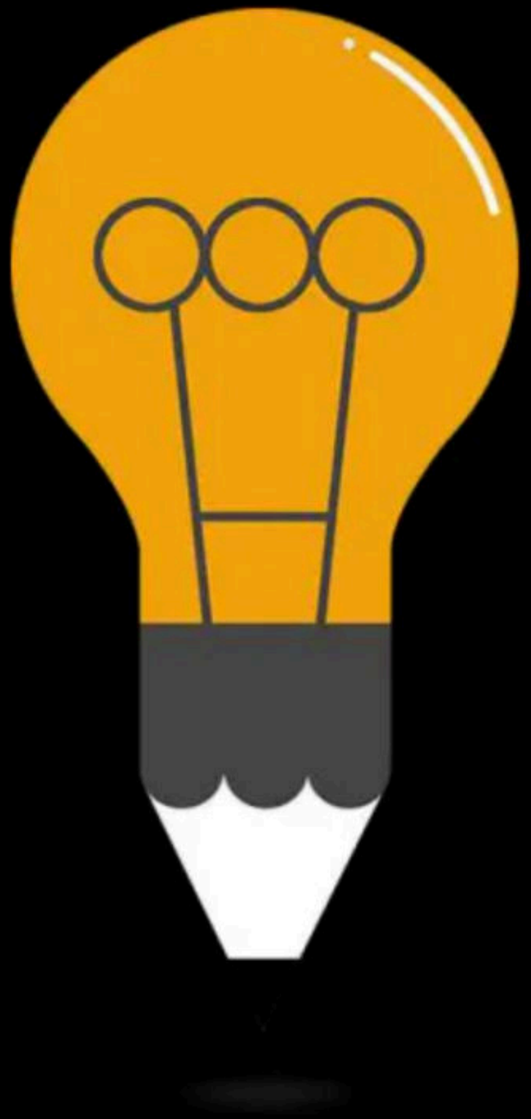
unacademy

Data Warehousing Basics - Part II

Complete Course on Data Warehousing

Vishvadeep Gothi • Lesson 2 • Nov 20, 2023

unacademy



Data Warehousing
Architecture & Schemas

By: Vishvadeep Gothi

unacademy

DBMS

A database-management system (DBMS) is a collection of interrelated data and a set of programs to access those data.

unacademy

Data Warehousing

In computing, a data warehouse, also known as an enterprise data warehouse, is a system used for reporting and data analysis and is considered a core component of business intelligence.

unacademy

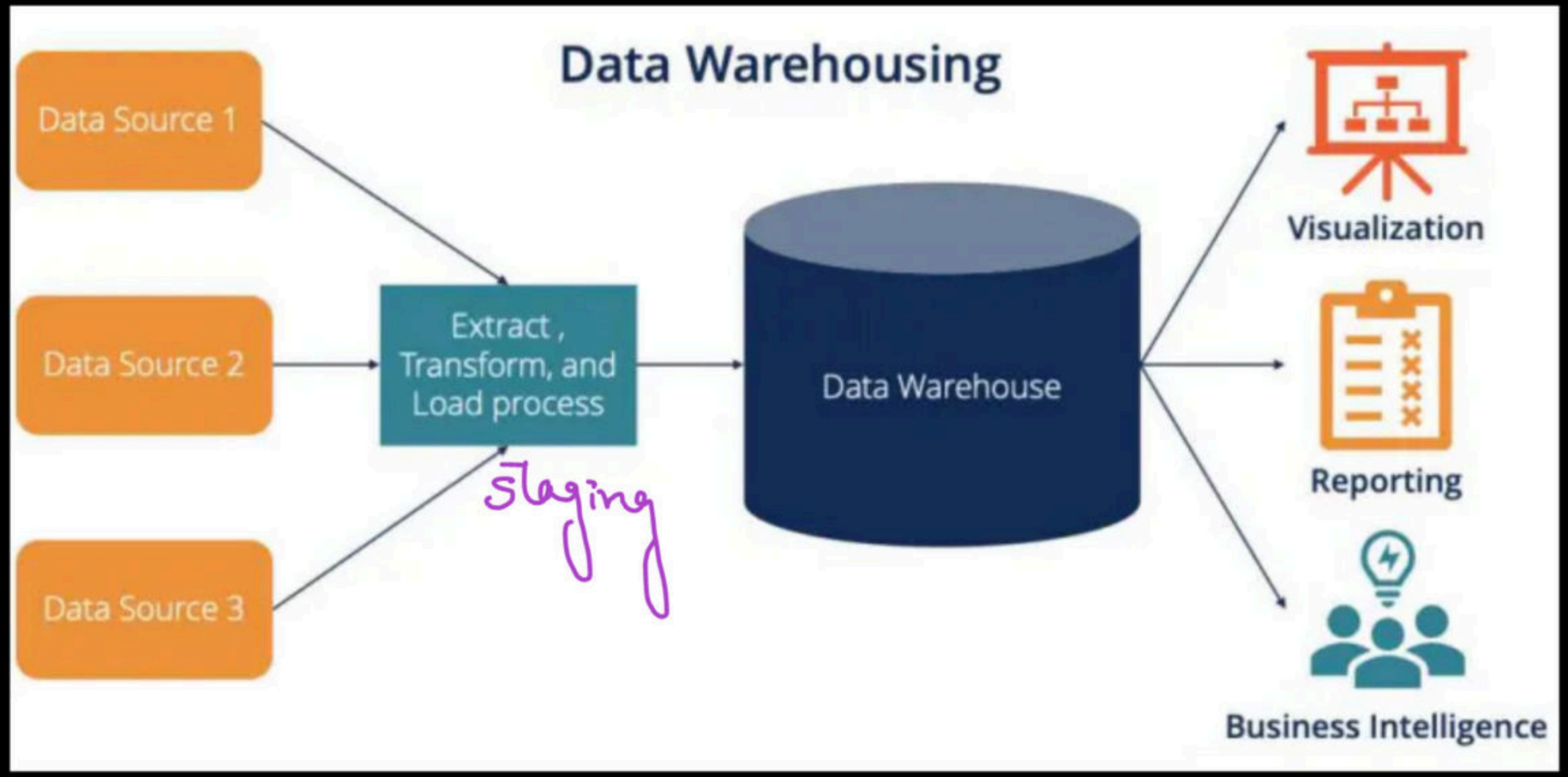
Data Warehousing

In computing, a data warehouse, also known as an enterprise data warehouse, is a system used for reporting and data analysis and is considered a core component of business intelligence.

Data warehouses are central repositories of integrated data from one or more disparate sources. They store current and historical data in one single place

unacademy

Data Warehousing



unacademy

ETL (Extract, Transform, Load)

Three-phase process where data is extracted, transformed and loaded into an output data container.

unacademy

ETL (Extract, Transform, Load)

Three-phase process where data is extracted, transformed and loaded into an output data container.

