SQL DATA WARHOUESE

Project

SQL Projects

Data Warehousing

"Organize, Structure, Prepare"

- ETL/ELT Processing
- Data Architecture
- Data Integration
- Data Cleansing
- Data Load
- Data Modeling

Exploratory Data Analysis (EDA)

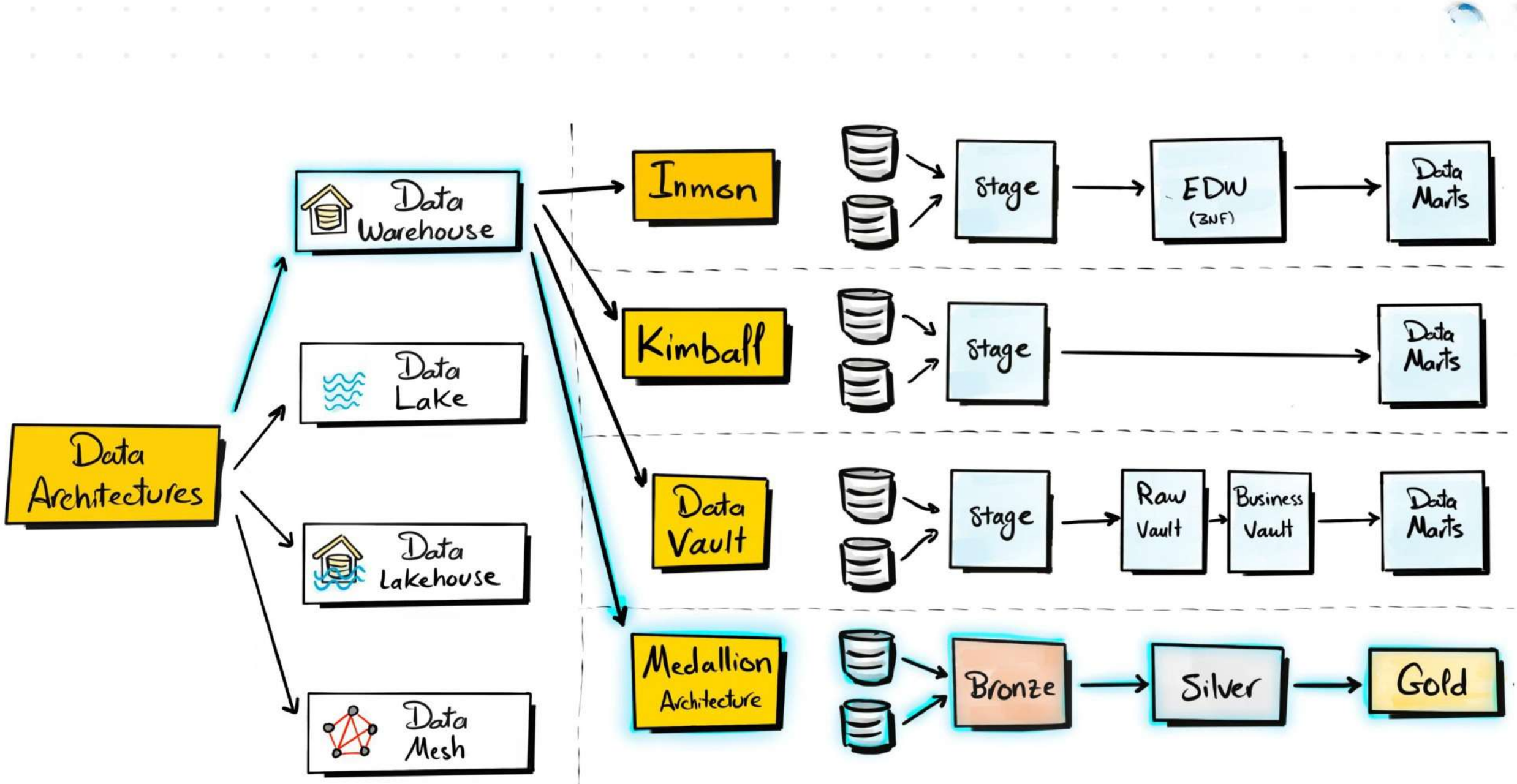
"Understand Data"

