Data Warehousing

Who am I?

Data Engineer @ TwentyToo AI

Ex Data Engineer @ GMC

Data Consultant @ Dynamu

Instructor @ Information Technology Institute (ITI) — Egypt

Instructor @ National Telecommunication Institute (NTI) — Egypt

Instructor @ CREATIVA — Egypt

Instructor @ DEPI — Egypt

Teaching Assistant @ Faculty of Artificial Intelligence, Cairo University

Coached students @ Faculty of Engineering, Helwan University

Lecturer @ Joseph Institute for Wireless Officers

in Linkedin.com/in/MohamedARoshdy

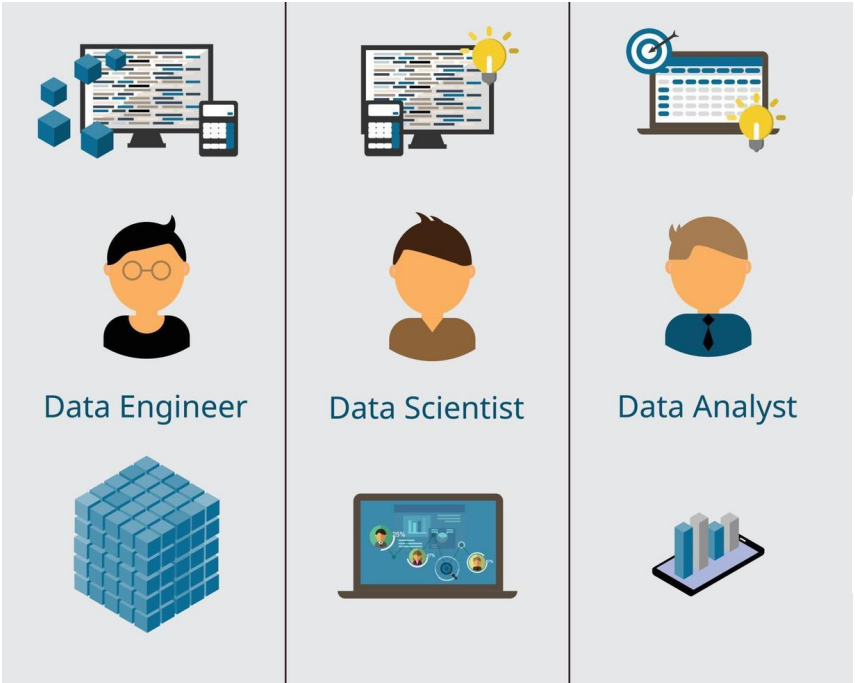
Agenda

- Introduction to the Program
- Understanding Data-related Job Roles
- Data Journey
- Data Engineer: Bridging TEch and business.
- OLTP Vs OLAP.
- History of Data Warehousing > Bill Inmon Vs Kimball.
- Dependent Vs Independent data mart.
- DWH Architectures.
- ETL VS ELT
- Surrogate vs Natural Key
- Preparing the Environment for Labs
- Q&A

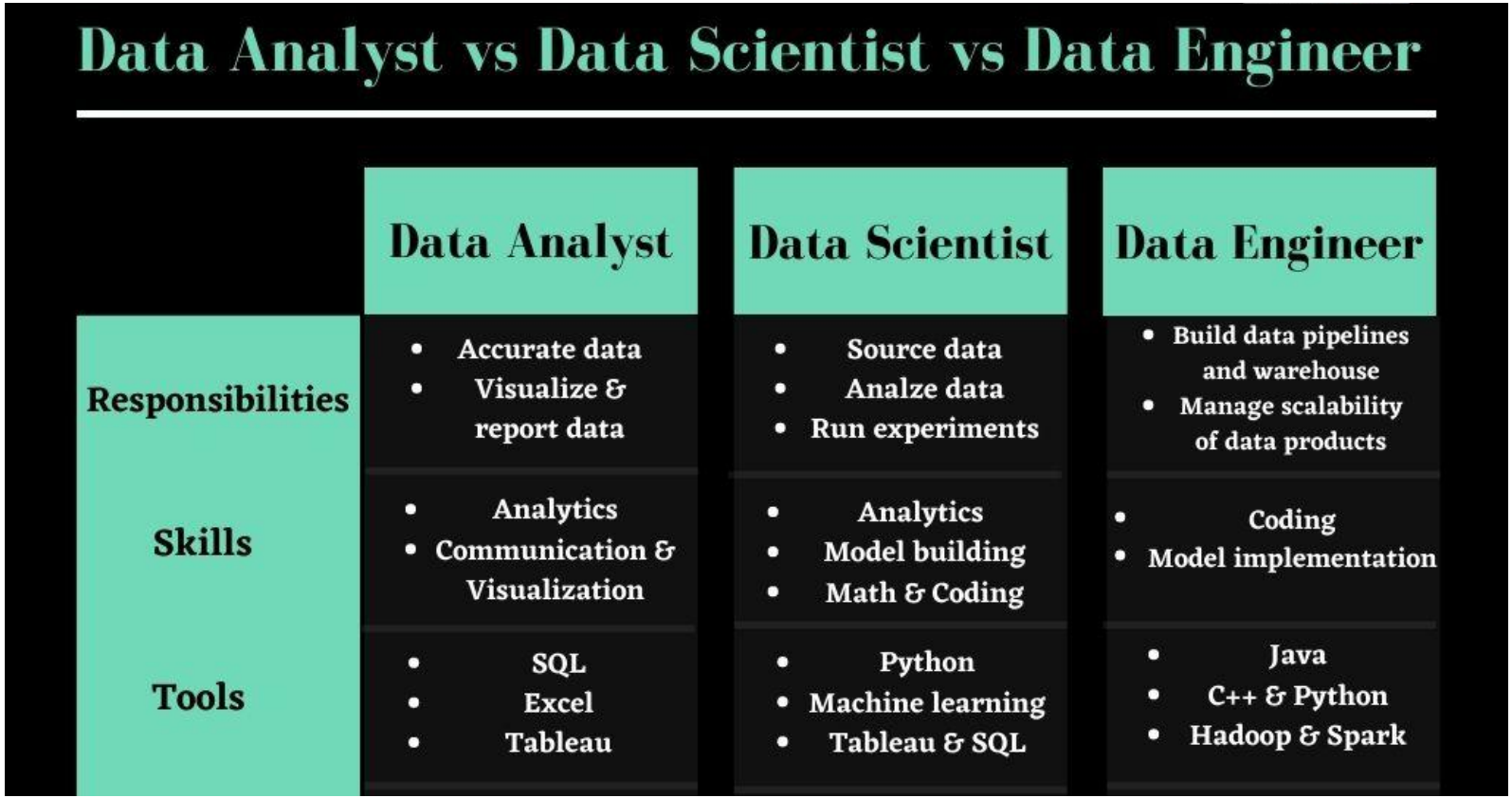
Icebreaker / Warm-up

What do you think a Data Engineer actually does?

Guess the Role



Overview of Data Roles

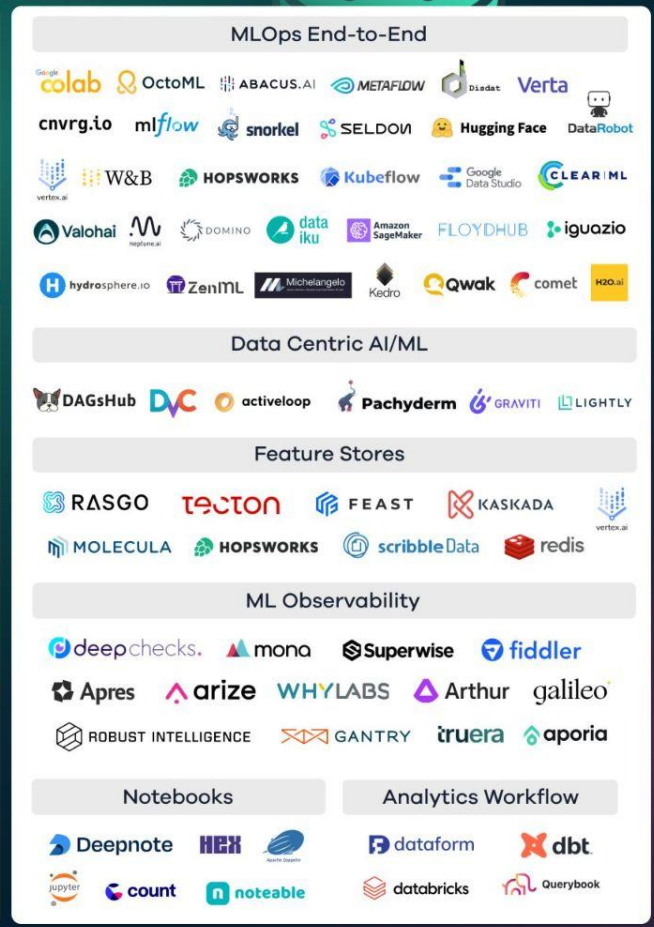


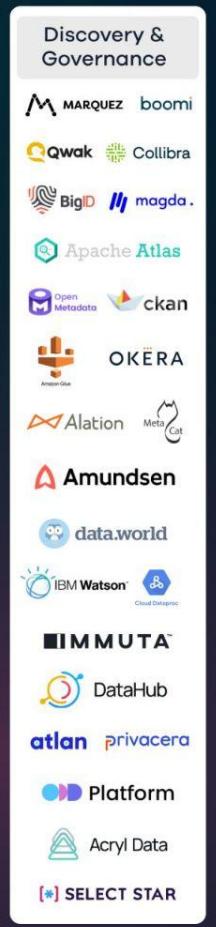
Be a Careful Data Engineer – It's a Mindset

"Data Engineering is not just about tools and coding — it's a mindset."

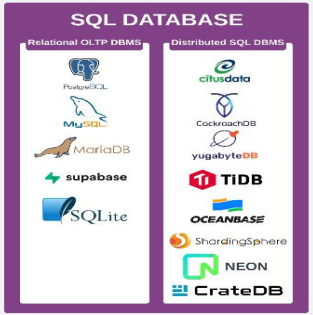
State of Data Engineering 2022 map

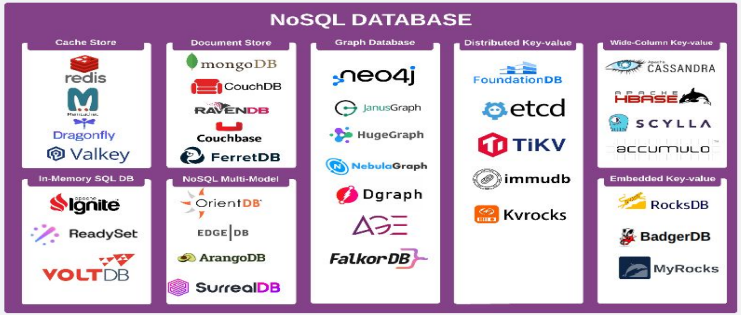
Presented by lakeFS

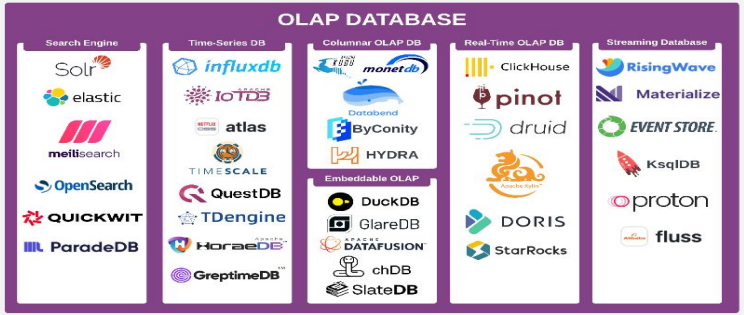


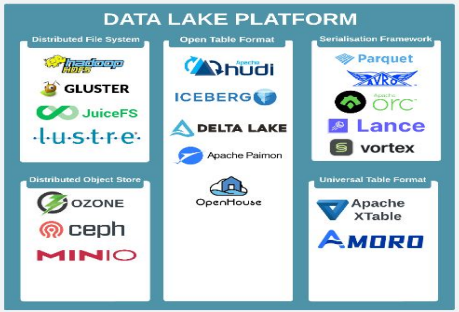


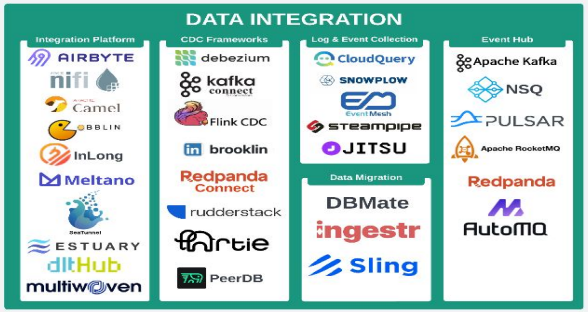
OPEN SOURCE DATA ENGINEERING LANDSCAPE 2025