Done by application software but can be done manually also

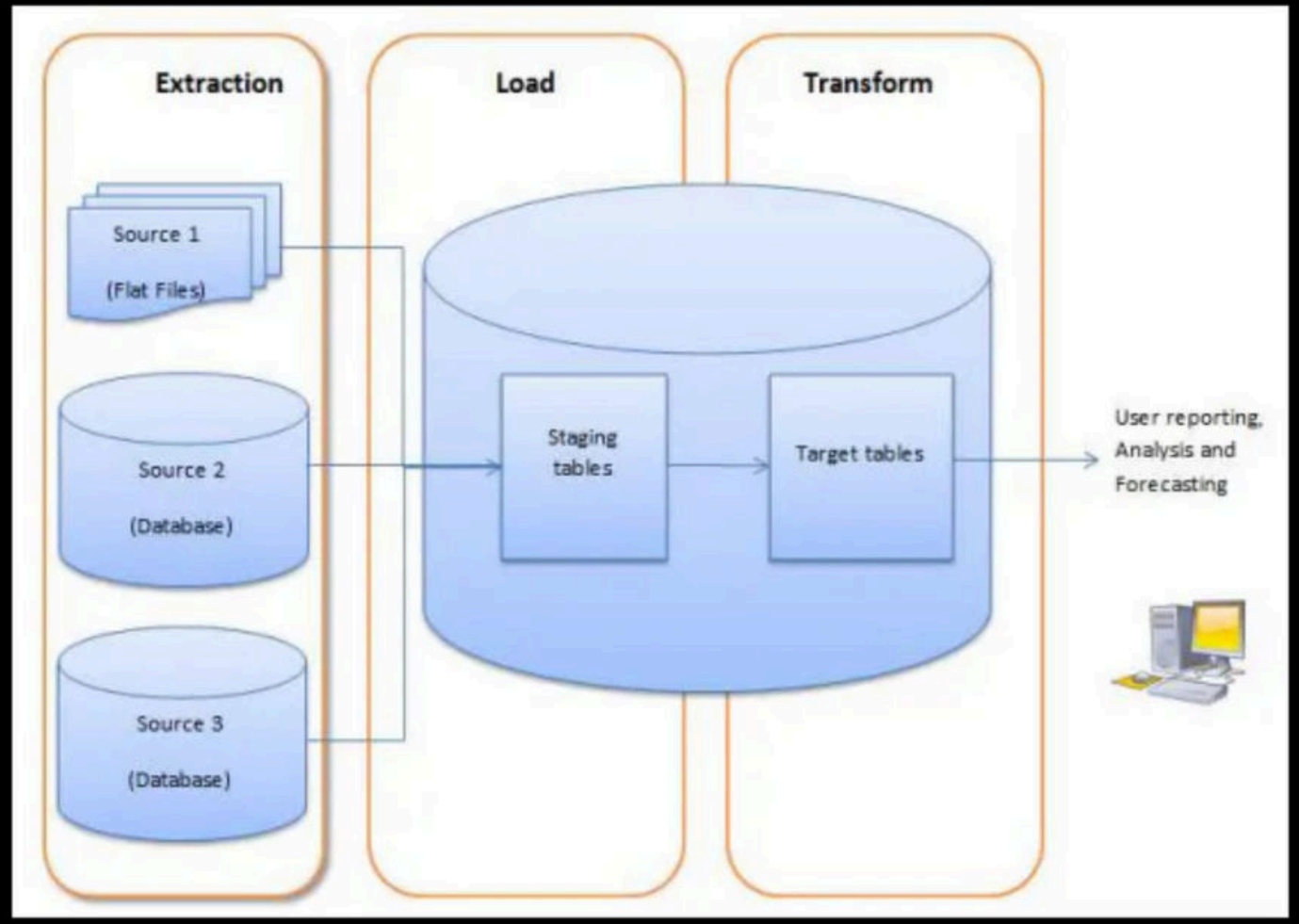
unacademy

Data Warehousing

- ETL Based: Uses Staging
- ELT Based: gets rid of a separate ETL tool for data transformation. Instead, it maintains a staging area inside the data warehouse itself.