- Basic Queries
- Data Profiling
- Simple Aggregations
- Subquery

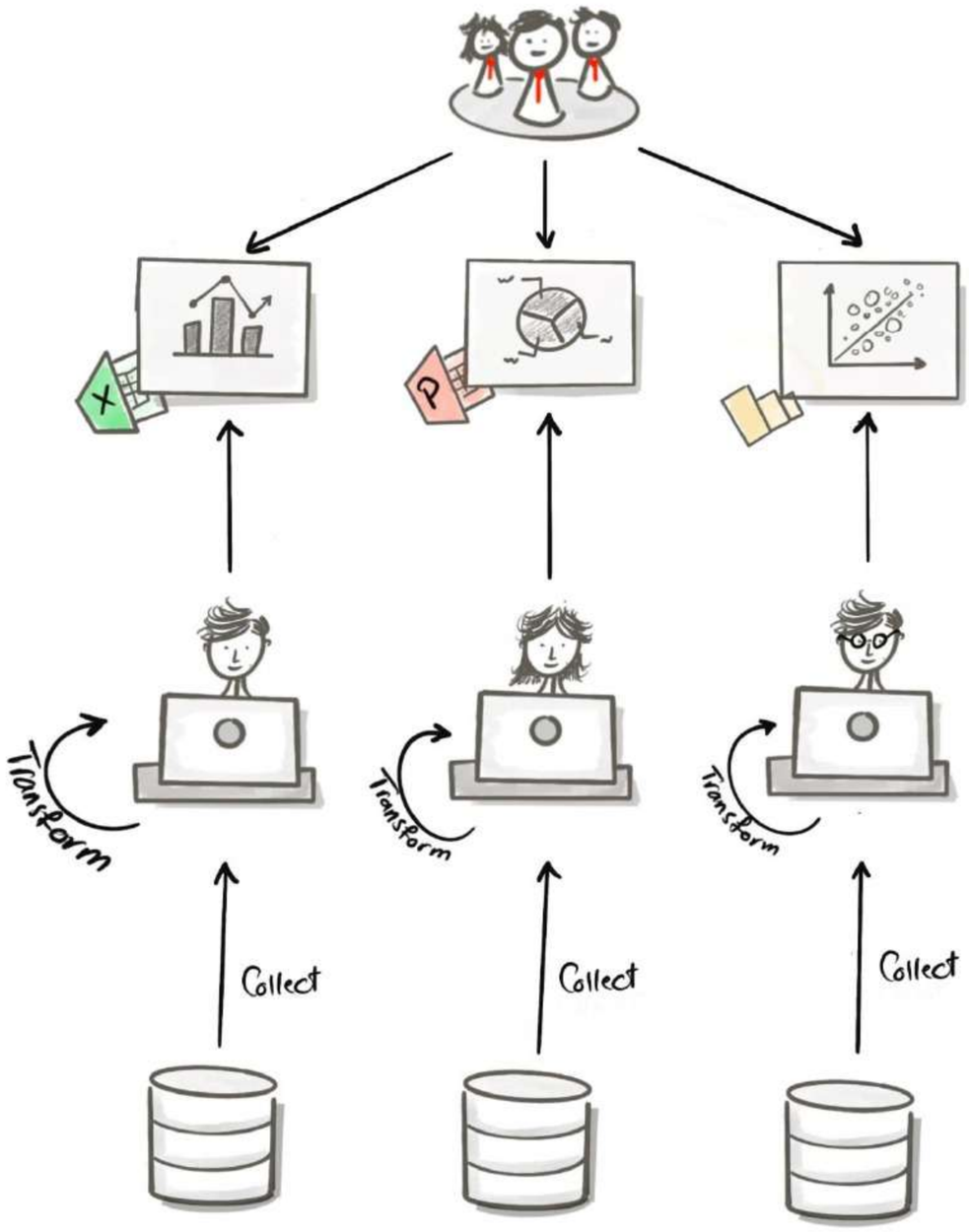
Advanced Data Analytics

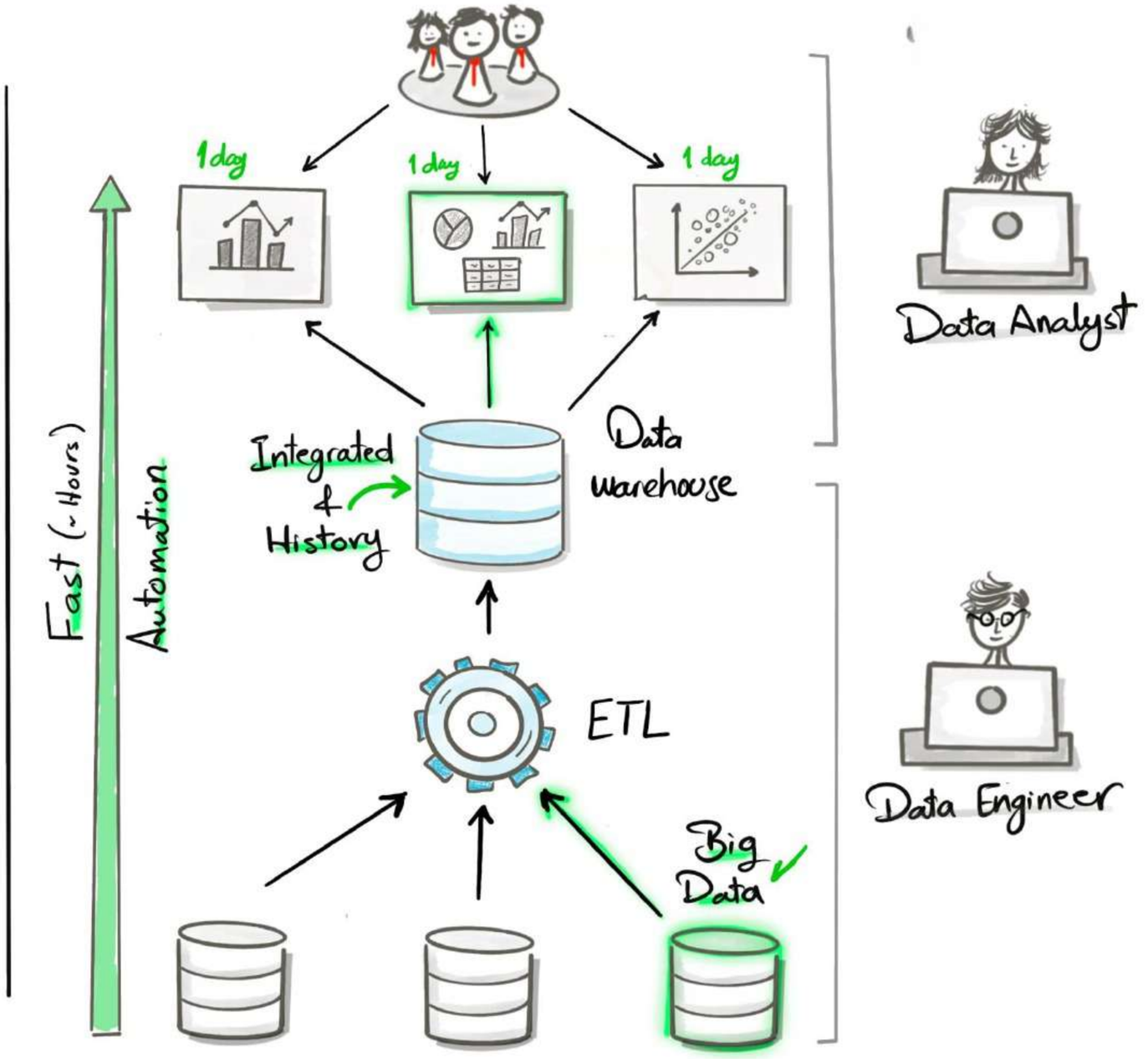
"Answer Business Questions"

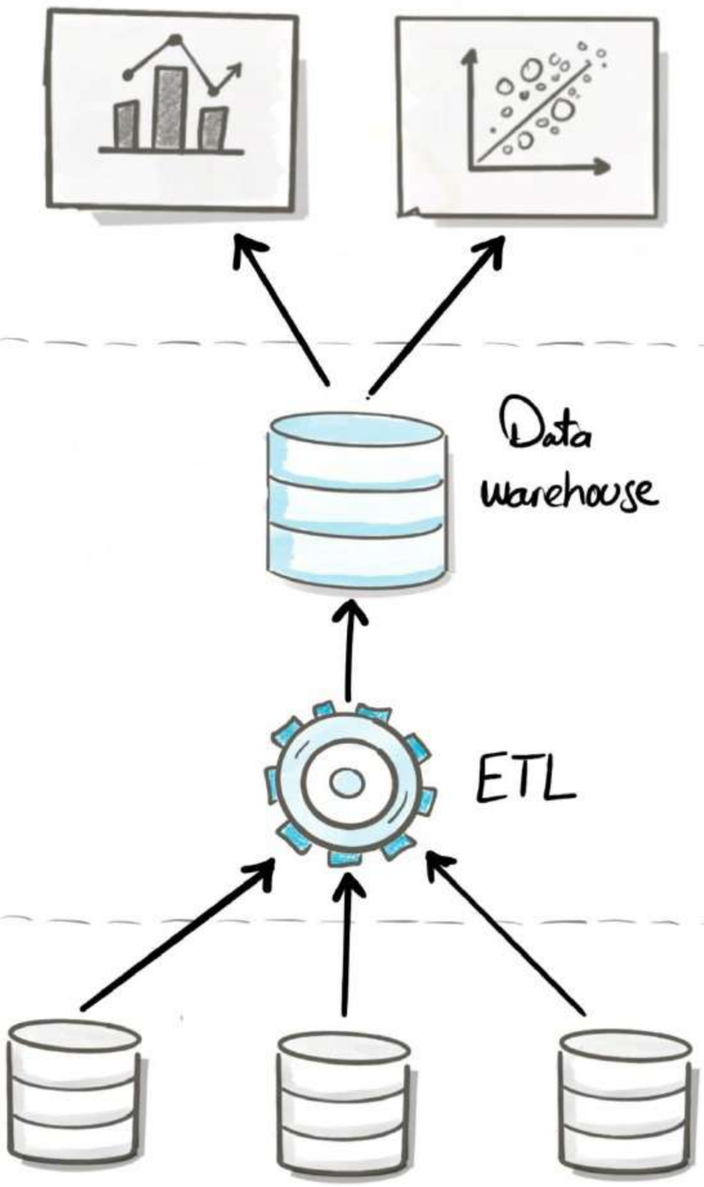
- Complex Queries
- Window Functions
- CTE
- Subqueries
- Reports

