

Data Warehouse fundamentals - a beginner's guide

You will learn about

Data warehouse

ETL / ELT

Facts

Snowflake Schema

Star Schema

Data mart

Databases

Slowly Changing
Dimensions

Dimensions

Dimensional Modelling

and much more...

handwritten by

Srihari Sridharan

Based on books and courses on Data Warehousing and Dimensional Modelling

DATA WAREHOUSE FUNDAMENTALS - A BEGINNER'S GUIDE

1. Data Warehousing concepts

a. Introduction to Data Warehousing concepts.

What is DWH? Not same as Database

DWH usage DB platform

↖ Data comes from elsewhere

↘ Possibly a dozen of them.

4 Rules of Data is copied.

- DWH →
- Integrated environment
 - Subject oriented
 - Time variant (historical data)
 - Non volatile (DWH remains stable as it is refreshed)
- Improvement in data quality.

6 Reasons to build DWH

Data driven decisions

Past

Present

Future

The unknown

BI

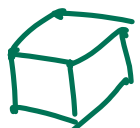
Business Intelligence

One stop shopping

Before DWH
We need to get or explore more sources.

BI + DWH = Best Value

③ Compare DWH and Data Lake

Traditional Data Warehousing
built on top of
RDBMS  → also on top of 

Data Lake Built on top of
Big Data

The 3V's

Differences

Volume	(amount of data)
Velocity	(takes new and changed data rapidly)
Variety	(structured and unstructured) Semi-structured.

