

Learn Data Warehouse in 1 Day

By Krishna Rungta

Copyright 2019 - All Rights Reserved – Krishna Rungta

ALL RIGHTS RESERVED. No part of this publication may be reproduced or transmitted in any form whatsoever, electronic, or mechanical, including photocopying, recording, or by any informational storage or retrieval system without express written, dated and signed permission from the author.

Table Of Content

[Chapter 1: What Is Data Warehousing? Types, Definition & Example](#)

1. [What is Data Warehousing?](#)
2. [History of Datawarehouse](#)
3. [How Datawarehouse works?](#)
4. [Types of Data Warehouse](#)
5. [General stages of Data Warehouse](#)
6. [Components of Data warehouse](#)
7. [Who needs Data warehouse?](#)
8. [What Is a Data Warehouse Used For?](#)
9. [Steps to Implement Data Warehouse](#)
10. [Best practices to implement a Data Warehouse](#)
11. [Why We Need Data Warehouse? Advantages & Disadvantages](#)
12. [The Future of Data Warehousing](#)
13. [Data Warehouse Tools](#)

[Chapter 2: Database vs Data Warehouse: Key Differences](#)

1. [What is Database?](#)
2. [What is a Data Warehouse?](#)
3. [Why use a Database?](#)
4. [Why Use Data Warehouse?](#)
5. [Characteristics of Database](#)
6. [Characteristics of Data Warehouse](#)
7. [Difference between Database and Data Warehouse](#)
8. [Applications of Database](#)
9. [Applications of Data Warehousing](#)
10. [Disadvantages of Database](#)
11. [Disadvantages of Data Warehouse](#)

12. [What Works Best for You?](#)

Chapter 3: Data Warehouse Architecture, Concepts and Components

1. [What is Data warehouse?](#)
2. [Characteristics of Data warehouse](#)
3. [Data Warehouse Architectures](#)
4. [Datawarehouse Components](#)
5. [Query Tools](#)
6. [Data warehouse Bus Architecture](#)
7. [Data warehouse Architecture Best Practices](#)

Chapter 4: ETL (Extract, Transform, and Load) Process

1. [What is ETL?](#)
2. [Why do you need ETL?](#)
3. [ETL Process in Data Warehouses](#)
4. [Step 1\) Extraction](#)
5. [Step 2\) Transformation](#)
6. [Step 3\) Loading](#)
7. [ETL tools](#)
8. [Best practices ETL process](#)

Chapter 5: ETL vs ELT: Must Know Differences

1. [What is ETL?](#)
2. [What is ELT?](#)
3. [Difference between ETL vs. ELT](#)

Chapter 6: What is Data Modelling? Conceptual, Logical, & Physical Data Models

1. [What is Data Modelling?](#)
2. [Why use Data Model?](#)
3. [Types of Data Models](#)
4. [Conceptual Model](#)
5. [Logical Data Model](#)
6. [Physical Data Model](#)
7. [Advantages and Disadvantages of Data Model](#)

Chapter 7: What is OLAP (Online Analytical Processing): Cube, Operations & Types

1. [What is Online Analytical Processing?](#)
2. [OLAP cube:](#)
3. [Basic analytical operations of OLAP](#)
4. [Types of OLAP systems](#)
5. [ROLAP](#)
6. [MOLAP](#)
7. [Hybrid OLAP](#)
8. [Advantages of OLAP](#)
9. [Disadvantages of OLAP](#)

Chapter 8: What is MOLAP? Architecture, Advantages, Example, Tools

1. [What is MOLAP?](#)
2. [MOLAP Architecture](#)
3. [Implementation considerations is MOLAP](#)
4. [MOLAP Advantages](#)
5. [MOLAP Disadvantages](#)
6. [MOLAP Tools](#)

Chapter 9: OLTP vs OLAP: What's the Difference?

1. [What is OLAP?](#)
2. [What is OLTP?](#)
3. [Example of OLAP](#)
4. [Example of OLTP system](#)
5. [Benefits of using OLAP services](#)
6. [Benefits of OLTP method](#)
7. [Drawbacks of OLAP service](#)
8. [Drawbacks of OLTP method](#)
9. [Difference between OLTP and OLAP](#)

Chapter 10: What is Dimensional Model in Data Warehouse?