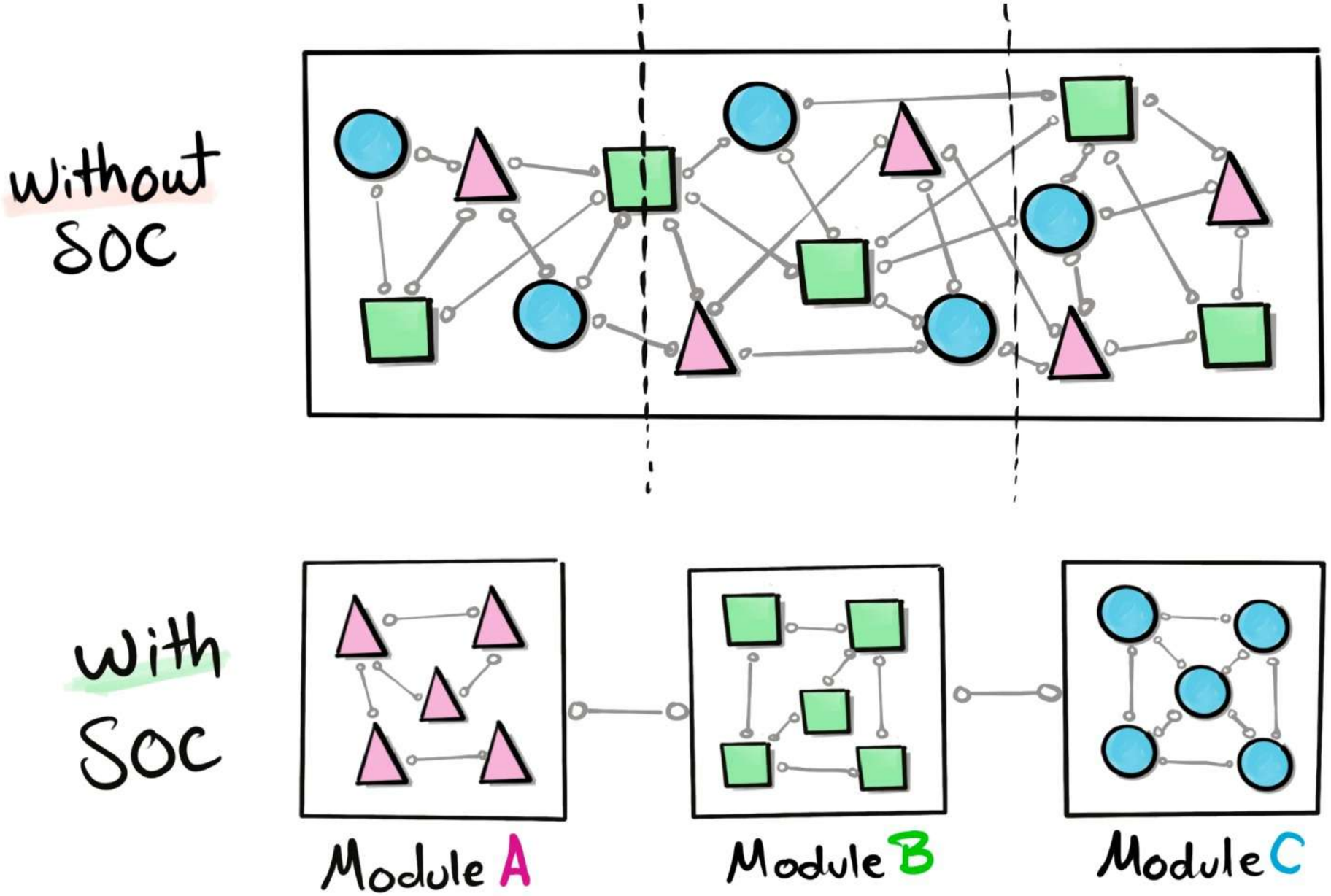

Data Architectures
Data Warehouse
Data Lake
Data Lakehouse
Data Mesh
Inmon
Kimball
Data Vault
Medallion Architecture
Stage
EDW (3NF)
Data Marts
Stage
Data Marts
Stage
Raw Vault
Business Vault
Data Marts
Bronze
Silver
Gold

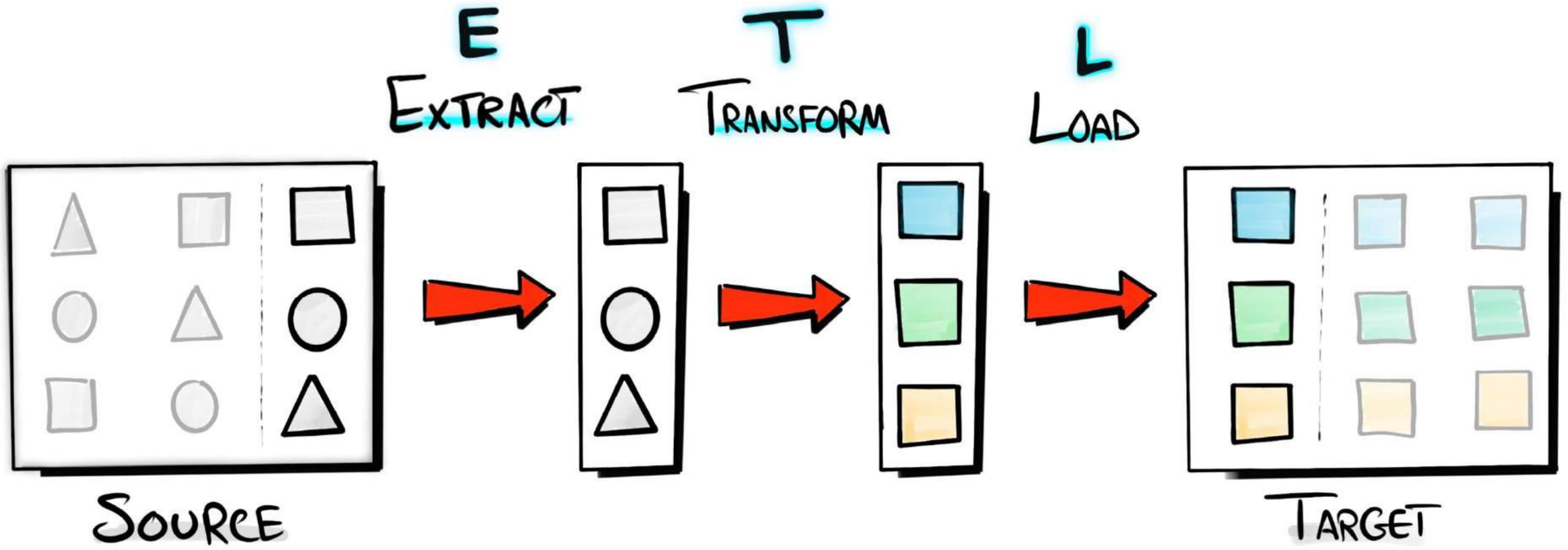
	Bronze Layer	Silver Layer	Gold Layer
Definition	Raw, unprocessed data as-is from sources	Clean & standardized data	Business-Ready data
Objective	Traceability & Debugging	(Intermediate Layer) Prepare Data for Analysis	Provide data to be consumed for reporting & Analytics
Object Type	Tables	Tables	Views
Load Method	**Full Load** (Truncate & Insert)	**Full Load** (Truncate & Insert)	None
Data Transformation	None (as-is)	- Data **Cleaning** - Data **Standardization** - Data **Normalization** - **Derived Columns** - Data **Enrichment**	- Data **Integration** - Data **Aggregation** - **Business Logic & Rules**
Data Modeling	None (as-is)	None (as-is)	- Start Schema - Aggregated Objects - Flat Tables
Target Audience	- Data Engineers	- Data Analysts - Data Engineers	- Data Analysts - Business Users

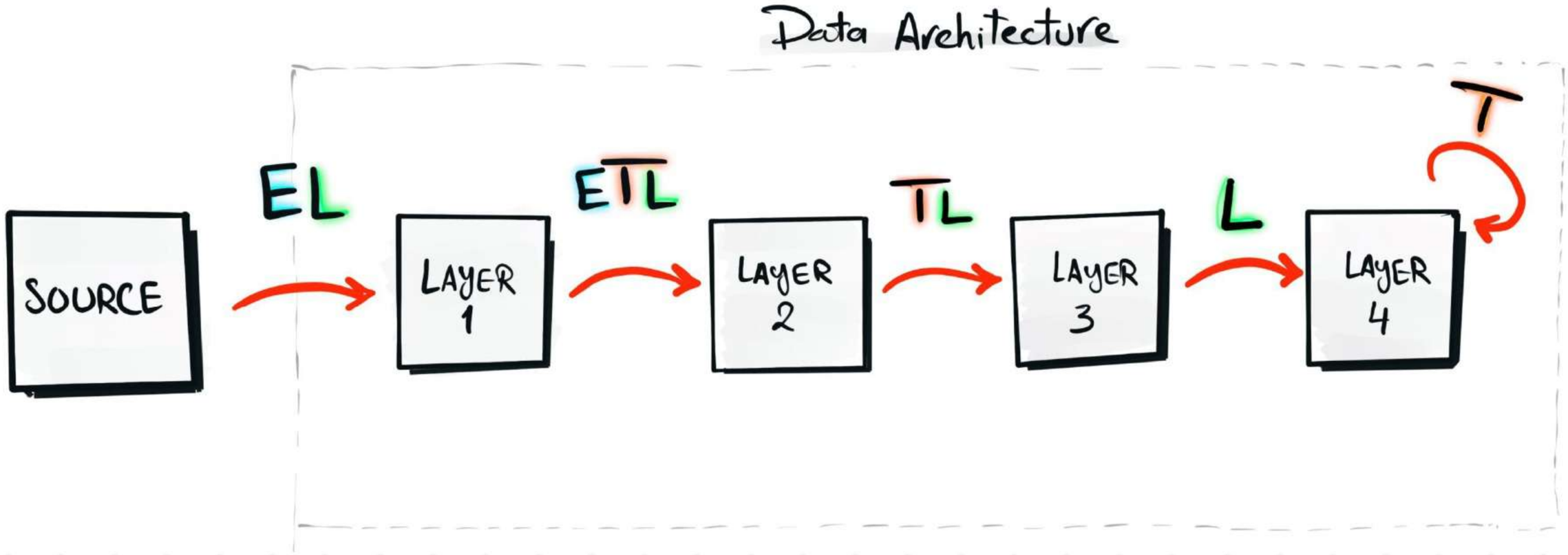

Transform
Collect
Transform
Collect
Transform
Collect


1 day
1 day
1 day
Data Analyst
Fast (~ Hours)
Automation
Integrated
&
History
Data
Warehouse
ETL
Big
Data
Data Engineer

