/
support@InterviewQuestionsAnswers.ORG](https://InterviewQuestionsAnswers.ORG/support@InterviewQuestionsAnswers.ORG)