1. Which of the following is/are DDL commands in SQL
Create and d) Alter

2. Which of the following is/are DML commands in SQL?
Update b) Delete

3. Full form of SQL is
b) Structured Query Language

4. Full form of DDL is
b) Data Definition Language

5. DML is
a) Data Manipulation Language

6. Which of the following statements can be used to create a table with column B int type and C float type?
C) Create Table A (B int,C float)

7. Which of the following statements can be used to add a column D (float type) to the table A created above?

8. Which of the following statements can be used to drop the column added in the above question?

9. Which of the following statements can be used to change the data type (from float to int ) of the column D of table A created in above questions

10. Suppose we want to make Column B of Table A as primary key of the table. By which of the following statements we can do it

## Q11. What is data-warehouse?

A data warehouse is a centralised repository that stores structured data (database tables, Excel sheets) and semi-structured data (XML files, webpages) for the purposes of reporting and analysis. The data flows in from a variety of sources, such as point-of-sale systems, business applications, and relational databases, and it is usually cleaned and standardised before it hits the warehouse. Because a data warehouse can store large amounts of information, it provides users with easy access to a wealth of historical data, which can be used for data mining, data visualisation, and other forms of business intelligence reporting.

Data warehouses serve as a central repository for storing and analysing information to make better informed decisions. An organization's data warehouse receives data from a variety of sources, typically on a regular basis, including transactional systems, relational databases, and other sources.

A data warehouse is a centralized storage system that allows for the storing, analysing, and interpreting of data in order to facilitate better decision-making. Transactional systems, relational databases, and other sources provide data into data warehouses on a regular basis.

A data warehouse is a type of data management system that facilitates and supports business intelligence (BI) activities, specifically analysis. Data warehouses are primarily designed to facilitate searches and analyses and usually contain large amounts of historical data.

A data warehouse can be defined as a collection of organizational data and information extracted from operational sources and external data sources. The data is periodically pulled from various internal applications like sales, marketing, and finance; customer-interface applications; as well as external partner systems. This data is then made available for decision-makers to access and analyze.

So what is data warehouse? For a start, it is a comprehensive repository of current and historical information that is designed to enhance an organization’s performance.

# Key Characteristics of Data Warehouse

The main characteristics of a data warehouse are as follows:

## Subject-Oriented

A data warehouse is subject-oriented since it provides topic-wise information rather than the overall processes of a business. Such subjects may be sales, promotion, inventory, etc. For example, if you want to analyze your company's sales data, you need to build a data warehouse that concentrates on sales. Such a warehouse would provide valuable information like 'who was your best customer last year?' or 'who is likely to be your best customer in the coming year?'

## Integrated

A data warehouse is developed by integrating data from varied sources into a consistent format. The data must be stored in the warehouse in a consistent and universally acceptable manner in terms of naming, format, and coding. This facilitates effective data analysis.

## Non-Volatile

Data once entered into a data warehouse must remain unchanged. All data is read-only. Previous data is not erased when current data is entered. This helps you to analyze what has happened and when.

## Time-Variant

The data stored in a data warehouse is documented with an element of time, either explicitly or implicitly. An example of time variance in Data Warehouse is exhibited in the Primary Key, which must have an element of time like the day, week, or month.

**Benefits of Data Warehouse**

- Improved data consistency
- Better business decisions
- Easier access to enterprise data for end-users
- Better documentation of data
- Reduced computer costs and higher productivity

# 12. What is the difference between OLTP VS OLAP?

Within the data science field, there are two types of data processing systems: online analytical processing (OLAP) and online transaction processing (OLTP). The main difference is that one uses data to gain valuable insights, while the other is purely operational. However, there are meaningful ways to use both systems to solve data problems.