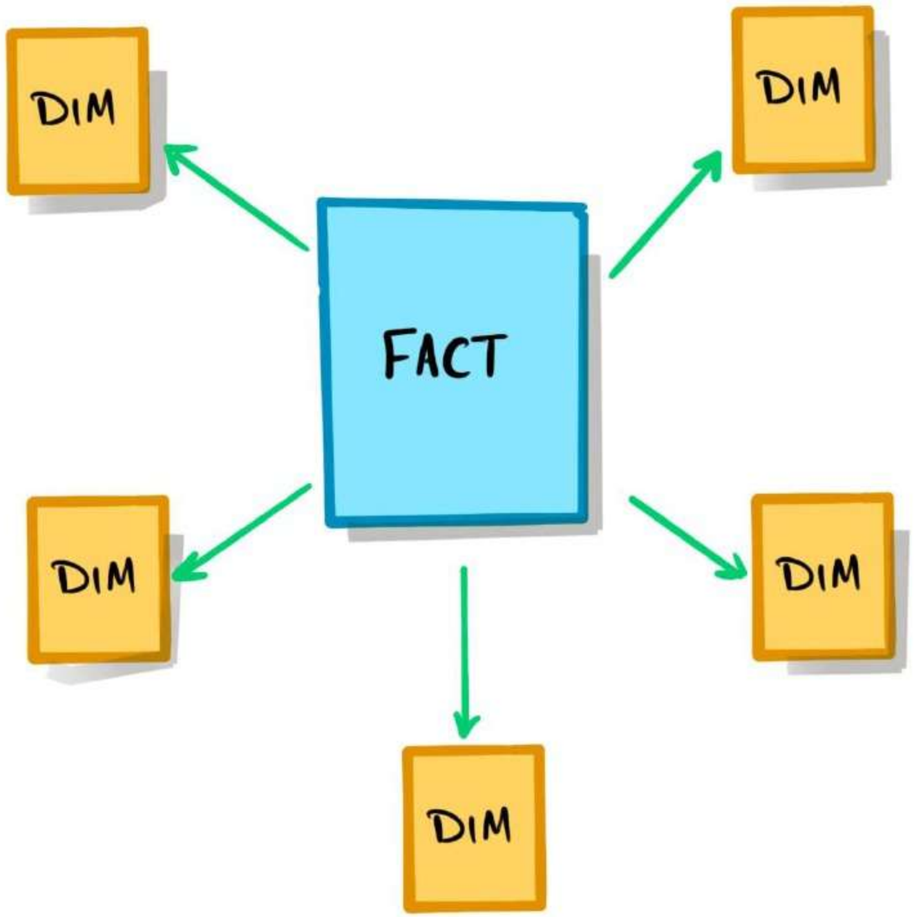

Data
Warehouse
ETL


Without SOC
With SOC
Module A
Module B
Module C

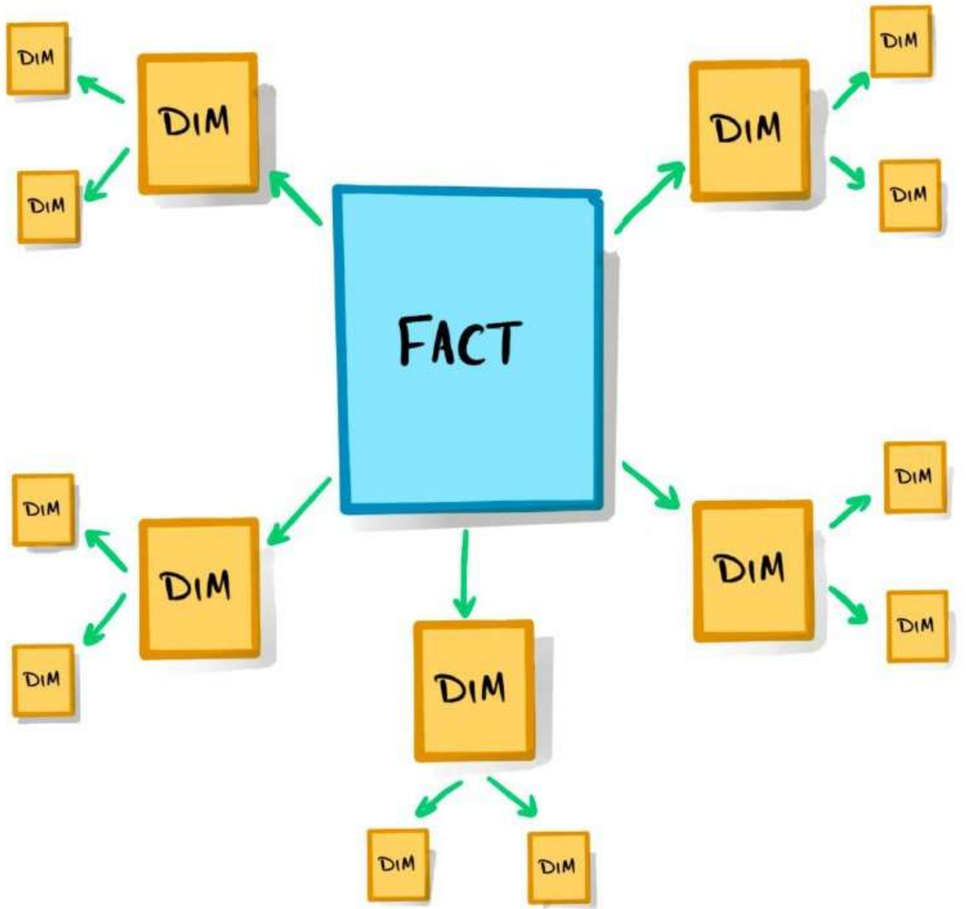

E
EXTRACT
T
TRANSFORM
L
LOAD
SOURCE
TARGET

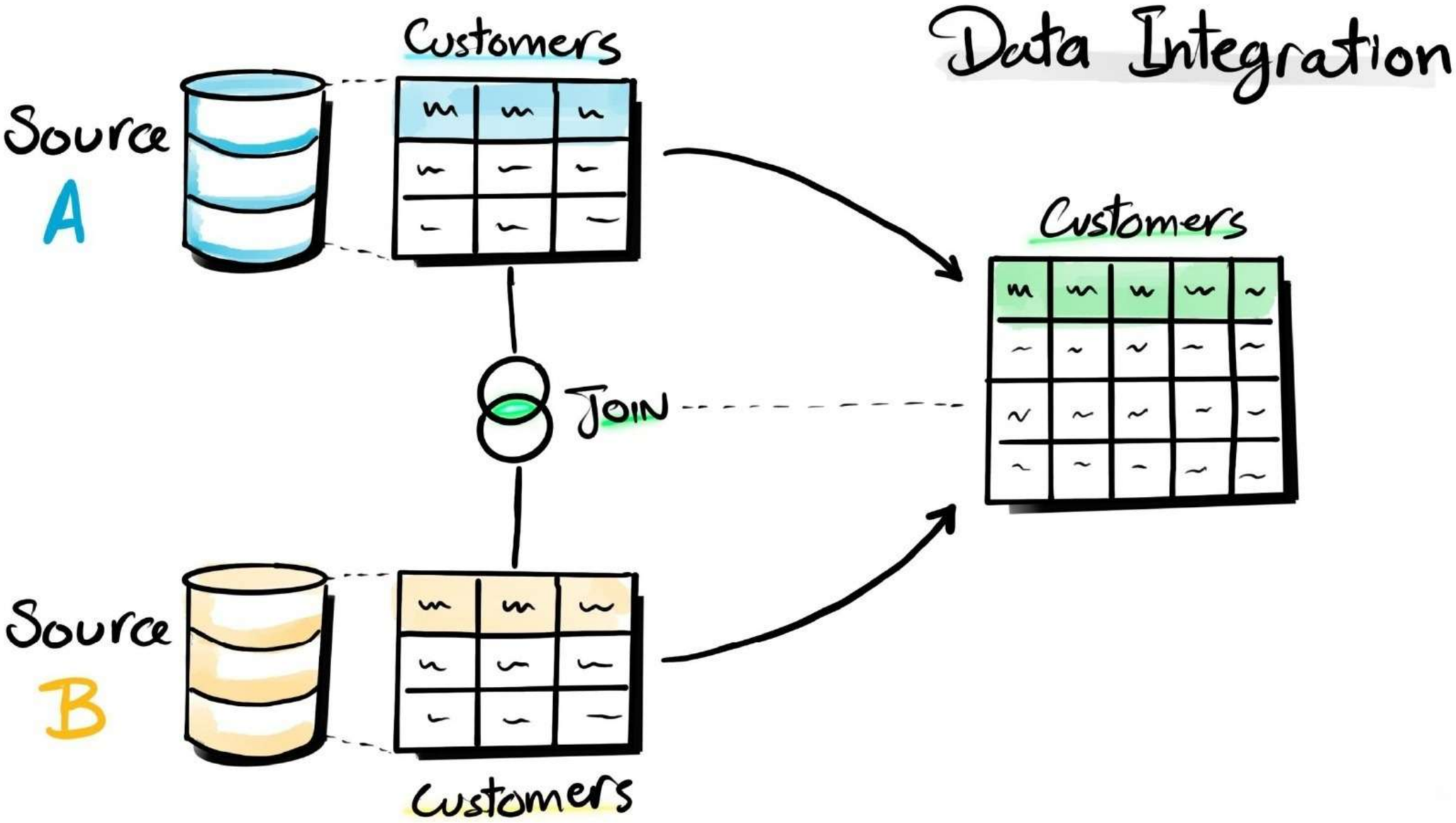