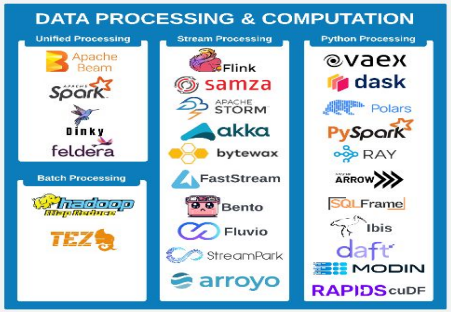


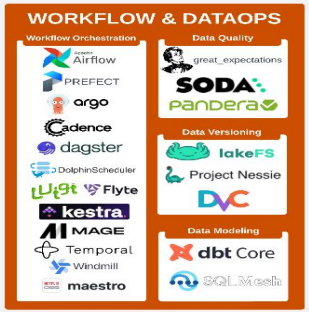


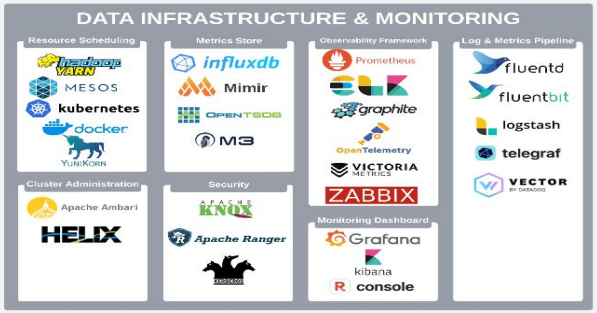


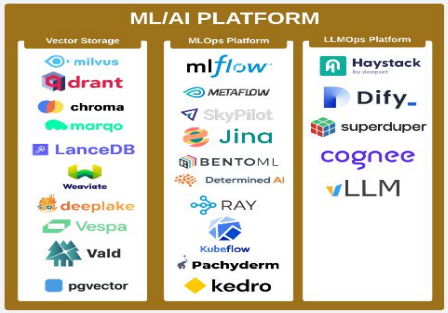


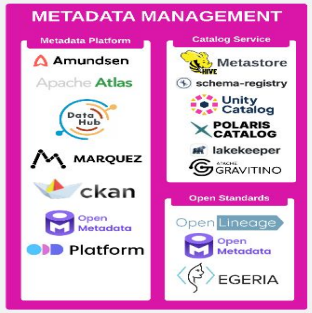


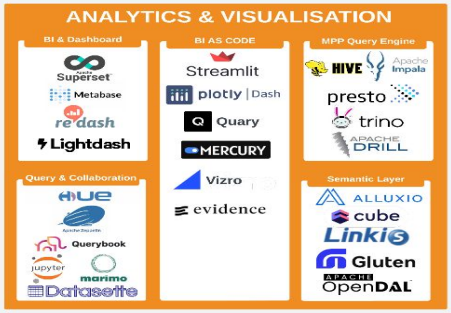












Whom You'll Work With!!
Data Engineer: Bridging TEch and business.

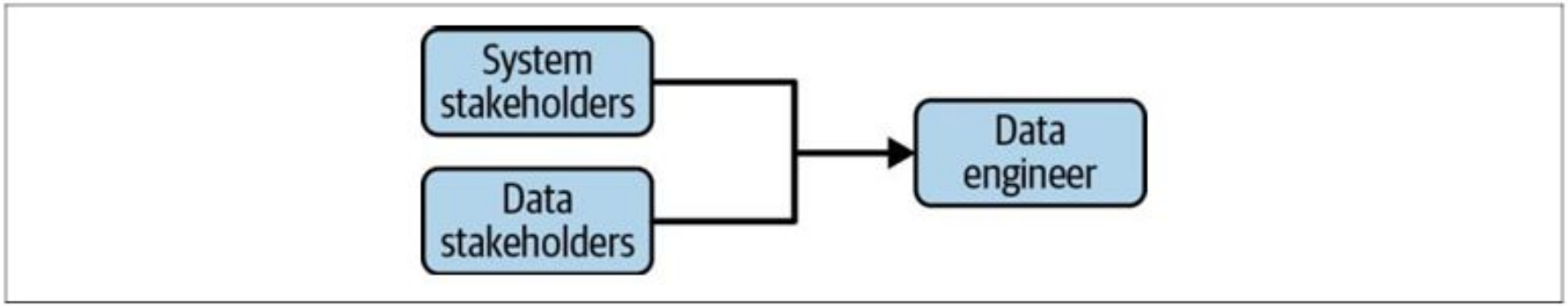


Figure 5-12. The data engineer's upstream stakeholders

A great Data Engineer !

Key Points:

A great Data Engineer understands the business as much as the technology.

Think beyond pipelines and storage — focus on delivering real value.

Bridge the gap between business needs and technical solutions.

Always ask: “How will this data solve a real problem?”

Data Journey

Files Era (Flat Files, CSV, Logs)

Data stored in files.

Simple but hard to integrate.

No standard structure → inconsistency problems.

From Files to Data Engineering – Supporting Analytics

Files Era (Flat Files, CSV, Logs)

Data stored in files.

Simple but hard to integrate.

No standard structure → inconsistency problems.

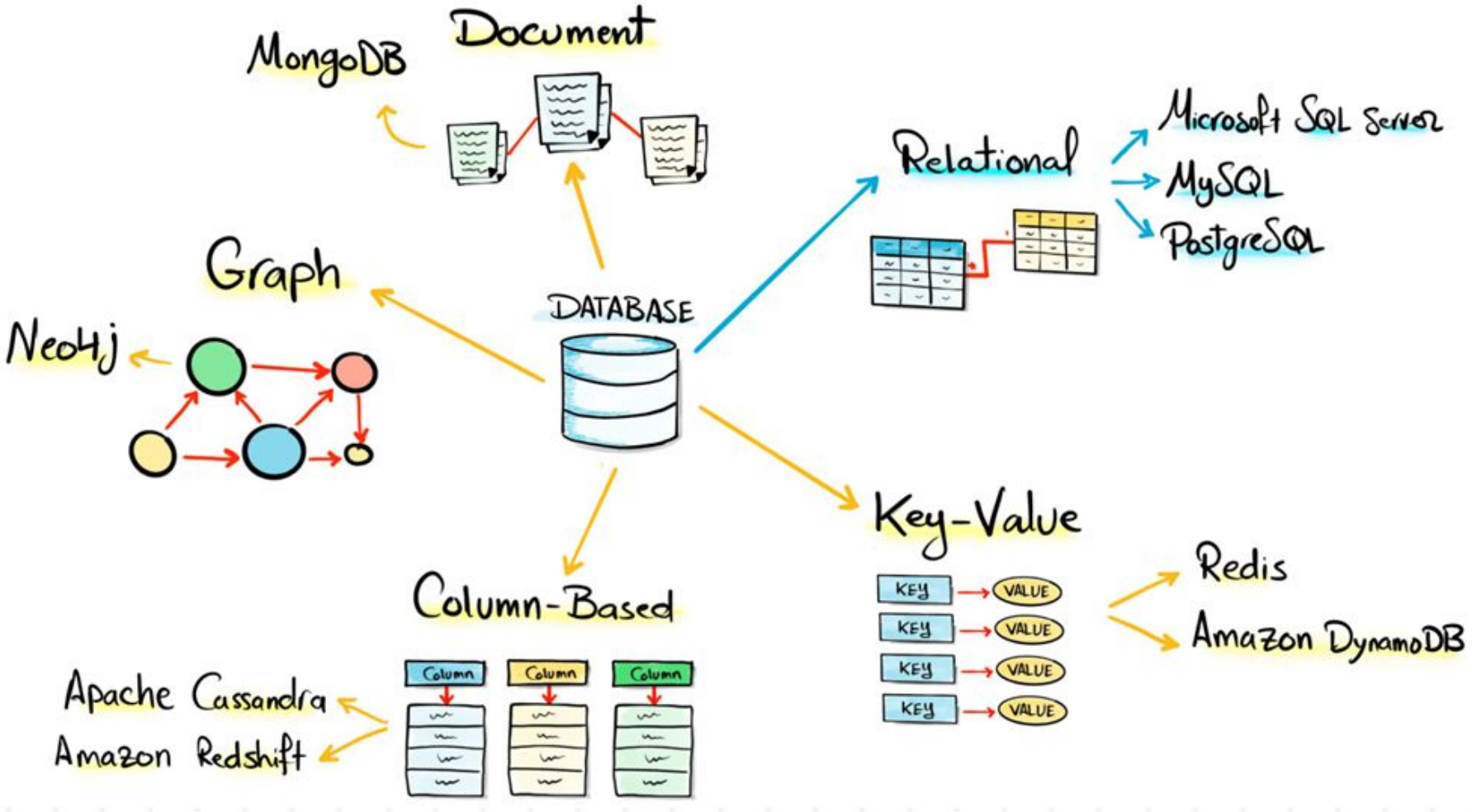
From Files to Data Engineering – Supporting Analytics

Databases Era (SQL Databases)

Solved integration & consistency issues.

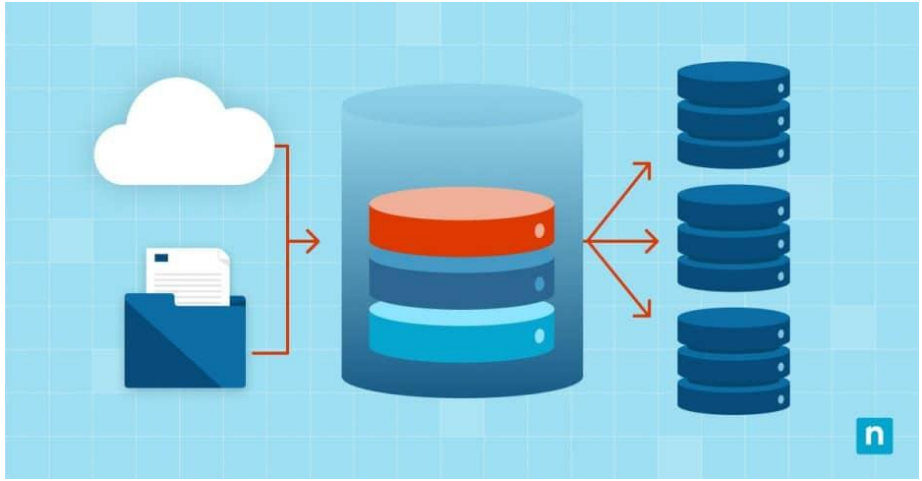
Structured schema, easy querying.

Great for operational systems, but limited for large-scale analytics.


