

Basics Data Warehouse Interview Questions And Answers Guide.



Global Guideline.

<http://www.globalguideline.com/>



Basics Data Warehouse Job Interview Preparation Guide.

Question # 1

Explain Differences between star and snowflake schemas?

Answer:-

star schema is a logical structure that can be arranged with fact and dimension tables in a star formation.it looks like a star with fact table at the core of the star and the dimension tables along the spikes of the star.the dimension model is therefore called a STAR SCHEMA.

SNOWFLAKING is a method of normalizing the dimension tables in a star schema.

Star Schema: can have less number of joins.

Snow flake: can have more number of joins.

Star Schema : has data redundancy, so the query performace is good.

Snow flake: is normalized, so does not has data redundancey. can have performance issues.

[Read More Answers.](#)

Question # 2

Explain the data type of the surrogate key?

Answer:-

Data type of the surrogate key is either integer or numeric or number

There is no data type for a Surrogate Key.

Requirement of a surrogate Key:UNIQUE

Recommended data type of a Surrogate key is NUMERIC.

[Read More Answers.](#)

Question # 3

What is a CUBE in datawarehousing concept?

Answer:-

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

[Read More Answers.](#)

Question # 4

Explain Why fact table is in normal form?

Answer:-

Basically the fact table consists of the Index keys of the dimension/ook up tables and the measures. so when ever we have the keys in a table .that itself implies that the table is in the normal form.

[Read More Answers.](#)

Question # 5

Explain What are the possible data marts in Retail sales?

Answer:-

Product information,sales information

Retailor_site_Dim, product_dim, Time_dim, Sales_staff_dim,Return_method_dim,

[Read More Answers.](#)

Question # 6

Explain degenerate dimension table?

Answer:-

Degenerate Dimensions : If a table contains the values, which r neither dimension nor measures is called degenerate dimensions.Ex : invoice id,empno

Degenerate dimensions are those without any attributes or measures



and they are included inside the fact table
Eg orderno

[Read More Answers.](#)

Question # 7

Explain hybrid slowly changing dimension?

Answer:-

Hybrid SCDs are combination of both SCD 1 and SCD 2.

It may happen that in a table, some columns are important and we need to track changes for them i.e capture the historical data for them whereas in some columns even if the data changes, we don't care.

For such tables we implement Hybrid SCDs, where in some columns are Type 1 and some are Type 2.

[Read More Answers.](#)

Question # 8

Explain conformed fact?

Answer:-

Conformed dimensions are the dimensions which can be used across multiple Data Marts in combination with multiple facts tables accordingly

IF A DIMENSION IS EQUALLY SHARED BETWEEN DIFFERENT DATAMARTS HAVING COMMON DATA VALUES IS CALLED CONFORMED DIMENSION.

[Read More Answers.](#)

Question # 9

Explain BUS Schema?

Answer:-

BUS Schema is composed of a master suite of confirmed dimension and standardized definition of facts.

In a BUS schema we would eventually have conformed dimensions and facts defined to be shared across all enterprise data marts. This way all Data Marts can use the conformed dimensions and facts without having them locally. This is the first step towards building an enterprise Data Warehouse from Kimball's perspective. For (e.g) we may have different data marts for Sales, Inventory and Marketing and we need common entities like Customer, Product etc to be seen across these data marts and hence would be ideal to have these as Conformed objects. The challenge here is that some times each line of business may have different definitions for these conformed objects and hence choosing conformed objects have to be designed with some extra care.

[Read More Answers.](#)

Question # 10

Explain What are the advantages data mining over traditional approaches?

Answer:-

Data Mining is used for the estimation of future. For example, if we take a company/business organization, by using the concept of Data Mining, we can predict the future of business interms of Revenue (or) Employees (or) Customers (or) Orders etc.

Traditional approaches use simple algorithms for estimating the future. But, it does not give accurate results when compared to Data Mining.

Traditional approach is also algorithmic perspective and its advantage is convert larger algorithms into smaller ones. In this approach entire software is procedure

[Read More Answers.](#)

Question # 11

Explain Difference between E-R Modeling and Dimensional Modeling?

Answer:-

Basic diff is E-R modeling will have logical and physical model. Dimensional model will have only physical model.

E-R modeling is used for normalizing the OLTP database design.

Dimensional modeling is used for de-normalizing the ROLAP/MOLAP design.

[Read More Answers.](#)

Question # 12

What is a Fact, Dimension, Measure?

