



Share on your Social
Media



Data Warehousing Tutorial

Published On: August 9, 2024

Data Warehousing Tutorial

Data from several different sources is integrated to create a data warehouse. In this data warehouse tutorial, we explore how it supports data science through analytical reporting and decision-making.

[**Download Data Warehousing Tutorial
PDF**](#)

Introduction to Data Warehousing

A subject-oriented, integrated, time-variant, and non-volatile data gathering is defined as a data warehouse. This information helps analysts make defensible decisions within a company. In this data warehousing tutorial, we cover the following:

- Overview of Data Warehousing

Featured Articles



[Click Here to Get Started](#) >>

100%
Placement
Assurance

AUTHORISED
CERTIFICATION
PARTNER

IBM

Related Courses at SLA

Related Posts



Tableau Developer Salary in Chennai

Published On: October 12, 2024

Introduction A Tableau Developer designs, develops, and maintains dashboards and visualizations using Tableau software. Key...

Quick Enquiry



- Concepts of Data Warehousing
- System Process of Data Warehousing
- Data Warehousing Architecture
- Understanding of OLAP

Overview of Data Warehousing

We can examine multidimensional data in a consolidated and generalized manner through the use of data warehouses.

- A data warehouse not only offers us a consolidated and generalized view of the data but also capabilities for online analytical processing (OLAP).
- We may analyze data in a multidimensional space more effectively and interactively with the use of these technologies.
- Data mining and data generalization are the outcomes of this approach.

Data Warehousing Interview Questions

Interpret a Data Warehouse

A data warehouse is a database that is maintained independently of the business's operational database. In a data warehouse, updates are not made frequently.

- Its integrated historical data aids in the organization's business analysis.
- Executives may better organize, comprehend, and apply their data to make strategic decisions with the aid of a data warehouse.
- Systems for data warehouses facilitate the integration of many application systems.
- Consolidated historical data analysis is facilitated by a data warehouse system.

Features of Data Warehouse

The following is a discussion of a data warehouse's main features:



VMware Tutorial for Cloud Computing Aspirants

Published On: October 12, 2024

VMware Tutorial for Cloud Computing Aspirants VMware software allows you to run a virtual machine...



VBA Macros Tutorial for Beginners

Published On: October 10, 2024

VBA Macros Tutorial for Beginners VBA macros are programs that automate repetitive operations in Microsoft...



VB.Net Tutorial for Beginners

Published On: October 10, 2024

VB.Net Tutorial for Beginners Visual Basic is an easy-to-learn programming language that is type-safe. Get...

- **Subject-Oriented:** Information on a subject, as opposed to the organization's ongoing operations, is provided by a data warehouse, which makes it subject-oriented.
- **Integrated:** Information from disparate sources, including relational databases and flat files, is combined to create a data warehouse.
- **Time Variant:** A data warehouse's collected data is associated with a specific time frame. From a historical perspective, the information found in a data warehouse is useful.
- **Non-volatile** data indicates that when new data is added to it, the old data is kept intact. The operational database is kept apart from the data warehouse.

Applications of Data Warehouse

Data warehouses are extensively employed in the subsequent domains:

- Financial services
- Banking services
- Consumer goods
- Retail sectors
- Controlled manufacturing

Concepts of Data Warehousing

A data warehouse's worth of data can be utilized with the aid of decision support tools. Executives can use the warehouse more swiftly and efficiently with the aid of this technology. They are able to collect information, evaluate it, and make judgments using the warehouse's contents.

Any of the following domains can make use of the data collected in a warehouse:

Fine-tuning Production Strategies: Product portfolio management and product repositioning can be achieved by comparing quarterly or annual sales data.

Consumer analysis involves examining the purchasing habits, purchasing timing, budgetary cycles, and other relevant information about the consumer.

Operation analysis: Correction of the environment and customer relationship management are further benefits of operations analysis and data warehousing.

[Data Warehousing Syllabus PDF](#)

Heterogeneous Database Integration

We have two methods to integrate heterogeneous databases:

Query-driven Approach

This is how different datasets are traditionally integrated. Using this method, wrappers and integrators were constructed on top of many heterogeneous databases. Another name for these integrators is mediators.