MongoDB
Document
Relational
Microsoft SQL Server
MySQL
PostgreSQL
Graph
Neo4j
DATABASE
Key-Value
Redis
Amazon DynamoDB
KEY
VALUE
Column-Based
Column
Apache Cassandra
Amazon Redshift

From Files to Data Engineering – Supporting Analytics

Data Warehouse Era

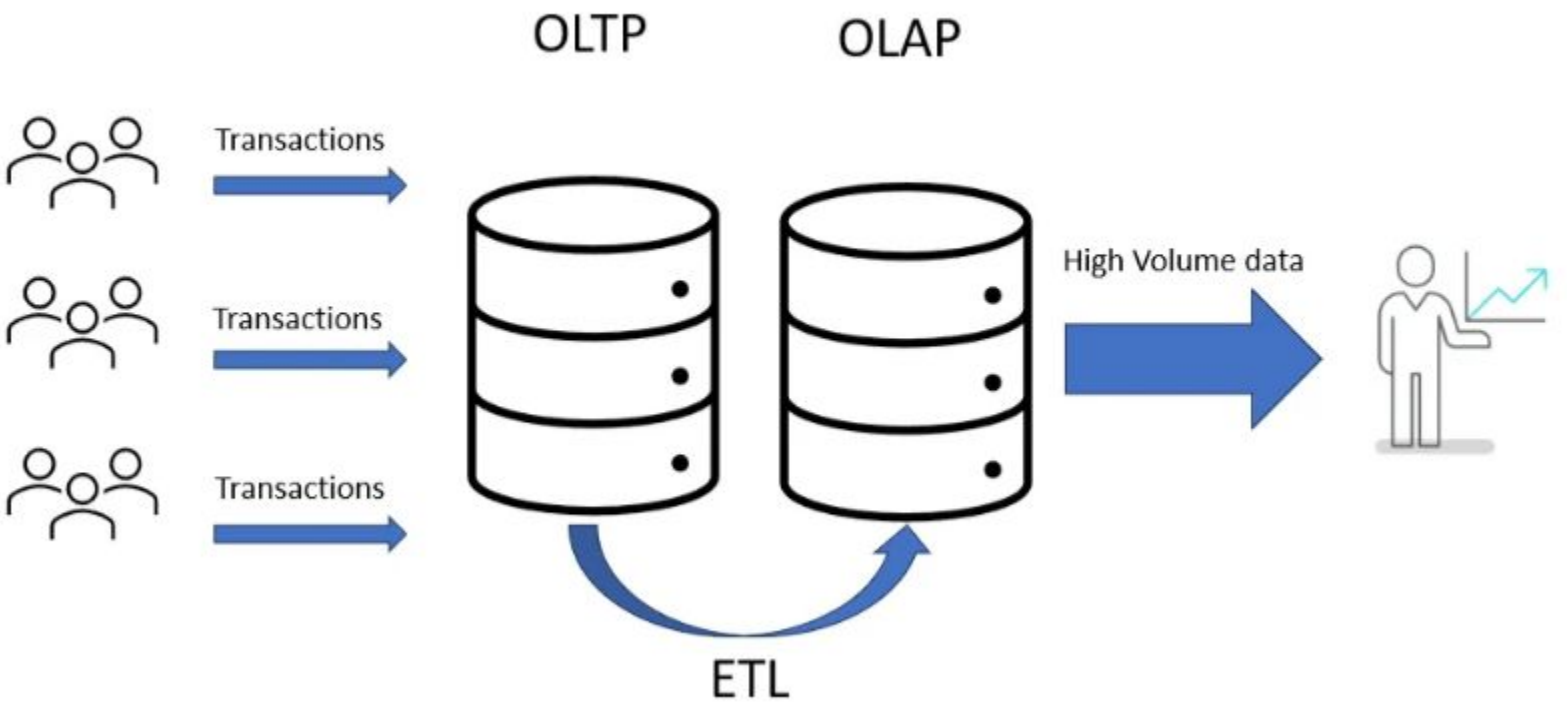


Designed for analytics, not operations.

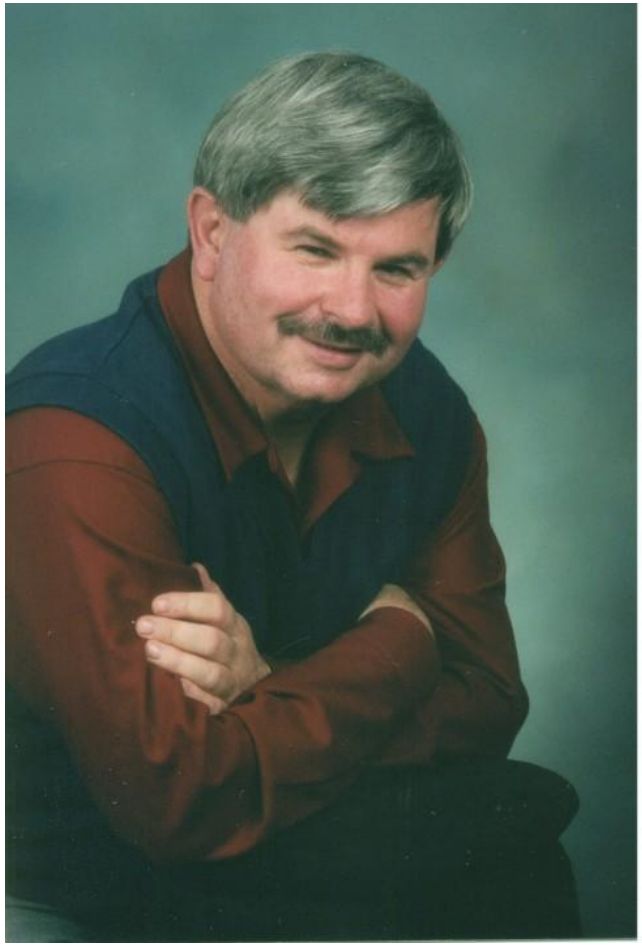
Integrates data from multiple sources.

Optimized for large-scale queries & historical analysis.

OLTP Vs OLAP


OLTP
OLAP
Transactions
Transactions
Transactions
High Volume data
ETL

History of Data Warehousing and “Bill Inmon”



Analytics India MAGAZINE

INTERVIEW

BILL INMON

FATHER OF DATA WAREHOUSE

History of Data Warehousing and “Bill Inmon”

1. Before Data Warehousing (Pre-1980s)

Data stored in separate operational (OLTP) systems.

Difficult to consolidate and analyze data across systems.

2. Birth of the Concept (1980s)

Bill Inmon introduced the idea of the Enterprise Data Warehouse (EDW).

Definition: "A subject-oriented, integrated, time-variant, and non-volatile collection of data to support decision-making."

History of Data Warehousing and “Bill Inmon”

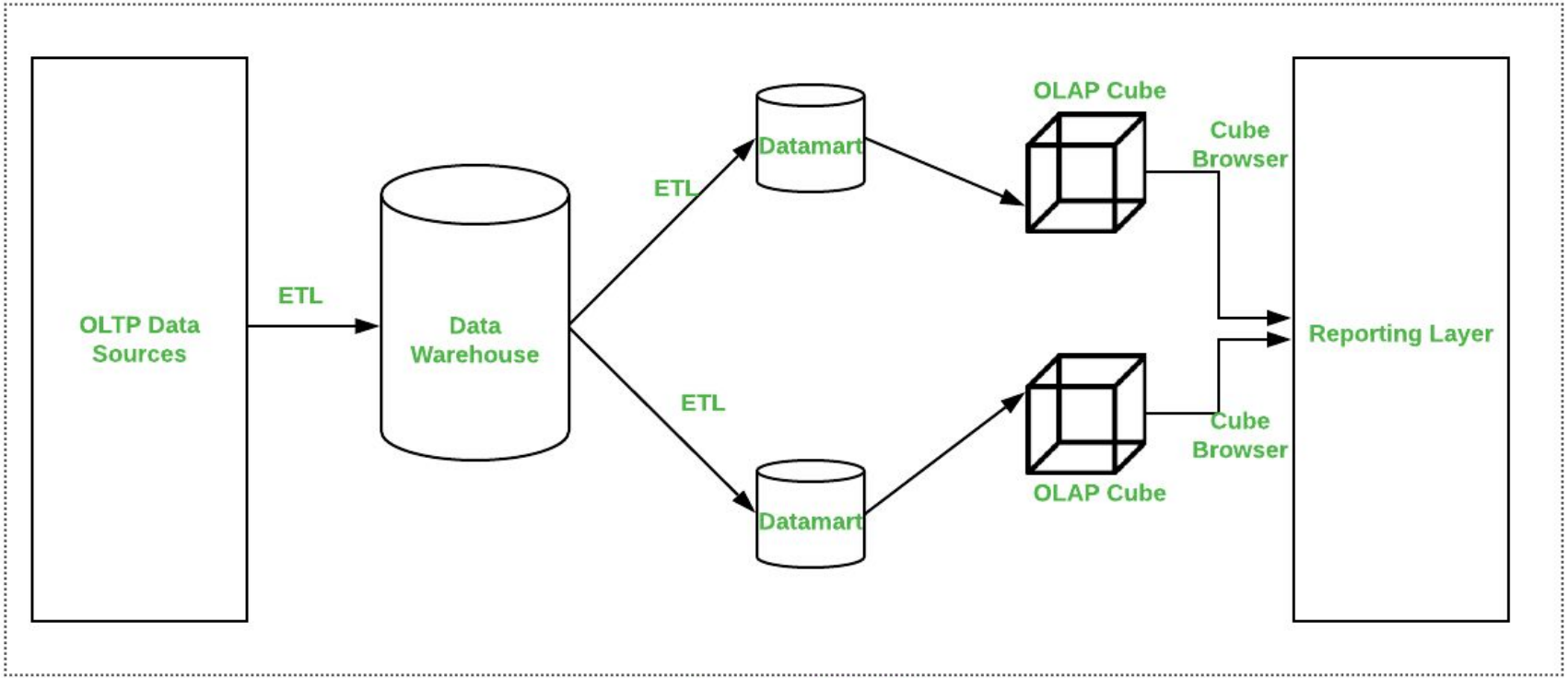
3. The Inmon Approach

Build EDW first (Normalized – often up to 3NF).

From EDW, derive Data Marts (Star or Snowflake) for analytics.

Pros: High data integration, single source of truth for the enterprise.