Answer:-

Fact: It is a measure, EX: Sales, Accounts etc.

Dimension: It deals with the details of the data. EX: Geography, Time etc.

Measure: It deals with the quantity of the data..

[Read More Answers.](#)

Question # 13

What is Data warehousing Hierarchy?

Answer:-

Hierarchies

Hierarchies are logical structures that use ordered levels as a means of organizing data. A hierarchy can be used to define data aggregation. For example, in a time dimension, a hierarchy might aggregate data from the month level to the quarter level to the year level. A hierarchy can also be used to define a navigational drill path and to establish a family structure.

Within a hierarchy, each level is logically connected to the levels above and below it. Data values at lower levels aggregate into the data values at higher levels. A dimension can be composed of more than one hierarchy. For example, in the product dimension, there might be two hierarchies--one for product categories and one



for product suppliers.

Dimension hierarchies also group levels from general to granular. Query tools use hierarchies to enable you to drill down into your data to view different levels of granularity. This is one of the key benefits of a data warehouse.

When designing hierarchies, you must consider the relationships in business structures. For example, a divisional multilevel sales organization.

Hierarchies impose a family structure on dimension values. For a particular level value, a value at the next higher level is its parent, and values at the next lower level are its children. These familial relationships enable analysts to access data quickly.

Levels

A level represents a position in a hierarchy. For example, a time dimension might have a hierarchy that represents data at the month, quarter, and year levels. Levels range from general to specific, with the root level as the highest or most general level. The levels in a dimension are organized into one or more hierarchies.

Level Relationships

Level relationships specify top-to-bottom ordering of levels from most general (the root) to most specific information. They define the parent-child relationship between the levels in a hierarchy.

Hierarchies are also essential components in enabling more complex rewrites. For example, the database can aggregate an existing sales revenue on a quarterly base to a yearly aggregation when the dimensional dependencies between quarter and year are known.

[Read More Answers.](#)

Question # 14

Explain What are the steps to build the data warehouse?

Answer:-

As far I know...

Gathering bussiness requiremnts

Identifying Sources

Identifying Facts

Defining Dimensions

Define Attributes

Redefine Dimensions & Attributes

Organise Attribute Hierarchy & Define Relationship

Assign Unique Identifiers

Additional conventions:Cardinality/Adding ratios

Gather requirements,

Identify Source tables,

Identify Destination tables,

Data modelling

Source to target matrix preparation,

ETL flow preparation and scheduling,

Reporting

[Read More Answers.](#)

Question # 15

Explain ER Diagram?

Answer:-

The Entity-Relationship (ER) model was originally proposed by Peter in 1976 [Chen76] as a way to unify the network and relational database views.

Simply stated the ER model is a conceptual data model that views the real world as entities and relationships. A basic component of the model is the Entity-Relationship diagram which is used to visually represents data objects.

Since Chen wrote his paper the model has been extended and today it is commonly used for database design For the database designer, the utility of the ER model is:

it maps well to the relational model. The constructs used in the ER model can easily be transformed into relational tables.

it is simple and easy to understand with a minimum of training. Therefore, the model can be used by the database designer to communicate the design to the end user.

In addition, the model can be used as a design plan by the database developer to implement a data model in a specific database management software.

[Read More Answers.](#)

Question # 16

what is junk dimension?

what is the difference between junk dimension and degenerated dimension?

Answer:-

Junk dimension: Grouping of Random flags and text Attributes in a dimension and moving them to a separate sub dimension.

Degenerate Dimension: Keeping the control information on Fact table ex: Consider a Dimension table with fields like order number and order line number and have 1:1 relationship with Fact table. In this case this dimension is removed and the order information will be directly stored in a Fact table inorder eliminate unnecessary joins while retrieving order information..

[Read More Answers.](#)

Question # 17

Explain What are the different types of data warehousing?

Answer:-

Types of data warehousing

1. Enterprise Data warehousing

2. ODS (Operational Data Store)

3. Data Mart

[Read More Answers.](#)

Question # 18

Explain Dimensional Modelling?



Answer:-

Dimensional Modelling is a design concept used by many data warehouse designers to build their datawarehouse. In this design model all the data is stored in two types of tables - Facts table and Dimension table. Fact table contains the facts/measurements of the business and the dimension table contains the context of measurements ie, the dimensions on which the facts are calculated.