Query-Driven Approach Process

- A metadata dictionary converts a query sent to the client side into a format suitable for each of the various heterogeneous sites involved.
- These inquiries are now sent to the nearby query processor.
- A global answer set incorporates the outcomes from diverse sources.

Limitations of a Query-Driven Approach

- A query-driven strategy necessitates intricate filtering and integration procedures.
- This strategy is incredibly ineffective.
- The cost of making repeated queries is high.
- Additionally, this method is highly costly for queries that need to be aggregated.

Update-driven Approach

The update-driven methodology used by modern data warehouse systems differs from the conventional methodology of the query-driven approach.

The information from several heterogeneous sources is merged ahead of time and kept in a warehouse when using an update-driven approach. You can conduct direct queries and analyses on this data.

Advantages of an Update-Driven Approach

The following are the benefits of this strategy:

- High performance is provided by this method.
- The data is pre-processed, merged, annotated, condensed, and reorganized in a semantic data repository.
- An interface is not necessary for query processing in order to handle data from local sources.

Functions of Data Warehouse Utilities and Tools

The following are the goals of data warehouse utilities and tools:

Data Extraction: It is the process of collecting information from various disparate sources.

Data Cleaning: It is the process of identifying and fixing mistakes in data.

Data Transformation: It entails transforming legacy data into warehouse format.

Data Loading: Sorting, summarizing, consolidating, ensuring integrity, and creating indices and partitions are all part of the data loading process.

Data Refreshing: It entails updating the warehouse with data from sources.

System Process of Data Warehousing

The operational databases will undergo a number of activities, and we have well-defined methods for providing a solution, like using normalized data and keeping tables short.

Process Flow in Data Warehouse

Four primary procedures are involved in creating a data warehouse:

- Extract and load the data.
- Cleaning and transforming the data.
- Backup and archive the data.
- Guiding users to the relevant data sources and handling their inquiries.

Extract and Load the Data

Data extraction uses the original systems' data. The extracted data is loaded into the data warehouse using data loading.

Controlling the Process: Determining when to begin data extraction and data consistency checks are key components of process control. It guarantees that the programs, logic modules, and tools are run in the right order and at the right time.

Loading the Data: The data is extracted, cleaned, and made consistent before being loaded into a temporary data store.

Clean and Transform Process

It is now time to perform cleaning and transforming after the data has been extracted and loaded into the temporary data store. The list of steps for cleaning and transforming is as follows:

- Clean and transform the loaded data into a structure.
- Partition the data
- Aggregation

Clean and transform the loaded data into a structure: Queries are sped up by cleaning and

modifying the loaded data. Making the data consistent is one way to achieve this.

- within of itself.
- Includes more information from the same data source.
- integrating information from different source systems.
- using the data that is currently in the warehouse.

The process of transforming entails giving the raw data a structure. Data structure lowers operating costs while improving query performance.

Partition of the Data: It will streamline data warehouse management and maximize hardware performance. Each fact table is now divided into several distinct sections.

Aggregation: To expedite frequently asked inquiries, aggregation is necessary. The basis for aggregation is the fact that the majority of frequently asked queries examine a portion or a collection of comprehensive data.

Warehouse Associate Salary

Backup and Archive the Data

Maintaining regular backups is essential to recovering data in the event of hardware, software, or data loss failure.

During the archiving process, outdated data is deleted from the system and kept in a way that makes it easily recoverable when needed.

Query Management Process

The following operations are carried out by this process:

- Controls the inquiries.
- Helps in accelerating the query execution time.

- Sends the queries to the data sources that will yield the best results.
- Makes certain that every system supply is utilized as efficiently as possible.
- Maintain real query profiles.

The warehouse management process uses the data produced by this process to choose which aggregations to produce. Usually, this process is not active when data is regularly loaded into the data warehouse.

Data Warehousing Architecture

The process of creating the general architecture of data communication, processing, and presentation that is available for end-client computing within the organization is known as a data warehouse architecture.