History of Data Warehousing and “Bill Inmon”



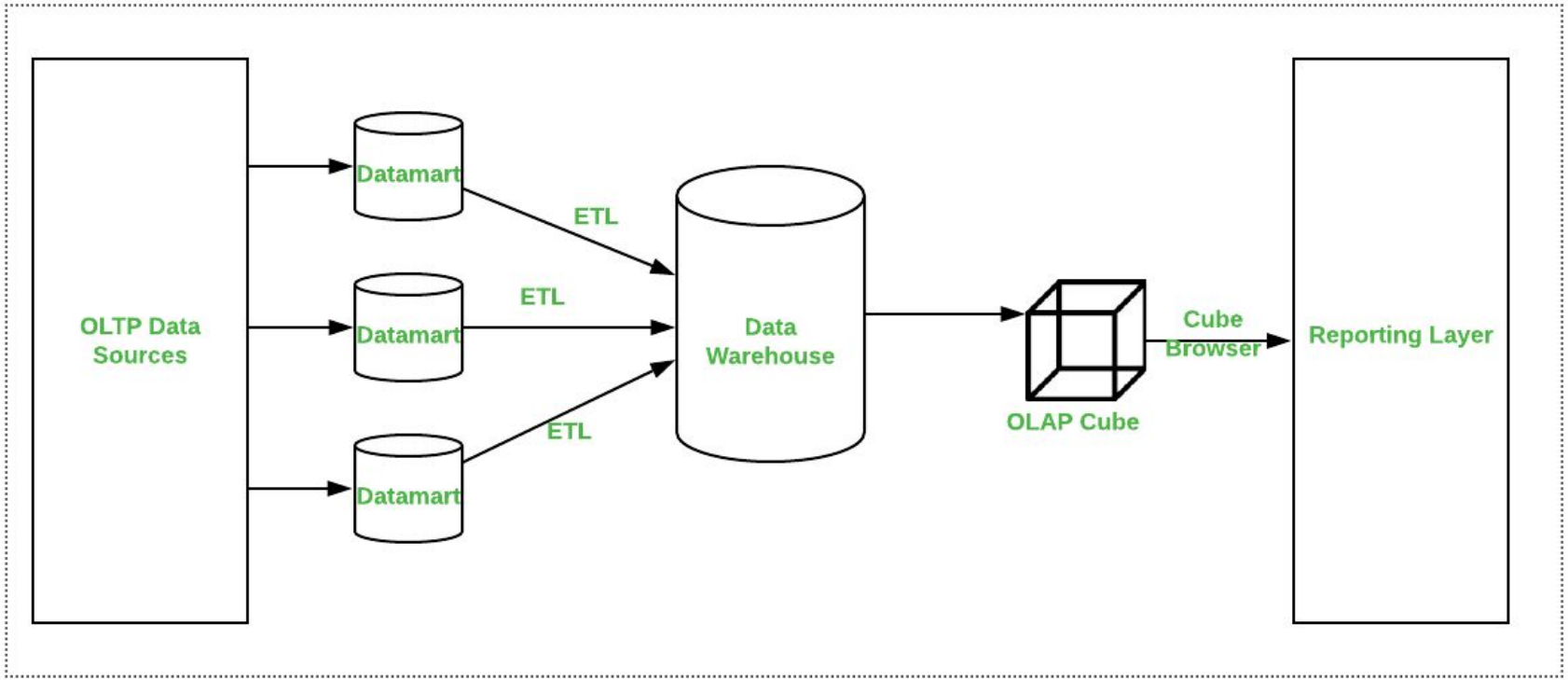
Inmon Model

After Inmon – The Kimball Approach (1990s)

Ralph Kimball proposed building Data Marts first (Star Schema).

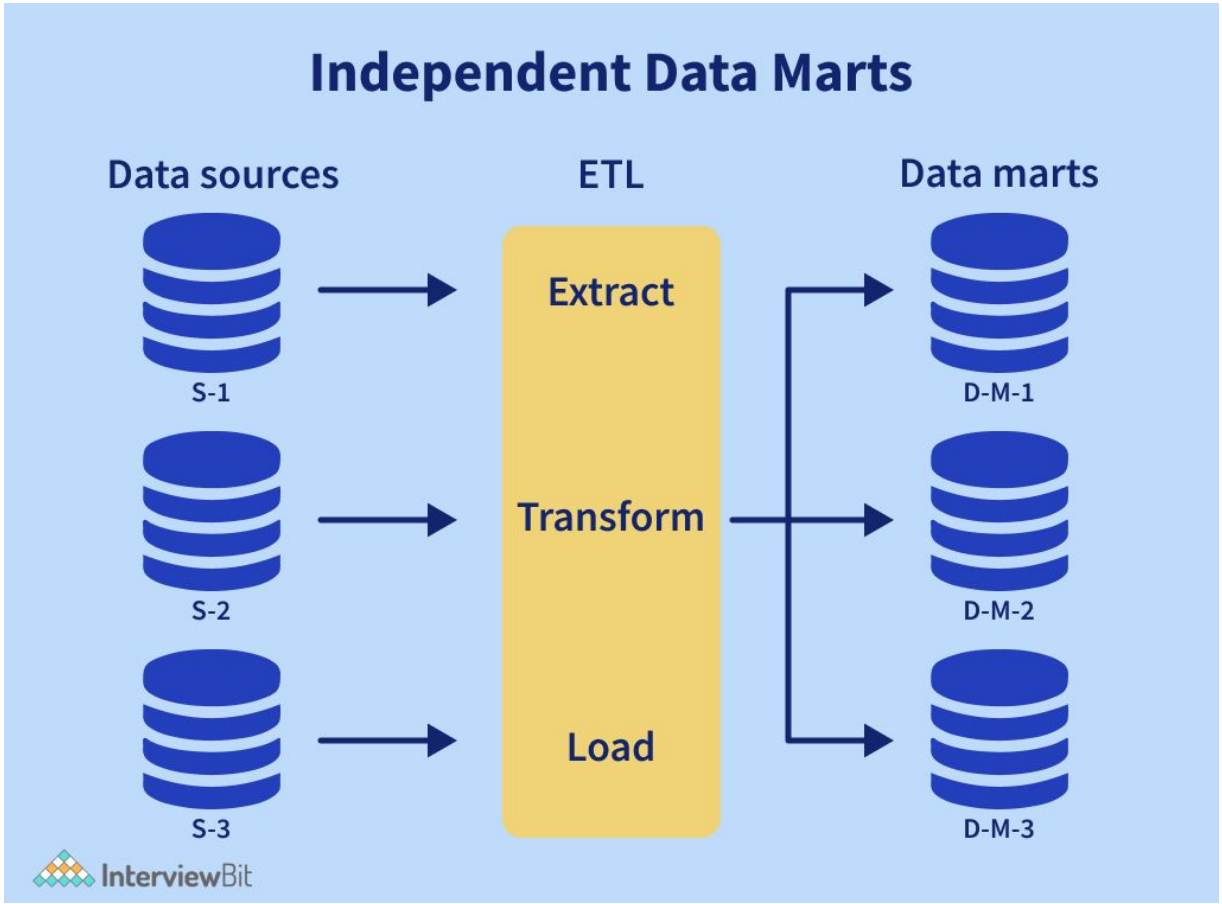
Faster analytics but less integrated at the start.

Kimball Approach

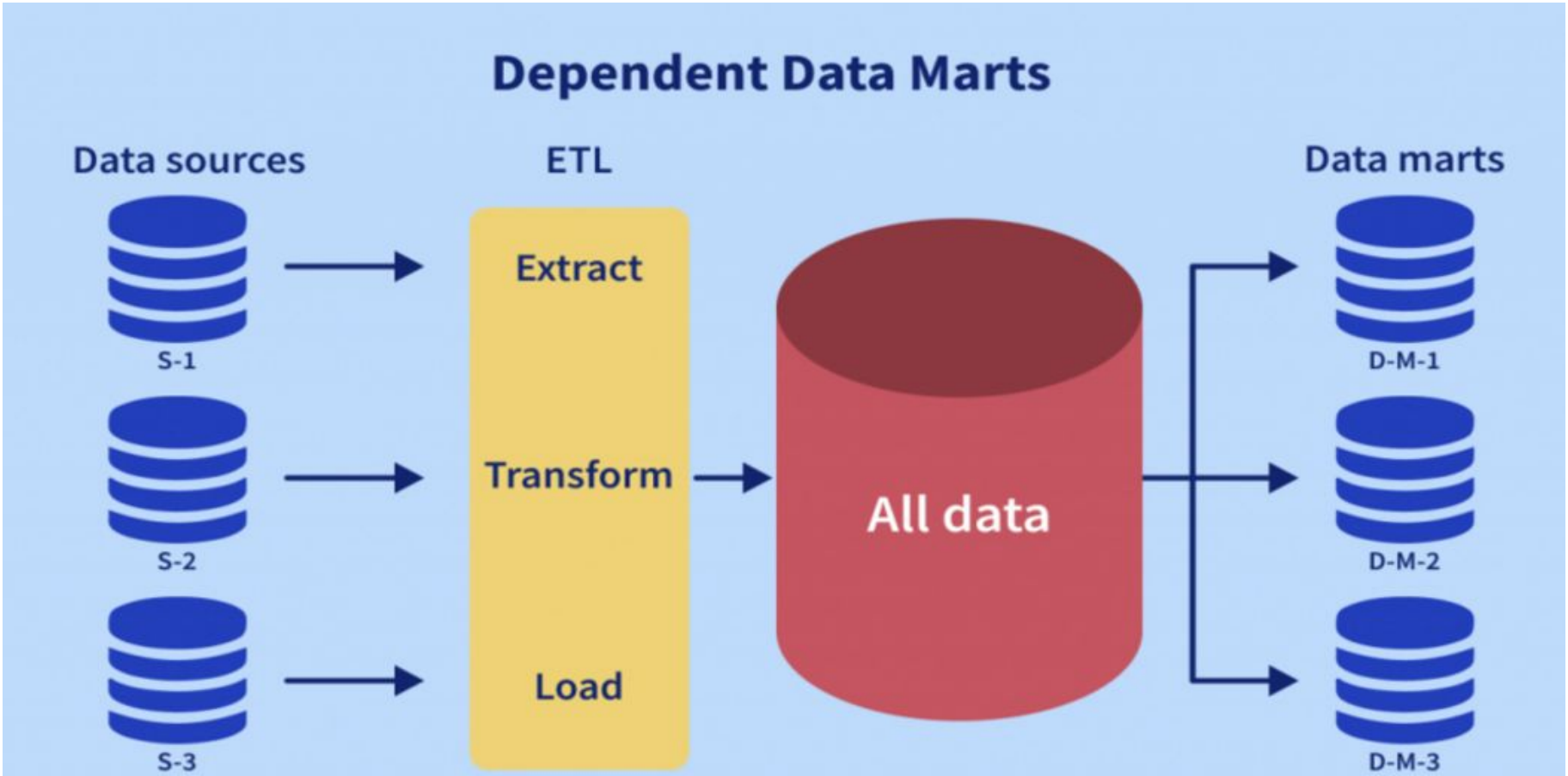


Kimball Model

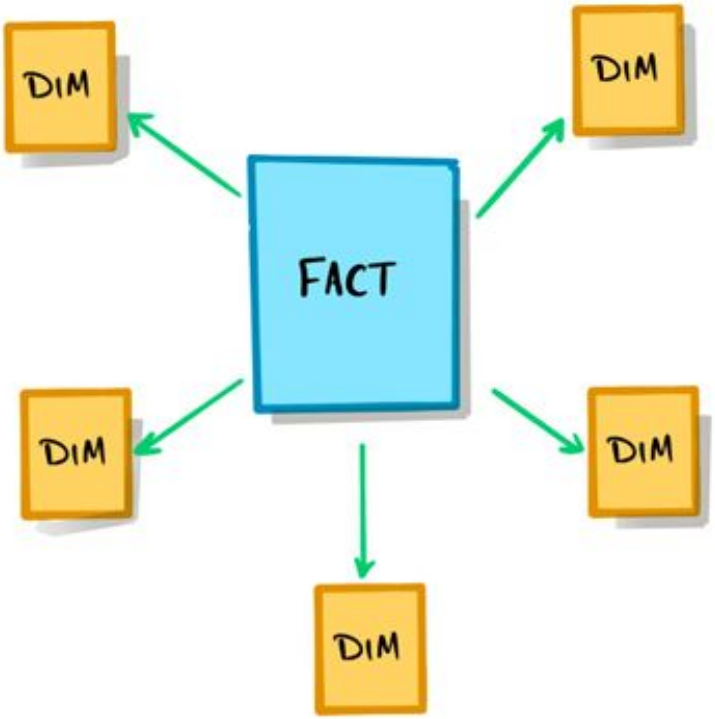
Dependent Vs Independent data mart "Which Approach?"