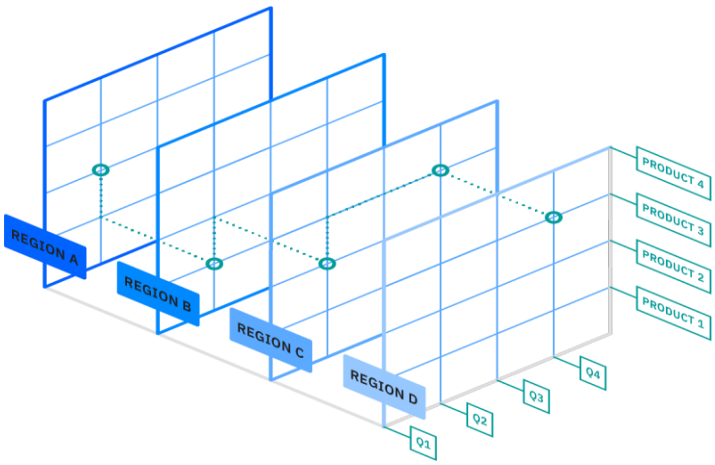
## OLAP

Online analytical processing (OLAP) is a system for performing multi-dimensional analysis at high speeds on large volumes of data. Typically, this data is from a data warehouse, data mart or some other centralized data store. OLAP is ideal for data mining, business intelligence and complex analytical calculations, as well as business reporting functions like financial analysis, budgeting and sales forecasting.

The core of most OLAP databases is the **OLAP cube**, which allows you to quickly query, report on and analyse multidimensional data. What's a data dimension? It's simply one element of a particular dataset. For example, sales figures might have several dimensions related to region, time of year, product models and more.

The OLAP cube extends the row-by-column format of a traditional relational database schema and adds layers for other data dimensions. For example, while the top layer of the cube might organize sales by region, data analysts can also "drill-down" into layers for sales by state/province, city and/or specific stores. This historical, aggregated data for OLAP is usually stored in a star schema or snowflake schema.

The following graphic shows the OLAP cube for sales data in multiple dimensions — by region, by quarter and by product:




## OLTP

Online transactional processing (OLTP) enables the real-time execution of large numbers of database transactions by large numbers of people, typically over the Internet. OLTP systems are behind many of our everyday transactions, from ATMs to in-store purchases to hotel reservations. OLTP can also drive non-financial transactions, including password changes and text messages.

OLTP systems use a relational database that can do the following:

- Process a large number of relatively simple transactions — usually insertions, updates and deletions to data.
- Enable multi-user access to the same data, while ensuring data integrity.
- Support very rapid processing, with response times measured in milliseconds.
- Provide indexed data sets for rapid searching, retrieval and querying.
- Be available 24/7/365, with constant incremental backups.

Many organizations use OLTP systems to provide data for OLAP. In other words, a combination of both OLTP and OLAP are essential in our data-driven world.

# The main difference between OLAP and OLTP: Processing type

The main distinction between the two systems is in their names: analytical vs. transactional. Each system is optimized for that type of processing.

OLAP is optimized for conducting complex data analysis for smarter decision-making. OLAP systems are designed for use by data scientists, business analysts and knowledge workers, and they support business intelligence (BI), data mining and other decision support applications.

OLTP, on the other hand, is optimized for processing a massive number of transactions. OLTP systems are designed for use by frontline workers (e.g., cashiers, bank tellers, hotel desk clerks) or for customer self-service applications (e.g., online banking, e-commerce, travel reservations).

