

# [By OnlineInterviewQuestions.com](http://OnlineInterviewQuestions.com)

## Data Warehousing Interview Questions

When it comes to career people see very few options in the ground level but when it comes to the cyber world, you can go beyond the expectation. Data warehouse or the DW serves as one of the initial checkpoints for many important business data which are of high demand. **Data warehouse interview questions** are one such area that has a great career opportunity. Here are a few questions that will help you find your dream job within Data warehouse field.

Read on the important **Data warehouse interview questions** and their answers.

### **Q1. What are the different types of data warehouses?**

There are three major types of data warehouses as-

- Enterprise data warehouse
- Operational data store
- Data mart

### **Q2. What are the fundamental elements of the data warehouse?**

A data warehouse consists of data which is obtained from data sources or in other words, external sources. The aim is to make the data available, searchable and valuable for business users. There are three fundamental elements are: -

- Various data sources like ERP, Excel, financial applications or CRM.
- A place where data is refined, sorted and put in order.
- A warehoused space where data is presented

### **Q3. What is Data Analytics in simple terms?**

Data analytics or simply DA is the science used for examining raw data with the purpose of concluding that information. This is mostly built to enable the Data Analytics

### **Q4. Can you state some of the innovations throughout history?**

This method was adopted back in the 1980s by IBM researchers Paul Murphy and Barry Devlin. They happen to put together business data warehouse in a 1988 paper, written by the duo.

William H. Inmon improved it further as data warehouse development, by the introduction of his book Building the Data Warehouse in 1992.

The data warehousing institute was founded in 1995 and the technology started growing. In 2002, Inmon introduced a new concept – data warehousing 2.0.

## **Q5. What are the design methods of data warehousing?**

Different design methods of data warehousing are

- **Top-down approach:** According to Inmon's methods, the data warehouse has to be built first. The data derived from the third party's external system is verified and finally combined into a normalized data model. The data stored in the data warehouse leads to further creation of data marts.
- **Bottom-up method:** According to Kimball's method, one should create the dimensional data marts first. Data obtained from the systems is passed to the staging area and then shaped into a star schema design. This data is at the end, processed and stored with the data marts and each of the marts focus on the individual business process.
- **Hybrid method:** This is an approach obtained from the combination of both the top-down and bottom-up methods. It means the pace of the bottom-up method is combined with the integration from top-down design.

## **Q6. Who needs data warehousing?**

- People who rely on a mass amount of data to make decisions.
- Users who wish to obtain information from multiple data sources using a customized, complex process.
- People who wish to access the data using simple technology can also, use it.
- People who wish to make decisions, based on a systematic approach.
- Users who want fast results out of a huge amount of data to be used in reports and grids or charts.
- Data warehousing is the first step toward discovering hidden patterns of data flows and groupings.

## **Q7. Name a few sectors which use data warehousing?**

Here is a list of the most common sectors which use data warehousing.

- **Airlines:** Used for the purpose such as crew assignment, analysis of profitability, frequent flyers program promotion, etc.
- **Banking:** Helps in the management of the available resources. Along with the management of market research, performance analysis of product and operations.
- **Healthcare:** This sector uses a data warehouse to strategize and predict outcomes. It also helps in

generating patient's treatment reports, medical aid services, share data with tie-in insurance companies, etc.

- **Public sector:** In this sector, data warehousing is used to gather intelligence. Government agencies use it to maintain and analyze tax records, health policy records, for every individual.
- **Investment and insurance sector:** In this sector, the warehouses come handy in analyzing various data patterns, customer trends, and tracking market movements.
- **Retail chain:** As far as the retail chain is considered, data warehousing helps in distribution and marketing. It also helps to track items, customer buying patterns, promotions and also used for determining pricing policy.
- **Telecommunication:** Telecommunication uses data warehousing for product promotions, sales decisions along with distribution decisions.
- **Hospitality industry:** This sector uses the warehouse to design and estimate its advertising and promotional campaigns. The main targets are the clients, based on their feedbacks along with travel patterns.

## **Q8. Give some difference between OLTP and OLAP.**

### **Difference between OLTP and OLAP**

#### **OLTP**

The transaction system that collects the business data is called as OLTP.

OLTP systems are usually optimized for INSERT and UPDATE operations, hence they are highly normalized in general.

#### **OLAP**

OLAP tends to report and analyze the system on that data.

When it comes to OLAP, systems are made denormalized for faster data retrieval through the operation of SELECT.

## **Q9. What is data mart in Data warehouse?**

Data marts are usually designed for just one unique subject area. The organization may have data pertaining to various departments such as Finance, Marketing, HR etc. Hence the data warehouse stores of each department need to be separate, which is solved by data marts. These can also be built on top of a data warehouse if needed.