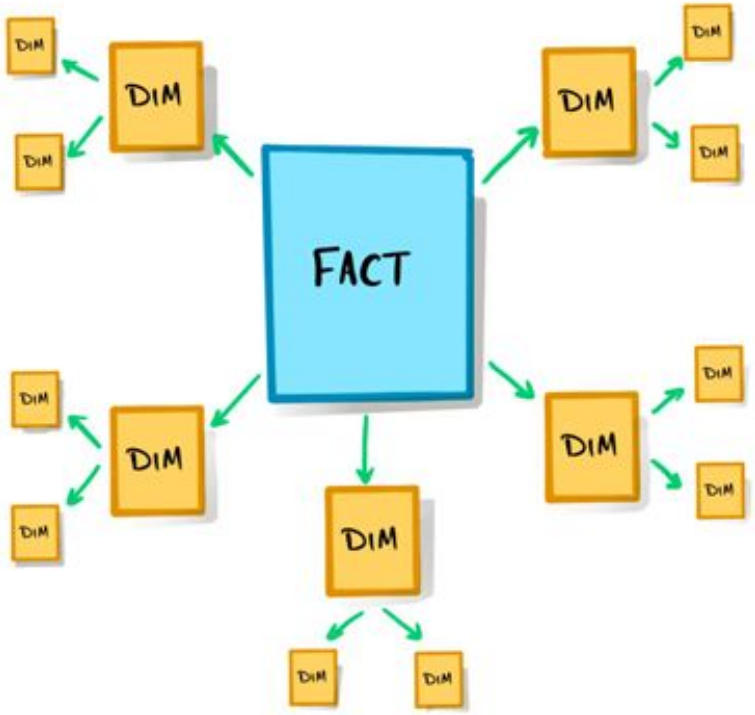
Dependent vs Independent data mart

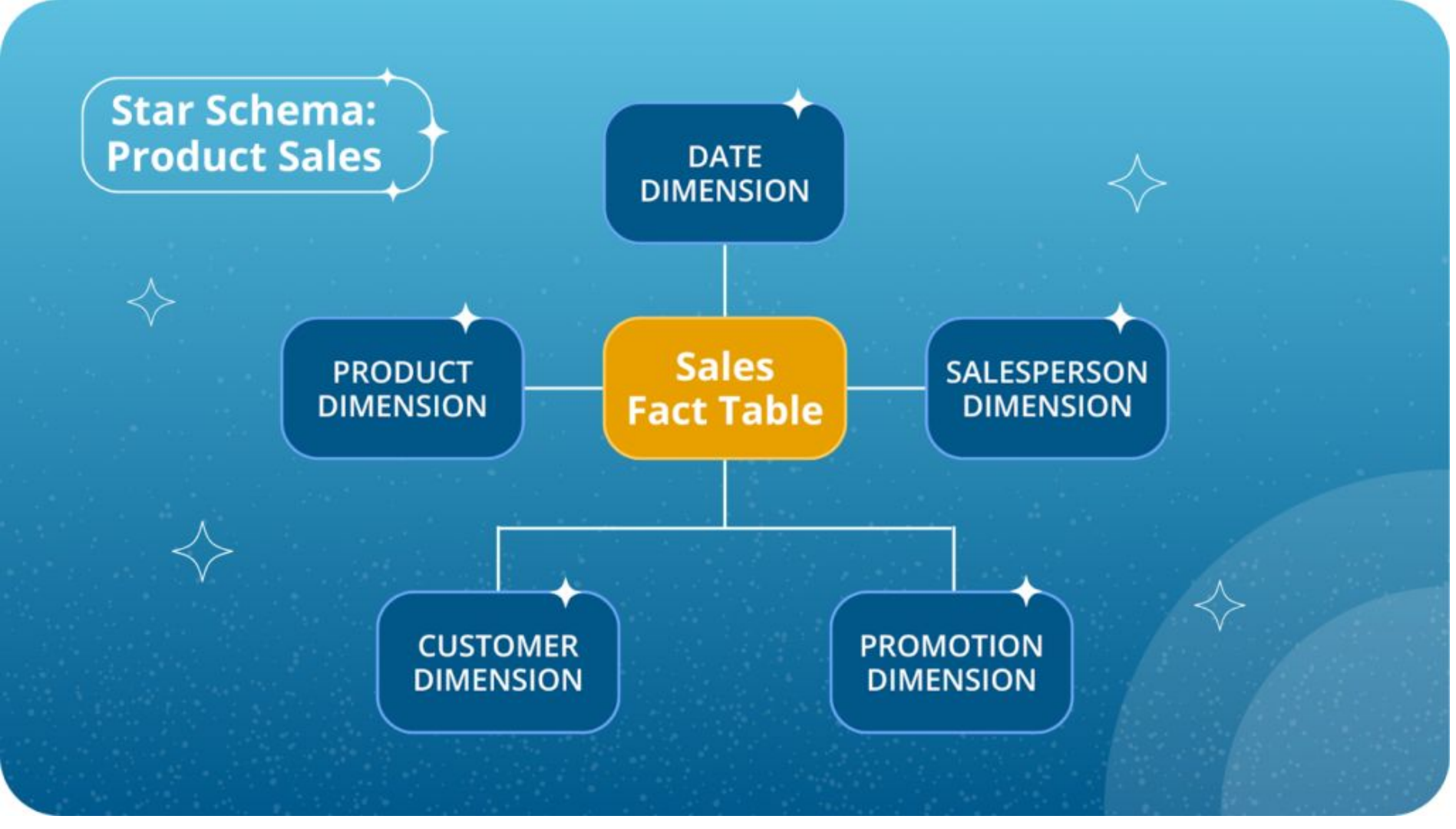


STAR SCHEMA

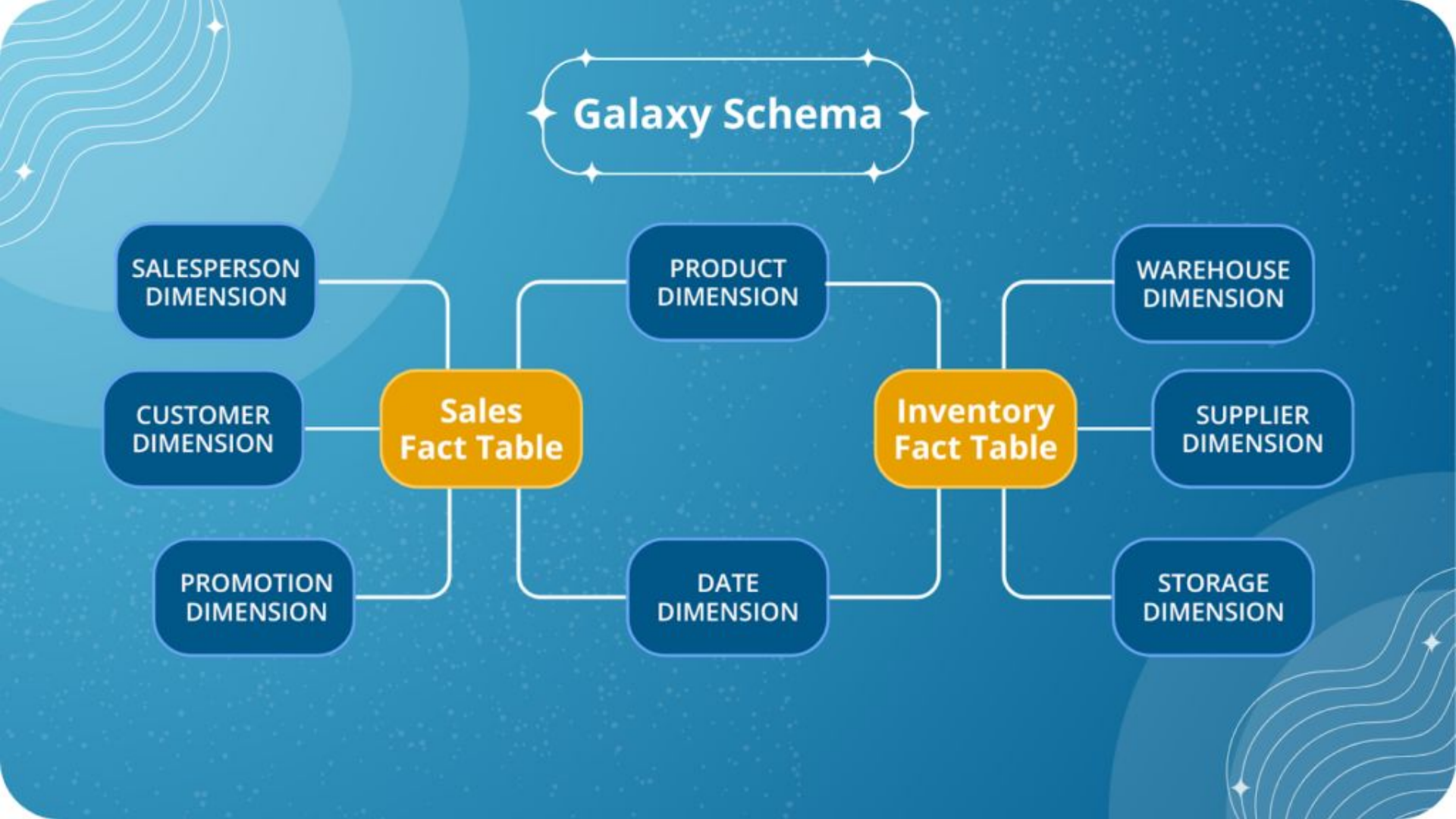


SNOWFLAKE SCHEMA




Star Schema:
Product Sales
DATE
DIMENSION
PRODUCT
DIMENSION
Sales
Fact Table
SALESPERSON
DIMENSION
CUSTOMER
DIMENSION
PROMOTION
DIMENSION


Snowflake Schema:
Product Sales
MONTH DIMENSION
YEAR DIMENSION
DATE DIMENSION
CATEGORY DIMENSION
REGION DIMENSION
PRODUCT DIMENSION
SALES FACT TABLE
SALESPERSON DIMENSION
BRAND DIMENSION
COMMISSION RATE DIMENSION
CUSTOMER DIMENSION
PROMOTION DIMENSION
INDUSRTY DIMENSION
COMPANY SIZE DIMENSION
PROMOTION TYPE DIMENSION
CHANNEL DIMENSION

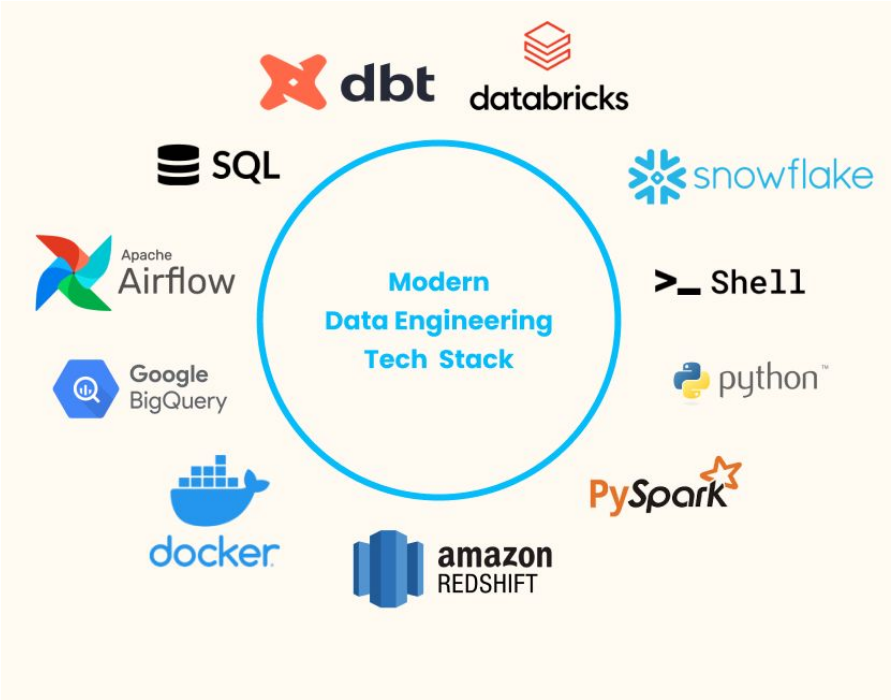

Galaxy Schema
SALESPERSON DIMENSION
CUSTOMER DIMENSION
PROMOTION DIMENSION
PRODUCT DIMENSION
DATE DIMENSION
Sales Fact Table
Inventory Fact Table
WAREHOUSE DIMENSION
SUPPLIER DIMENSION
STORAGE DIMENSION