Data Architecture
SOURCE
EL
LAYER 1
ETL
LAYER 2
TL
LAYER 3
L
LAYER 4
T

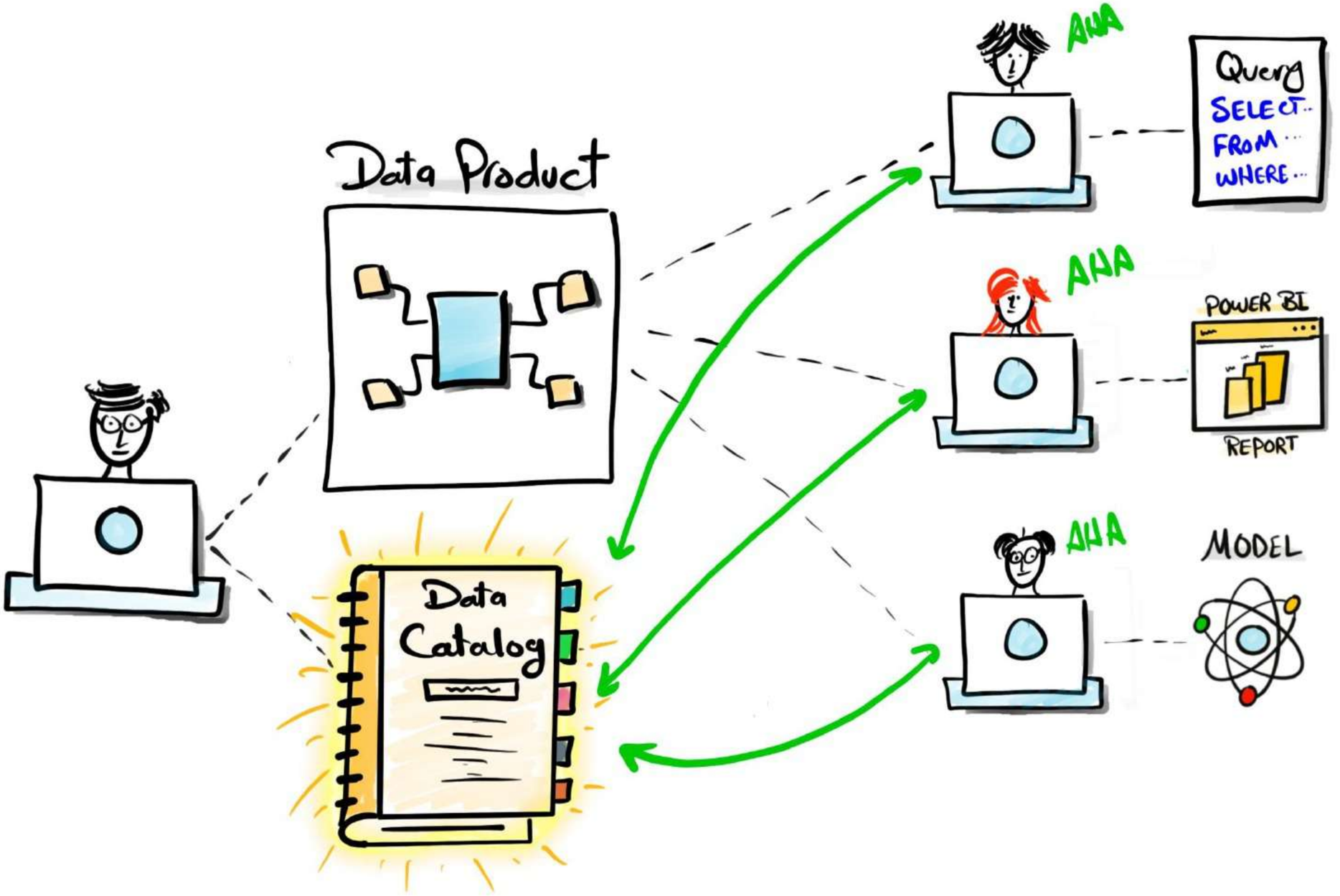
STAR SCHEMA

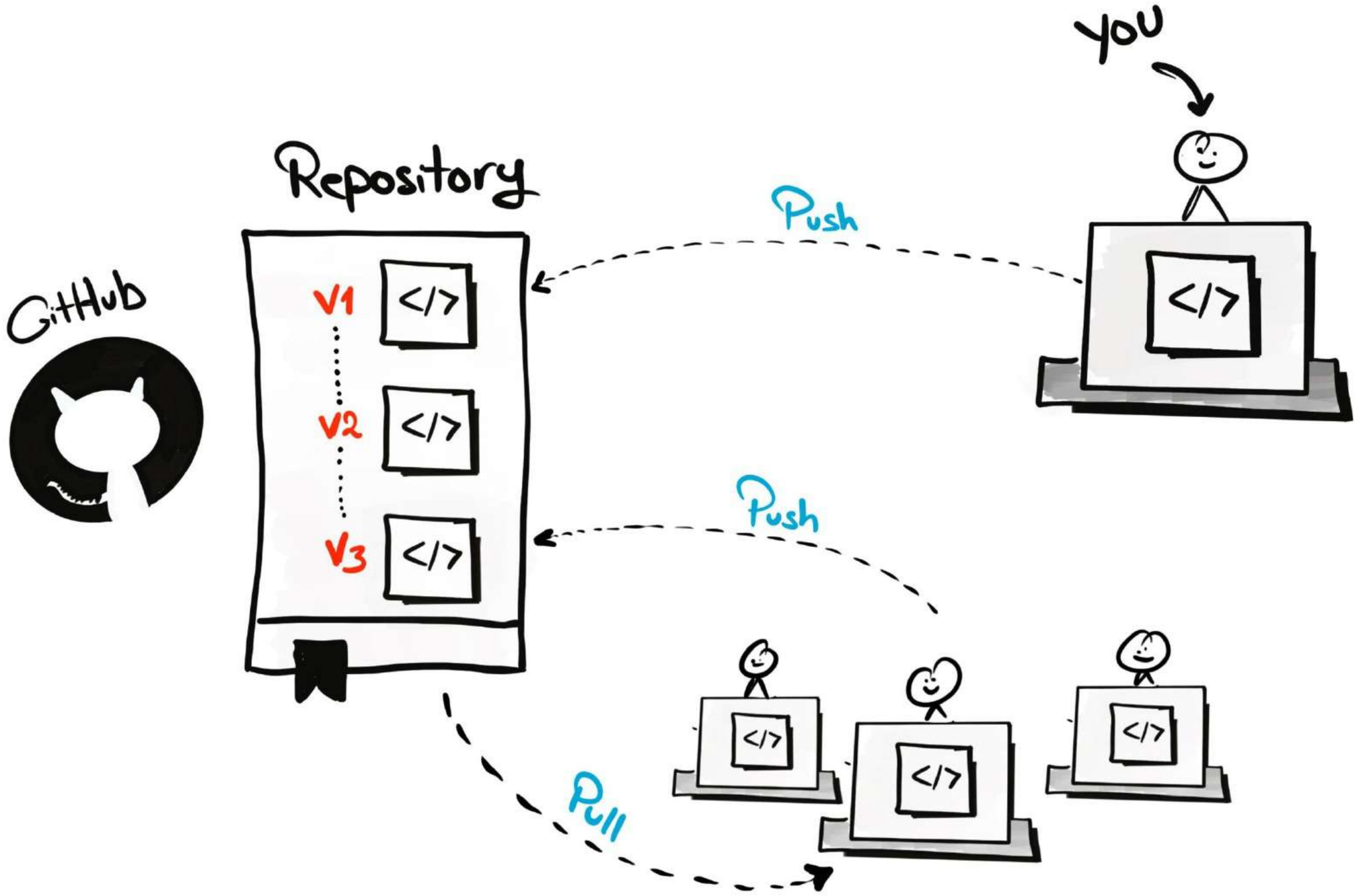


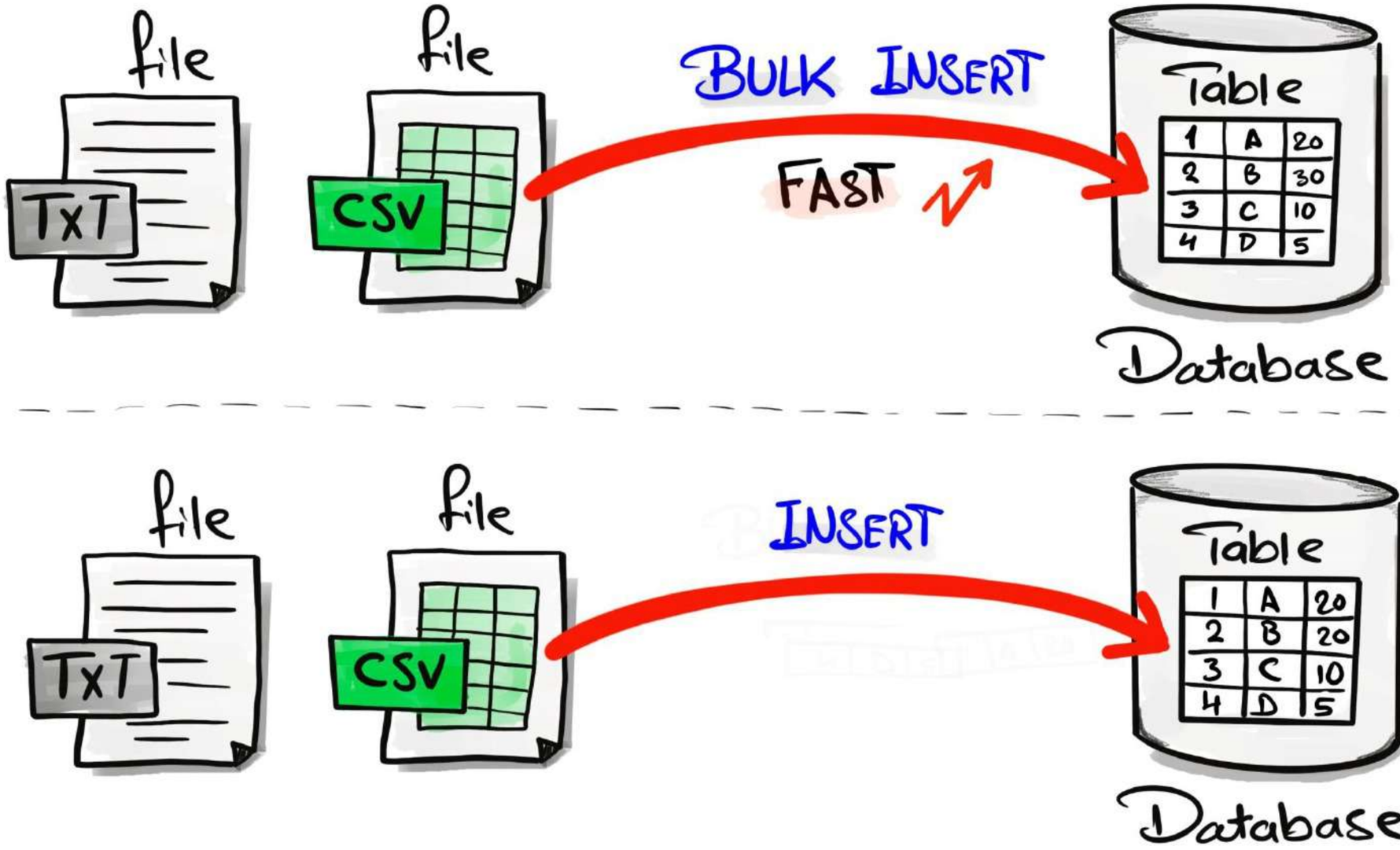