The architectures of data warehouses vary based on the circumstances around a business. There are three typical architectures:

- Basic
- With Staging Area
- With Staging Area and Data Marts

Basic Data Warehousing Architecture

It contains the following:

- **Operational System:** A data warehousing technique called “operational system” designates a system that handles an organization’s daily transactions.
- **Flat Files:** Transactional data is stored in a flat file system, which is a collection of files with unique names for each file.
- **Metadata:** The essential information about data is condensed into meta data, which can facilitate locating and interacting with specific instances of data.
- **Lightly and highly summarized data:** All of the preset, highly summarized (aggregated) data

created by the warehouse manager is stored in this section of the data warehouse.

- **End-User Access Tools:** A data warehouse's main objective is to give business management the information they need to make strategic decisions. These customers employ end-client access tools to communicate with the warehouse.

Data Warehouse Architecture: With Staging Area

Before putting your operational data into the warehouse, we need to clean and process it.

A temporary location where a record from source systems is copied is called the Data Warehouse Staging Area.

For operational methods originating from several source systems, a staging area streamlines the process of data cleansing and consolidation. This is particularly useful for enterprise data warehouses, which integrate all pertinent enterprise data.

Data Warehouse Architecture: With Staging Area and Data Marts

We could want to alter the architecture of our warehouse to accommodate different organizational groups.

Data Marts can be added to do this. A data mart is a subset of a data warehouse that contains information for analysis and reporting on a particular division, department, or business activity, such as sales, payroll, manufacturing, etc.

Types of Data Warehousing Architecture

Below are the three major types of data warehousing architecture:

Single Tier Architecture

In actuality, single-tier architecture is not frequently employed. It eliminates redundant data with the

intention of storing the least amount of data possible to accomplish this.

- This architecture's weakness is that it doesn't satisfy the need to keep analytical and transactional processes separate.
- Once the middleware has interpreted the analysis queries, they are agreed to operational data.
- Queries have an impact on transactional workloads in this way.

Two-Tier Architecture

It comprises four consecutive data flow stages, even though it is commonly referred to as a two-layer architecture to emphasize a separation between physically available sources and data warehouses:

- Source layer
- Data staging
- Data warehouse layer
- Analysis

Three-Tier Architecture

The source layer, which contains several source systems, the reconciliation layer, and the data warehouse layer, which contains both data warehouses and data marts, make up the three-tier architecture. The reconciliation layer sits between the source data and the data warehouse.

- The reconciled layer's primary benefit is that it establishes an enterprise-wide uniform reference data model.
- It distinguishes between the issues pertaining to data warehouse population and source data extraction and integration.
- Particularly beneficial are the extensive, enterprise-wide systems that adopt this architecture.

Understanding of OLAP (On-Line

Analytical Processing)

OLAP facilitates the multidimensional examination of company data and offers advanced data modeling, trend analysis, and complex estimating capabilities.

With fast, consistent, interactive access to a wide range of views of data that have been transformed from raw information to reflect the true dimensionality of the enterprise.

They will be perceived by the clients, analysts, managers, and executives who can obtain insight into information through the classification of software technology. It is known as OLAP.

Data Warehousing Training

How Does OLAP Operate?

OLAP is based on a very basic premise. Most queries, including aggregation, joining, and grouping, which are sometimes exceedingly difficult to run over tabular databases, are pre-calculated by it.

The process of 'building' or 'processing' the OLAP cube involves the calculation of these queries. The data will have been updated by the time end users report for work after this overnight process.

Types of OLAP Servers

There exist four varieties of OLAP servers.

- Relational OLAP (ROLAP)
- Multidimensional OLAP (MOLAP)
- Hybrid OLAP (HOLAP)
- Specialized SQL Servers

Relational OLAP

Relational backend servers and client front-end tools are positioned in front of ROLAP servers. ROLAP makes use of relational or extended-relational

DBMS for the storing and management of warehouse data.

ROLAP consists of the following:

- Application of navigation logic for aggregates.
- DBMS back-end optimization for everyone.
- Extra resources and equipment.

Multidimensional OLAP

MOLAP provides multidimensional representations of data through array-based multidimensional storage engines.

- If the data set is sparse, multidimensional data stores may have low storage utilization.
- To manage dense and sparse data sets, many MOLAP servers have two tiers of data storage representation.

Hybrid OLAP

ROLAP and MOLAP are combined to create a hybrid OLAP. It provides faster MOLAP computation and more scalability for ROLAP.