Dimensional Modelling is a design concept used by many data warehouse designers to build their datawarehouse. In this design model all the data is stored in two types of tables - Facts table and Dimension table. Fact table contains the facts/measurements of the business and the dimension table contains the context of measurements ie, the dimensions on which the facts are calculated.

[Read More Answers.](#)

Question # 19

Explain can a dimension table contains numeric values?

Answer:-

Yes dimension can have numeric values, that is surrogate Key which holds numeric value for unique identification of records in the dimension

[Read More Answers.](#)

Question # 20

Explain the difference between view and materialized view?

Answer:-

View - store the SQL statement in the database and let you use it as a table. Everytime you access the view, the SQL statement executes.

Materialized view - stores the results of the SQL in table form in the database. SQL statement only executes once and after that everytime you run the query, the stored result set is used. Pros include quick query results.

Views: At run time, the query will be executed against the database.

Materialized views: The data for the materialized view query will be generated at compile time.

Mviews can be created by the following ways:

1. Immediate - mview will be created along with data.

2. Deferred - Mview structure alone will be created. Data will be populated only when you refresh the mview.

We have the option of refreshing the mviews. It means when the data in the master table used in the mview query changes, the refreshing of mviews helps to get the updated (new) data for the mview.

Mview will behave very much like a table. At run time, data will be retrieved from the result set just as retrieved from a table. The retrieval time will be very fast unlike the views.

[Read More Answers.](#)

Question # 21

Explain VLDB?

Answer:-

The perception of what constitutes a VLDB continues to grow. A one terabyte database would normally be considered to be a VLDB.

[Read More Answers.](#)

Question # 22

Explain the data types present in both what happens if we implement view in the designer report?

Answer:-

Three different data types: Dimensions, Measure and Detail.

View is nothing but an alias and it can be used to resolve the loops in the universe.

[Read More Answers.](#)

Question # 23

Explain What are the methodologies of Data Warehousing?

Answer:-

Every company has methodology of their own. But to name a few SDLC Methodology, AIM methodology are standardly used. Other methodologies are AMM, World class methodology and many more.

There are only two methodologies in building DW they are

1. Top down

2. Bottom-up

[Read More Answers.](#)

Question # 24

Explain Difference between Snowflake and Star Schema. What are situations where Snowflake Schema is better than Star Schema to use and when the opposite is true?

Answer:-

Star schema contains the dimension tables mapped around one or more fact tables.

It is a denormalised model.

No need to use complicated joins.

Queries results fastly.

Snowflake schema

It is the normalised form of Star schema.

contains indepth joins ,bcas the tables r splitted in to many pieces.We can easily do modification directly in the tables.

We hav to use comlicated joins ,since we hav more tables .

There will be some delay in processing the Query .



in star schema look like a centrally locate fact table and surrounded by dimension tables . its look like a star thats why people called as a starschema
in star schema dimension tables are de_normalised but fact table is normalised table
in snow flake schema dimension tables are splitted one or more tables
dimension tables are quit bit a table bit space
here dimension tables are normalised
here having the more no of joins
so the performance degrades
as per the client requirement we used star or snow flake schema
client may ask like data normalised or de_normalised

[Read More Answers.](#)

Question # 25

Explain clearly how to explain any (sales) project in interview.actually feom where report developer work starts?pls reply as soon as possible?

Answer:-

if you are a Report developer

- 1,you have to specify the front end and back end tool used for creating the reports
- 2.Then you have to tell the purpose of the project..what you are going to acheive using the reports.
- 3.Then you can explain the backend part which is important.FOR example,you have to tell what are all the facts and dimension going to be used
4. Once the facts and dimension are identified yo might want to restructure the fact and dimension using the views.Also have to decide on the schemas you are going to use whether it s an snowflake schema or an star schema.
- 5,Once the schema has been finalised we might want to include certain KPI(Key Performance Indicators) into it.
- 6,Once the cube has been ready now we are into the deployment of the cube.
- 7,After the deployment has been done sucessfully now we have to use our front end tool such as proclarity and feed the cube into it.
- 8,Once the cube has been feed into the proclarity we can create various reports that can helps us for the business.In our case i had identified the top 10 customers in sales , successful saples period over time by using trend analysis and using ranking KPI for rank the customer.

[Read More Answers.](#)

Question # 26

Explain What is the main difference between Inmon and Kimball philosophies of data warehousing?

Answer:-

Both differed in the concept of building teh datawarehouse..

