



# Basic of Data Engineering

David Samanyaporn  
*Credit: Baraa*

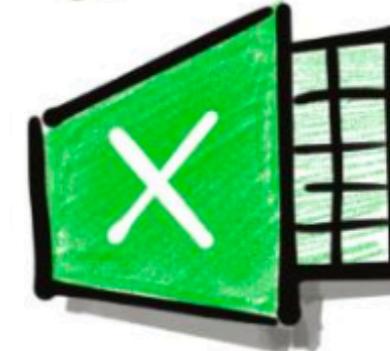
Big Data & Variety

# VARIETY

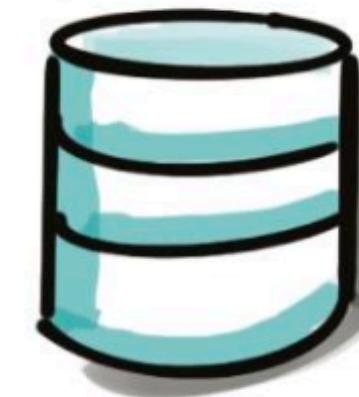
Structured



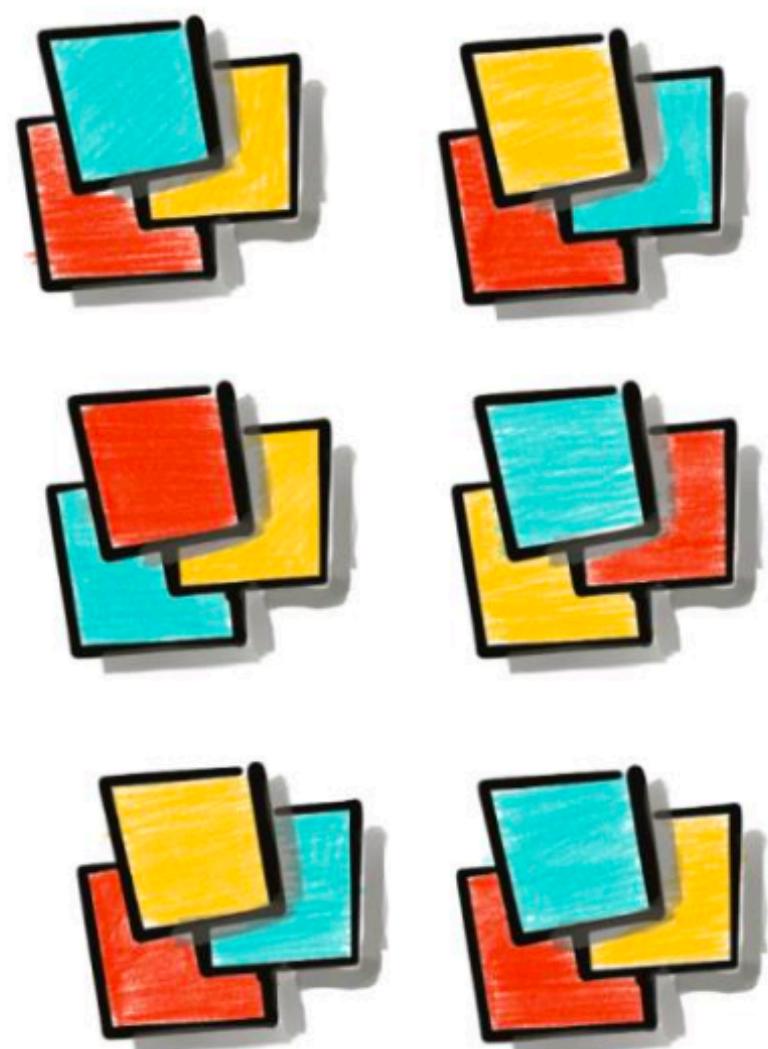
EXCEL



DATABASE



Semi- Structured



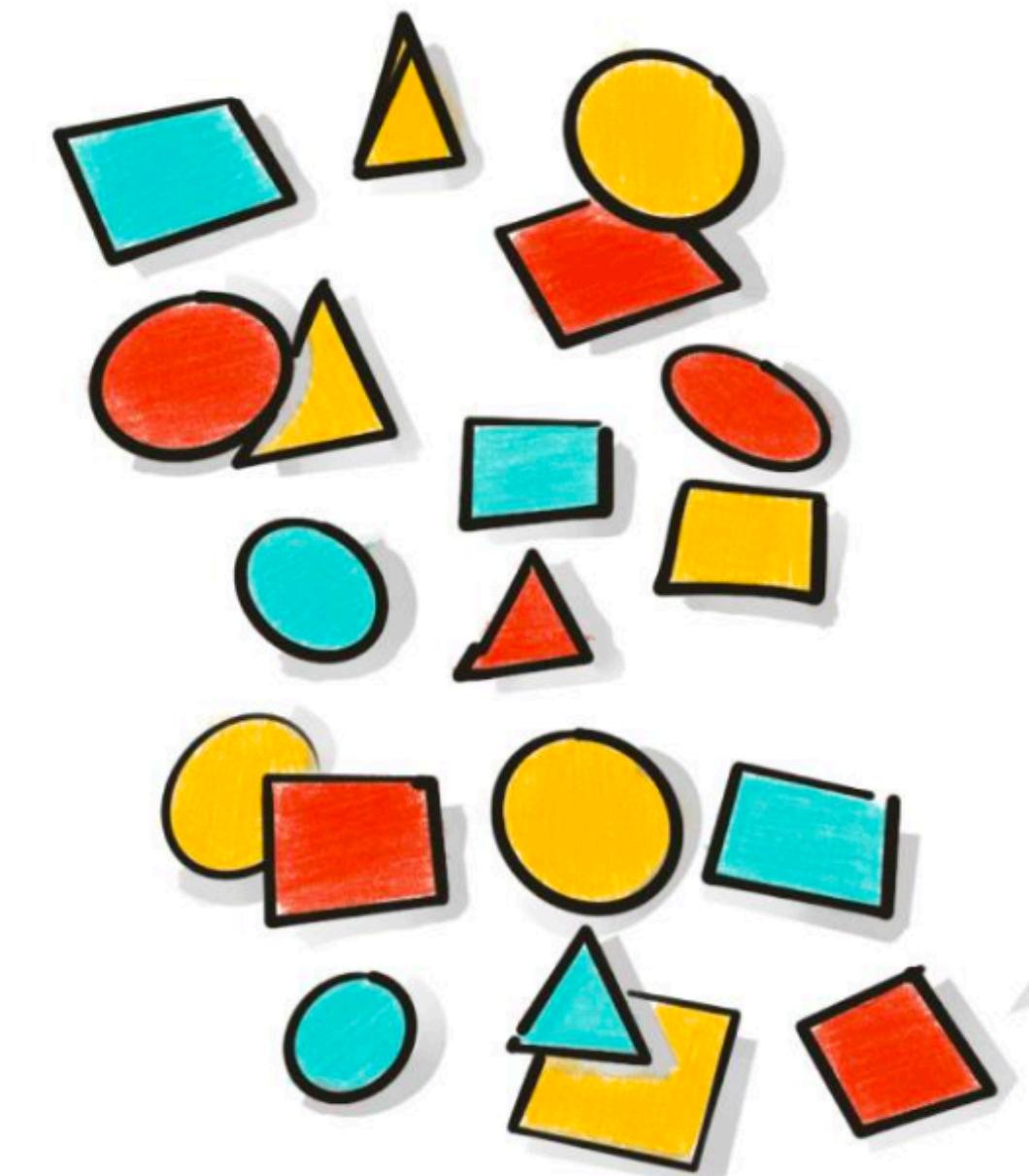
LOGs



WEBSITE



unstructured



VIDEO

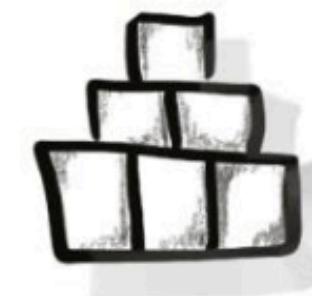


IMAGE

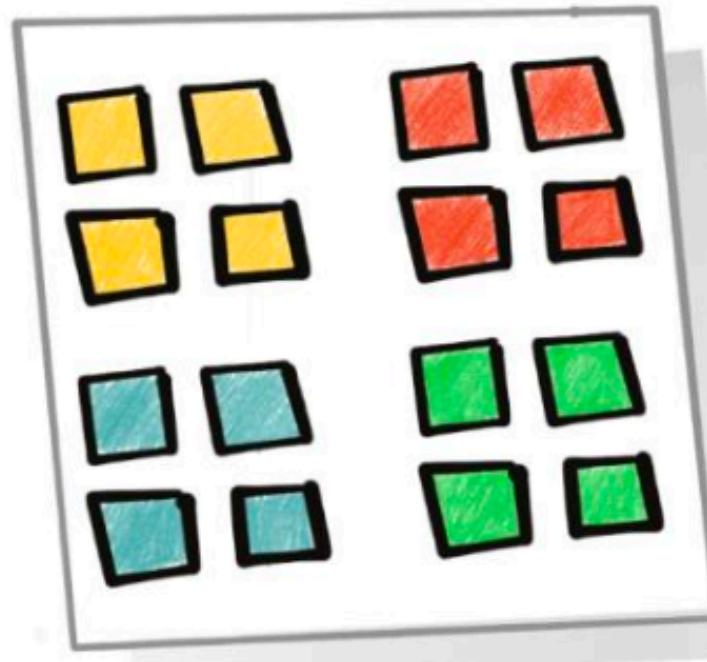


# Big Data & Velocity

# VELOCITY

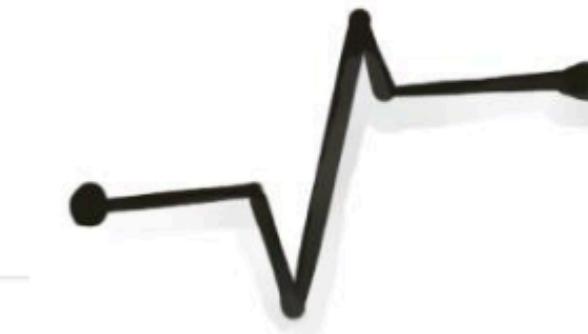