SNOWFLAKE SCHEMA




Data Integration
Customers
Source
A
JOIN
Customers
Source
B
Customers

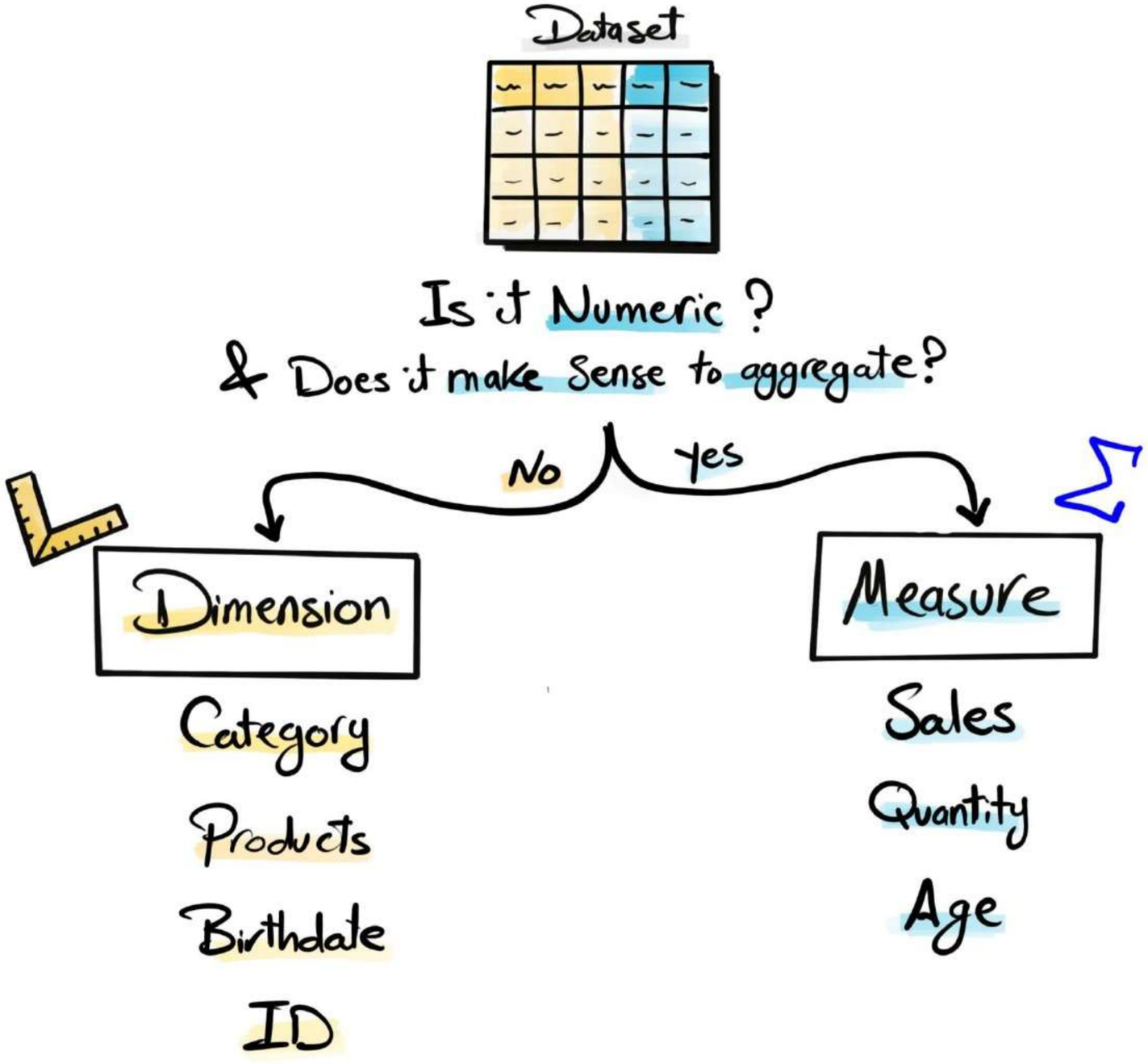

Data Product
Data Catalog
AHA
Query
SELECT...
FROM...
WHERE...
AHA
POWER BI
REPORT
AHA
MODEL