DWH $\xrightarrow{\text{next gen}}$ DL Some companies have both DWH & DL

Blurring Differences

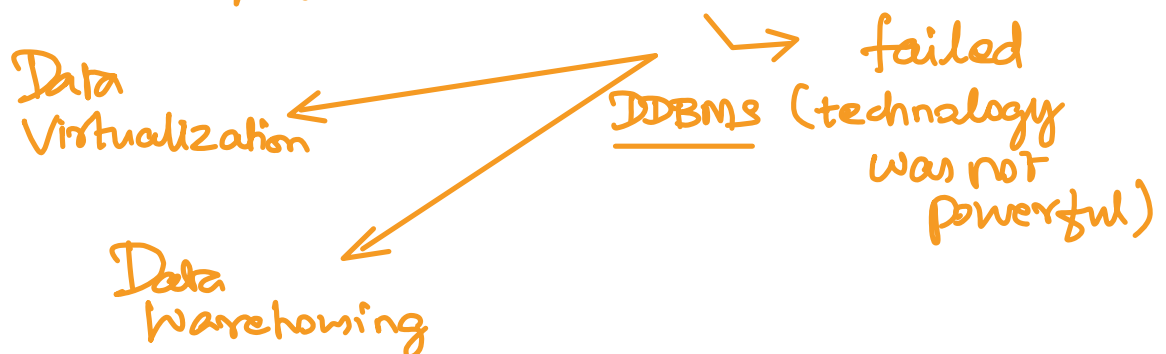
BI + DWH/DL = Business Value

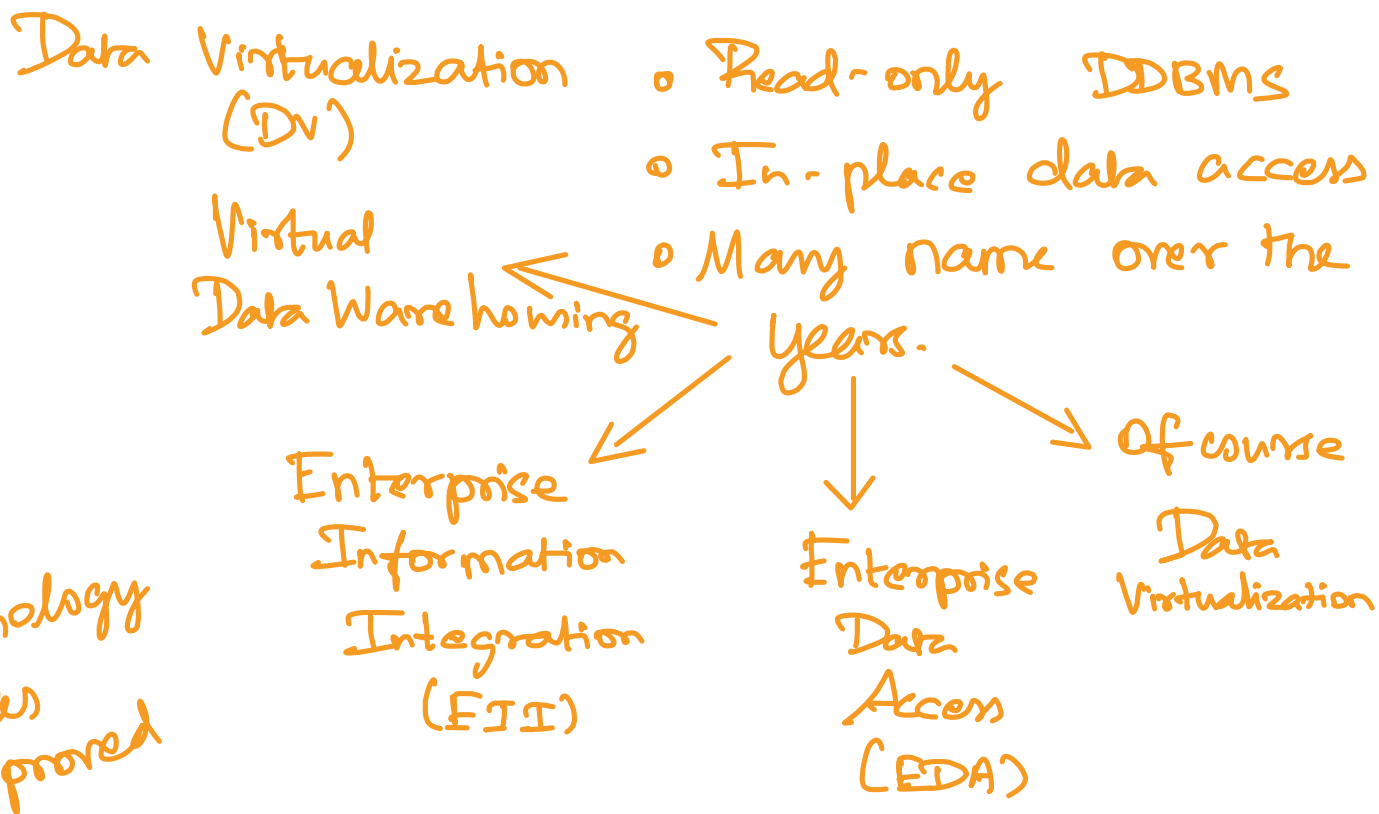
④ DWH and Data Virtualization

DWH History

1980's era extract files.

late 1980's Distributed DBMS



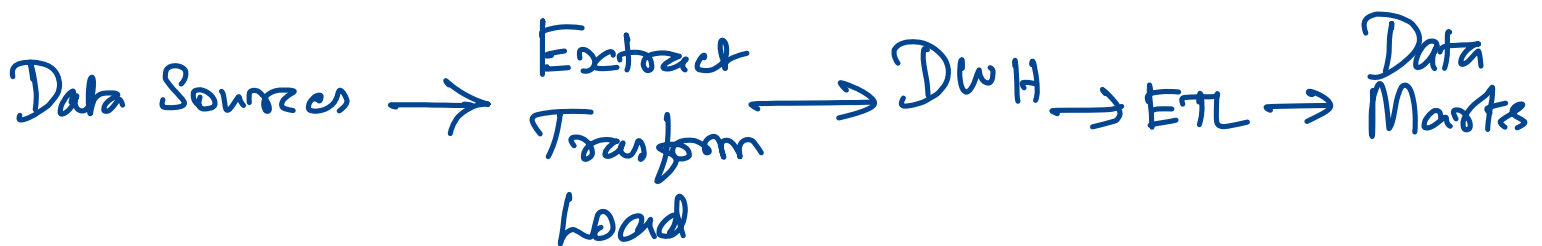


- DV use cases
- Simple transformations
 - Smaller number of data sources
 - Reduced response time

BI + DWH/DL/DV = Best Value.

© Simple End to End DWH Environment

Typical DWH



© Suppliers → Wholesalers → Retailers