BATCH

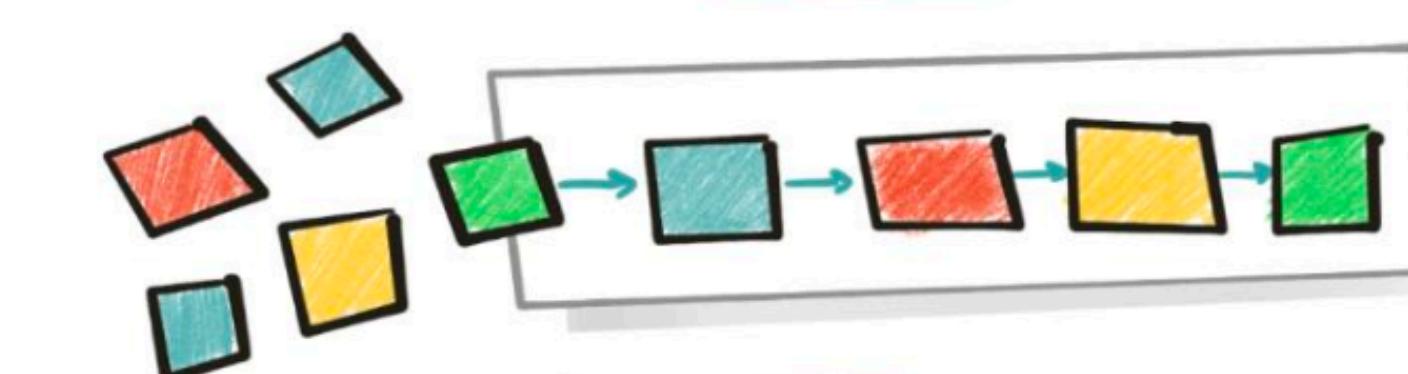


once a day

Hours → Mins



STREAM

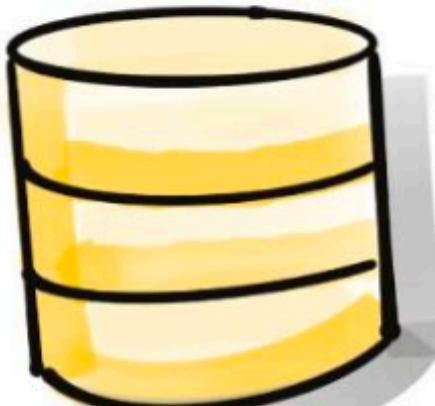


24

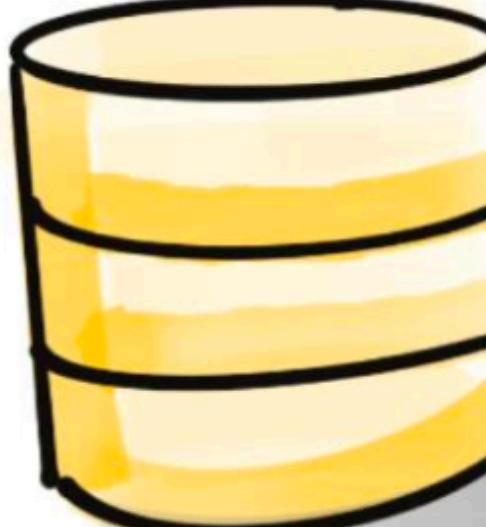
SEC → MSEC

# Big Data & Volume

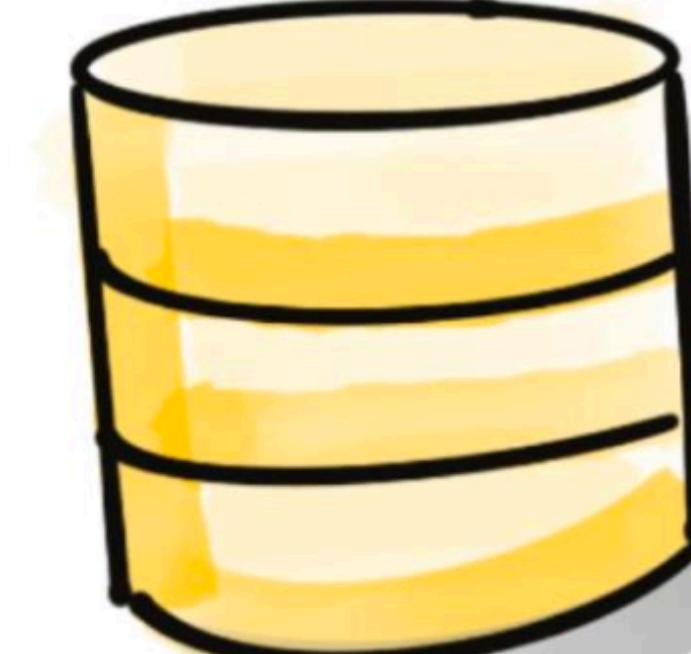
# VOLUME



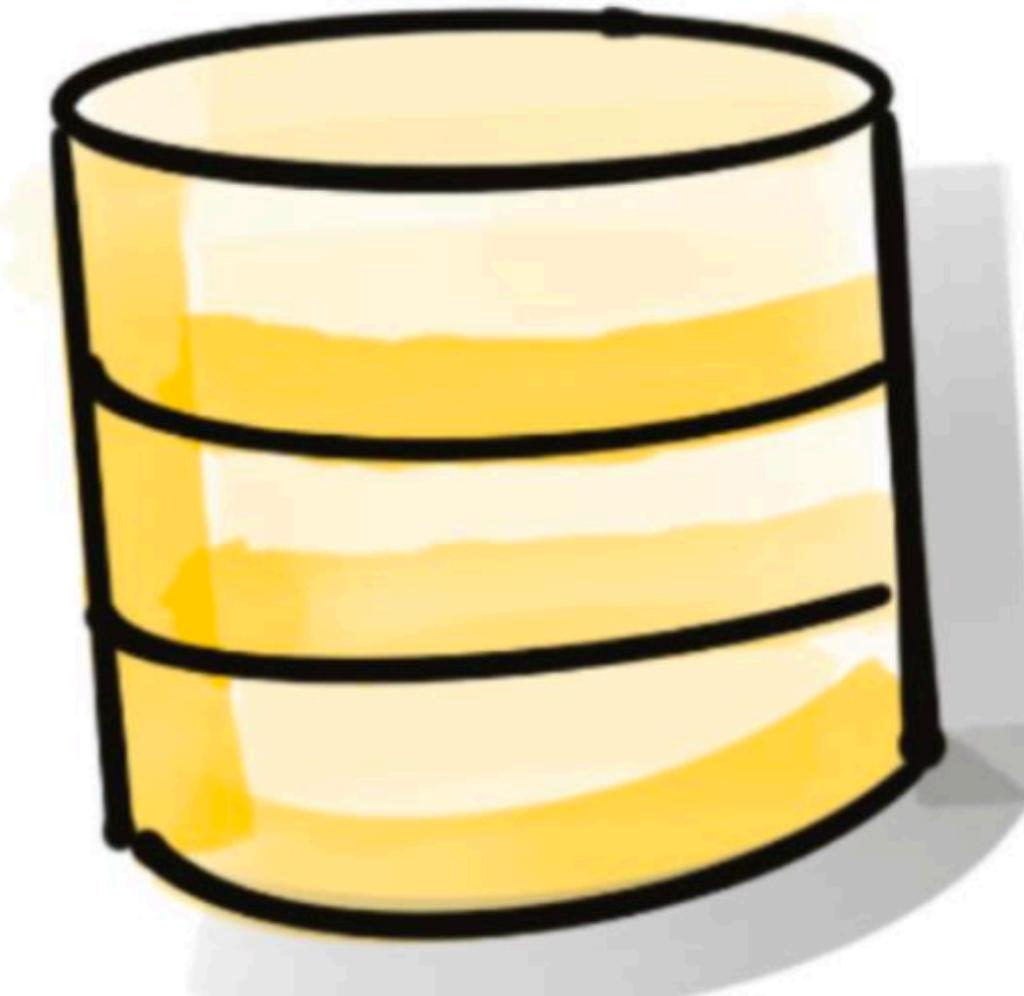
MB



GB



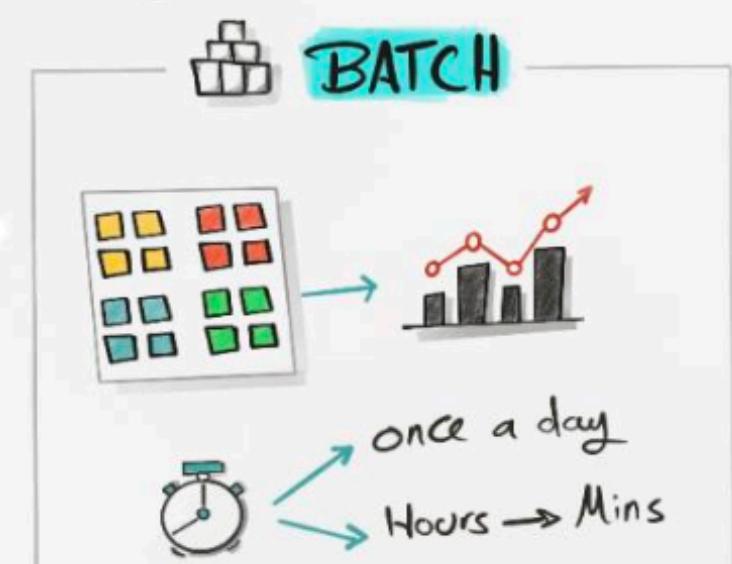
TB



PB

# Big Data

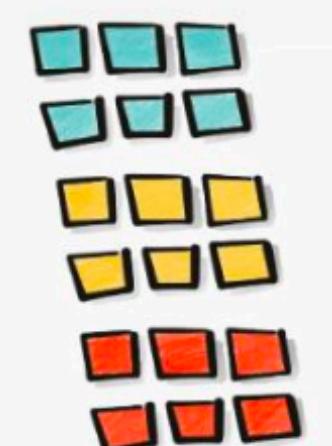