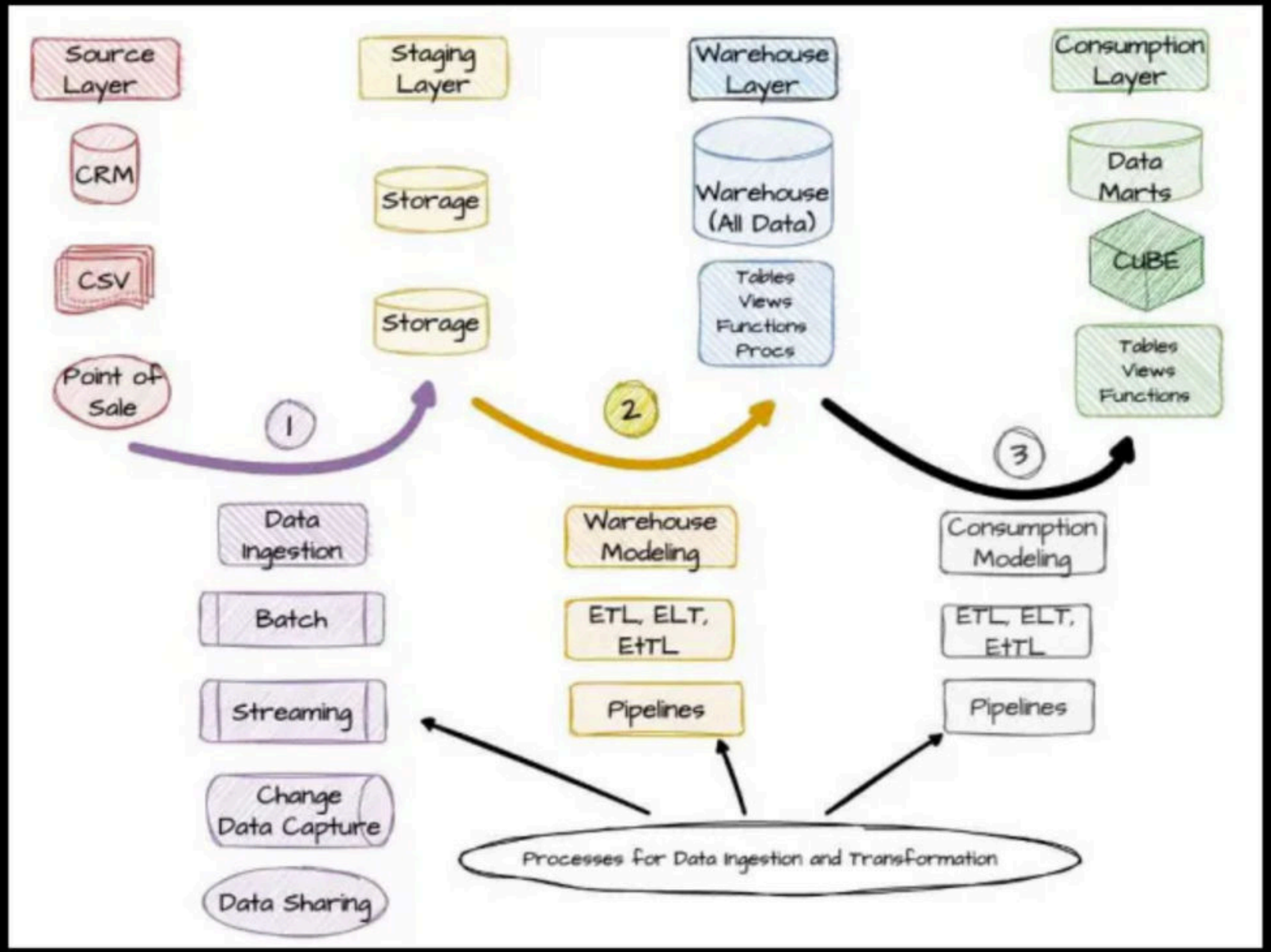
unacademy

ELT Based Data Warehousing



unacademy

Data Warehousing Architecture



unacademy

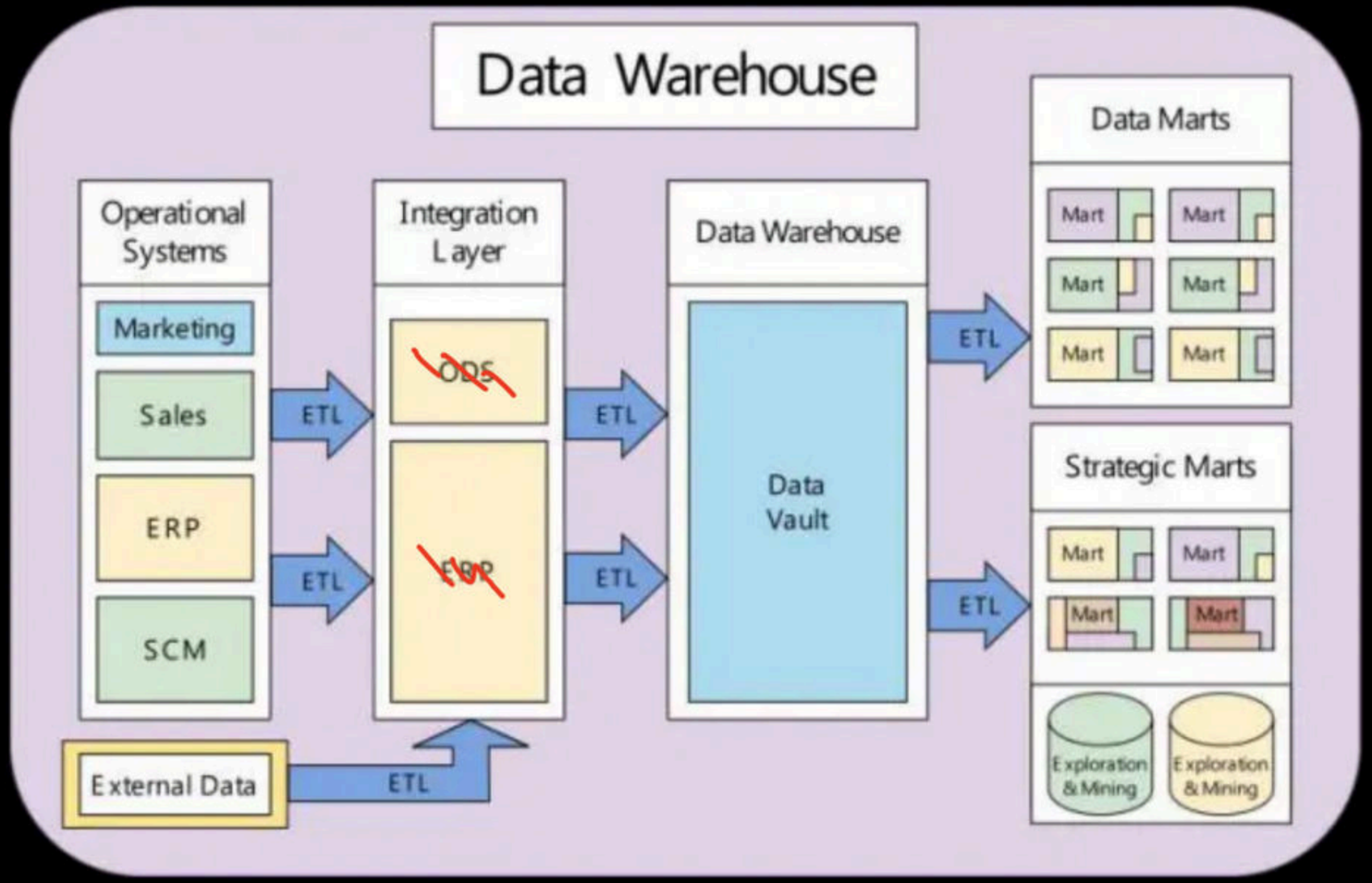
Data Mart

Simple form of a data warehouse, focused on a single subject (or functional area), hence they draw data from a limited number of sources such as sales, finance or marketing.

Data marts are often built and controlled by a single department within an organization.

unacademy

Data Mart



Data Warehousing vs Data Mart

Data warehouse:

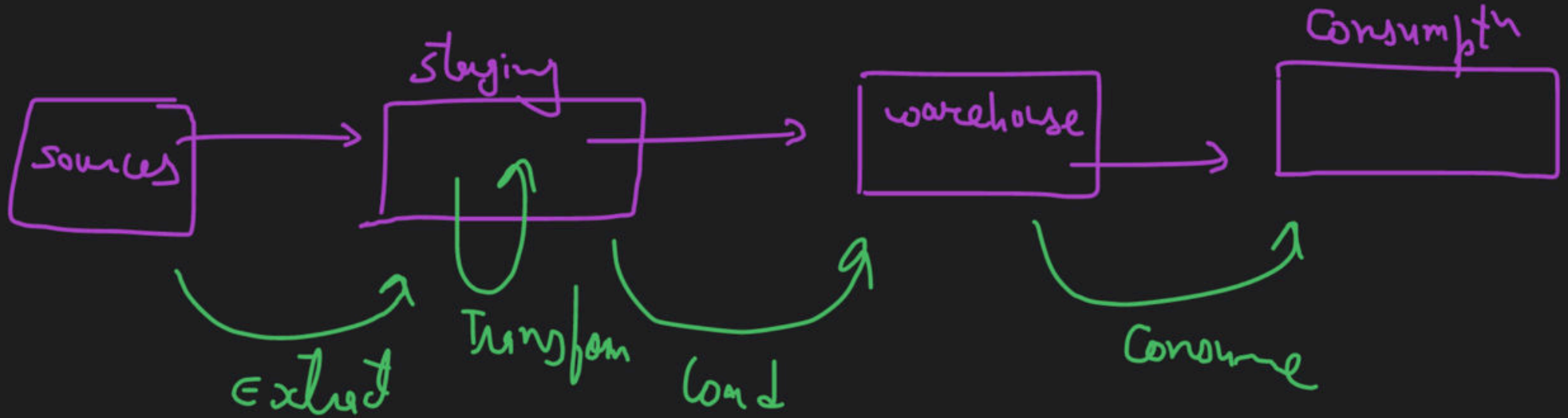
- Holds multiple subject areas
- Holds very detailed information
- Works to integrate all data sources
- Does not necessarily use a dimensional model but feeds dimensional models

Data mart:

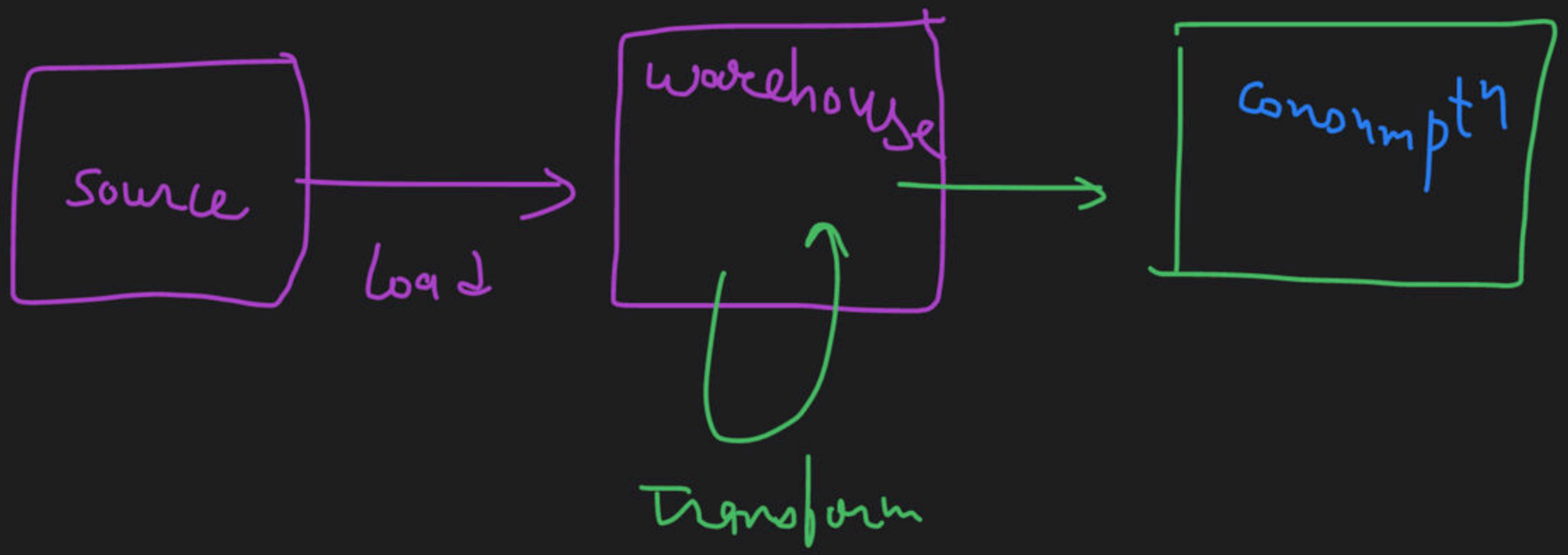
- Often holds only one subject area- for example, Finance, or Sales
- May hold more summarized data (although it may hold full detail)
- Concentrates on integrating information from a given subject area or set of source systems
- Is built focused on a dimensional model using a star schema.