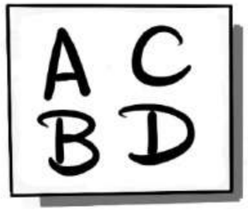

you
Repository
Push
GitHub
V1
V2
V3
Push
Pull


file
TXT
file
CSV
BULK INSERT
FAST
Table
1 A 20
2 B 30
3 C 10
4 D 5
Database
file
TXT
file
CSV
INSERT
Table
1 A 20
2 B 20
3 C 10
4 D 5
Database

SQL DATA Analytics

Project

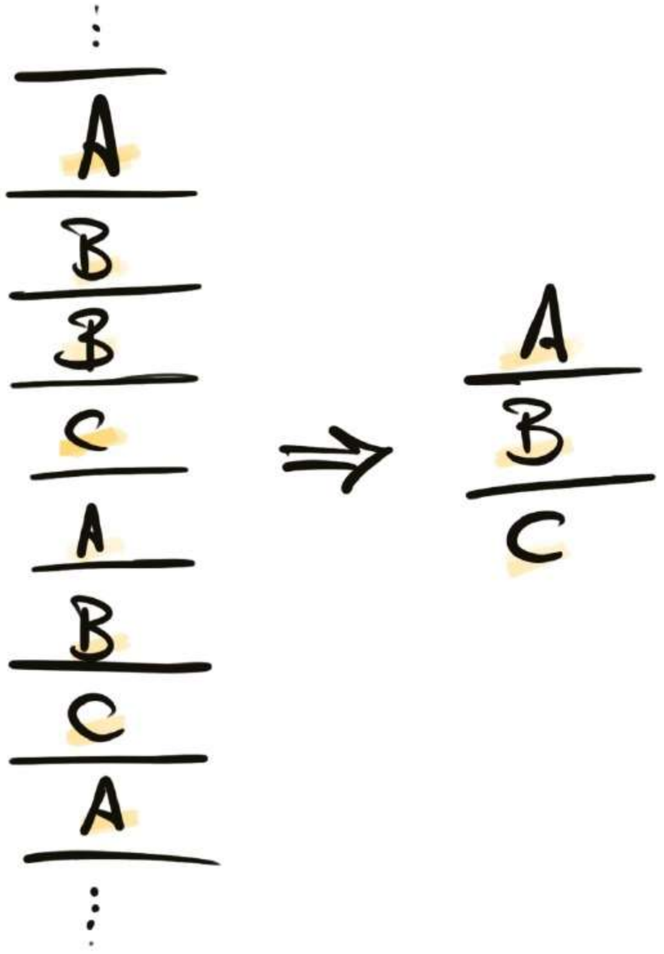

Dataset
Is it Numeric ?
& Does it make Sense to aggregate?
No
Yes
Dimension
Category
Products
Birthdate
ID
Measure
Sales
Quantity
Age

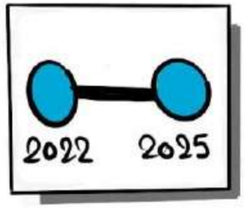


Dimensions Exploration

```
DISTINCT [Dimension]

DISTINCT Country
DISTINCT Category
DISTINCT Product
```





Date Exploration

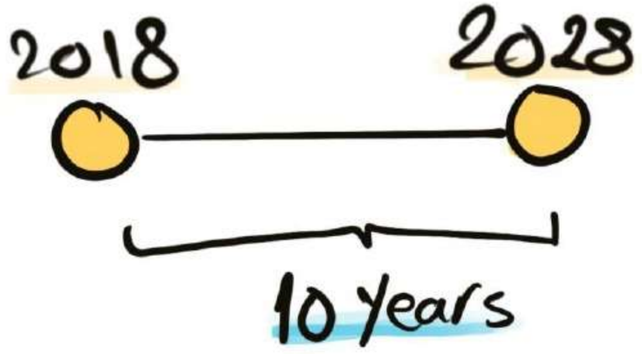
MIN/MAX [Date Dimension]

MIN Order_date

MAX Create_date

MIN Birthdate

2019
2020
2018
2018
2022
2023
2023
2028
2022



DATEDIFF

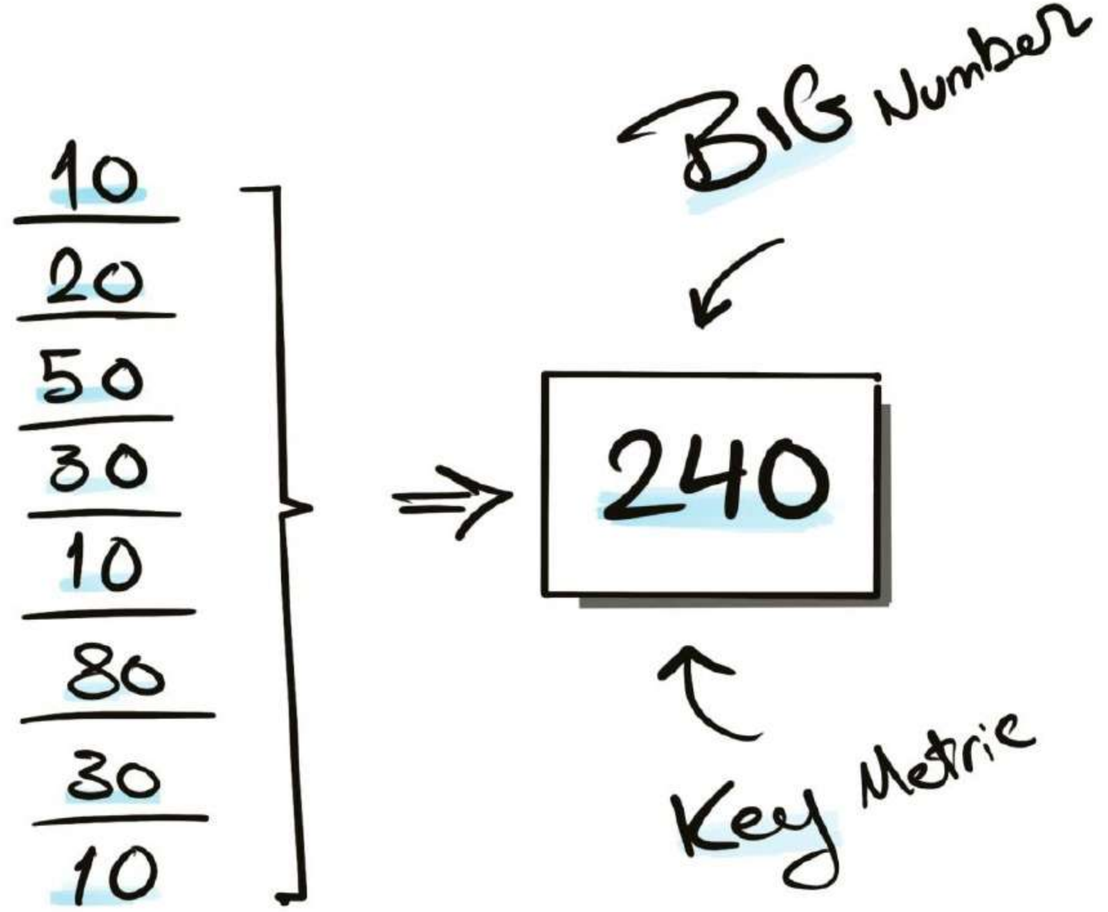
999 Measures Exploration

$\sum$ [Measure]

SUM (Sales)

AVG (Price)