# VELOCITY



# BIG DATA

# VARIETY

## Structured



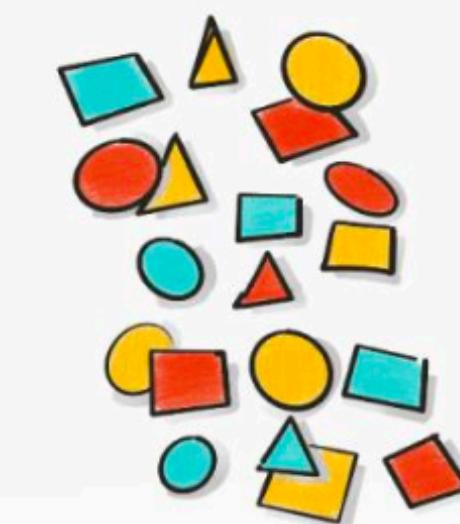
EXCEL DATABASE

## Semi-Structured



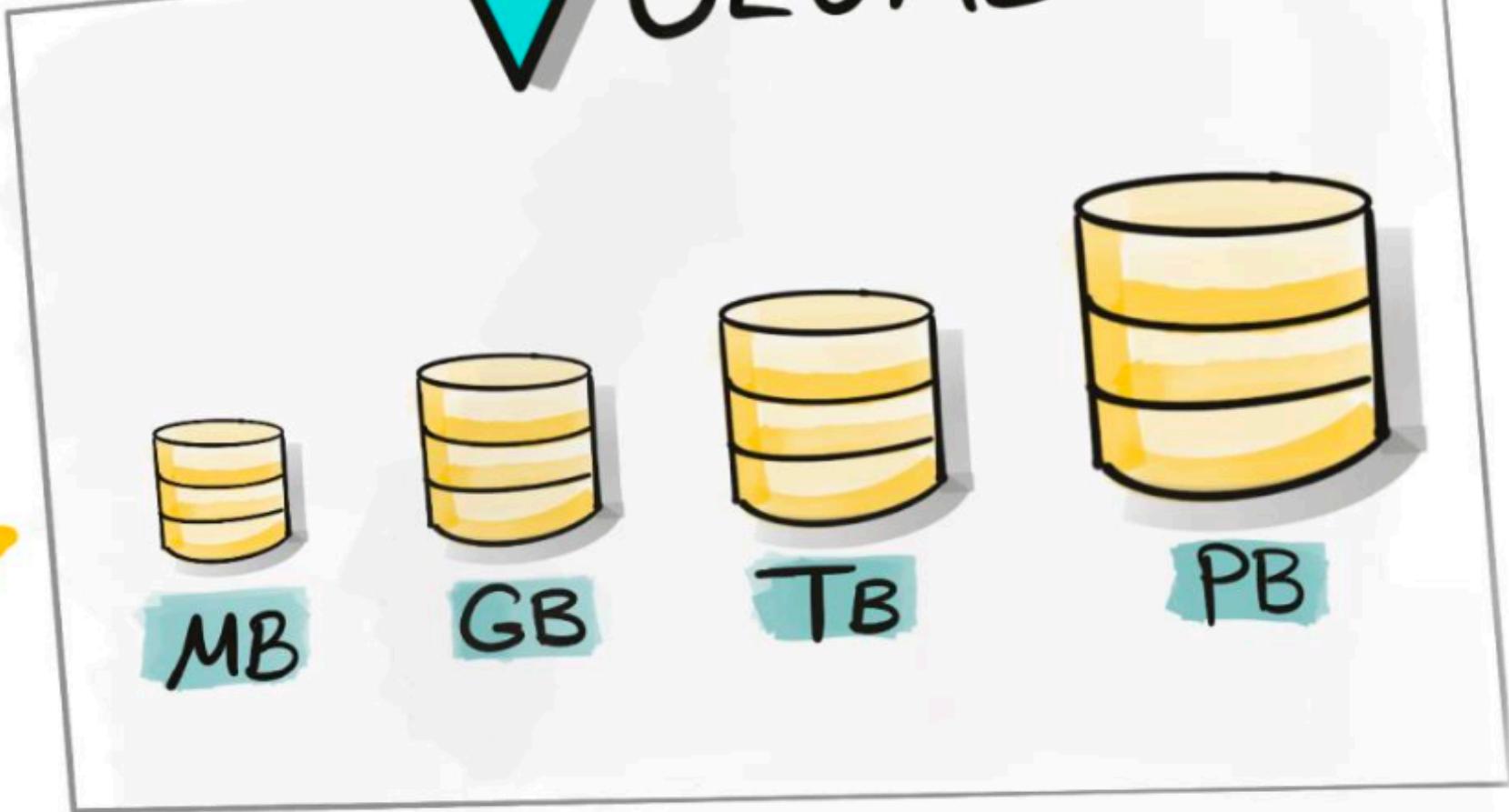
LOGS WEBSITE  
XML

## unstructured



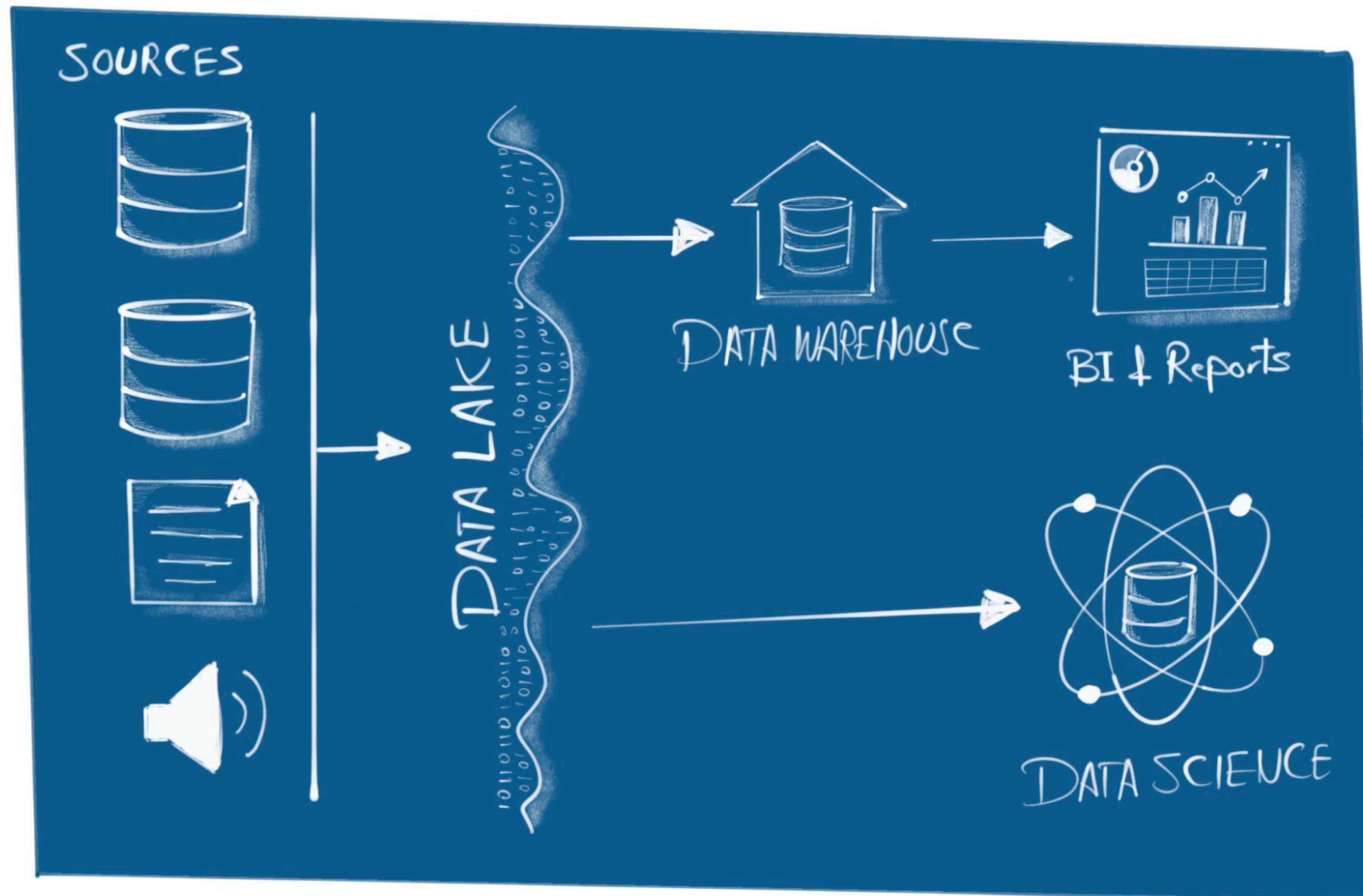
VIDEO IMAGE  
IMAGE

# VOLUME

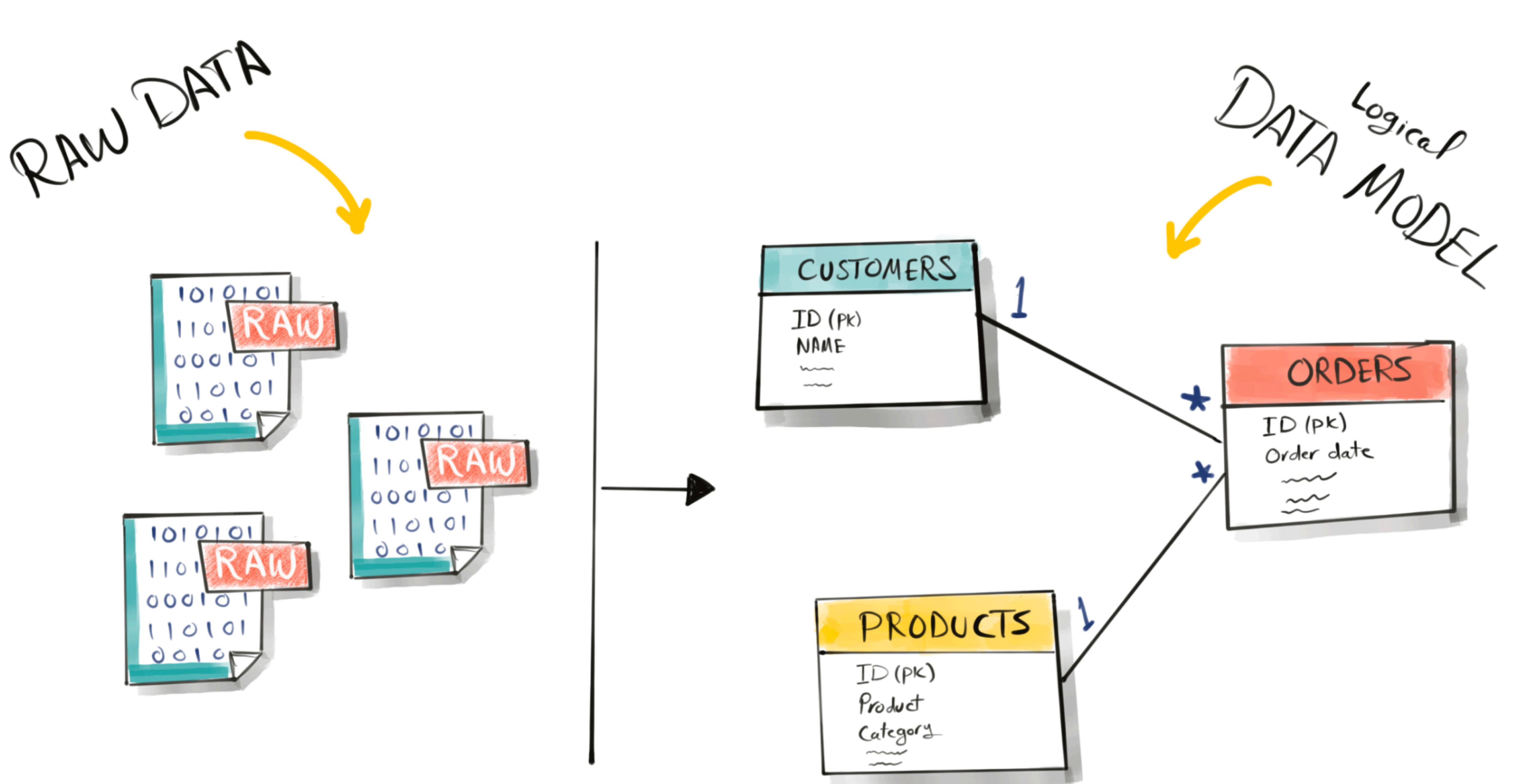


# Data Architecture

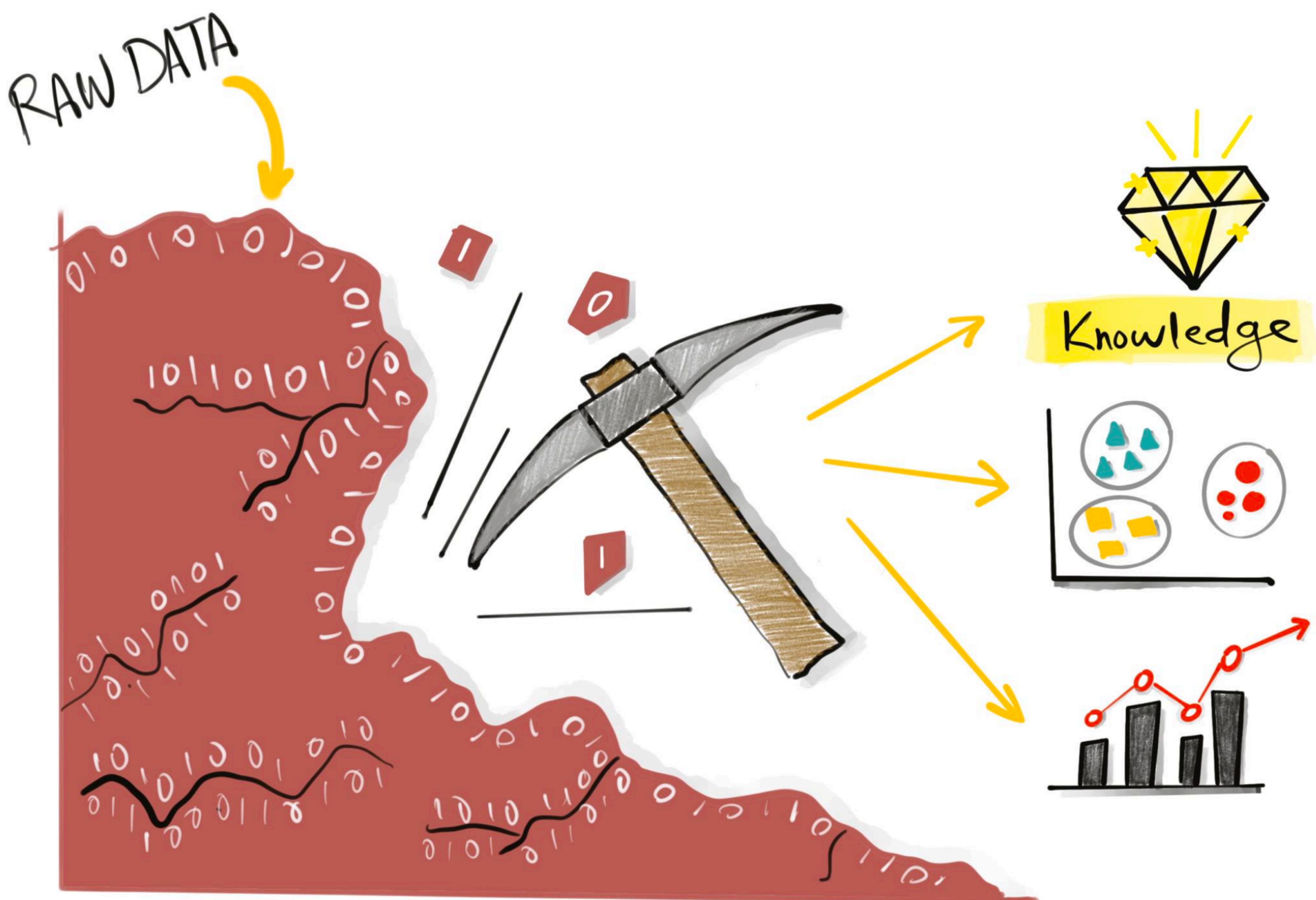
Blueprint



# Data Modeling



# Data Mining



# Data Visualization

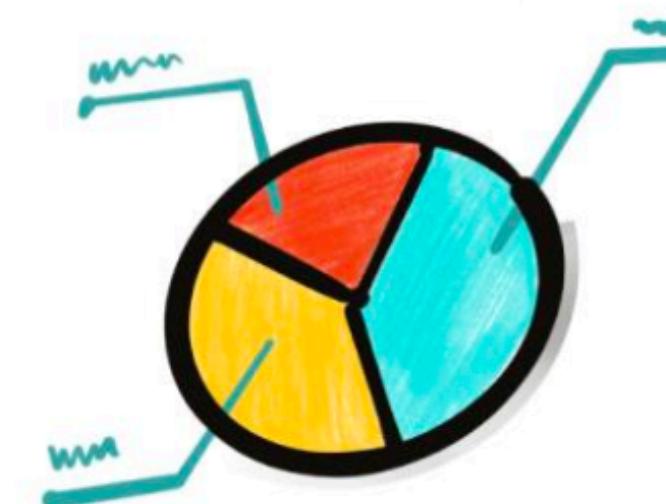
RAW DATA



101101010010101010