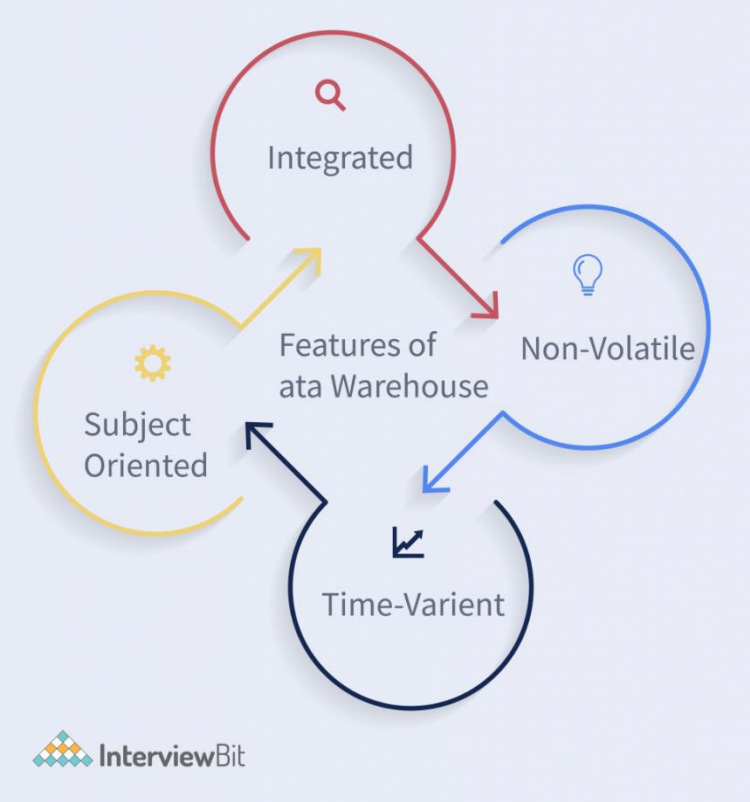
- **Focus:** OLAP systems allow you to extract data for complex analysis. To drive business decisions, the queries often involve large numbers of records. In contrast, OLTP systems are ideal for making simple updates, insertions and deletions in databases. The queries typically involve just one or a few records.
- **Data source**: An OLAP database has a multi-dimensional schema, so it can support complex queries of multiple data facts from current and historical data. Different OLTP databases can be the source of aggregated data for OLAP, and they may be organized as a data warehouse. OLTP, on the other hand, uses a traditional DBMS to accommodate a large volume of real-time transactions.