1. [What is Dimensional Model?](#)
2. [Elements of Dimensional Data Model](#)
3. [Steps of Dimensional Modelling](#)
4. [Rules for Dimensional Modelling](#)
5. [Benefits of dimensional modeling](#)

Chapter 11: Star and Snowflake Schema in Data Warehousing

1. [What is Multidimensional schemas?](#)
2. [What is a Star Schema?](#)
3. [What is a Snowflake Schema?](#)
4. [Star Vs Snowflake Schema: Key Differences](#)
5. [What is a Galaxy schema?](#)
6. [What is Star Cluster Schema?](#)

Chapter 12: Data Mart Tutorial: What is Data Mart, Types & Example

1. [What is Data Mart?](#)
2. [Why do we need Data Mart?](#)

3. [Type of Data Mart](#)
4. [Steps in Implementing a Datamart](#)
5. [Best practices for Implementing Data Marts](#)
6. [Advantages and Disadvantages of a Data Mart](#)

Chapter 13: Data Warehouse vs Data Mart: Know the Difference

1. [What is Data Warehouse?](#)
2. [What is Data Mart?](#)
3. [Differences between Data Warehouse and Data Mart](#)

Chapter 14: What is Data Lake? It's Architecture

1. [What is Data Lake?](#)
2. [Why Data Lake?](#)
3. [Data Lake Architecture](#)
4. [Key Data Lake Concepts](#)
5. [Maturity stages of Data Lake](#)
6. [Best practices for Data Lake Implementation:](#)
7. [Difference between Data lakes and Data warehouse](#)
8. [Benefits and Risks of using Data Lake](#)

Chapter 15: Data Lake vs Data Warehouse: What's the Difference?

1. [What is Data Warehouse?](#)
2. [What is Data Lake?](#)
3. [Data Warehouse Concept:](#)
4. [Data Lake Concept:](#)
5. [Key Difference between the Data Lake and Data Warehouse](#)

Chapter 16: What is Business Intelligence? Definition & Example

1. [What is Business Intelligence?](#)
2. [Why is BI important?](#)
3. [How Business Intelligence systems are implemented?](#)
4. [Examples of Business Intelligence System used in Practice](#)
5. [Four types of BI users](#)
6. [Advantages of Business Intelligence](#)
7. [BI System Disadvantages](#)
8. [Trends in Business Intelligence](#)

Chapter 17: Data Mining Tutorial: Process, Techniques, Tools, EXAMPLES

1. [What is Data Mining?](#)
2. [Types of Data](#)
3. [Data Mining Implementation Process](#)
4. [Business understanding:](#)
5. [Data understanding:](#)
6. [Data preparation:](#)
7. [Data transformation:](#)
8. [Modelling](#)
9. [Data Mining Techniques](#)
10. [Challenges of Implementation of Data mine:](#)
11. [Data mining Examples:](#)
12. [Data Mining Tools](#)
13. [Benefits of Data Mining:](#)
14. [Disadvantages of Data Mining](#)
15. [Data Mining Applications](#)

Chapter 18: DataStage Tutorial: Beginner's Training

1. [What is DataStage?](#)
2. [DataStage Overview](#)

3. [DataStage Components and Architecture](#)
4. [Pre-requisite for Datastage tool](#)
5. [Download and Installation InfoSphere Information Server](#)
6. [Process flow of Change data in a CDC Transaction stage Job.](#)
7. [Setting Up SQL Replication](#)
8. [Creating the SQL Replication objects](#)
9. [Creating the definition files to map CCD tables to DataStage](#)
10. [Starting Replication](#)
11. [How to create Projects in Datastage tool](#)
12. [How to import replication Jobs in Datastage and QualityStage Designer](#)
13. [Creating a data connection from DataStage to the STAGEDB database](#)
14. [Importing table definitions from STAGEDB into DataStage](#)
15. [Setting properties for the DataStage jobs](#)
16. [Compiling and running the DataStage jobs](#)
17. [Testing integration between SQL Replication and DataStage](#)

Chapter 19: Difference between Data Mining and Data Warehouse

1. [What is Data warehouse?](#)
2. [What Is Data Mining?](#)
3. [Data Mining Vs Data Warehouse: Key Differences](#)
4. [Why use Data Warehouse?](#)
5. [Why use Data mining?](#)