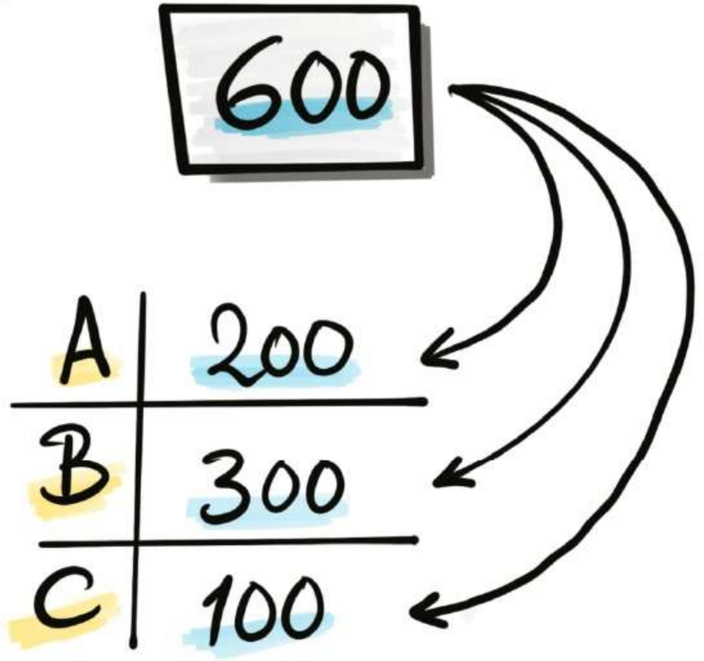
SUM (Quantity)

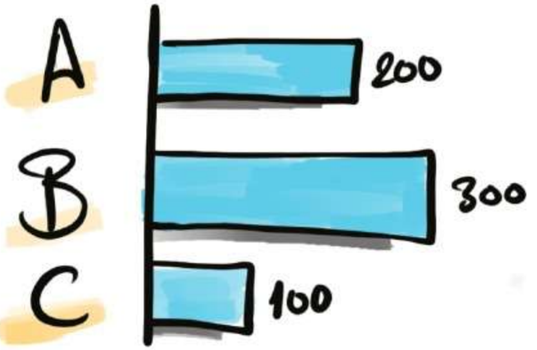


Magnitude

$\sum$ [Measure] By [Dimension]

- Total Sales By Country
- Total Quantity By Category
- Average Price By Product
- Total Orders By Customer





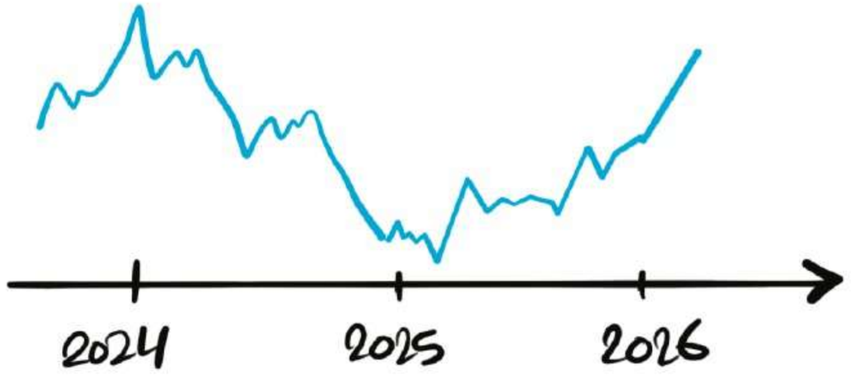
Change-Over-Time Trends

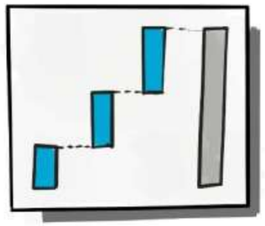
$\sum$ [Measure] By [Date Dimension]

Total Sales By Year

Average Cost By Month

2024	300
2025	100
2026	200





Cumulative Analysis

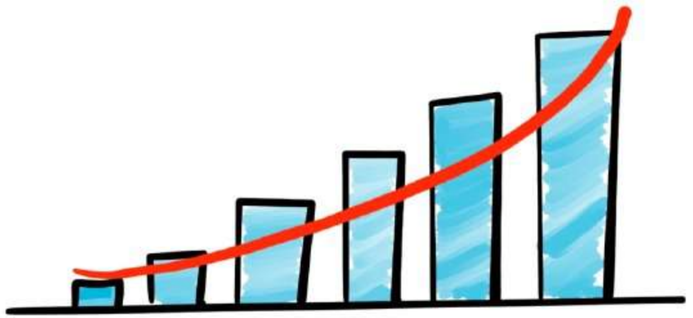
Cumulative

∑ [Cumulative Measure] By [Date Dimension]

Running Total Sales By Year

Moving Average of Sales By Month

2024	300	300
2025	100	400
2026	200	600



WINDOW FUNCTIONS

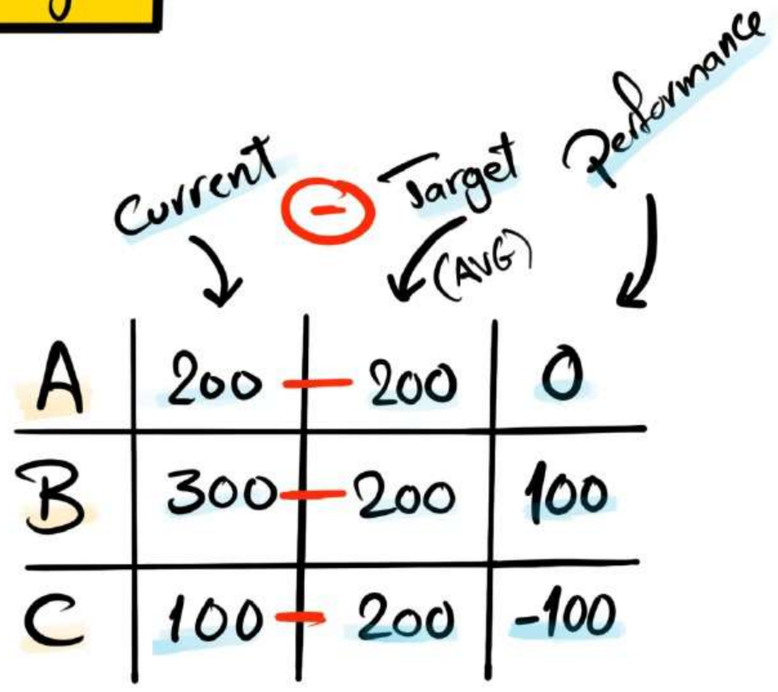
Performance Analysis

Current [Measure] – Target [Measure]

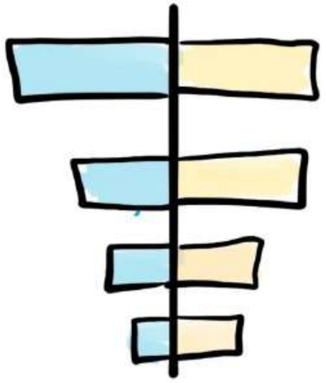
Current Sales – Average Sales

Current Year Sales – Previous Year Sales

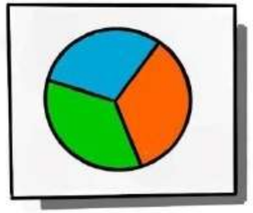
Current Sales – lowest Sales



A	200	200	0
B	300	200	100
C	100	200	-100



WINDOW FUNCTIONS



Part-to-Whole

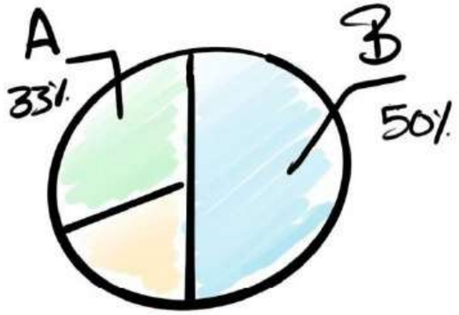
Proportional Analysis

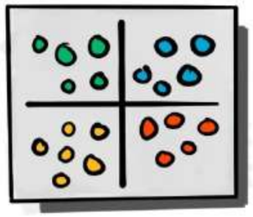
([Measure] / Total [Measure]) * 100 By [Dimension]

(Sales / Total Sales) * 100 By Category

(Quantity / Total Quantity) * 100 By Country

A	200	33%
B	300	50%
C	100	17%





Data Segmentation

[Measure] By [Measure]

Total Products By Sales Range

Total Customers By Age

CASE WHEN STATEMENT