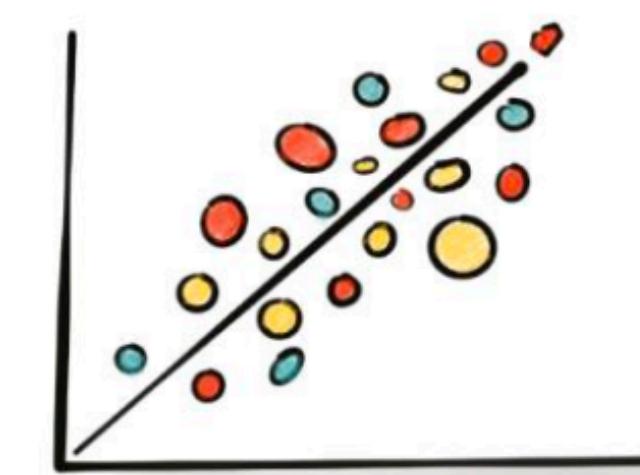
BARS



PIE

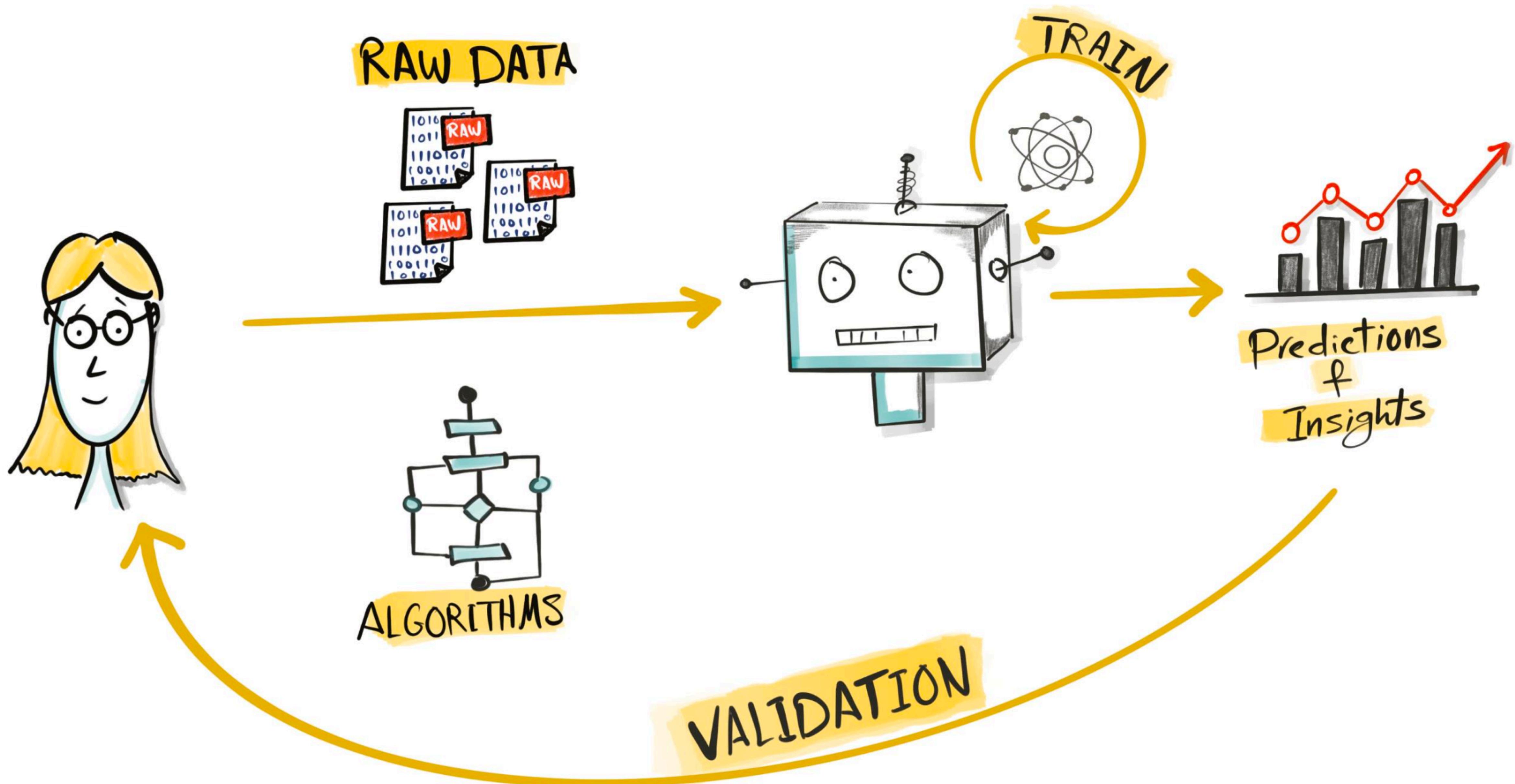


TREE



PLOTS

# Machine Learning

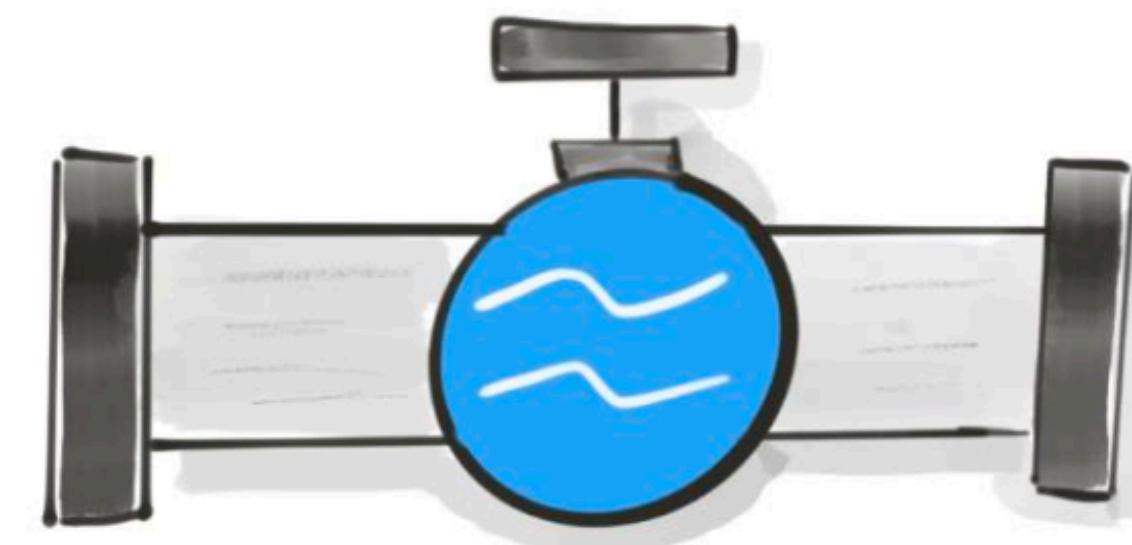


# Data Engineering

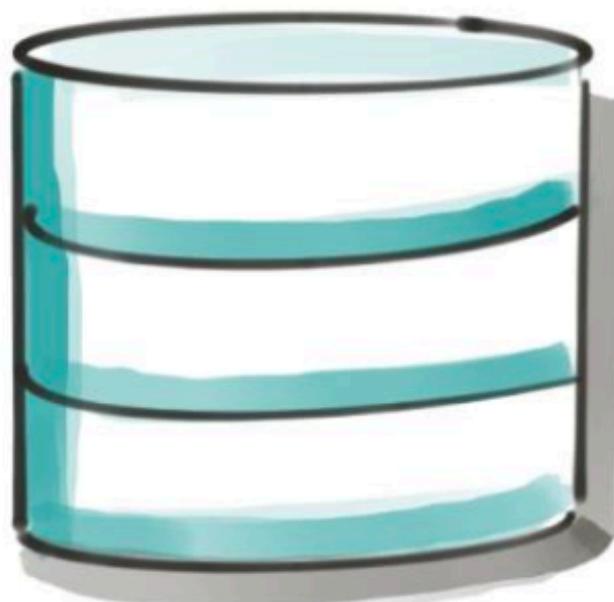
## DATA STORAGE



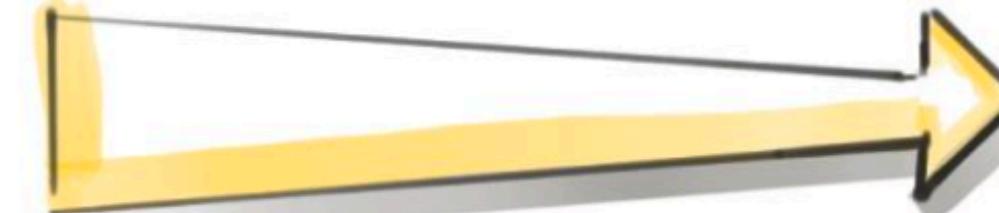
## DATA PIPELINE



SOURCE



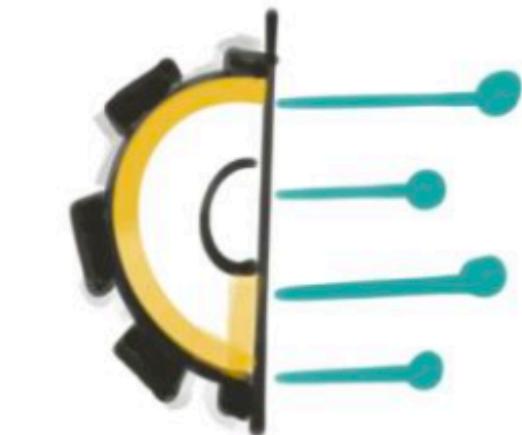
EXTRACT



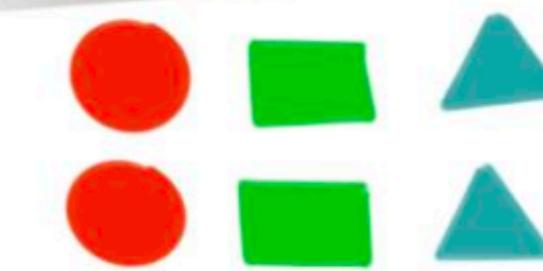
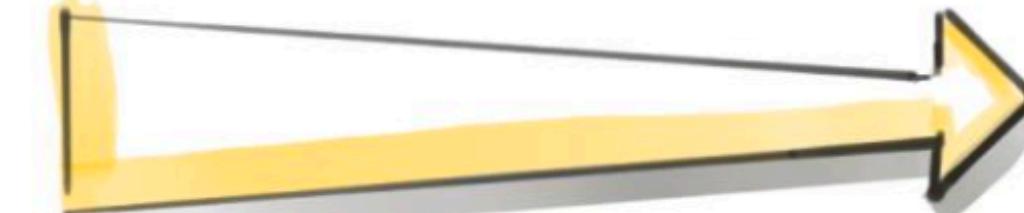
10110101011010  
1101010110110