Chapter 20: Difference Between Fact Table and Dimension Table

1. [Fact Table:](#)
2. [Dimension table:](#)
3. [Difference between Dimension table vs. Fact table](#)

4. [Type of facts](#)
5. [Types of Dimensions:](#)

Chapter 21: Difference between Information and Data

1. [What is Data?](#)
2. [What is Information?](#)
3. [Data Vs. Information](#)
4. [DIKW \(Data Information Knowledge Wisdom\)](#)

Chapter 22: Teradata Tutorial: Learn Basics for Beginners

1. [What is Teradata?](#)
2. [History of Teradata:](#)
3. [Why Teradata?](#)
4. [Features of Teradata:](#)
5. [Teradata – Architecture](#)
6. [MPP vs. SMP](#)
7. [Teradata Warehouse Product Suite](#)
8. [Applications of Teradata:](#)
9. [Difference between Teradata and other RDBMS](#)

Chapter 1: What Is Data Warehousing?

What is Data Warehousing?

A data warehousing is defined as a technique for collecting and managing data from varied sources to provide meaningful business insights. It is a blend of technologies and components which aids the strategic use of data.

It is electronic storage of a large amount of information by a business which is designed for query and analysis instead of transaction processing. It is a process of transforming data into information and making it available to users in a timely manner to make a difference.

The decision support database (Data Warehouse) is maintained separately from the organization's operational database. However, the data warehouse is not a product but an environment. It is an architectural construct of an information system which provides users with current and historical decision support information which is difficult to access or present in the traditional operational data store.

The data warehouse is the core of the BI system which is built for data analysis and reporting.

You may know that a 3NF-designed database for an inventory system may have tables related to each other. For example, a report on

current inventory information can include more than 12 joined conditions. This can quickly slow down the response time of the query and report. A data warehouse provides a new design which can help to reduce the response time and helps to enhance the performance of queries for reports and analytics.

Data warehouse system is also known by the following name:

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



History of Datawarehouse

The Datawarehouse benefits users to understand and enhance their organization's performance. The need to warehouse data evolved as computer systems became more complex and needed to handle increasing amounts of Information. However, Data Warehousing is a not a new thing.

Here are some key events in evolution of Data Warehouse-

- 1960- Dartmouth and General Mills in a joint research project, develop the terms dimensions and facts.
- 1970- A Nielsen and IRI introduces dimensional data marts for retail sales.
- 1983- Tera Data Corporation introduces a database management system which is specifically designed for decision support
- Data warehousing started in the late 1980s when IBM worker Paul Murphy and Barry Devlin developed the Business Data Warehouse.
- However, the real concept was given by Inmon Bill. He was considered as a father of data warehouse. He had written about a variety of topics for building, usage, and maintenance of the warehouse & the Corporate Information Factory.

How Datawarehouse works?

A Data Warehouse works as a central repository where information arrives from one or more data sources. Data flows into a data

warehouse from the transactional system and other relational databases.

Data may be:

1. Structured
2. Semi-structured
3. Unstructured data

The data is processed, transformed, and ingested so that users can access the processed data in the Data Warehouse through Business Intelligence tools, SQL clients, and spreadsheets. A data warehouse merges information coming from different sources into one comprehensive database.

By merging all of this information in one place, an organization can analyze its customers more holistically. This helps to ensure that it has considered all the information available. Data warehousing makes data mining possible. Data mining is looking for patterns in the data that may lead to higher sales and profits.

Types of Data Warehouse

Three main types of Data Warehouses are:

1. Enterprise Data Warehouse:

Enterprise Data Warehouse is a centralized warehouse. It provides decision support service across the enterprise. It offers a unified approach for organizing and representing data. It also provide the ability to classify data according to the subject and give access according to those divisions.

2. Operational Data Store:

Operational Data Store, which is also called ODS, are nothing but data store required when neither Data warehouse nor OLTP systems support organizations reporting needs. In ODS, Data warehouse is refreshed in real time. Hence, it is widely preferred for routine activities like storing records of the Employees.

3. Data Mart:

A data mart is a subset of the data warehouse. It specially designed for a particular line of business, such as sales, finance, sales or finance. In an independent data mart, data can collect directly from sources.

General stages of Data Warehouse

Earlier, organizations started relatively simple use of data warehousing. However, over time, more sophisticated use of data warehousing begun.