From Files to Data Engineering – Supporting Analytics

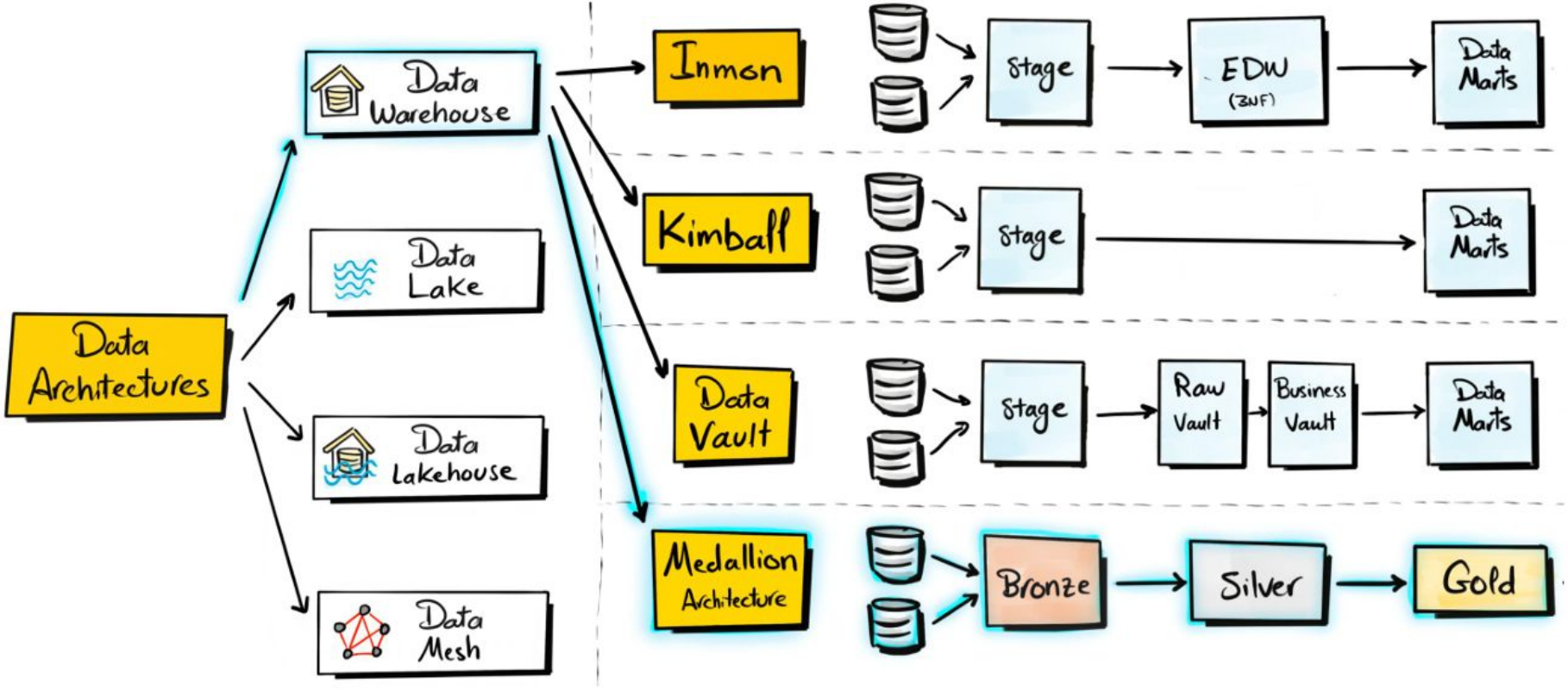
Modern Data Engineering

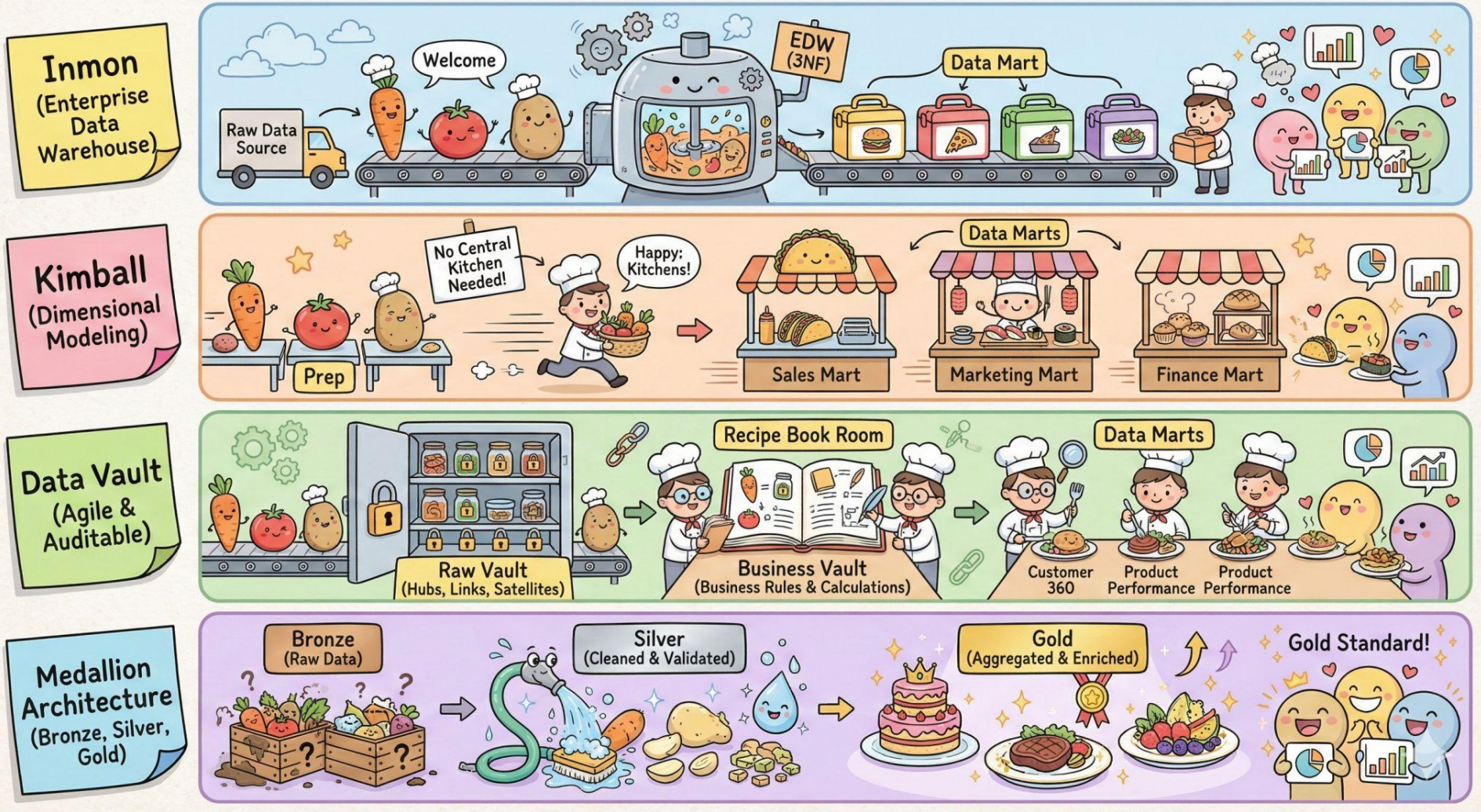
Combines databases, NoSQL, data lakes, and warehouses.
Builds pipelines for clean, integrated, analytics-ready data.

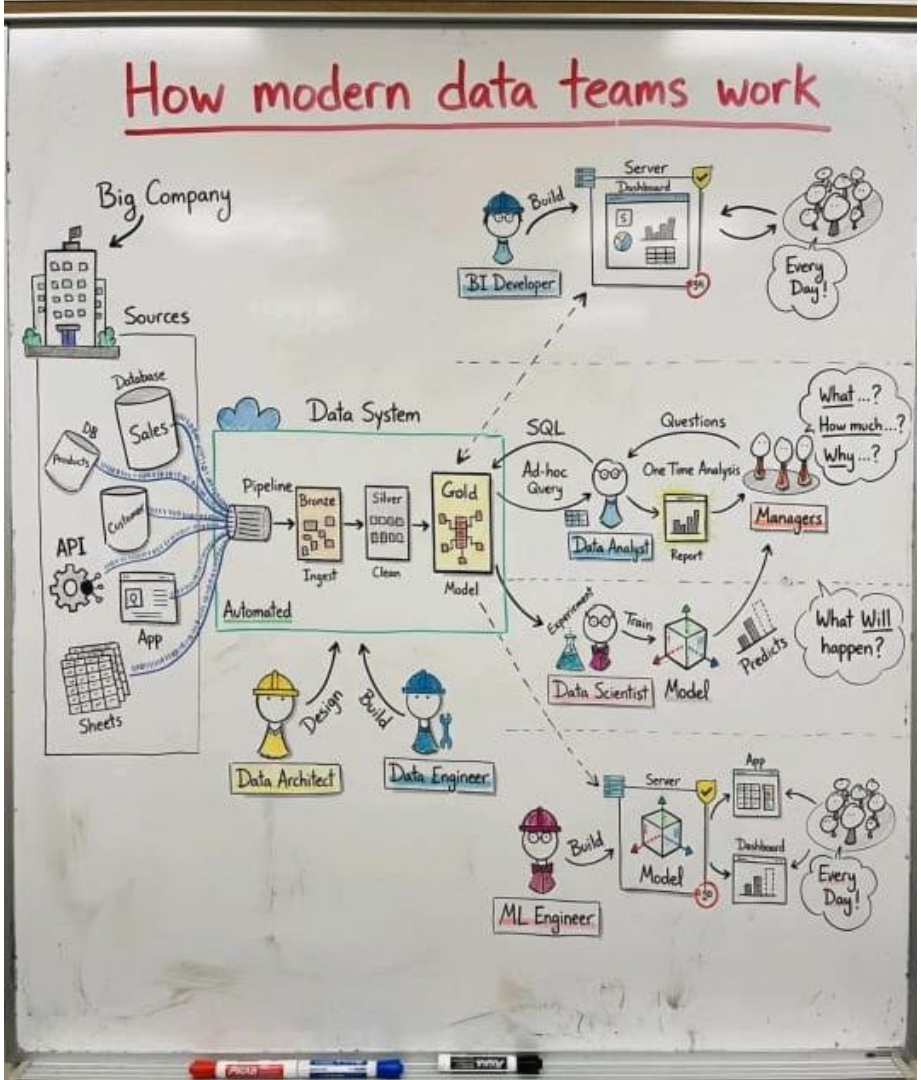
1980s: Bill Inmon → EDW (3NF)
1990s: Kimball → Star Schema
2000s+: Hybrid & Cloud DW