- **Processing time**: In OLAP, response times are orders of magnitude slower than OLTP. Workloads are read-intensive, involving enormous data sets. For OLTP transactions and responses, every millisecond counts. Workloads involve simple read and write

operations via SQL (structured query language), requiring less time and less storage space.

- **Availability**: Since they don't modify current data, OLAP systems can be backed up less frequently. However, OLTP systems modify data frequently, since this is the nature of transactional processing. They require frequent or concurrent backups to help maintain data integrity.

## 13. What are the various characteristics of data-warehouse?

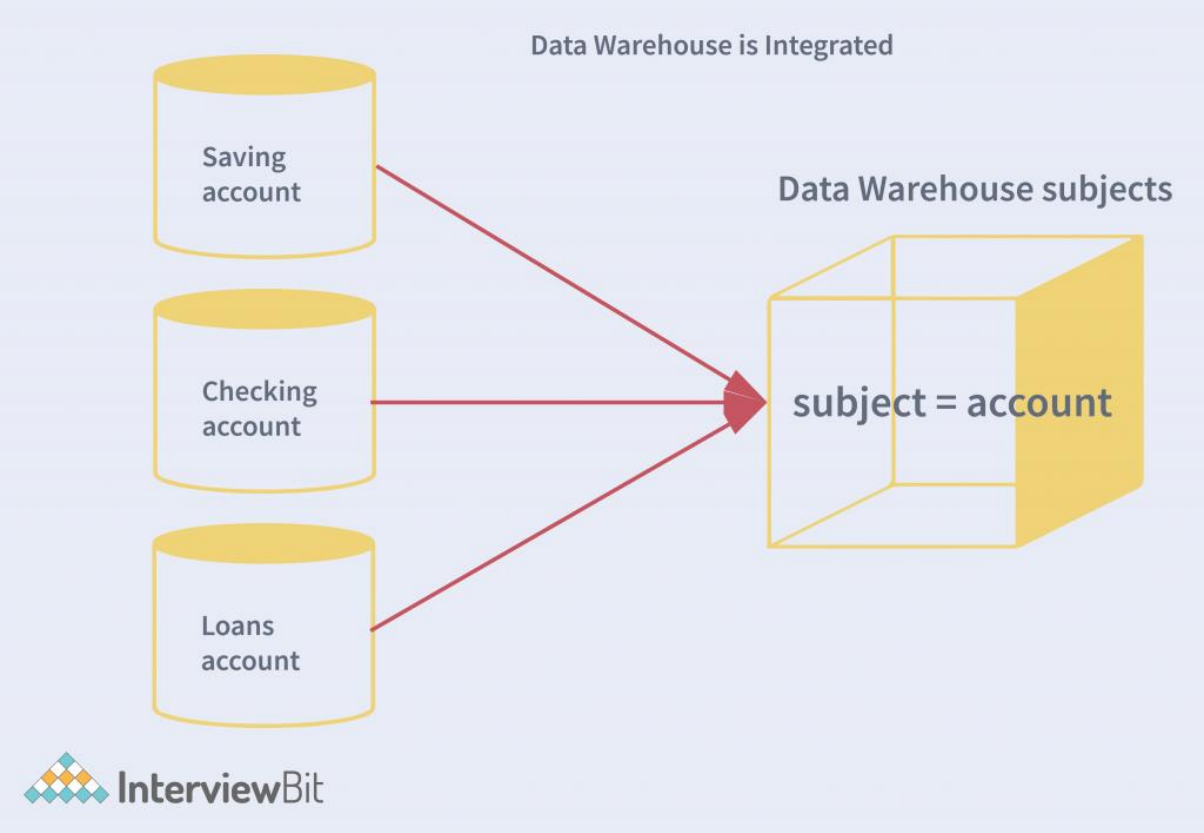
Here are a few caracteristics of data warehouses:




**A Data Warehouse is Subject-Oriented**

A data warehouse provides information on the topic rather than the current operations of organizations. It is subject-oriented and does not mainly concentrate on ongoing processes. A data warehouse aids in creating emphasized models and analytical reports. This is in turn used in decision-making processes. It provides a brief description of the concerned subject and filters all information that does not contribute to the decision-making processes.

Subject-oriented data warehouse (SODW) is a type of data warehouse that is designed to support complex event processing (CEP) and similar applications. The subject-oriented data warehouse is designed to deliver information based on user-defined topics. The information in a subject-oriented data warehouse is structured based on user-defined topics, where a topic is a set of related data that is of interest to a specific business user. This is in contrast to a traditional data warehouse which is designed to support online analytical processing (OLAP) queries.

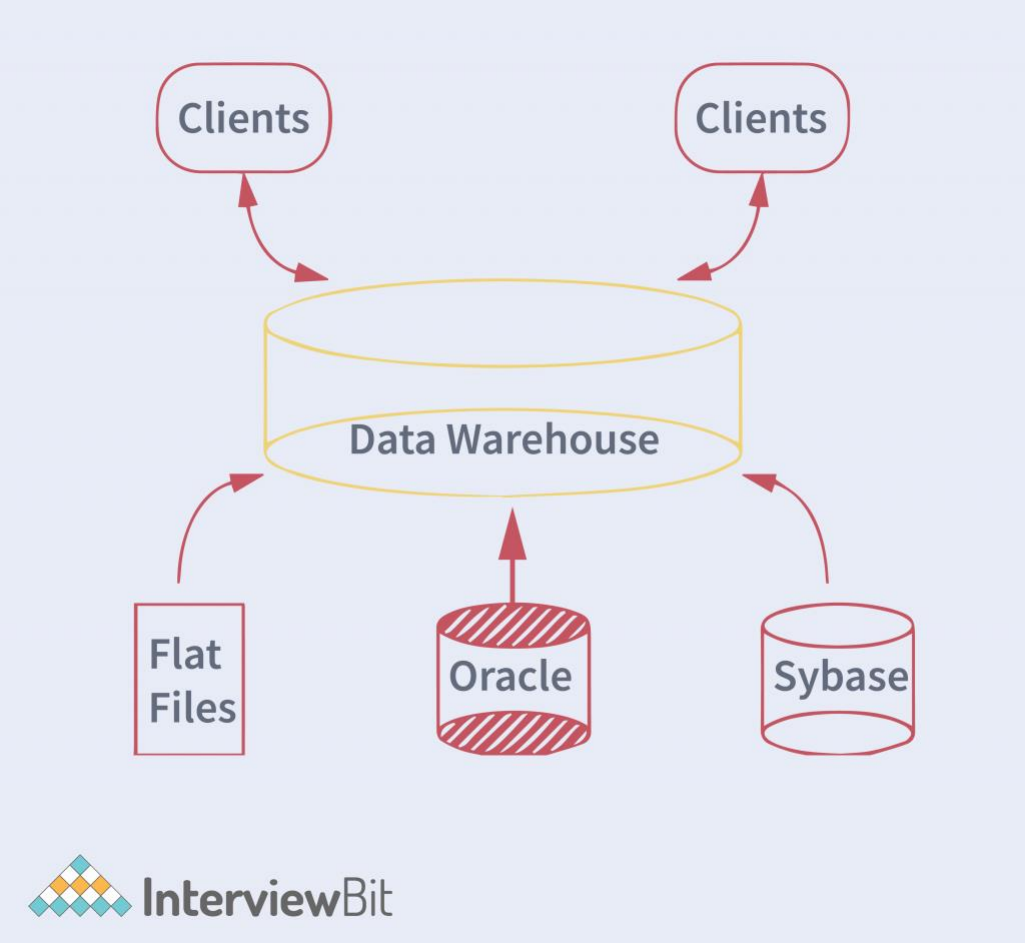



To provide information, data cubes are created from a subset of the attributes in the datasets in the warehouse. A cube is a schema used to store and analyze data from a data warehouse. It is also in contrast to a star schema which is a cube where the dimension tables are connected to the fact table through a single table.