According to Kimball ...

Kimball views data warehousing as a constituency of data marts. Data marts are focused on delivering business objectives for departments in the organization. And the data warehouse is a conformed dimension of the data marts. Hence a unified view of the enterprise can be obtain from the dimension modeling on a local departmental level.

Inmon beliefs in creating a data warehouse on a subject-by-subject area basis. Hence the development of the data warehouse can start with data from the online store. Other subject areas can be added to the data warehouse as their needs arise. Point-of-sale (POS) data can be added later if management decides it is necessary.

i.e.,

Kimball--First DataMarts--Combined way ---Datawarehouse

Inmon---First Datawarehouse--Later----Datamarts

[Read More Answers.](#)

Question # 27

Explain a linked cube?

Answer:-

Linked cube in which a sub-set of the data can be analysed into great detail. The linking ensures that the data in the cubes remain consistent.

[Read More Answers.](#)

Question # 28

What is the main difference between schema in RDBMS and schemas in DataWarehouse?

Answer:-

RDBMS Schema

- * Used for OLTP systems
- * Traditional and old schema
- * Normalized
- * Difficult to understand and navigate
- * Cannot solve extract and complex problems
- * Poorly modelled

DWH Schema

- * Used for OLAP systems
- * New generation schema
- * De Normalized
- * Easy to understand and navigate
- * Extract and complex problems can be easily solved
- * Very good model

[Read More Answers.](#)

Question # 29

What is static and local variable?

Answer:-

Static variable is not created on function stack but is created in



the initialized data segment and hence the variable can be shared across the multiple call of the same function. Usage of static variables within a function is not thread safe.

On the other hand local variable or auto variable is created on function stack and valid only in the context of the function call and is not shared across function calls.

[Read More Answers.](#)

Question # 30

What is metadata in context of a Datawarehouse and how it is important?

Answer:-

Meta data is the data about data; Business Analyst or data modeler usually capture information about data - the source (where and how the data is originated), nature of data (char, varchar, nullable, existence, valid values etc) and behavior of data (how it is modified / derived and the life cycle) in data dictionary a.k.a metadata. Metadata is also presented at the Datamart level, subsets, fact and dimensions, ODS etc. For a DW user, metadata provides vital information for analysis / DSS.

[Read More Answers.](#)

Question # 31

Suggest a good author for Hyperion in data warehousing?

Answer:-

Ralph Kimball

[Read More Answers.](#)

Question # 32

What is incremental loading?

What is batch processing?

What is crass reference table?

What is aggregate fact table?

Answer:-

Incremental loading means loading the ongoing changes in the OLTP.

Aggregate table contains the [measure] values ,aggregated /grouped/summed up to some level of hierarchy.

Batch Processing means executing more than one session in single run at the same time. we can execute these session in 2 ways :

linear: excecuting one after another

parralel: executing more than one session at a time

[Read More Answers.](#)

Question # 33

What is the datatype of the surrogate key?

Answer:-

Datatype of the surrogate key is either integer or numeric.it,s always generated by system because surrogate key works as primary key.surrogate key help us to distinguish the information about the data and store the data history.

[Read More Answers.](#)

Question # 34

What is surrogate key? where we use it explain with example?

Answer:-

surrogate key is a substitution for the natural primary key.

It is just a unique identifier or number for each row that can be used for the primary key to the table. The only requirement for a surrogate primary key is that it is unique for each row in the table.

Data warehouses typically use a surrogate, (also known as artificial or identity key), key for the dimension tables primary keys. They can use Infa sequence generator, or Oracle sequence, or SQL Server Identity values for the surrogate key.

It is useful because the natural primary key (i.e. Customer Number in Customer table) can change and this makes updates more difficult.

Some tables have columns such as AIRPORT_NAME or CITY_NAME which are stated as the primary keys (according to the business users) but ,not only can these change, indexing on a numerical value is probably better and you could consider creating a surrogate key called, say, AIRPORT_ID. This would be internal to the system and as far as the client is concerned you may display only the AIRPORT_NAME.

2. Adapted from response by Vincent on Thursday, March 13, 2003

Another benefit you can get from surrogate keys (SID) is :

Tracking the SCD - Slowly Changing Dimension.