The following are general stages of use of the data warehouse:

Offline Operational Database:

In this stage, data is just copied from an operational system to another server. In this way, loading, processing, and reporting of the copied data do not impact the operational system's performance.

Offline Data Warehouse:

Data in the Datawarehouse is regularly updated from the Operational Database. The data in Datawarehouse is mapped and transformed to

meet the Datawarehouse objectives.

Real time Data Warehouse:

In this stage, Data warehouses are updated whenever any transaction takes place in operational database. For example, Airline or railway booking system.

Integrated Data Warehouse:

In this stage, Data Warehouses are updated continuously when the operational system performs a transaction. The Datawarehouse then generates transactions which are passed back to the operational system.

Components of Data warehouse

Four components of Data Warehouses are:

Load manager: Load manager is also called the front component. It performs with all the operations associated with the extraction and load of data into the warehouse. These operations include transformations to prepare the data for entering into the Data warehouse.

Warehouse Manager: Warehouse manager performs operations associated with the management of the data in the warehouse. It performs operations like analysis of data to ensure consistency, creation of indexes and views, generation of denormalization and aggregations, transformation and merging of source data and archiving and baking-up data.

Query Manager: Query manager is also known as backend component. It performs all the operation operations related to the management of user queries. The operations of this Data warehouse components are direct queries to the appropriate tables for scheduling the execution of queries.

End-user access tools:

This is categorized into five different groups like 1. Data Reporting 2. Query Tools 3. Application development tools 4. EIS tools, 5. OLAP tools and data mining tools.

Who needs Data warehouse?

Data warehouse is needed for all types of users like:

- Decision makers who rely on mass amount of data Users
- who use customized, complex processes to obtain information from multiple data sources.
- It is also used by the people who want simple technology to access the data
- It also essential for those people who want a systematic approach for making decisions.
- If the user wants fast performance on a huge amount of data which is a necessity for reports, grids or charts, then Data warehouse proves useful.
- Data warehouse is a first step If you want to discover 'hidden patterns' of data-flows and groupings.

What Is a Data Warehouse Used For?

Here, are most common sectors where Data warehouse is used:

Airline:

In the Airline system, it is used for operation purpose like crew assignment, analyses of route profitability, frequent flyer program promotions, etc.

Banking:

It is widely used in the banking sector to manage the resources available on desk effectively. Few banks also used for the market research, performance analysis of the product and operations.

Healthcare:

Healthcare sector also used Data warehouse to strategize and predict outcomes, generate patient's treatment reports, share data with tie-in insurance companies, medical aid services, etc.

Public sector:

In the public sector, data warehouse is used for intelligence gathering. It helps government agencies to maintain and analyze tax records, health policy records, for every individual.

Investment and Insurance sector:

In this sector, the warehouses are primarily used to analyze data patterns, customer trends, and to track market movements.

Retain chain:

In retail chains, Data warehouse is widely used for distribution and marketing. It also helps to track items, customer buying pattern, promotions and also used for determining pricing policy.

Telecommunication:

A data warehouse is used in this sector for product promotions, sales decisions and to make distribution decisions.

Hospitality Industry:

This Industry utilizes warehouse services to design as well as estimate their advertising and promotion campaigns where they want to target clients based on their feedback and travel patterns.

Steps to Implement Data Warehouse

The best way to address the business risk associated with a Datawarehouse implementation is to employ a three-prong strategy as below

1. **Enterprise strategy:** Here we identify technical including current architecture and tools. We also identify facts, dimensions, and attributes. Data mapping and transformation is also passed.
2. **Phased delivery:** Datawarehouse implementation should be phased based on subject areas. Related business entities like booking and billing should be first implemented and then integrated with each other.
3. **Iterative Prototyping:** Rather than a big bang approach to implementation, the Datawarehouse should be developed and tested iteratively.

Here, are key steps in Datawarehouse implementation along with its deliverables.