3- Tier architecture

ETL based warehouse



2-Tier architecture :-

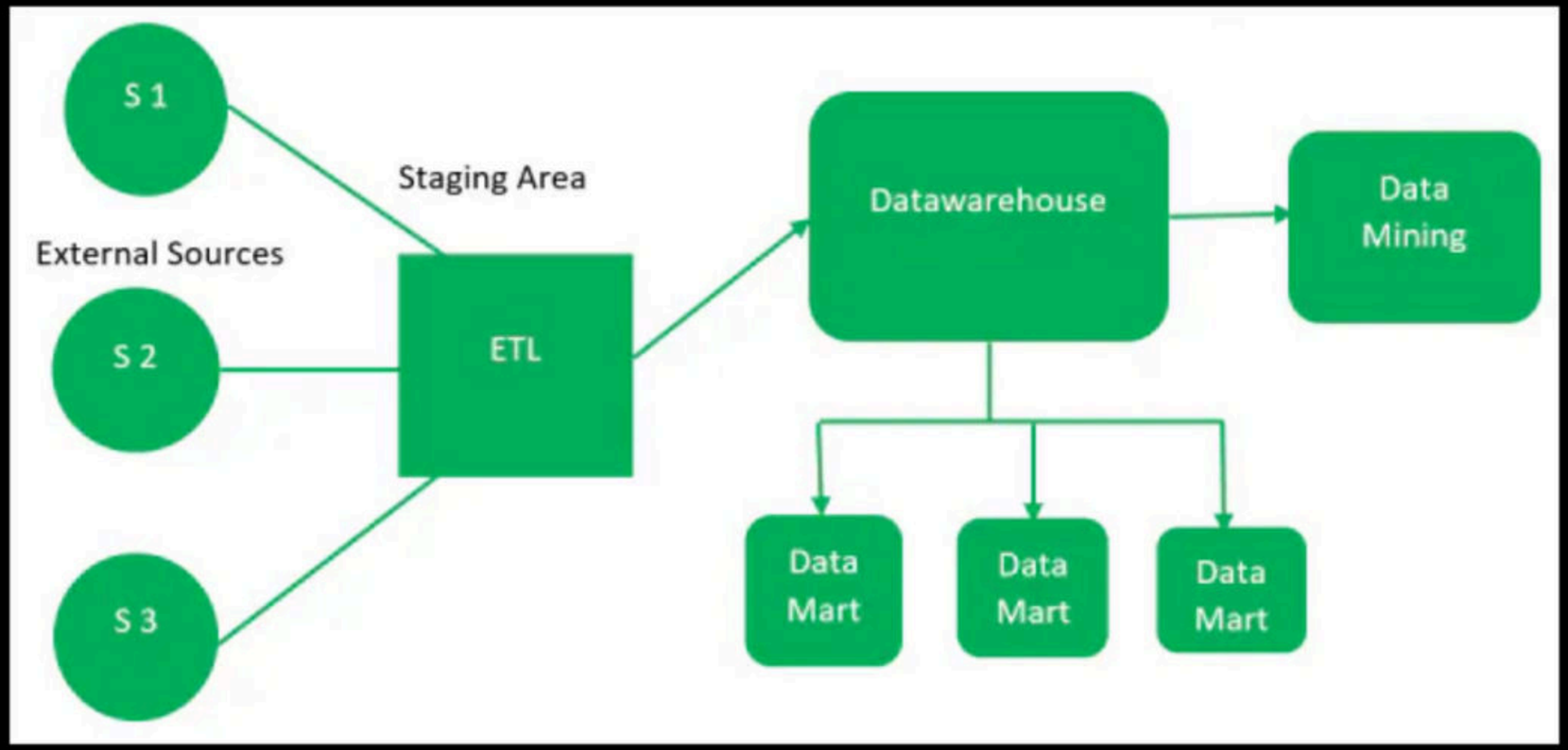


Data Warehousing Architecture

- Top-down approach
- Bottom-up approach

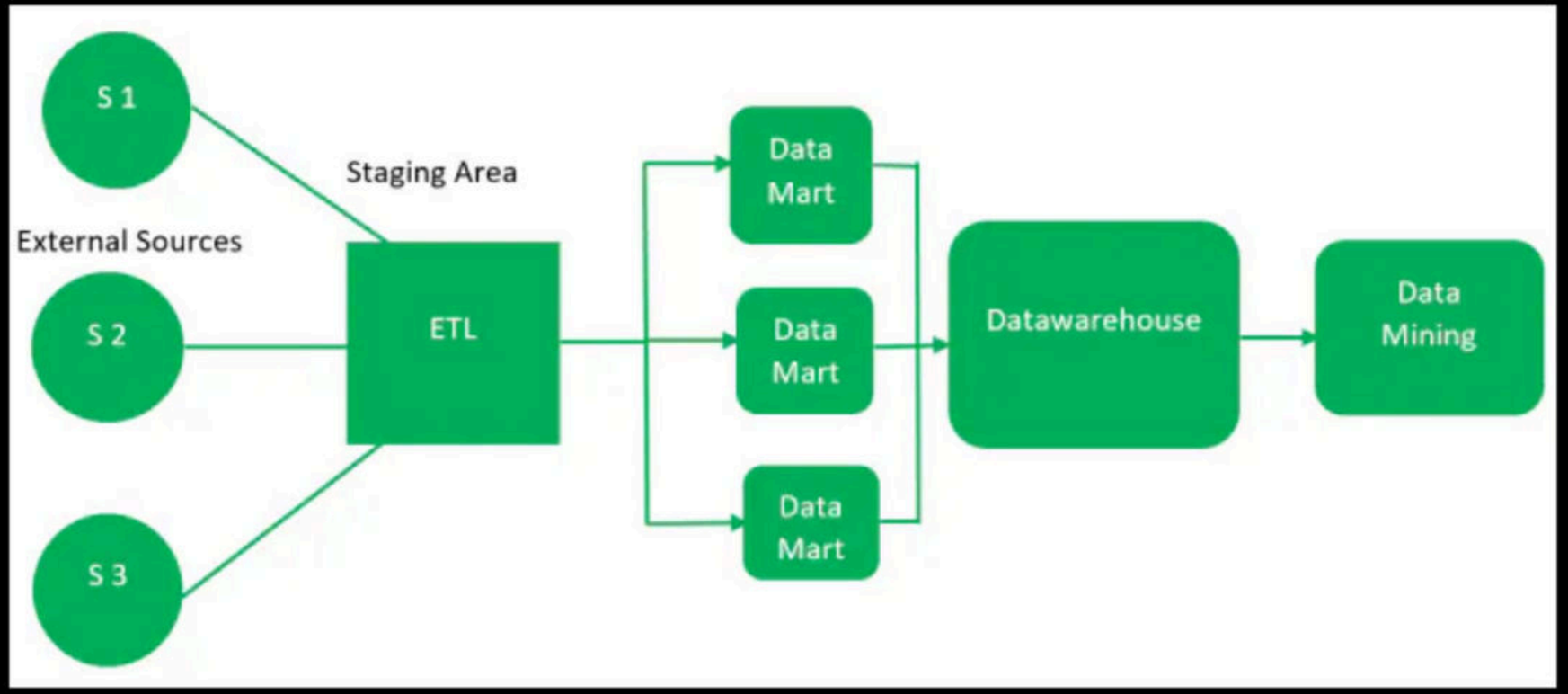
unacademy

Top-down approach



unacademy

Bottom-up approach



Online Transaction Processing (OLTP)

Type of data processing that consists of executing a number of transactions occurring concurrently.

on Database

Examples:

Online banking, shopping, order entry, or sending text messages

---

Database Transactn:-

series of db access statements

Online Analytical Processing (OLAP)

Software for performing multidimensional analysis at high speeds on large volumes of data from a data warehouse, data mart, or some other unified, centralized data store

OLAP cubes

Collection of data in multidimensional format

unacademy

Year wise financials

	Mar 2012	Mar 2013	Mar 2014	Mar 2015	Mar 2016	Mar 2017	Mar 2018	Mar 2019	Mar 2020	Mar 2021	Mar 2022	Mar 2023
Sales +	38	47	64	90	124	289	411	536	625	723	906	1,073
Expenses +	35	44	61	97	130	291	387	481	534	601	770	892
Operating Profit	**3**	**3**	**3**	**-7**	**-6**	**-2**	**24**	**54**	**91**	**123**	**136**	**182**
OPM %	7%	6%	5%	-8%	-4%	-1%	6%	10%	15%	17%	15%	17%
Other Income +	0	0	0	2	1	1	-1	1	2	4	11	10
Interest	1	1	2	1	0	3	4	5	6	4	1	1
Depreciation	1	1	1	1	1	14	17	17	20	24	31	34
Profit before tax	**1**	**1**	**1**	**-7**	**-6**	**-18**	**2**	**34**	**67**	**98**	**114**	**156**
Tax %	33%	32%	40%	0%	0%	10%	-14%	20%	29%	28%	26%	15%
Net Profit +	**1**	**1**	**0**	**-7**	**-6**	**-16**	**3**	**27**	**48**	**71**	**84**	**132**

unacademy

OLTP vs OLAP

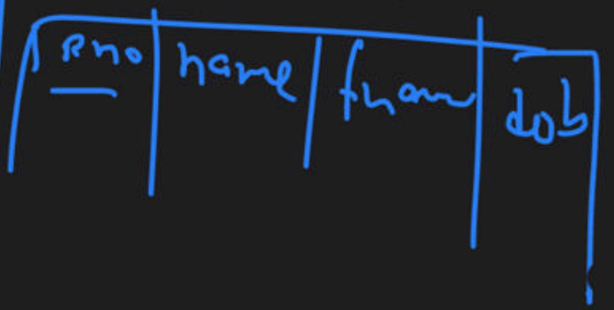
OLTP systems	OLAP systems
Enable the real-time execution of large numbers of database transactions by large numbers of people	Usually involve querying many records (even all records) in a database for analytical purposes
Require lightning-fast response times	Require response times that are orders of magnitude slower than those required by OLTP
Modify small amounts of data frequently and usually involve a balance of reads and writes	Do not modify data at all; workloads are usually read-intensive
Use indexed data to improve response times	Store data in columnar format to allow easy access to large numbers of records
Require frequent or concurrent database backups	Require far less frequent database backup
Require relatively little storage space	Typically have significant storage space requirements, because they store large amounts of historical data
Usually run simple queries involving just one or a few records	Run complex queries involving large numbers of records

Relational Database schema

design of d.b.

4 tables

student



Books

Datawarehouse Design Schema

Description, represented by objects such as tables and indexes, of how data relates logically within a data warehouse

unacademy

Datawarehouse Schema

1. Star Schema
2. Snowflake Schema
3. Fact Constellation Schema

normalizat n

unacademy

Dimension Table Table

Dimension Tables hold the descriptive information for all related fields that are included in the fact table's records.