HOLAP servers have the capacity to store vast amounts of detailed data. The aggregations are separated in the MOLAP store.

Specialized SQL Servers

In a read-only setting, specialized SQL servers offer enhanced query language and query processing support for SQL queries across star and snowflake schemas.

OLAP Operations

We shall talk about OLAP operations with multidimensional data since OLAP servers are predicated on multidimensional views of data.

The list of OLAP operations is as follows:

- Roll-up

- Drill-down
- Slice and dice
- Pivot (rotate)

Roll-up

Roll-up can execute aggregation on a data cube via the subsequent methods:

- By moving up the dimension's idea hierarchy
- Through the process of dimension reduction

How Roll-up Works:

- Climbing up a notion hierarchy for the dimension location is how roll-up is done.
- "Street < city < province < country" was the initial hierarchy of concepts.
- As the data rolls up, it is aggregated by moving up the location hierarchy, from the municipal level to the national level.
- Instead of grouping the data by country, it groups it by city.
- One or more of the data cube's dimensions are eliminated during roll-up.

Drill-down

It is carried out in one of the two methods listed below:

- By decreasing the level of a dimension's idea hierarchy
- By adding a fresh perspective.

How does drill-down work?

- Stepping down a concept hierarchy for the dimension time is how drill-down is carried out.
- The hierarchy of concepts was initially "day < month < quarter < year."
- Drilling down causes the time dimension to drop from the quarter to the month level.
- One or more dimensions from the data cube are added during the drill-down process.
- It moves the data from less detailed to more

detailed information.

Slice

A new sub-cube is produced by the slice operation, which takes one specific dimension from a given cube.

How does slice work?

- In this case, Slice is carried out for the dimension "time" according to the formula $\text{time} = \text{"Q1"}$.
- It will choose one or more dimensions to create a new sub-cube.

Dice

A dice generates a new sub-cube by choosing two or more dimensions from an existing cube.

Three dimensions are involved in the dice operation on the cube based on the following selection criteria:

(location = "Toronto" or "Vancouver")

(time = "Q1" or "Q2")

(item = "Mobile" or "Modem")

Pivot

Rotation is another name for the pivot operation. To present the data in a different way, it rotates the axes that are visible.

Conclusion

We have covered the important aspects in this data warehousing tutorial. Unveil the opportunities in this field by enrolling in our [**data warehousing training in Chennai.**](#)

Share on your Social
Media



Softlogic Academy

Softlogic Systems

KK Nagar [Corporate Office]

No.10, PT Rajan Salai, K.K. Nagar, Chennai
– 600 078.

Landmark: Karnataka Bank Building

Phone: [+91 86818 84318](tel:+918681884318)

Email: enquiry@softlogicsys.in

Map: [Google Maps Link](#)

OMR

No. E1-A10, RTS Food Street
92, Rajiv Gandhi Salai (OMR),
Navalur, Chennai – 600 130.

Landmark: Adj. to AGS Cinemas

Phone: [+91 89256 88858](tel:+918925688858)

Email: info@softlogicsys.in

Map: [Google Maps Link](#)

Courses

Python

Software Testing

Full Stack Developer

Java

Power BI

Clinical SAS

Data Science

Embedded

Cloud Computing

Navigation

[About Us](#)

[Blog Posts](#)

[Careers](#)

[Contact](#)

[Placement Training](#)

[Corporate Training](#)

[Hire With Us](#)

[Job Seekers](#)

[SLA's Recently Placed Students](#)

[Reviews](#)

[Sitemap](#)

Important Links

[Disclaimer](#)

[Privacy Policy](#)

[Terms and Conditions](#)

Social Media Links



Review Sources

[Google](#)

[Trustpilot](#)

[Glassdoor](#)

[Mouthshut](#)

[Sulekha](#)

Hardware and Networking

VBA Macros

Mobile App Development

DevOps

Justdial

Ambitionbox

Indeed

Software Suggest

Sitejabber

Copyright © 2024 – Softlogic
Systems. All Rights Reserved

SLA™ is a trademark of Softlogic Systems, Chennai.
Unauthorised use prohibited.