Step	Tasks	Deliverables
1	Need to define project scope	Scope Definition
2	Need to determine business needs	Logical Data Model
3	Define Operational Datastore requirements	Operational Data Store Model
4	Acquire or develop Extraction tools	Extract tools and Software
5	Define Data Warehouse Data requirements	Transition Data Model
6	Document missing data	To Do Project List
7	Maps Operational Data Store to Data Warehouse	D/W Data Integration Map
8	Develop Data Warehouse Database design	D/W Database Design
9	Extract Data from Operational Data Store	Integrated D/W Data Extracts
10	Load Data Warehouse	Initial Data Load
11	Maintain Data Warehouse	On-going Data Access and Subsequent Loads

Best practices to implement a Data Warehouse

- Decide a plan to test the consistency, accuracy, and integrity of the data.
- The data warehouse must be well integrated, well defined and time stamped.
- While designing Datawarehouse make sure you use right tool, stick to life cycle, take care about data conflicts and ready to learn you're your mistakes.
- Never replace operational systems and reports
- Don't spend too much time on extracting, cleaning and loading data.
- Ensure to involve all stakeholders including business personnel in Datawarehouse implementation process. Establish that Data warehousing is a joint/ team project. You don't want to create Data warehouse that is not useful to the end users.
- Prepare a training plan for the end users.

Why We Need Data Warehouse? Advantages & Disadvantages

Advantages of Data Warehouse:

- Data warehouse allows business users to quickly access critical data from some sources all in one place.
- Data warehouse provides consistent information on various cross-functional activities. It is also supporting ad-hoc reporting and query.

- Data Warehouse helps to integrate many sources of data to reduce stress on the production system.
- Data warehouse helps to reduce total turnaround time for analysis and reporting.
- Restructuring and Integration make it easier for the user to use for reporting and analysis.
- Data warehouse allows users to access critical data from the number of sources in a single place. Therefore, it saves user's time of retrieving data from multiple sources.
- Data warehouse stores a large amount of historical data. This helps users to analyze different time periods and trends to make future predictions.

Disadvantages of Data Warehouse:

- Not an ideal option for unstructured data.
- Creation and Implementation of Data Warehouse is surely time confusing affair.
- Data Warehouse can be outdated relatively quickly
- Difficult to make changes in data types and ranges, data source schema, indexes, and queries.
- The data warehouse may seem easy, but actually, it is too complex for the average users.
- Despite best efforts at project management, data warehousing project scope will always increase.
- Sometime warehouse users will develop different business rules.
- Organisations need to spend lots of their resources for training and Implementation purpose.

The Future of Data Warehousing

- Change in **Regulatory constraints** may limit the ability to combine source of disparate data. These disparate sources may include unstructured data which is difficult to store.
- As the **size** of the databases grows, the estimates of what constitutes a very large database continue to grow. It is complex to build and run data warehouse systems which are always increasing in size. The hardware and software resources available today do not allow to keep a large amount of data online. **Multimedia data** cannot be easily manipulated as text data, whereas textual information can be retrieved by the relational software available today. This could be a research subject.

Data Warehouse Tools

There are many Data Warehousing tools available in the market. Here, are some most prominent one:

1. MarkLogic:

MarkLogic is useful data warehousing solution that makes data integration easier and faster using an array of enterprise features. This tool helps to perform very complex search operations. It can query different types of data like documents, relationships, and metadata.

<http://developer.marklogic.com/products>

2. Oracle:

Oracle is the industry-leading database. It offers a wide range of choice of data warehouse solutions for both on-premises and in the cloud. It helps to optimize customer experiences by increasing operational efficiency.

<https://www.oracle.com/index.html>

3. Amazon RedShift:

Amazon Redshift is Data warehouse tool. It is a simple and cost-effective tool to analyze all types of data using standard SQL and existing BI tools. It also allows running complex queries against petabytes of structured data, using the technique of query optimization.

https://aws.amazon.com/redshift/?nc2=h_m1

Here is a complete list of useful Datawarehouse Tools.

Conclusion:

- The data warehouse works as a central repository where information is coming from one or more data sources.
- Three main types of Data warehouses are Enterprise Data Warehouse, Operational Data Store, and Data Mart.
- General state of a datawarehouse are Offline Operational Database, Offline Data Warehouse, Real time Data Warehouse and Integrated Data Warehouse.

- Four main components of Datawarehouse are Load manager, Warehouse Manager, Query Manager, End-user access tools
- Datawarehouse is used in diverse industries like Airline, Banking, Healthcare, Insurance, Retail etc.
- Implementing Datawarehouse is a 3 prong strategy viz. Enterprise strategy, Phased delivery and Iterative Prototyping.
- Data warehouse allows business users to quickly access critical data from some sources all in one place.

Buy Now \$9.99