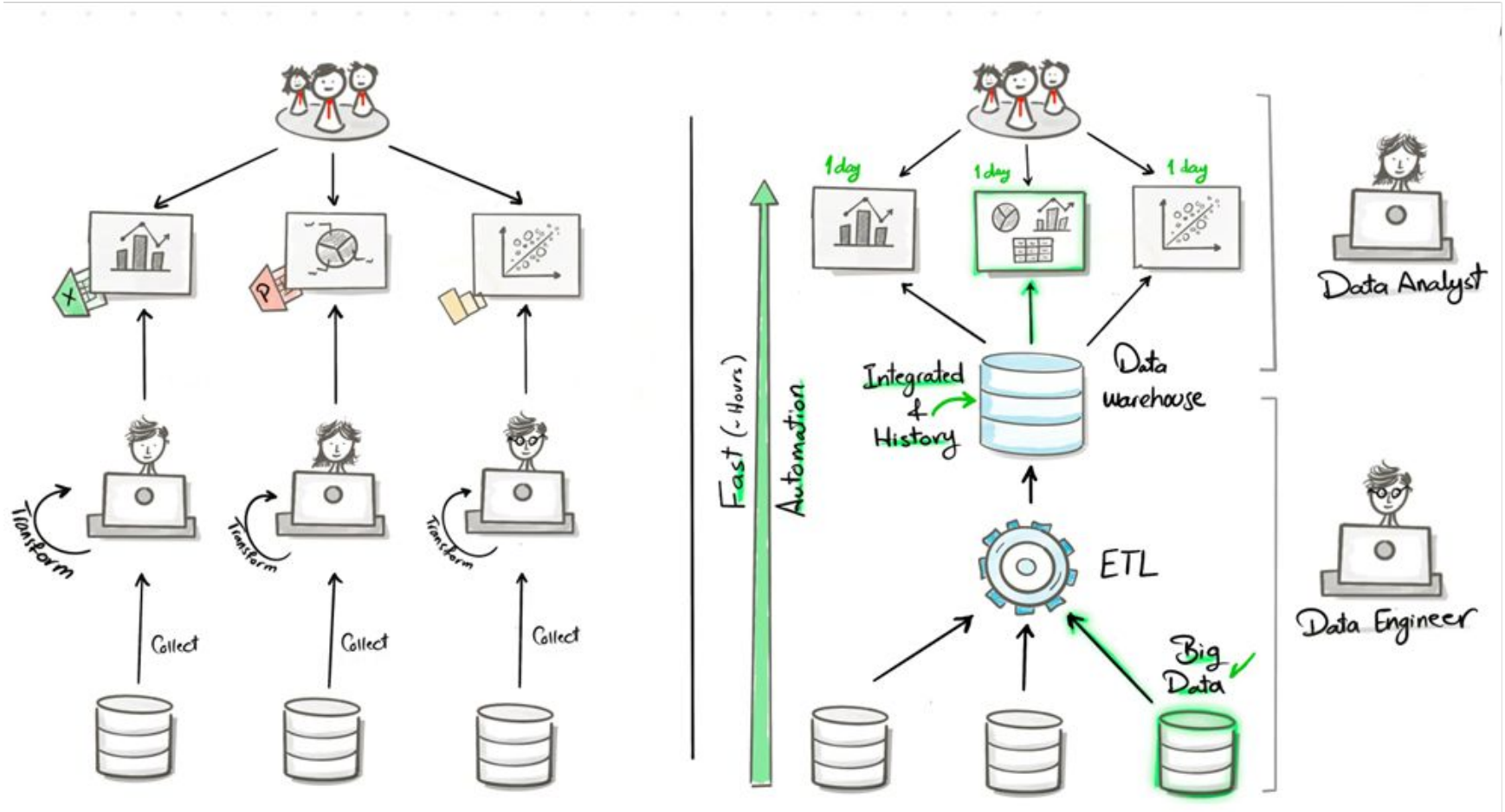
Data Architectures



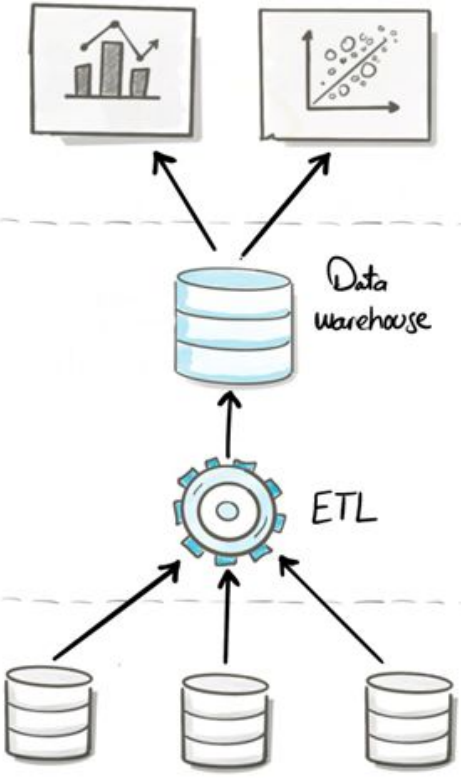

Inmon
(Enterprise Data Warehouse)
Raw Data Source
Welcome
EDW (3NF)
Data Mart
Kimball
(Dimensional Modeling)
No Central Kitchen Needed!
Happy; Kitchens!
Prep
Data Marts
Sales Mart
Marketing Mart
Finance Mart
Data Vault
(Agile & Auditable)
Raw Vault
(Hubs, Links, Satellites)
Recipe Book Room
Business Vault
(Business Rules & Calculations)
Data Marts
Customer 360
Product Performance
Product Performance
Medallion Architecture
(Bronze, Silver, Gold)
Bronze
(Raw Data)
Silver
(Cleaned & Validated)
Gold
(Aggregated & Enriched)
Gold Standard!


How modern data teams work
Big Company
Sources
Database
Sales
DB
Products
Customer
API
App
Sheets
Data System
Pipeline
Bronze
Ingest
Silver
Clean
Gold
Model
Automated
Data Architect
Design
Data Engineer
Build
BI Developer
Build
Server
Dashboard
Every Day!
SQL
Ad-hoc Query
Questions
One Time Analysis
Data Analyst
Report
Managers
What ...?
How much...?
Why ...?
Experiment
Train
Data Scientist
Model
Predicts
What Will happen?
ML Engineer
Build
Server
Model
App
Dashboard
Every Day!

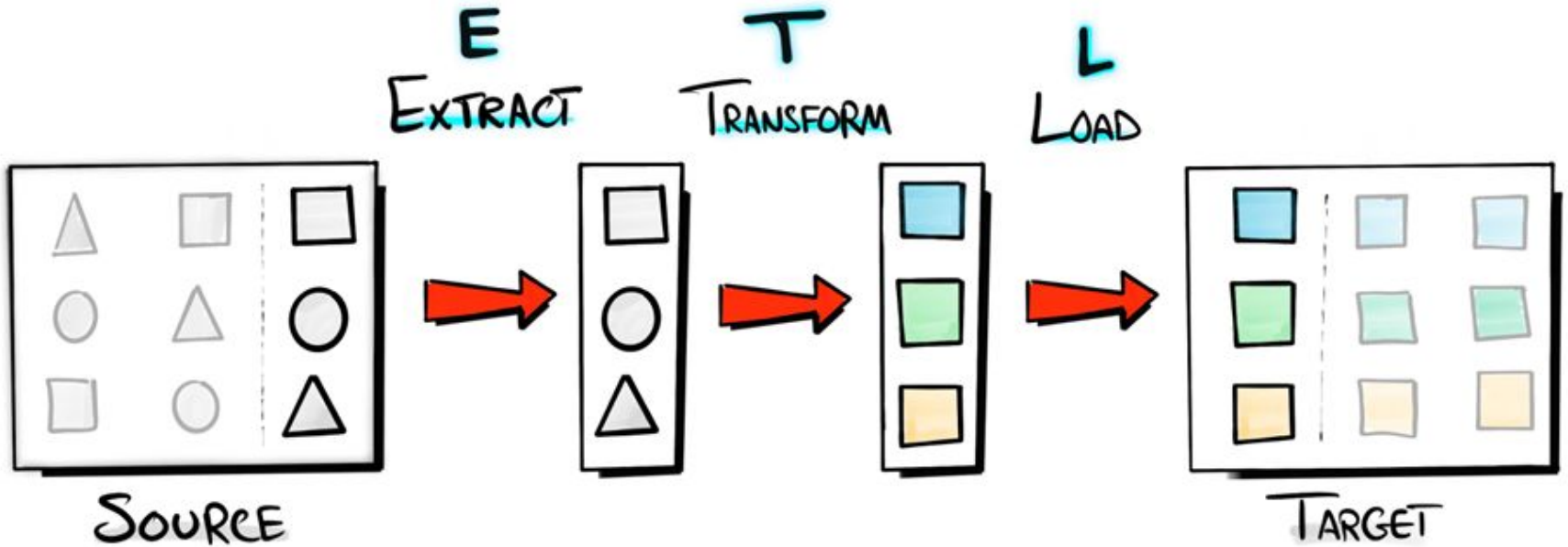
Why we need DWH !



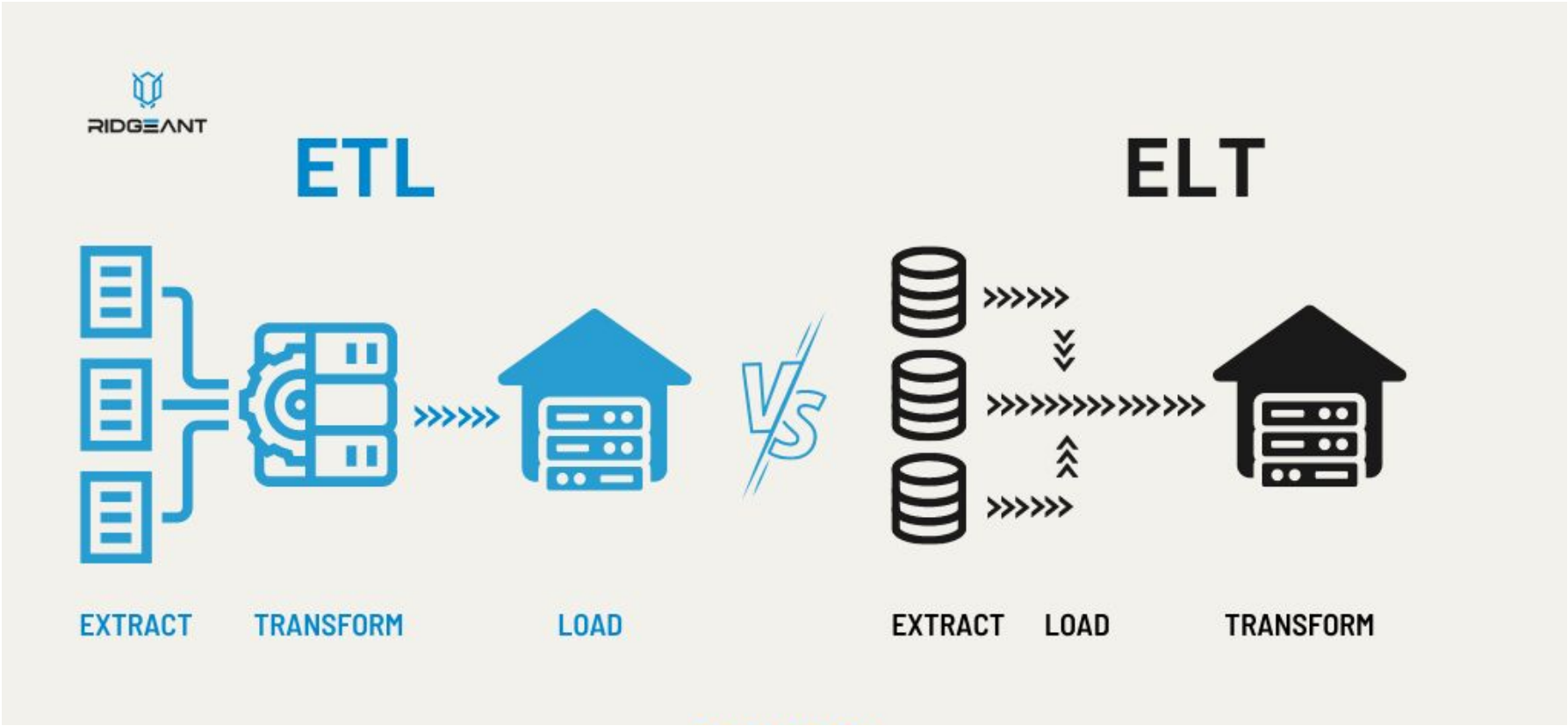
Why we need DWH !



ETL Process



ETL VS ELT Guess??