ETL

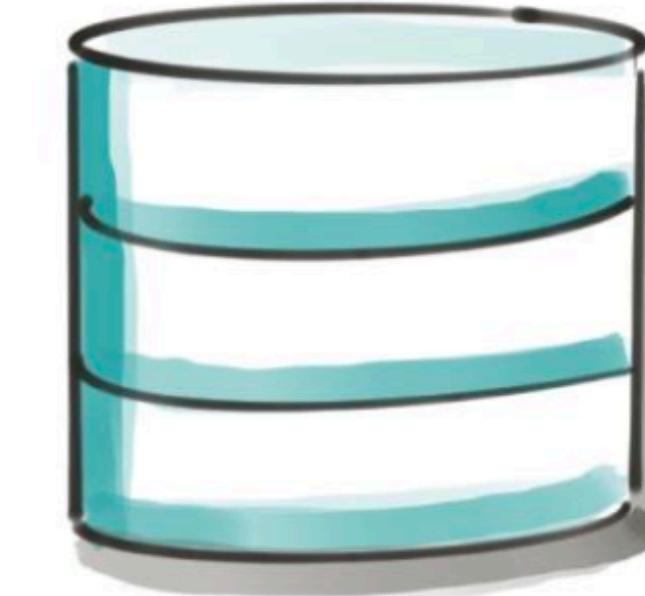
TRANSFORM



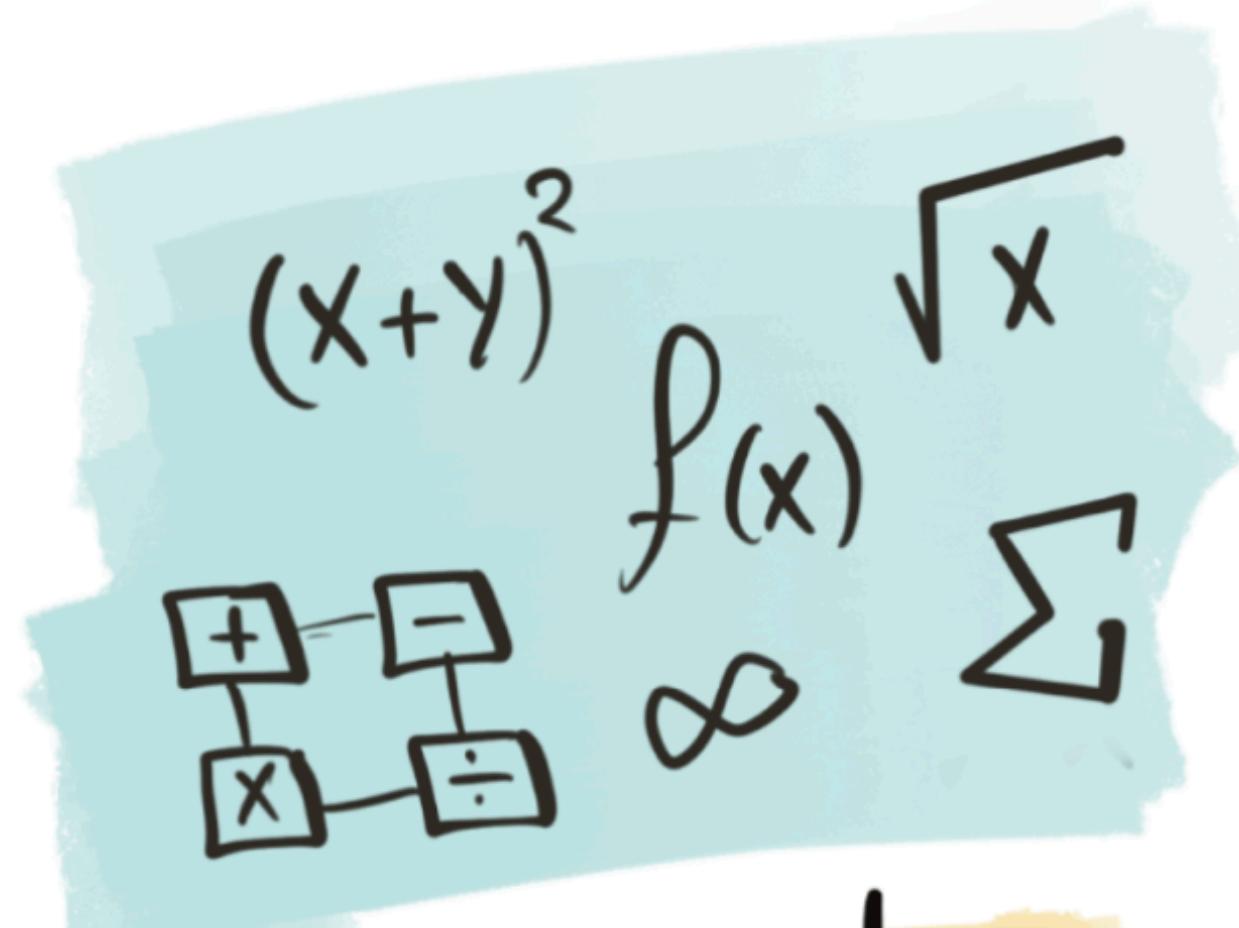
LOAD



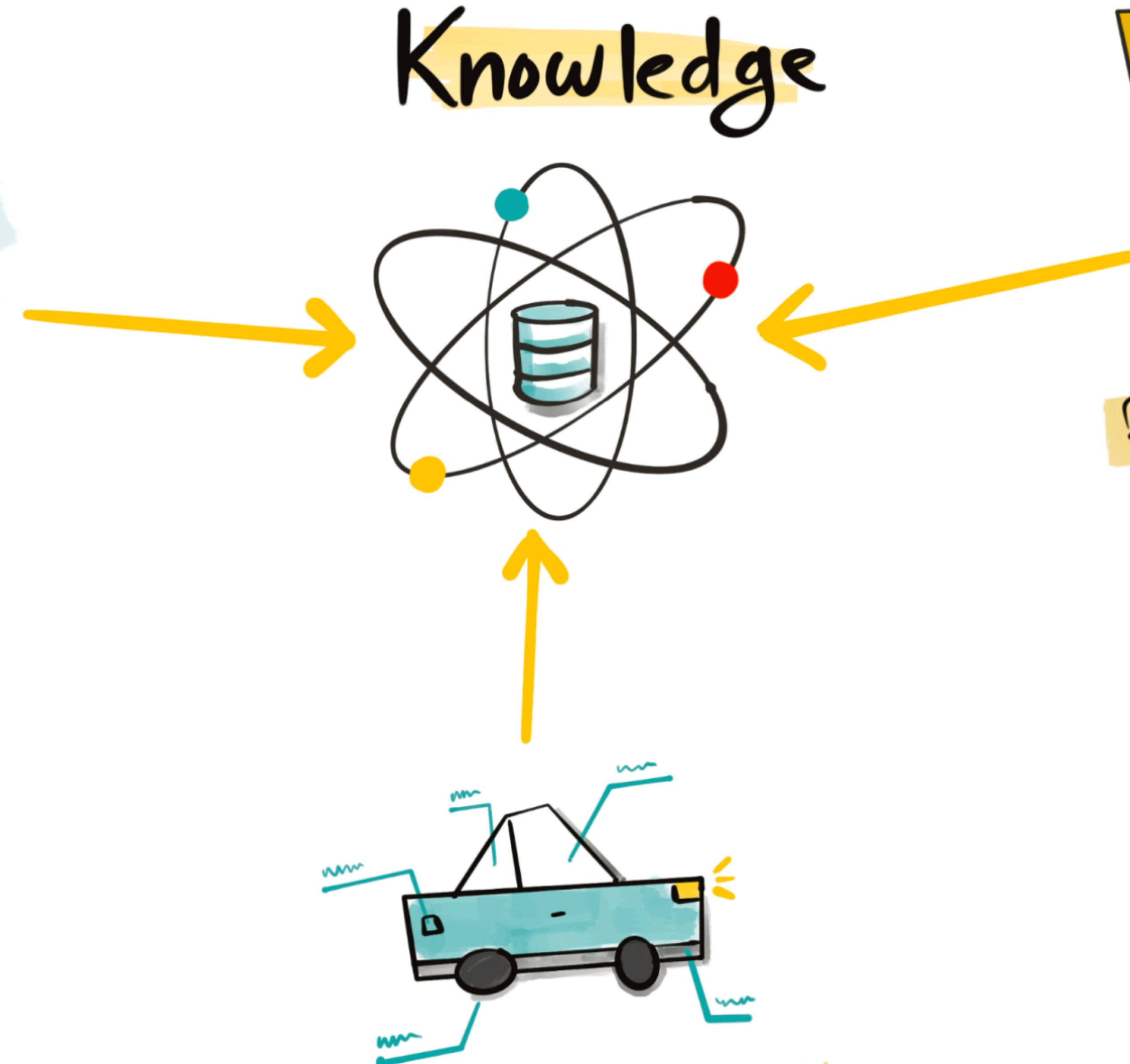
TARGET



Data Science



Mathematics



Domain Knowledge



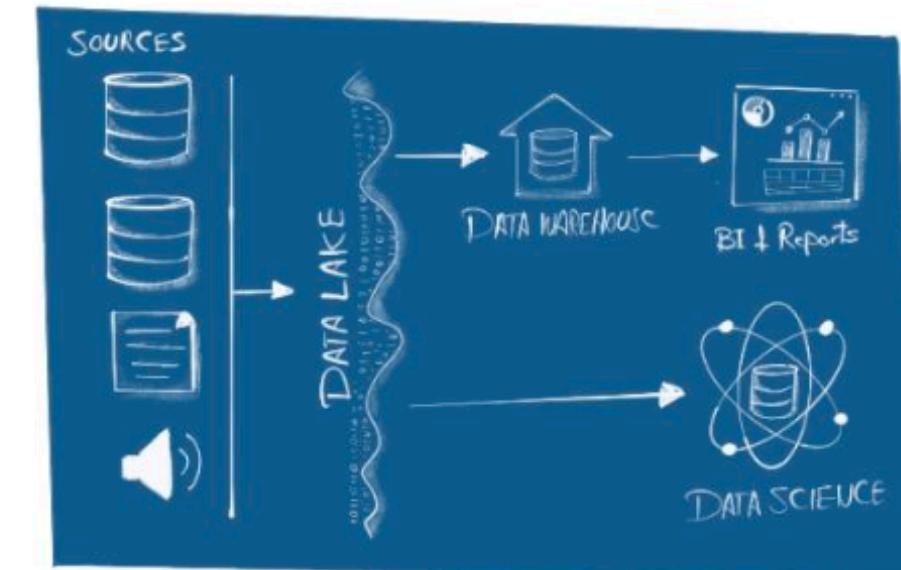
Programming  
Languages

# Data Buzzwords

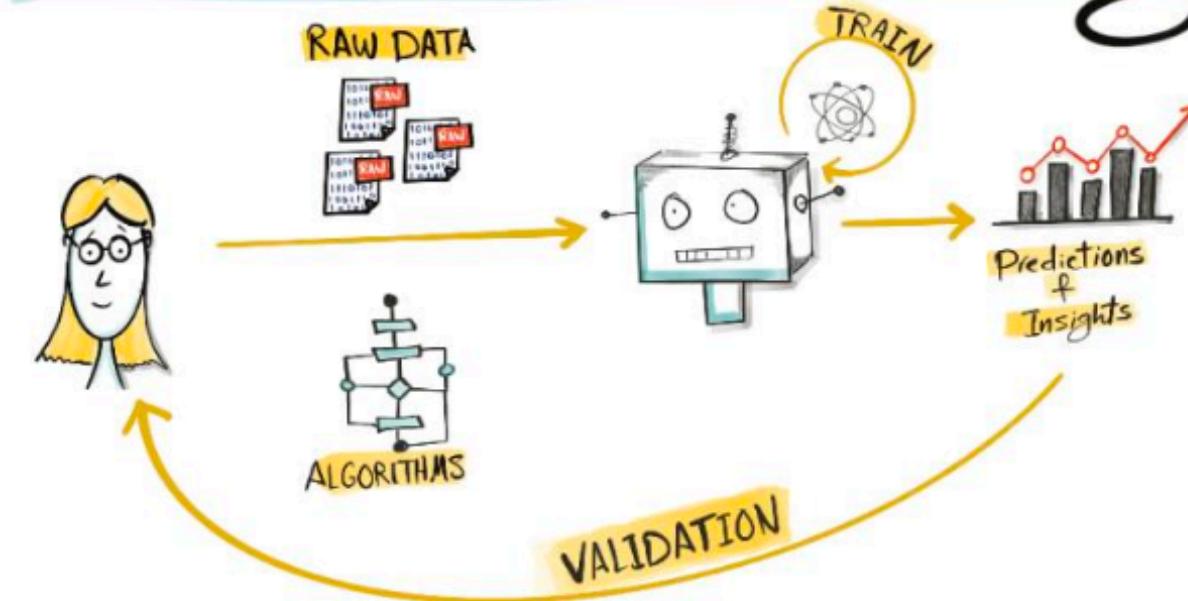
## DATA Mining



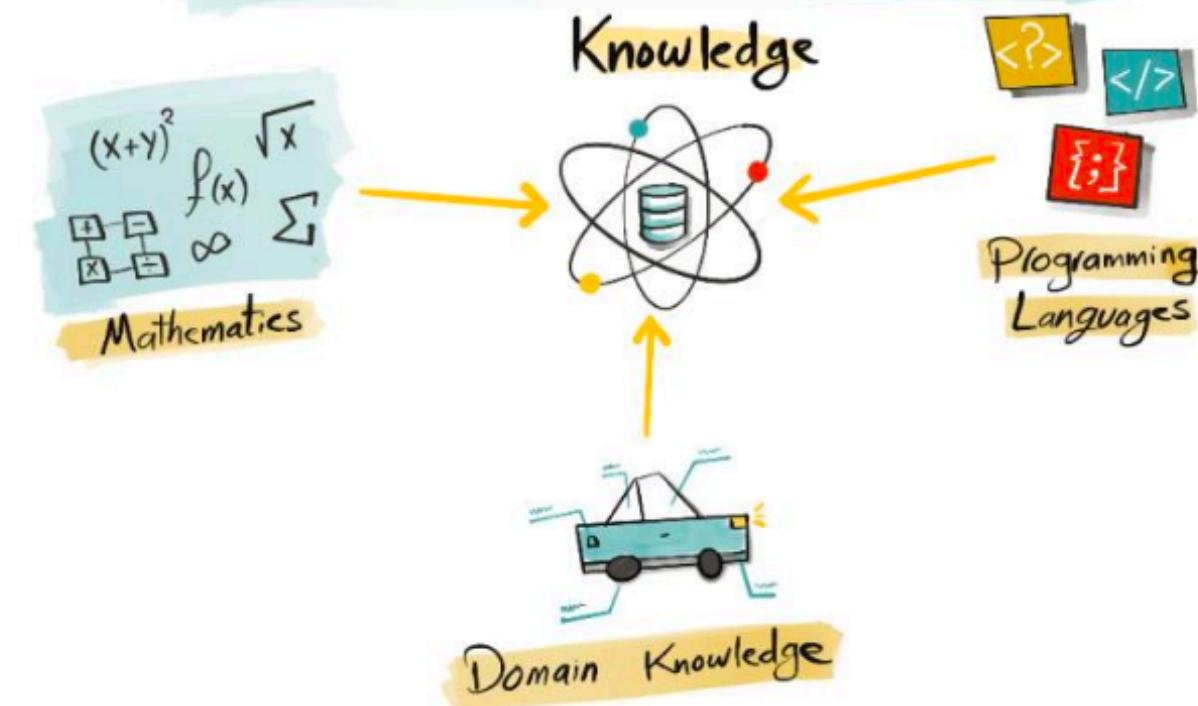
## DATA Architecture



## Machine Learning



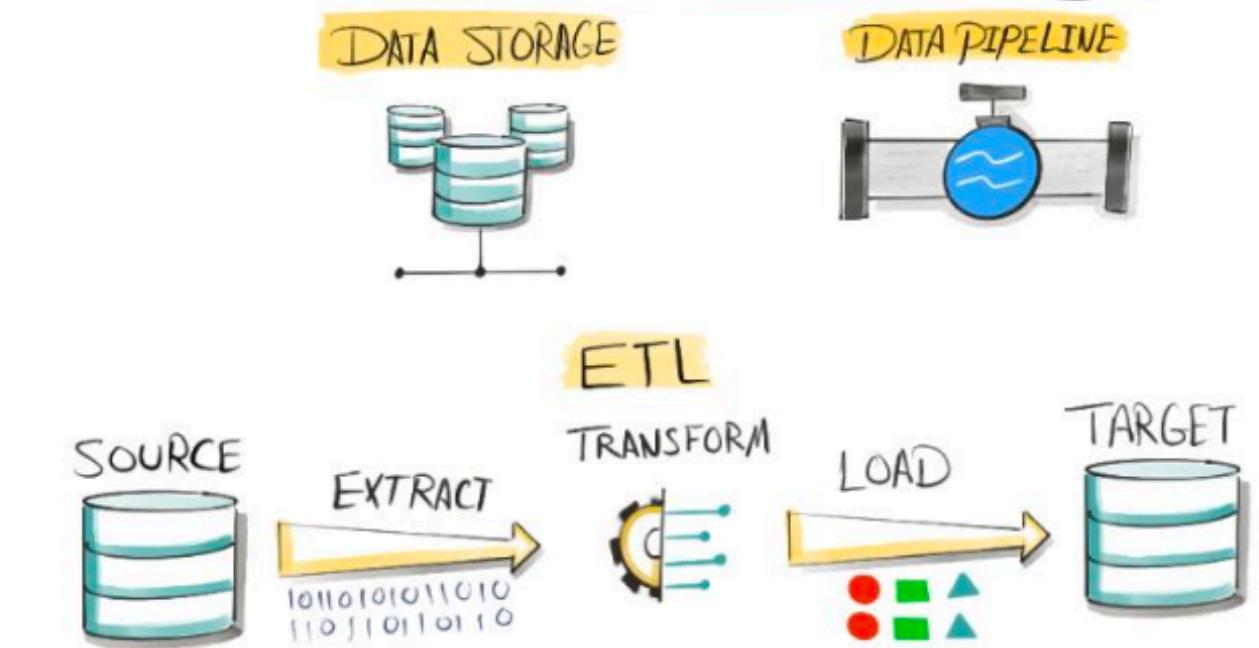