## **Q10. What are the reasons to use a chameleon method in data warehousing?**

The hierarchical clustering of the algorithm that overcomes all limitations of the base models and methods that are present in the data warehousing in combination is called the Chameleon. This method operates as a sparse graph that has nodes, that is a Chameleon can represent data items and edges representing the need of the data items.

Chameleon representation is the one in the data warehouse that allows a large dataset to create and operate successfully. The method finds the clusters that can be used in the dataset using the two-phase algorithm.

### **Q11. What are the steps of implementing data warehousing?**

There are three steps, which would help address the business risk associated.

- **Enterprise strategy**

Technical requirement is identified here, including the current architecture and tools. Facts, dimensions and attributes are also identified. It also includes data mapping and transformation.

- **Phased delivery**

Data warehouse implementation demands to be phased as per the subject areas. Any kind of related business entities such as booking and billing needs to be implemented prior to integration with each other.

- **Iterative prototyping**

The data warehouse needs to developed and tested iteratively and, does need a big approach for the implementation.

### **Q12. What are the other names given to data warehousing?**

Data warehouse system can also be referred to as:

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

### **Q13. What are cluster analysis in Data Warehousing?**

Cluster analysis is mostly used to define an object without a class label. It helps in analyzing all the data that is present in the data warehouse. It can compare the cluster with another already running cluster. It also performs assigning tasks to set some of the objects into the groups.

Cluster analysis includes all the information and knowledge around other fields like the machine learning, image analysis, pattern recognition, and bio-informatics and helps in performing the iterative process of knowledge discovery that is used with pre-processing and other parameters

### **Q14. What are various warehousing tools?**

The prominent tools of warehousing are given as below: -

- MarkLogic
- Oracle
- Amazon RedShift

### Q15. What according to you might be the future of data warehousing?

The size of the databases continues to grow, which might actually be a problem in the future. The present data warehousing system would not be able to support such a huge data in the future.

Regulatory constraints are changing too, which might lead to loss of ability in combining a source of data. This can lead to unstructured data which is quite difficult to store.

### Q16. What is data lake storage?

A **data lake storage** is a store where a huge amount of raw data is kept. It is only when the data is needed that it is brought out. Files are the storage facility of a hierarchical data warehouse while a data lake stores data by making use of a flat architecture. Every data element in a data lake is given a particular identifier. It is also tagged with metadata tags. Information or data in the data lake is queried by business ventures when there is a need. The queried data is now analyzed to solve a problem or answer a question.

### Q17. What is data reconciliation?

There are possible mistakes when data is migrating from one source to another. Such mistakes are seen during transformation logic and mapping. These errors have led to many issues like incorrect values, missing values and records, records duplication and others. Data reconciliation was necessary as a result of this.

**Data reconciliation** is a data verifying process during as data migration. The data reconciliation process involves the comparison of target data with source data to make sure data is transferred by the migration architecture. Data reconciliation has other benefits apart from making use of different mathematical models to extract and process reliable information.

### Q18. What is Dimensional Model in Data Warehouse?

A **dimensional model** is a model that is developed to read numeric information, summarise the information and analyzing it. This numeric information includes counts, balance, and weight in a warehouse. This model is a data structure technique made to function more efficiently for Data warehousing tools. A man known as Ralph Kimball is responsible for the development of the concept of Dimensional Modelling. The dimension modeling is made up of dimension and fact tables. On the other hand, relational models focus on adding, deleting and

updating in an online transfer system (real-time). The dimensional and relation models are both used in any system involved in the data warehouse.

### **Q19. What is dimension table in Warehouse?**

A **dimension table** is a table in a star schema of a data warehouse. Data warehouses are built using dimensional data models which consist of fact and dimension tables. Dimension tables are used to describe dimensions; they contain dimension keys, values and attributes. They are typically small, ranging from a few to several thousand rows. Occasionally dimensions can grow fairly large, however. For example, a large credit card company could have a customer dimension with millions of rows. Dividing a data warehouse project into dimensions provides structured information for reporting purposes. When you create a dimension, you logically create a structure for your projects. This dimension table can be utilized across for reports and it's about re-usability. If there are any changes to be made, it is evident that only a particular table will get affected. When a company wants to create a report, they can read the data from the dimension table since the table consists of the necessary information.

### **Q20. What is Data warehousing?**

**Data warehousing** is a process of integrating data from different sources. It supports analytical reporting, structured and/or ad hoc queries, and decision making.

Please Visit [OnlineInterviewquestions.com](http://OnlineInterviewquestions.com) to download more pdfs