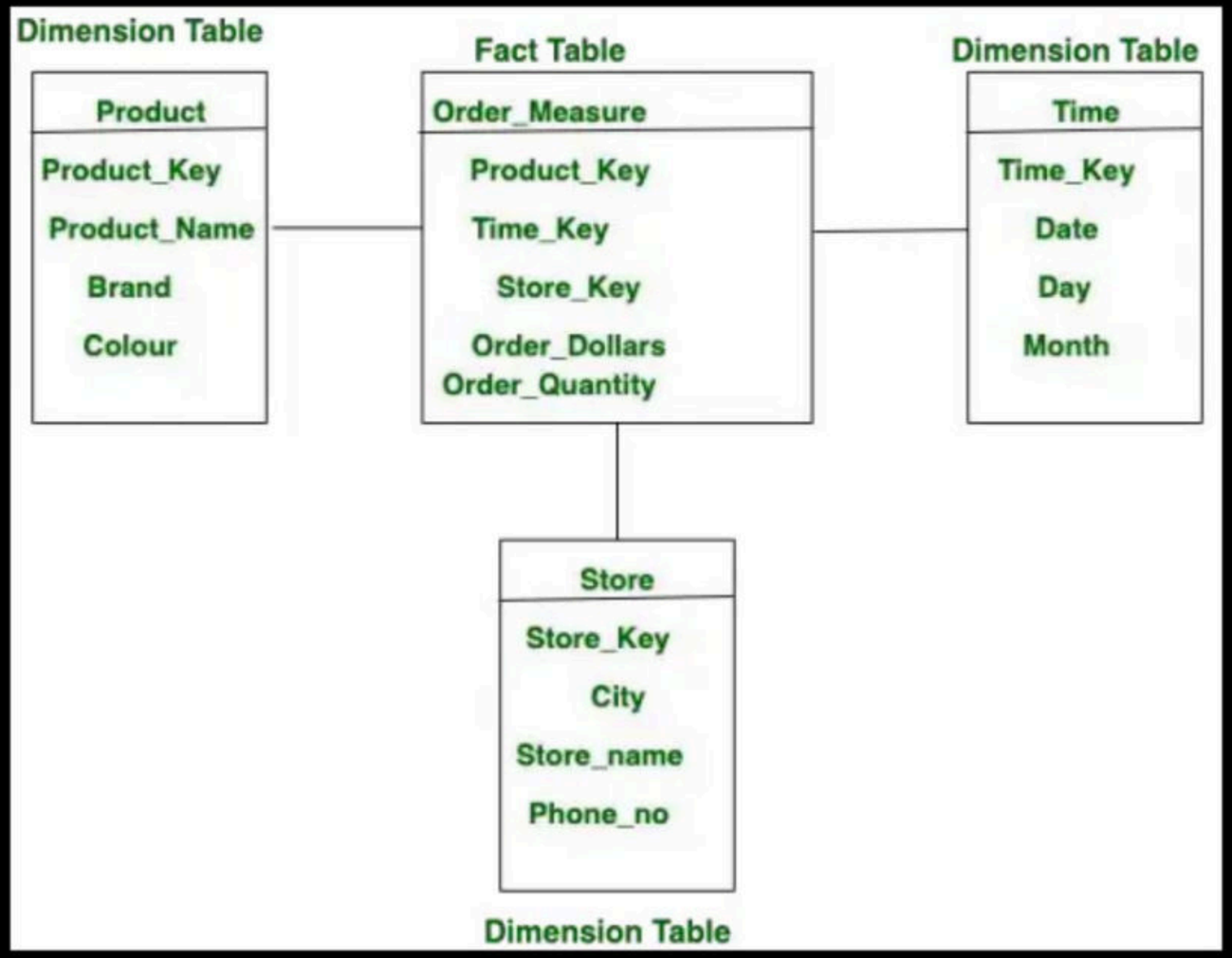
unacademy

Fact Table

A fact table aggregates metrics, measurements, or facts about business processes

unacademy

Fact Table



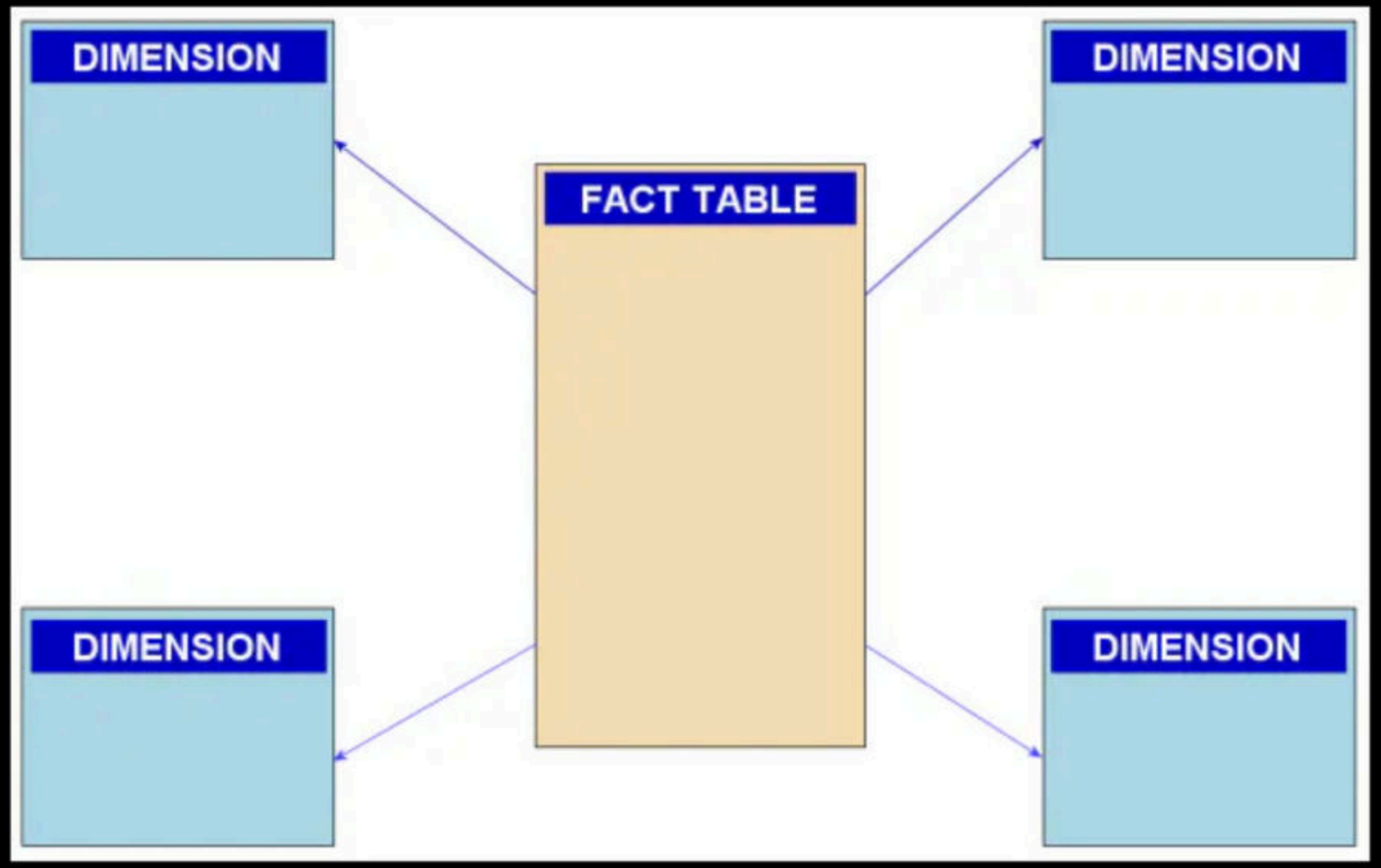
unacademy

Star Schema

Consists of one or more fact tables referencing any number of dimension tables

unacademy

Star Schema

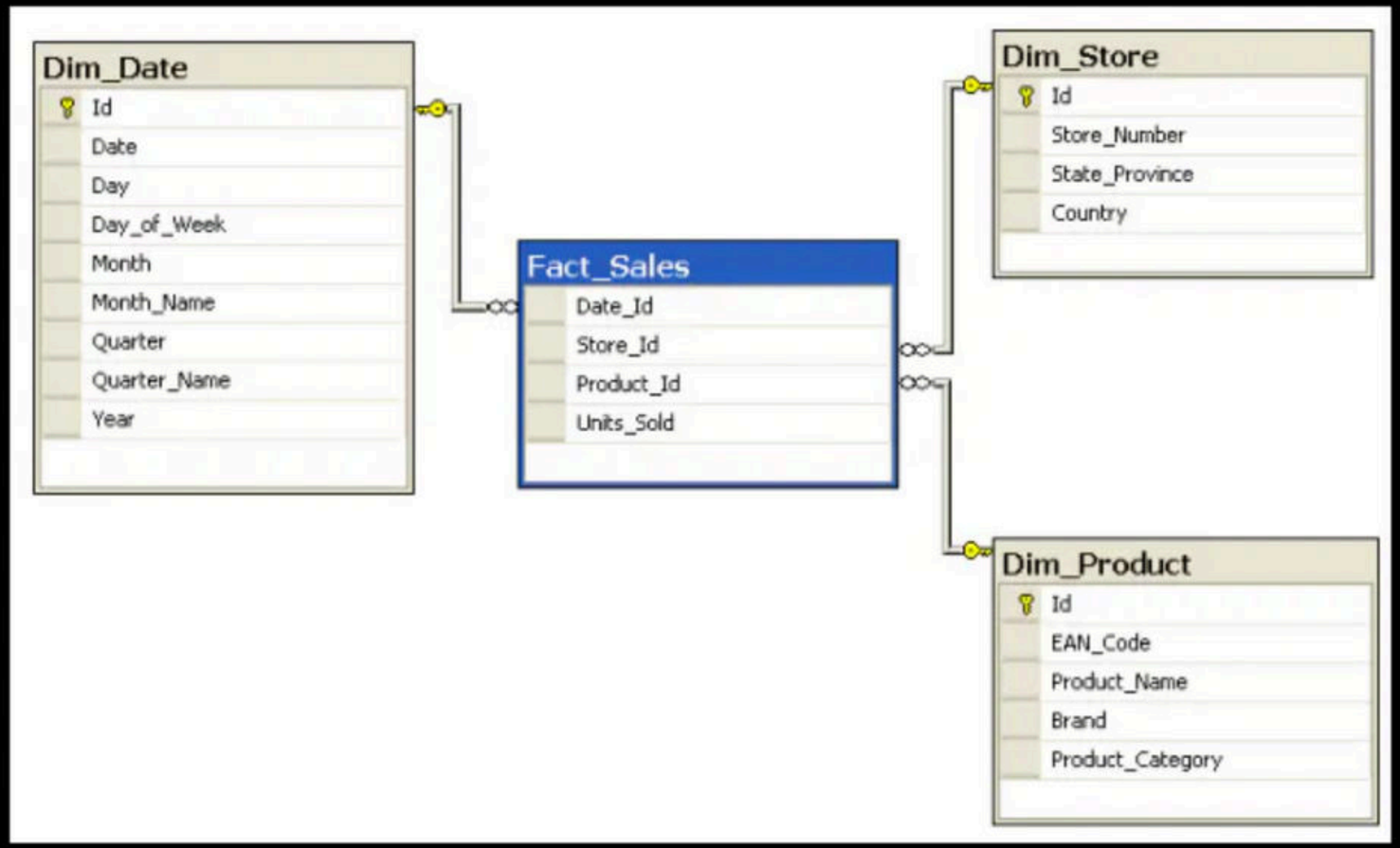


unacademy

Star Schema

unacademy

Star Schema



unacademy

Star Schema

- Star schemas are denormalized
- The primary key in the dimension table is joined to the fact table by the foreign key
- Each dimension in the star schema maps to one dimension table