## DATA Science



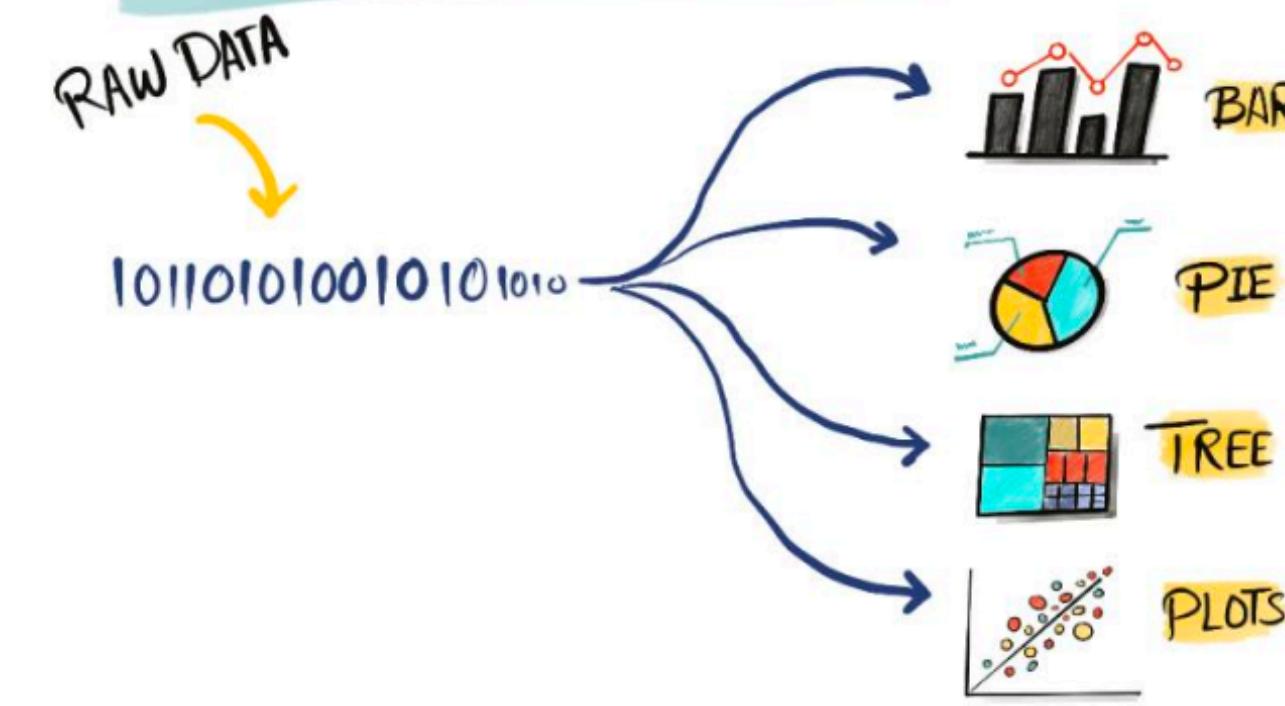
01011010001100101  
1100101

## RAW DATA

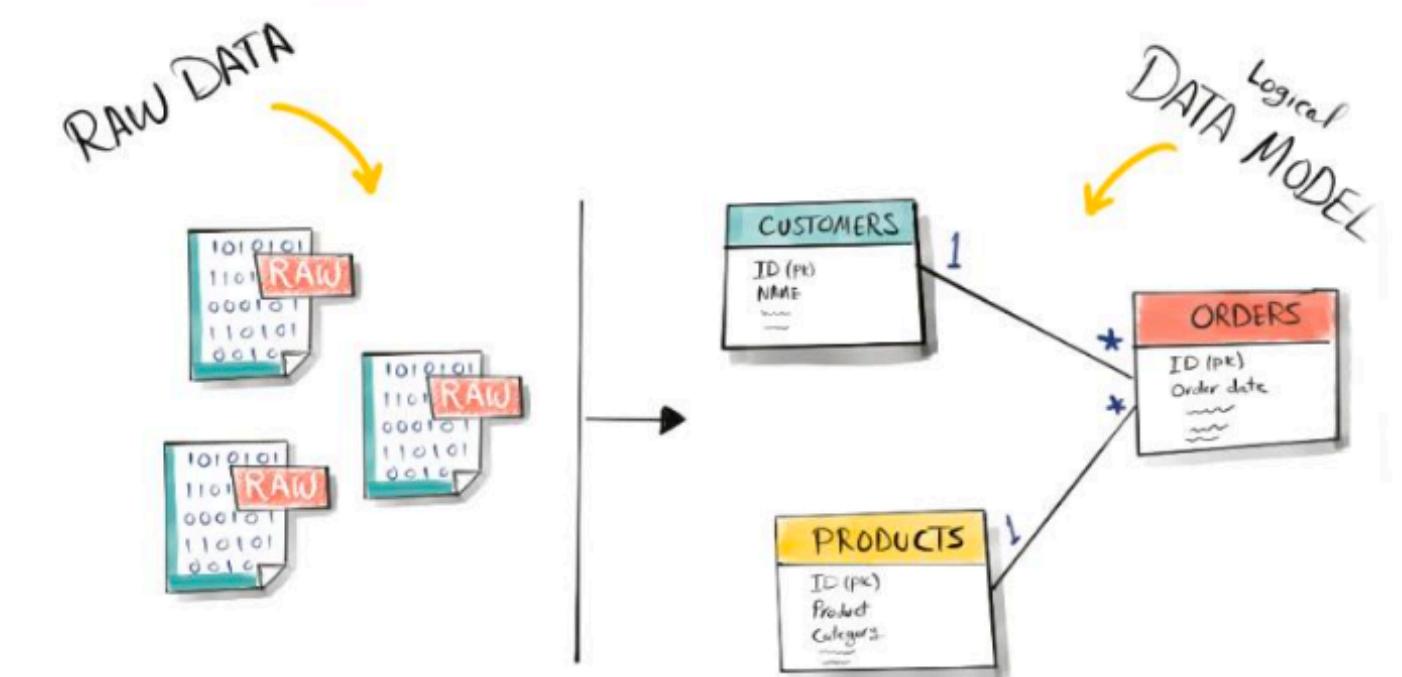
## DATA Engineering



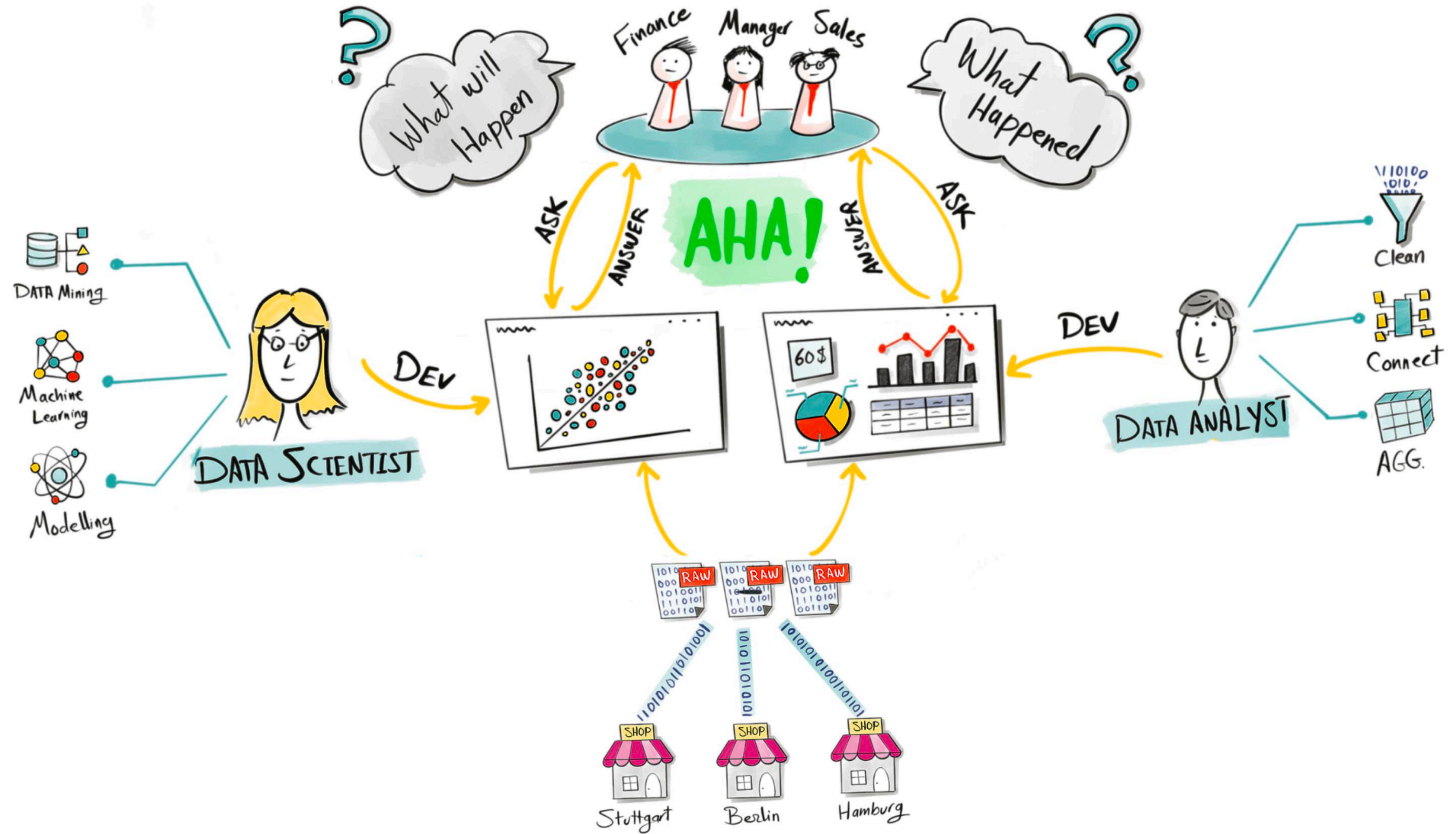
## DATA Visualization



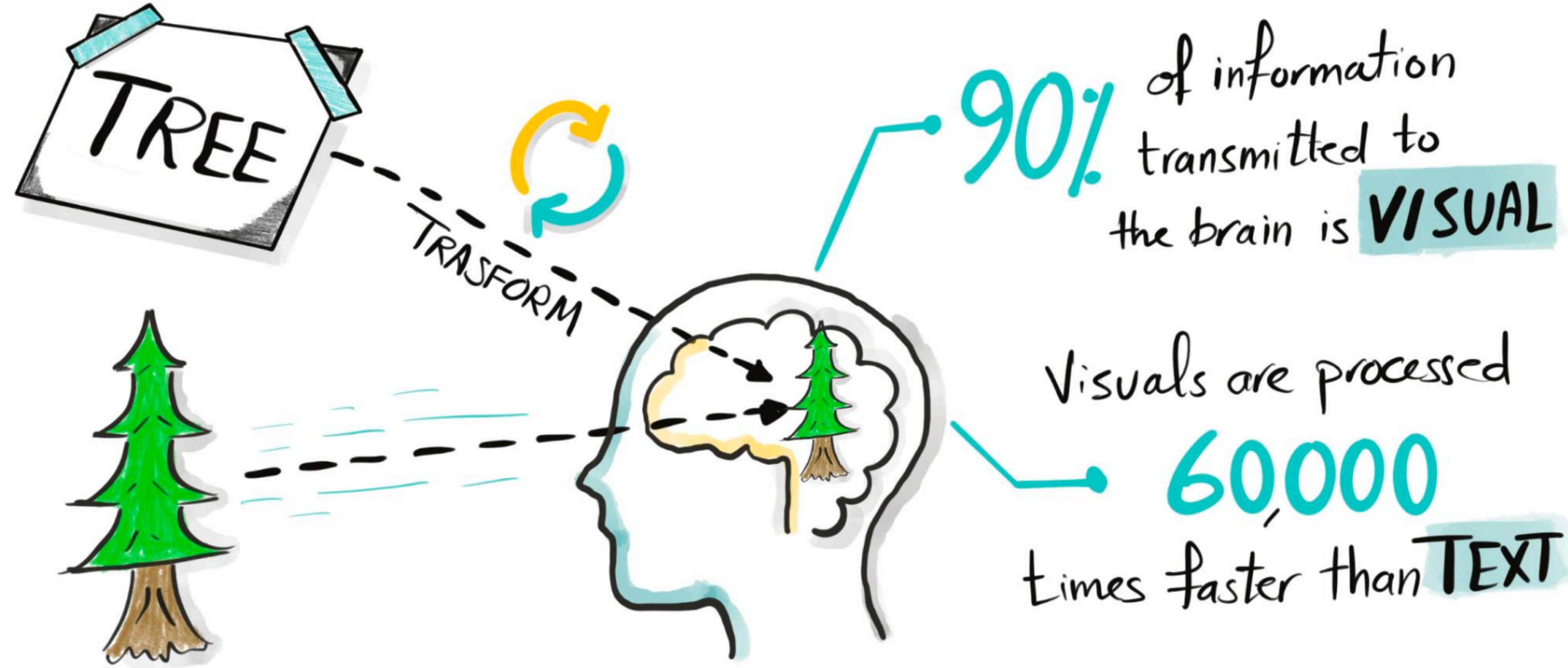
## DATA Modeling



# Business Intelligence (BI)



Power of Visual

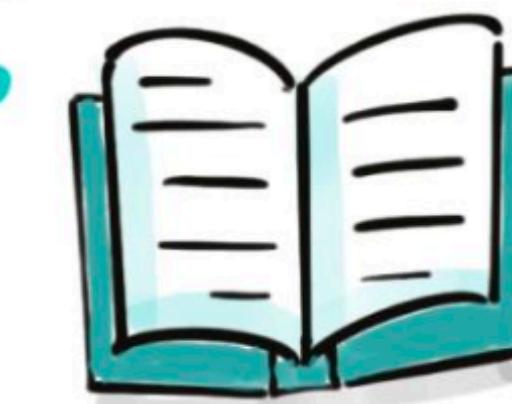


We remember ...