Let me give you a simple, classical example:

On the 1st of January 2002, Employee 'E1' belongs to Business Unit 'BU1' (that's what would be in your Employee Dimension). This employee has a turnover allocated to him on the Business Unit 'BU1' But on the 2nd of June the Employee 'E1' is muted from Business Unit 'BU1' to Business Unit 'BU2.' All the new turnover have to belong to the new Business Unit 'BU2' but the old one should Belong to the Business Unit 'BU1.'

If you used the natural business key 'E1' for your employee within your datawarehouse everything would be allocated to Business Unit 'BU2' even what actually belongs to 'BU1.'

If you use surrogate keys, you could create on the 2nd of June a new record for the Employee 'E1' in your Employee Dimension with a new surrogate key.

This way, in your fact table, you have your old data (before 2nd of June) with the SID of the Employee 'E1' + 'BU1.' All new data (after 2nd of June) would take the SID of the employee 'E1' + 'BU2.'

You could consider Slowly Changing Dimension as an enlargement of your natural key: natural key of the Employee was Employee Code 'E1' but for you it becomes Employee Code + Business Unit - 'E1' + 'BU1' or 'E1' + 'BU2.' But the difference with the natural key enlargement process, is that you might not have all part of your new key within your fact table, so you might not be able to do the join on the new enlarge key -> so you need another id.

A surrogate key is a system generated sequential number which acts as a primary key.

[Read More Answers.](#)

Question # 35

Explain the definition of normalized and denormalized view and what are the differences between them?



Answer:-

Normalization is the process of removing redundancies.
Denormalization is the process of allowing redundancies.
Normalization is the process of removing redundancies.
Denormalization is the process of allowing redundancies.

[Read More Answers.](#)

Question # 36

Explain What are the various ETL tools in the Market?

Answer:-

Various ETL tools used in market are:

Informatica

Data Stage

Oracle Warehouse Bulider

Ab Initio

Data Junction

BusinessObjects DataIntegrator is another ETL tool.

[Read More Answers.](#)

Question # 37

What are the Data Marts?

Answer:-

A data mart is a collection of tables focused on specific business group/department. It may have multi-dimensional or normalized. Data marts are usually built from a bigger data warehouse or from operational data.

[Read More Answers.](#)

Data Warehouse Most Popular Interview Topics.

- 1 : [Data Warehousing Frequently Asked Interview Questions and Answers Guide.](#)
- 2 : [Ab Initio Frequently Asked Interview Questions and Answers Guide.](#)
- 3 : [Informatica Data Warehousing Frequently Asked Interview Questions and Answers Guide.](#)
- 4 : [ColdFusion Frequently Asked Interview Questions and Answers Guide.](#)
- 5 : [Cognos Data Warehouse Frequently Asked Interview Questions and Answers Guide.](#)
- 6 : [Business Objects Frequently Asked Interview Questions and Answers Guide.](#)
- 7 : [Micro Strategy Frequently Asked Interview Questions and Answers Guide.](#)
- 8 : [BO Designer Frequently Asked Interview Questions and Answers Guide.](#)
- 9 : [Extract, transform, load \(ETL\) Frequently Asked Interview Questions and Answers Guide.](#)
- 10 : [Warehouse DataStage Frequently Asked Interview Questions and Answers Guide.](#)

About Global Guideline.

Global Guideline is a platform to develop your own skills with thousands of job interview questions and web tutorials for fresher's and experienced candidates. These interview questions and web tutorials will help you strengthen your technical skills, prepare for the interviews and quickly revise the concepts. Global Guideline invite you to unlock your potentials with thousands of [Interview Questions with Answers](#) or begin a tutorial right away, such as [HTML](#), [XML](#), [XSLT](#), [Cascading Style Sheet \(CSS\)](#), [Search Engine Optimization \(SEO\)](#), [JavaScript](#), [Structure Query Language \(SQL\)](#), [Database Articles](#), [Web Hosting Guide](#) and much more. Learn the most common technologies [Interview Questions and Answers](#). We will help you to explore the resources of the World Wide Web and develop your own skills from the basics to the advanced. Here you will learn anything quite easily and you will really enjoy while learning. Global Guideline will help you to become a professional and Expert, well prepared for the future.

* This PDF was generated from <http://www.GlobalGuideline.com> at **January 29th, 2017**

* If any answer or question is incorrect or inappropriate or you have correct answer or you found any problem in this document then don't hesitate feel free and [e-mail us](#) we will fix it.

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

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

Best Of Luck.

Global Guideline Team
<http://www.globalguideline.com>
Support@globalguideline.com