Lambda Architecture

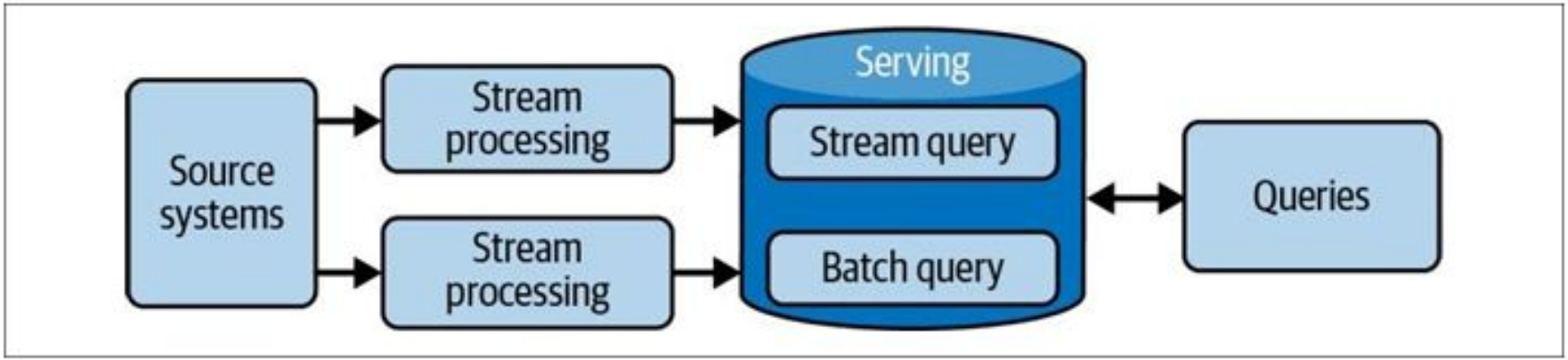


Figure 3-14. Lambda architecture

Kappa Architecture

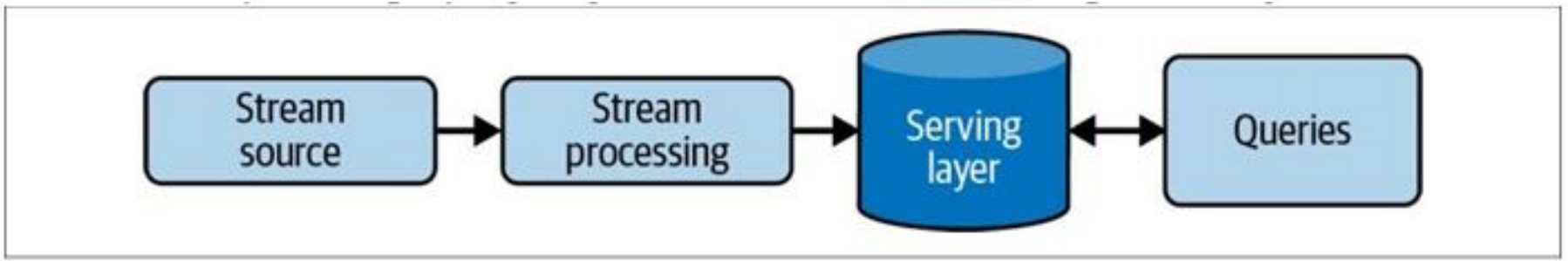
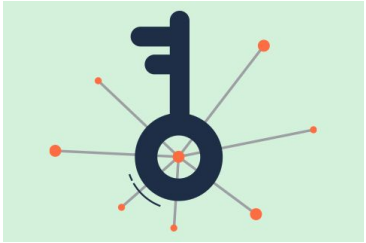


Figure 3-15. Kappa architecture

Surrogate vs Natural Key



It depend on several factors.

- The natural of the data.
- Database (DWH) platform.
- The group who uses this data.

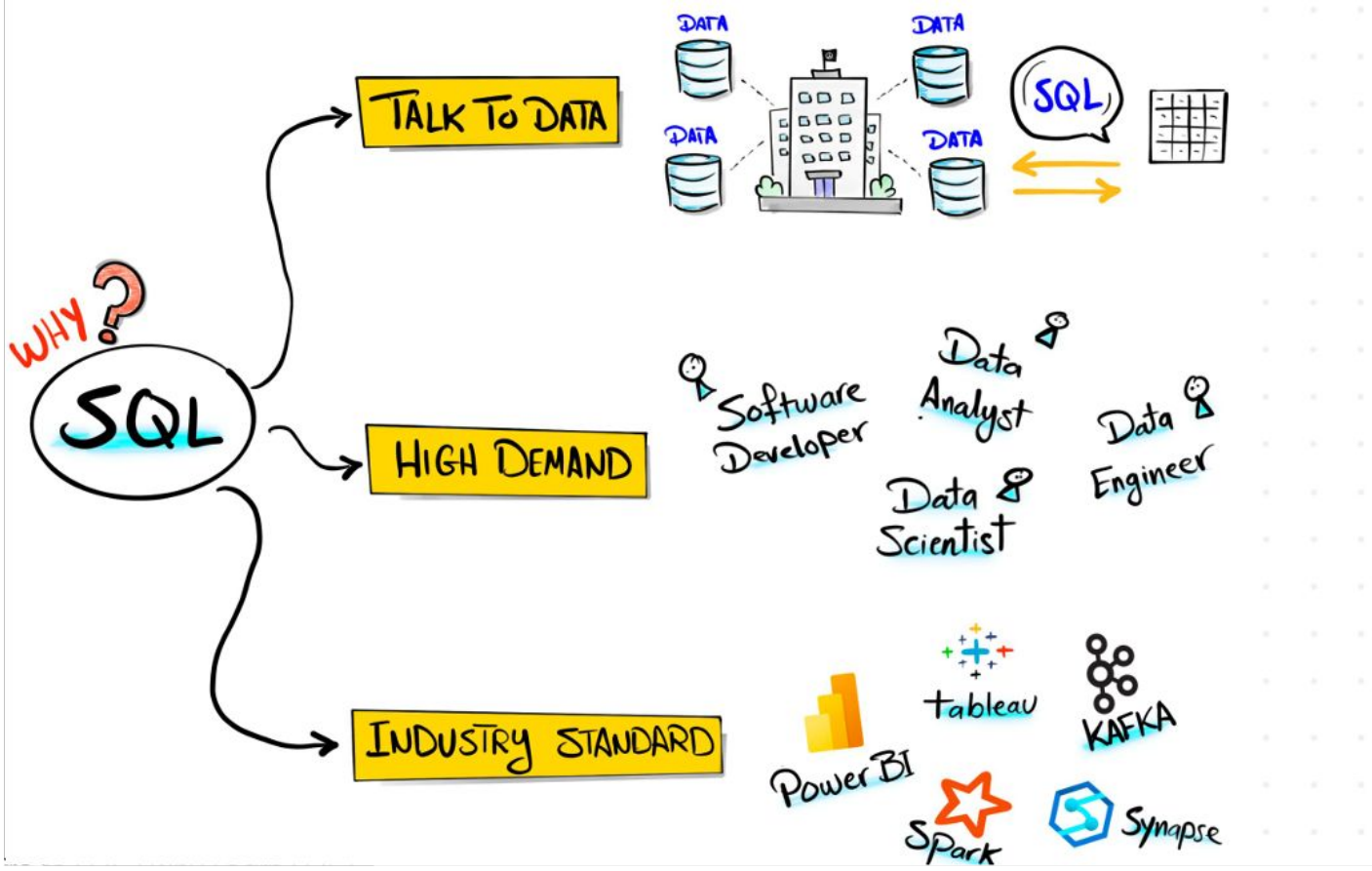
Do we need to remove the natural key to use surrogate key ?

- NO, we will keep both in the table and treat the surrogate key as primary key.

Reasons to Learn SQL

Data Engineer

Talk To Data!


WHY?
SQL
TALK TO DATA
HIGH DEMAND
INDUSTRY STANDARD
DATA
DATA
DATA
DATA
SQL
Software Developer
Data Analyst
Data Engineer
Data Scientist
tableau
KAFKA
PowerBI
Spark
Synapse

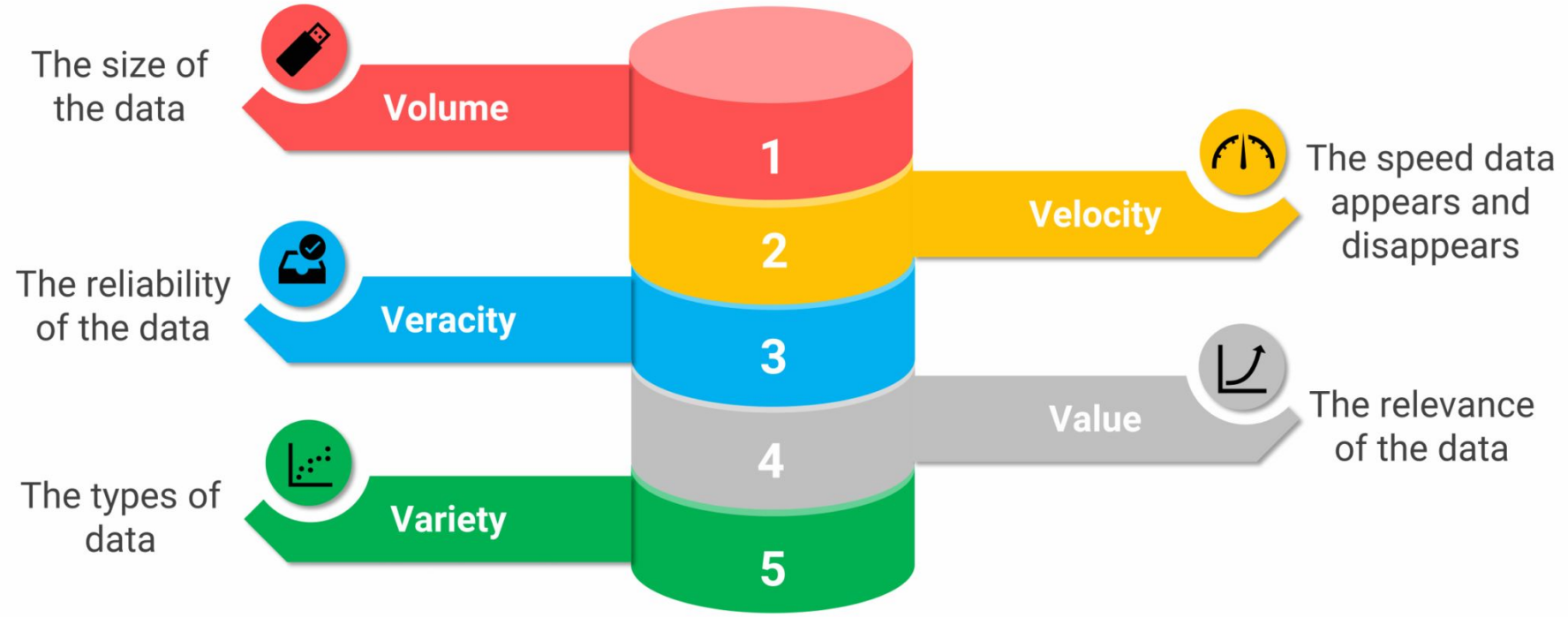
When Does Data Become "Big Data"

The 3 V's of big data

Big data is a collection of data from various sources, often characterized by what's become known as the 3 V's: *volume*, *velocity* and *variety*.

Volume	Velocity	Variety
The amount of data from myriad sources.	The speed at which big data is generated.	The types of data: structured, semistructured, unstructured.

The 5 Vs of Big Data



SQL Server Overview

- What is SQL Server?
- Editions (Developer, Express, Enterprise)
- Why we use it in this course

Installation Steps Overview

- Download SQL Server (Developer Edition)
- Download SQL Server Management Studio (SSMS)
- Custom Installation Configurations.


Any
QUESTIONS?