Let us understand this with an example. A company wants to learn more about its sales data. To support their needs a data warehouse that concentrates on sales is built. The architecture of the data warehouse is trained to gather sales-related information such as the company's revenue growth, products that sold more or less than others, industrial comparisons, number of regular customers, etc. Using this data warehouse, the company can derive insights about the best customers of the month, the product that was highly sold in a month, etc.

**Data Warehouses Support Integration**

Integration goes hand in hand with the previous characteristics; subject orientation. A data warehouse is capable of combining data from various sources such as a mainframe, relational databases, flat files, etc.




These warehouses are designed to store and organize data from both transactional and analytical data sources of various types. A data warehouse is typically built for and by the business users in an organization. These users are usually more concerned with how data is used rather than how it is stored. An integrated data warehouse is a central place for reporting, analysis, and business intelligence. This involves a multidimensional data model and a metadata repository. It is also called an enterprise data warehouse or knowledge data warehouse.

The data warehouse adopts a variety of data integration techniques including;

- ETL (Extract Transform and Load): data from different sources are gathered, transformed, and loaded into the warehouse.
- ELT( Extract Load and Transform): data is first loaded as it is into the big data system and transformed later for specific analytical use cases.
- Data-capture changes: identifies real-time changes in the databases and updates them to the data warehouse.
- Data Replication: data in one database is replicated to others to ensure that the information is not lost if any disaster occurs.

- Data virtualization: data from the different systems are combined virtually to create a unified view at a single warehouse. For example, a sales-related data warehouse combines customer purchase details and product information in one place, instead of two separate windows.

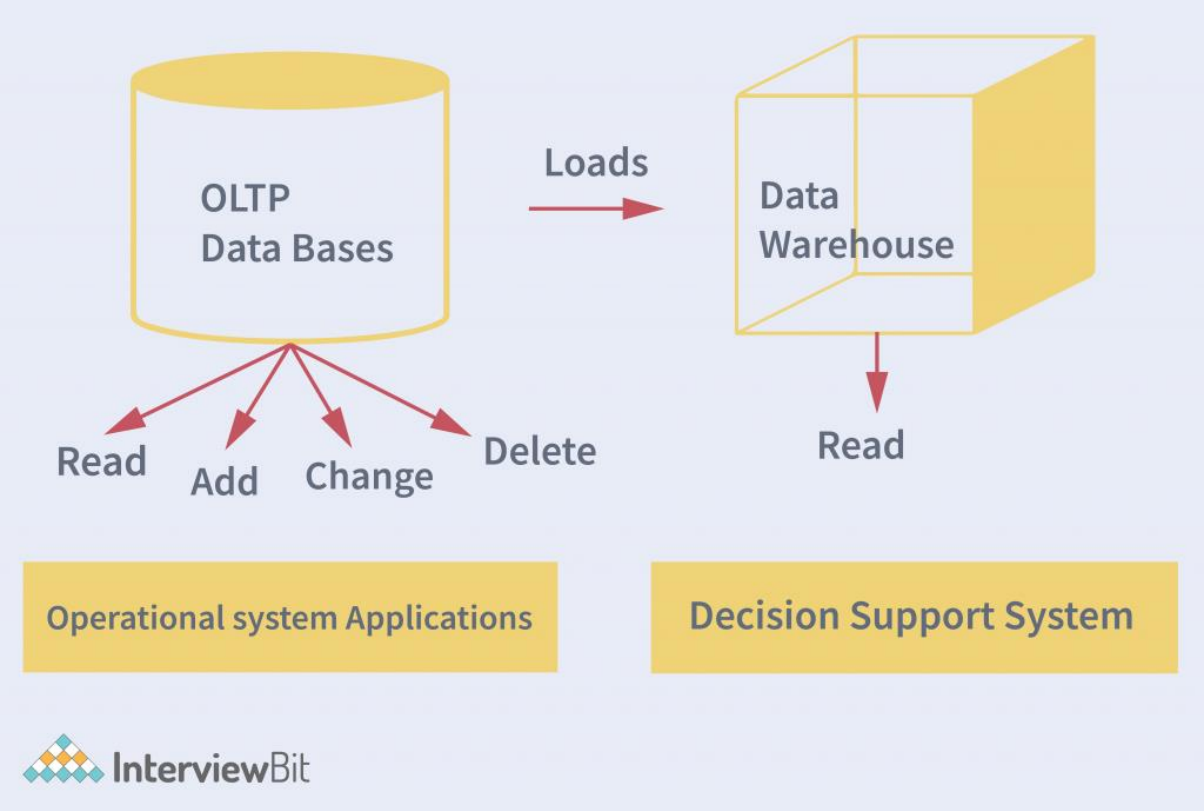
The data warehouse resolves issues like naming conflicts and inconsistencies among units of measurement. The naming conventions, format, coding, measurements, encoding specifications, etc must be consistent throughout the data warehouse. A standard unit of measurement from the different databases for all the similar data is established. This helps in identifying various aspects of the data at-hand quicker. These features of the data warehouse support robust data analysis.