- Dimension tables within a star scheme are not to be connected directly
- Star schema creates denormalized dimension tables

unacademy

Star Schema

Benefits:

- Simpler queries
- Simplified business reporting logic
- Query performance gains
- Fast aggregations
- Feeding cubes

unacademy

Snowflake Schema

logical arrangement of tables in a multidimensional database such that the entity relationship diagram resembles a snowflake shape

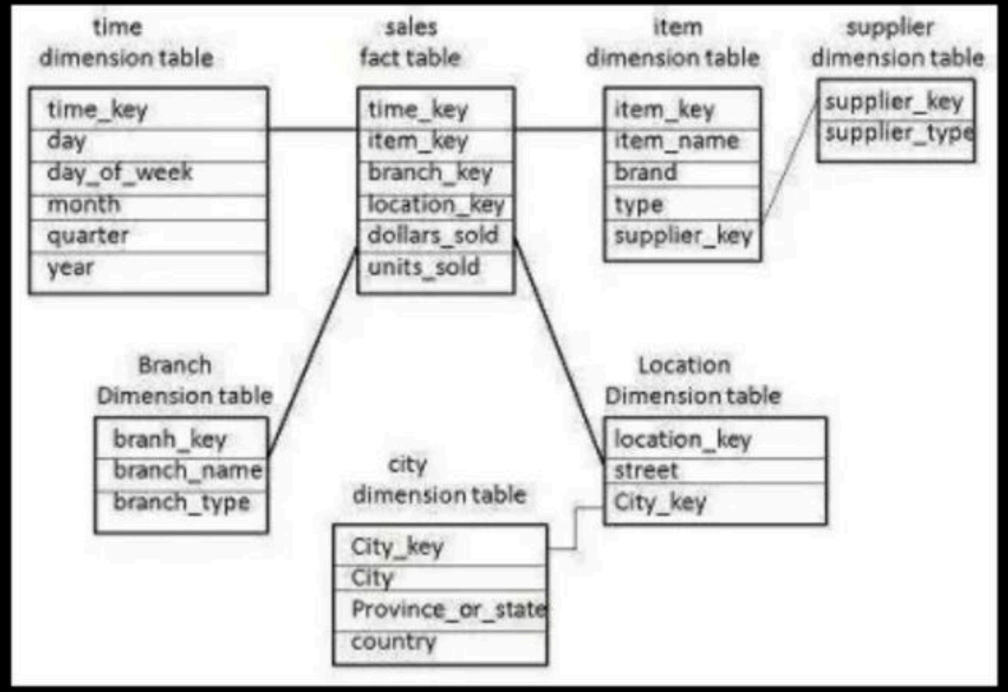
unacademy

Snowflake Schema

Represented by centralized fact tables which are connected to multiple dimensions. "Snowflaking" is a method of normalizing the dimension tables in a star schema