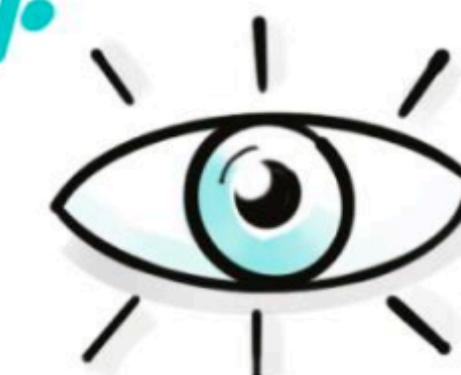


10%

20%



80%



# Benefits of Data Viz

Increase Engagement

Discover Unknown Facts

Faster & Smarter Decisions

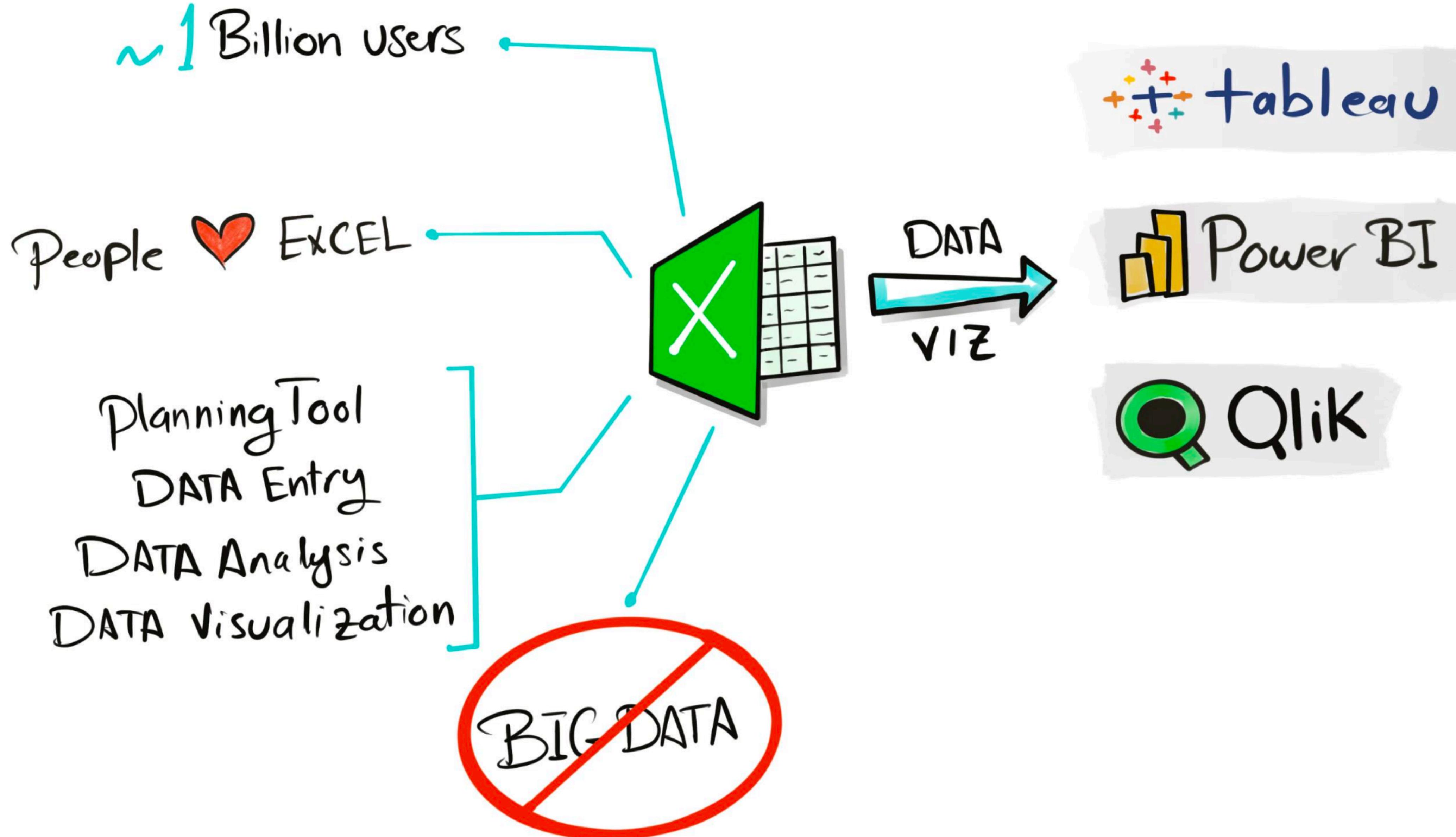
Telling a Story

Seeing Big Picture

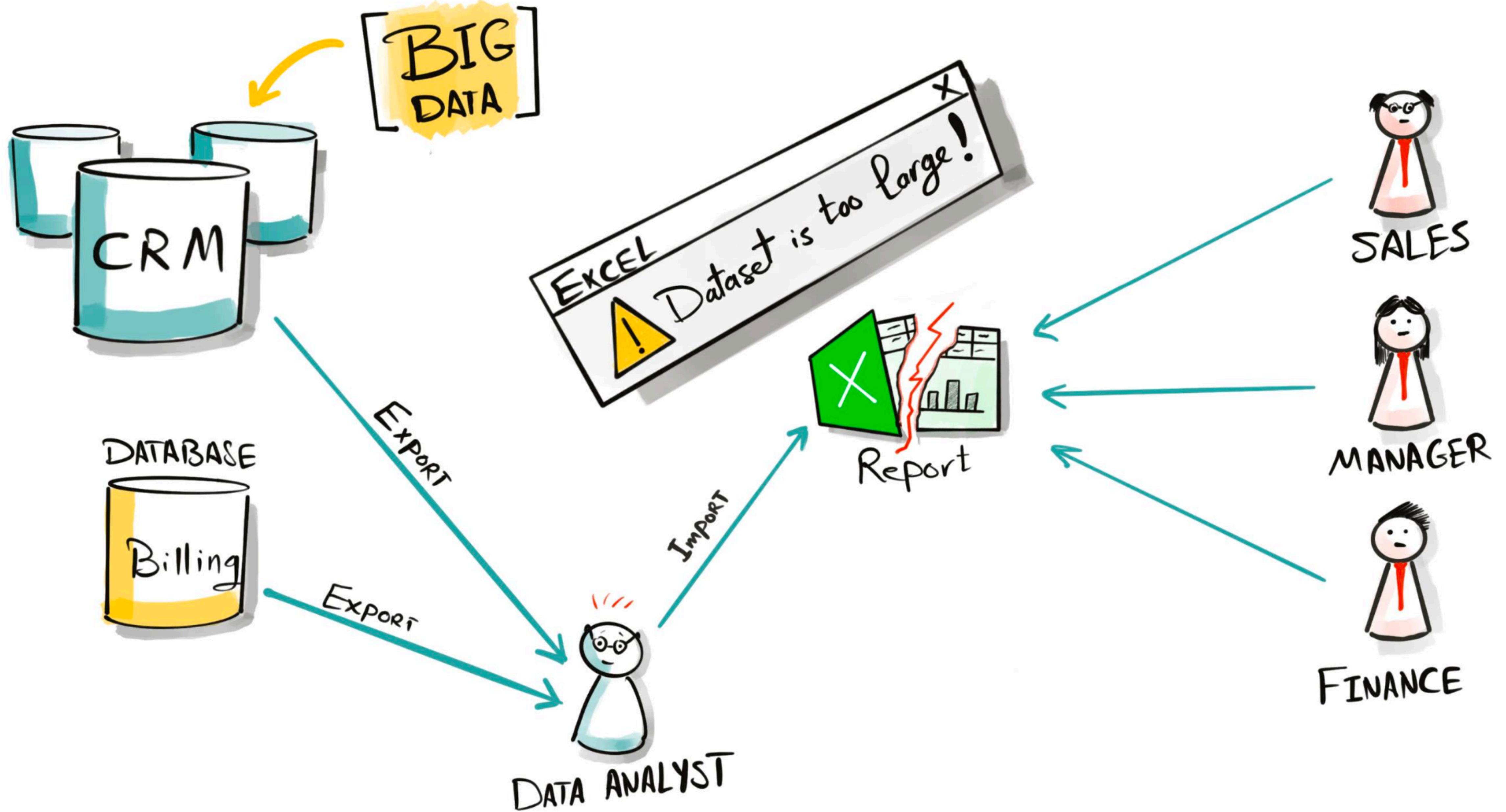
Tracking Trends

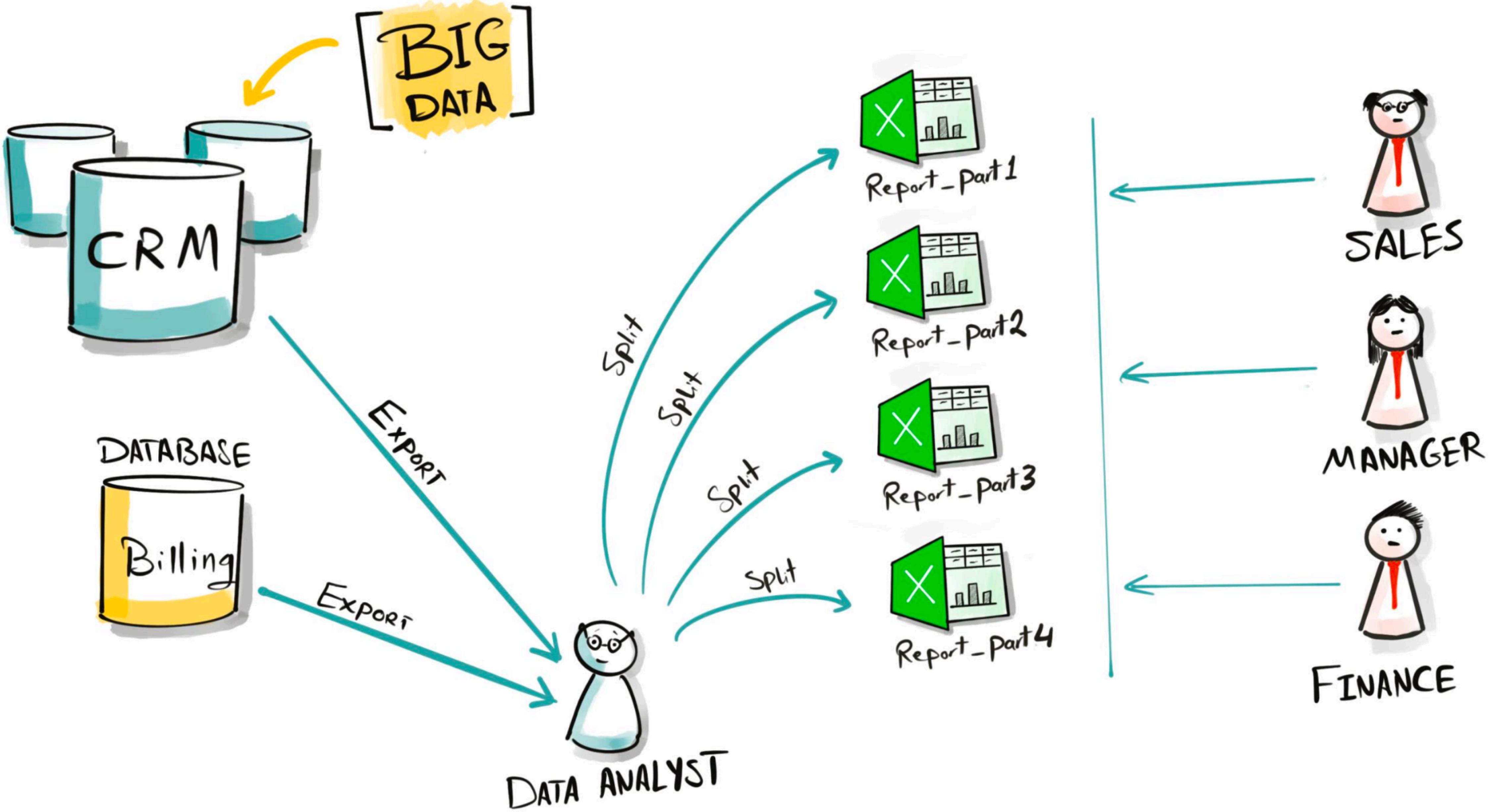


BI with Excel

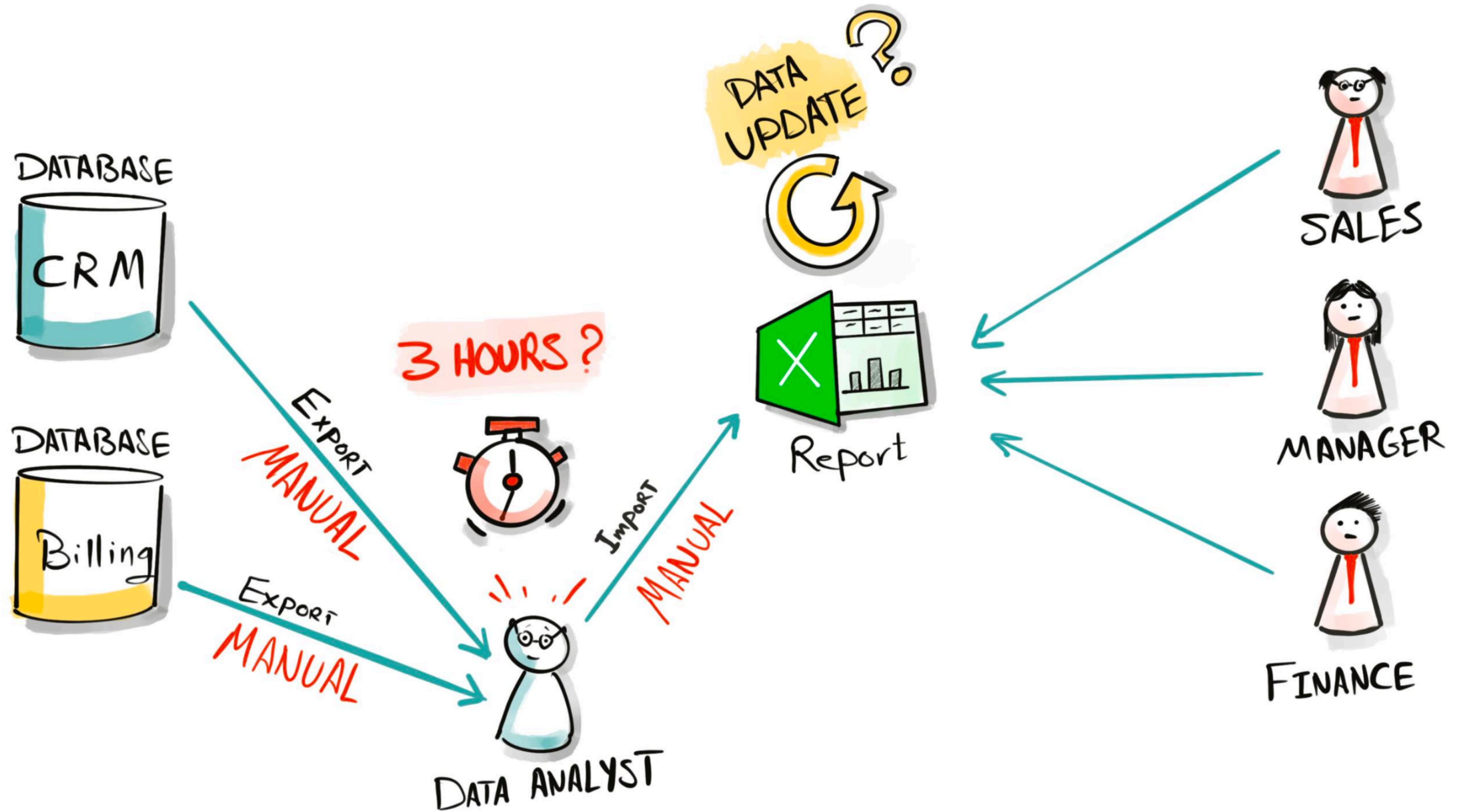


BI with Excel | Big Data



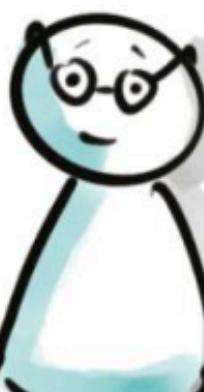


# BI with Excel | Data Load





DATA ANALYST



DASHBOARD

DEVELOP

EXTRACT  
AUTOMATIC

EXTRACT  
AUTOMATIC



SALES

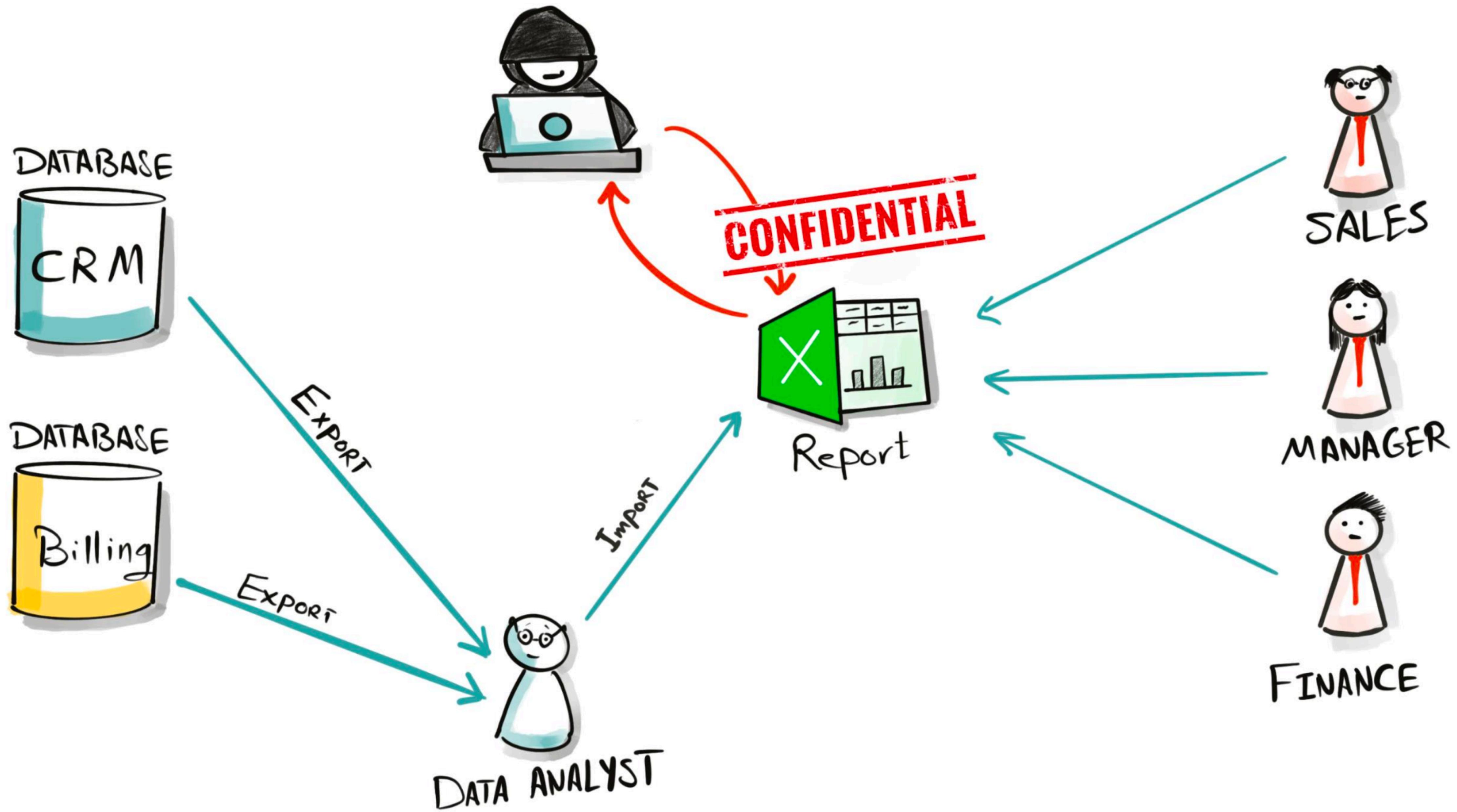


MANAGER



FINANCE

BI with Excel | Security



DATABASE



DATABASE



EXTRACT

EXTRACT

DEVELOP

DATA ANALYST



SALES



MANAGER