The integrated data warehouse has two benefits. The first benefit is that it eliminates data redundancy by having one version of the truth across the enterprise. The second benefit is that it allows for ad hoc reporting.

Let us understand the benefits with an example. An e-commerce company X provides a mobile app to its users. It uses Google ads and Facebook Ads to acquire new users. The company now wants to increase its advertising budget, but they are unsure whether to spend more on Facebook Ads or Google Ads. They need to know the types of users they have acquired through both channels to make an informed decision. When the user acquisition data via these channels are synchronized in a data warehouse, it helps the company to analyze and compare both the channels using BI tools, and come up with relevant strategies.

## Data Warehouses are Non-Volatile

Data is a company’s wealth as it can be manipulated in several ways to gain insights on many aspects. We never know when a dataset that is ignored and deleted would come in handy for a crucial analytics report. To support this cause, data warehouses are non-volatile, which means that any prior data will not be erased upon the entry of new data in the warehouse. This is done by omitting functions of an operational application environment such as deleting, updating, and inserting.



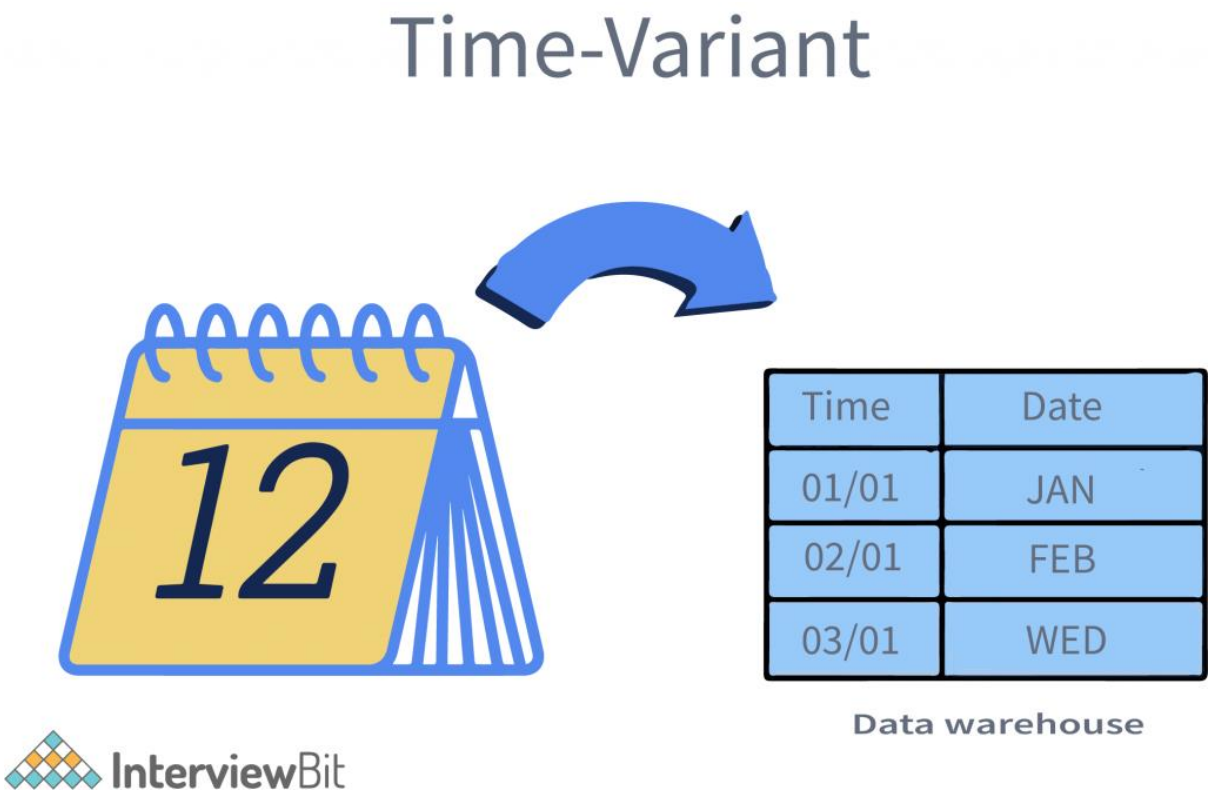

As a result, transaction process, recovery, and competitiveness control mechanisms are not strictly needed.

Data in a data warehouse is subject to the same standards of quality and consistency as data used in the business. Generally, data in a data warehouse is more current and more accurate than data in operational systems. Data warehouses can be used for historical, current, and predictive analysis. They are also well suited for ad hoc queries and applications that perform real-time analytics.

Let us understand the benefits with an example. A retail company Y wants to find out their revenue growth for the past 2 years. With the help of non-volatile data warehouses, they can pull the data directly from their warehouses, instead of going through multiple random accounting files.

**Data in Warehouses are Predictable with Time Intervals**

In the previous characteristics of data warehouses, we read that all data is to be retained for better analytical purposes, but wouldn't it bombard the system? To answer this question, data warehouses have another special feature. The data in a warehouse is maintained via different intervals of time such as hourly, weekly, monthly, annually, etc.




In comparison to operational systems supporting OLTP(online transaction processes), time limits for data warehouses are wide-ranging. The data comprises elements of time either implicitly or explicitly, thus supporting the non-volatility features of data warehouses.

We need to remember that a time-variant data warehouse is a data warehouse that changes with time. The data in the warehouse is transformed as frequently as possible so that the existing data is a true reflection of current business transactions. The original data is retained, but the way the data transformation is carried out is based on the current needs.