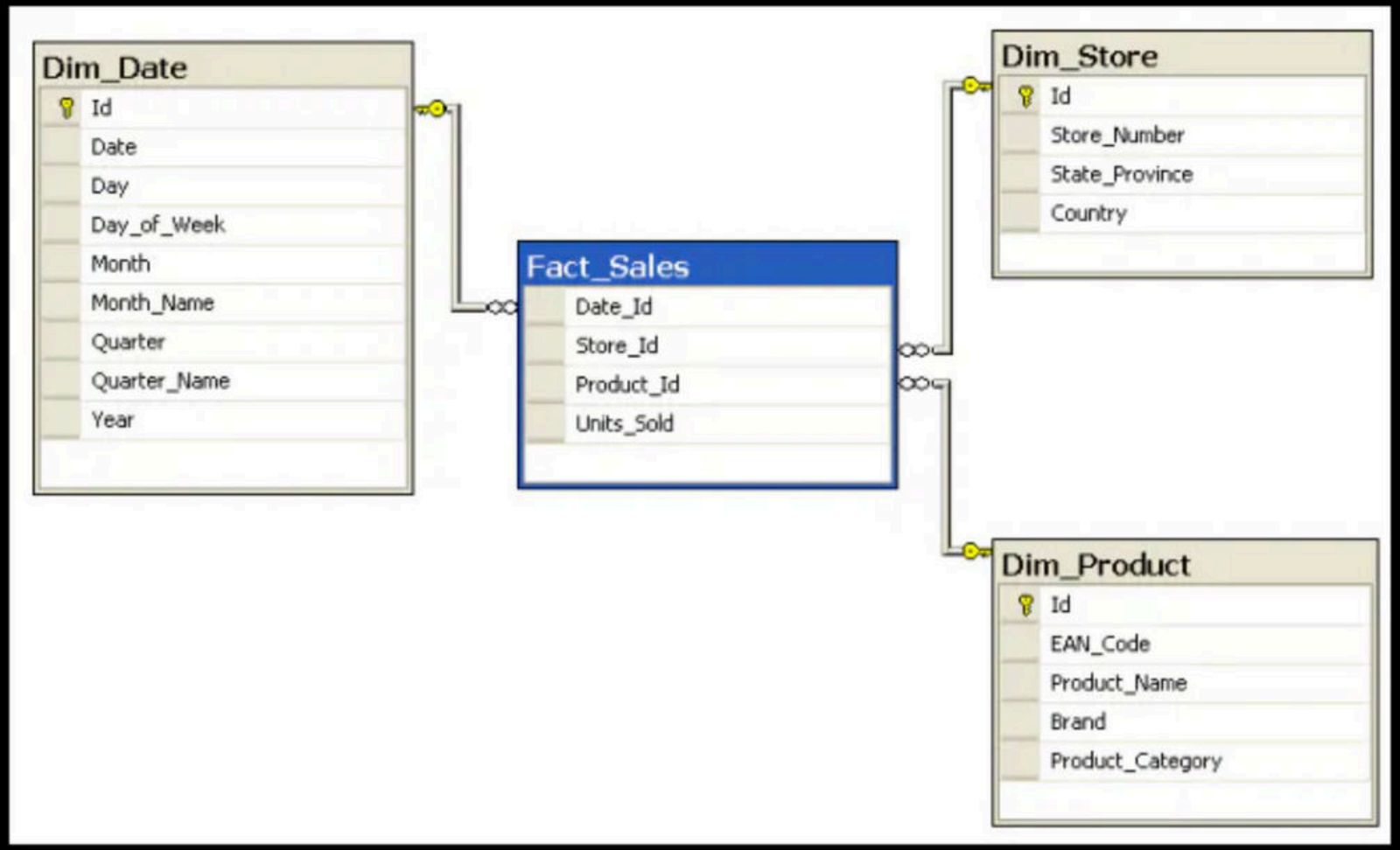
unacademy

Snowflake Schema



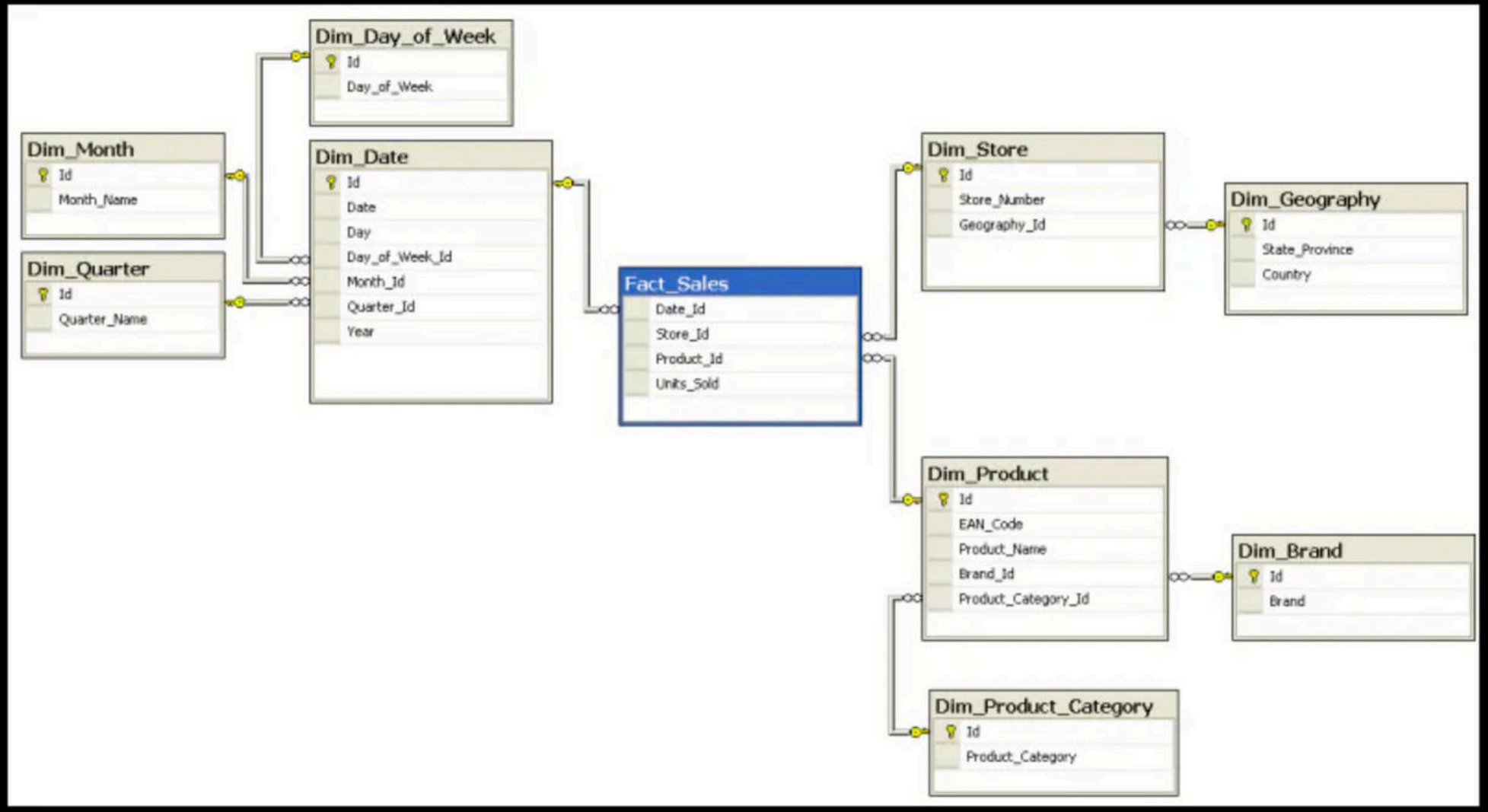
unacademy

Star Schema



unacademy

Snowflake Schema



unacademy

Snowflake Schema

Advantages:

- Some OLAP multidimensional database modeling tools are optimized for snowflake schemas.
- Normalizing attributes results in storage savings, the tradeoff being additional complexity in source query join

Disadvantages:

- Additional levels of attribute normalization adds complexity to source query joins
- Poor Performance when browsing the joins

unacademy

Happy Learning.!