## 14. What is Star-Schema?

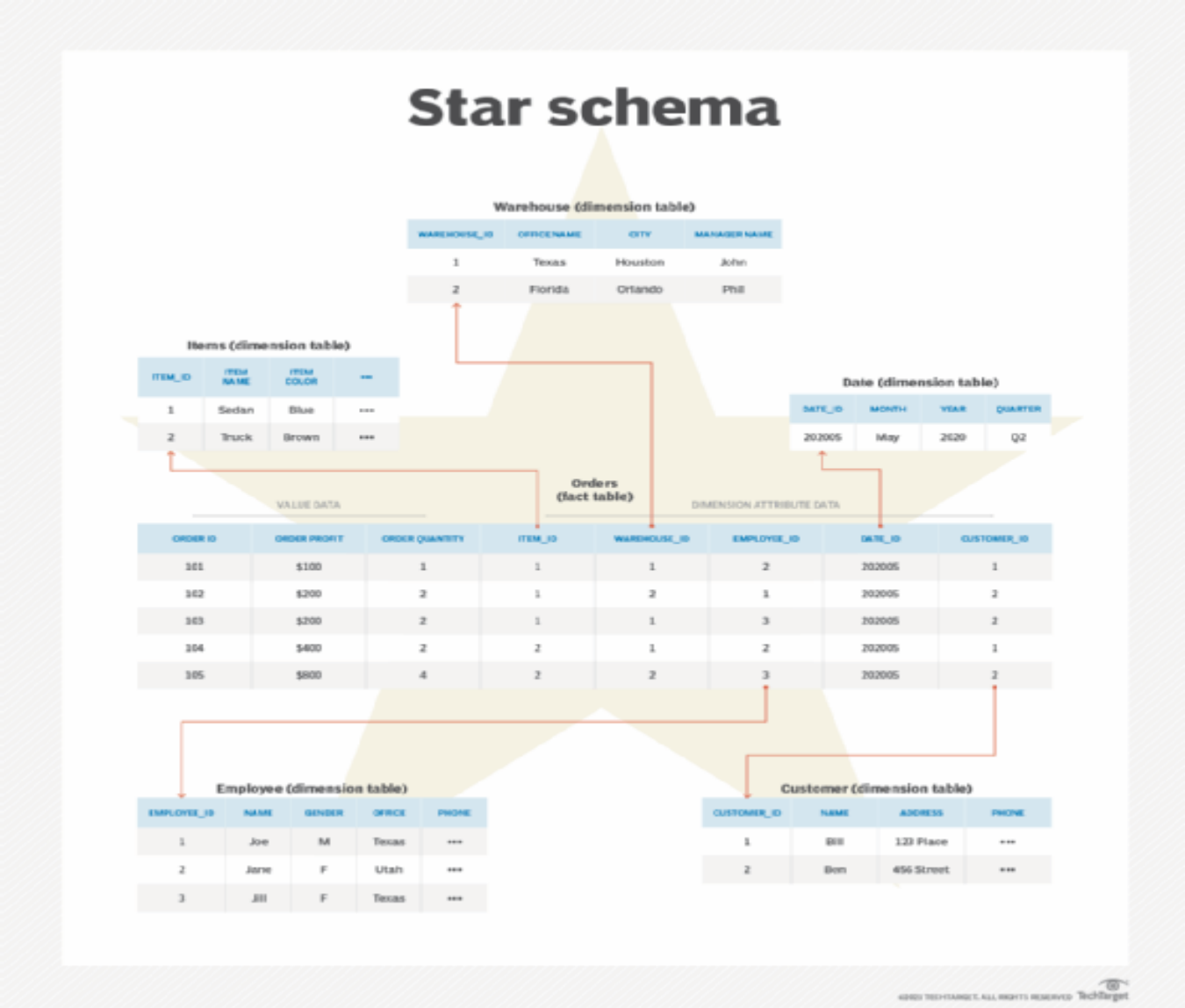
A star schema is a conference for constructing the data into dimension tables, fact tables, and materialized views. All data is saved in columns, and metadata is needed to identify the columns that function as multidimensional objects.

A star schema is a relational schema where a relational schema whose design defines a multidimensional data model. The star schema is the explicit data warehouse schema. It is referred to as star schema because the entity-relationship diagram of this schemas reproduces a star, with points, diverge from the main table. The middle of the schema includes a high fact table, and the star is the dimension table.

**Dimension Tables −** A star schema saves all of the data about a dimension in a single table. Each level of a hierarchy is defined by a column or column set in the dimension table. A dimension object can be used to describe the hierarchical relationship between two columns (or column sets) that defines two levels of a hierarchy; without a dimension object, the hierarchical relationships are represented only in metadata. Attributes are saved in columns of the dimension tables.

**Fact Tables −** Measures are saved in fact tables. Fact tables include a composite primary key, which is composed of multiple foreign keys (one for each dimension table) and a column for each measure that uses these dimensions.

**Materialized Views** − Aggregate data is computed based on the hierarchical relationships represented in the dimension tables. These aggregates are saved in independent tables, known as summary tables or materialised views. Oracle offers extensive support for materialised views, containing automatic refresh and query rewrite.




Optimized for querying large data sets, star schemas are primarily used in data warehouses and data marts for BI and analytics, among other applications.

15. What do you mean by SETL?

SETL provides two basic aggregate data types: *unordered sets*, and *sequences* (the latter also called *tuples*). The elements of sets and tuples can be of any arbitrary type, including sets and tuples themselves. *Maps* are provided as sets of *pairs* (i.e., tuples of length 2) and can have arbitrary domain and range types. Primitive operations in SETL include set membership, union, intersection, and power set construction, among others.

SETL provides quantified Boolean expressions constructed using the universal and existential quantifiers of first-order predicate logic.

SETL provides several iterators to produce a variety of loops over aggregate data structures.