# Row-Level-Security (RLS)

UserName	Country	Sales
Baraa	USA	9000
Baraa	Germany	8000
Maria	USA	6500
John	France	10000
John	USA	7500
Peter	Spain	5900

All Rows in Sales

UserName	Country	Sales
Baraa	USA	9000
Baraa	Germany	8000

Baraa's View

UserName	Country	Sales
John	France	10000
John	USA	7500

John's View

BI with Excel | Row-level-Security

# BI with EXCEL



**CONFIDENTIAL**

DATABASE



DATABASE



[NEED-TO-KNOW]

DATA ANALYST

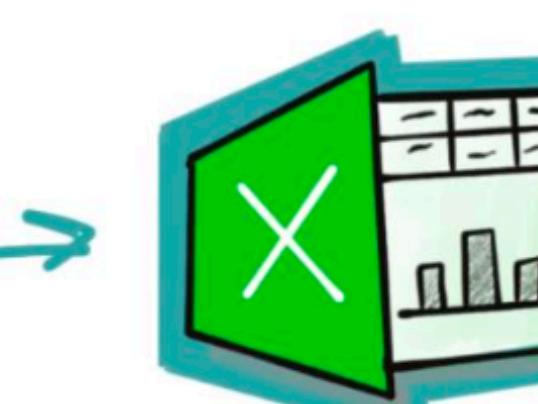
EXPORT

EXPORT

Import

Import

Import



Sales - Report

Manager - Report

Finance - Report

SALES

MANAGER

FINANCE

**CONFIDENTIAL**

DATABASE



DATABASE



[NEED-TO-KNOW]

EXTRACT

EXTRACT

DASHBOARD

DEVELOP

DATA ANALYST



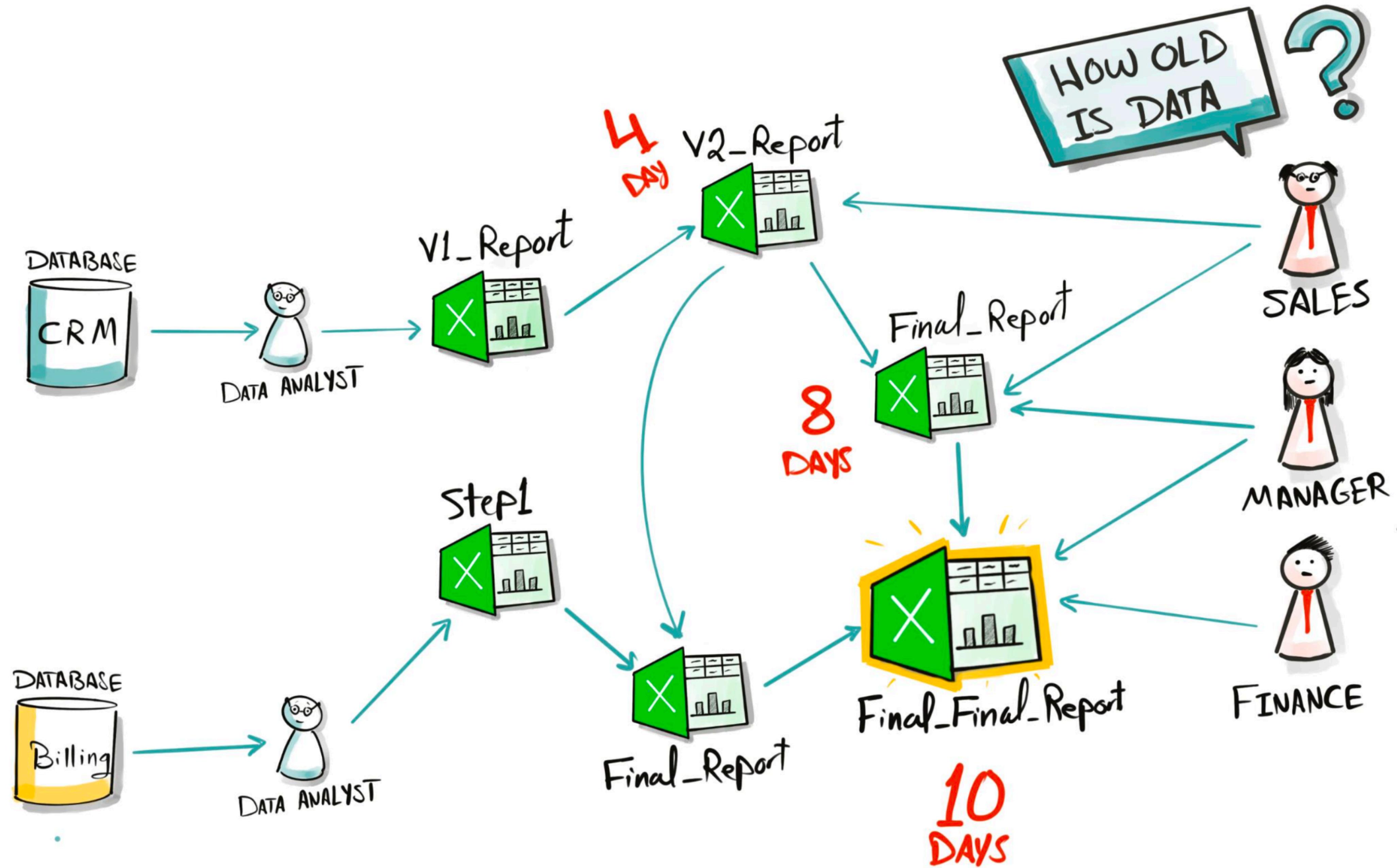
SALES



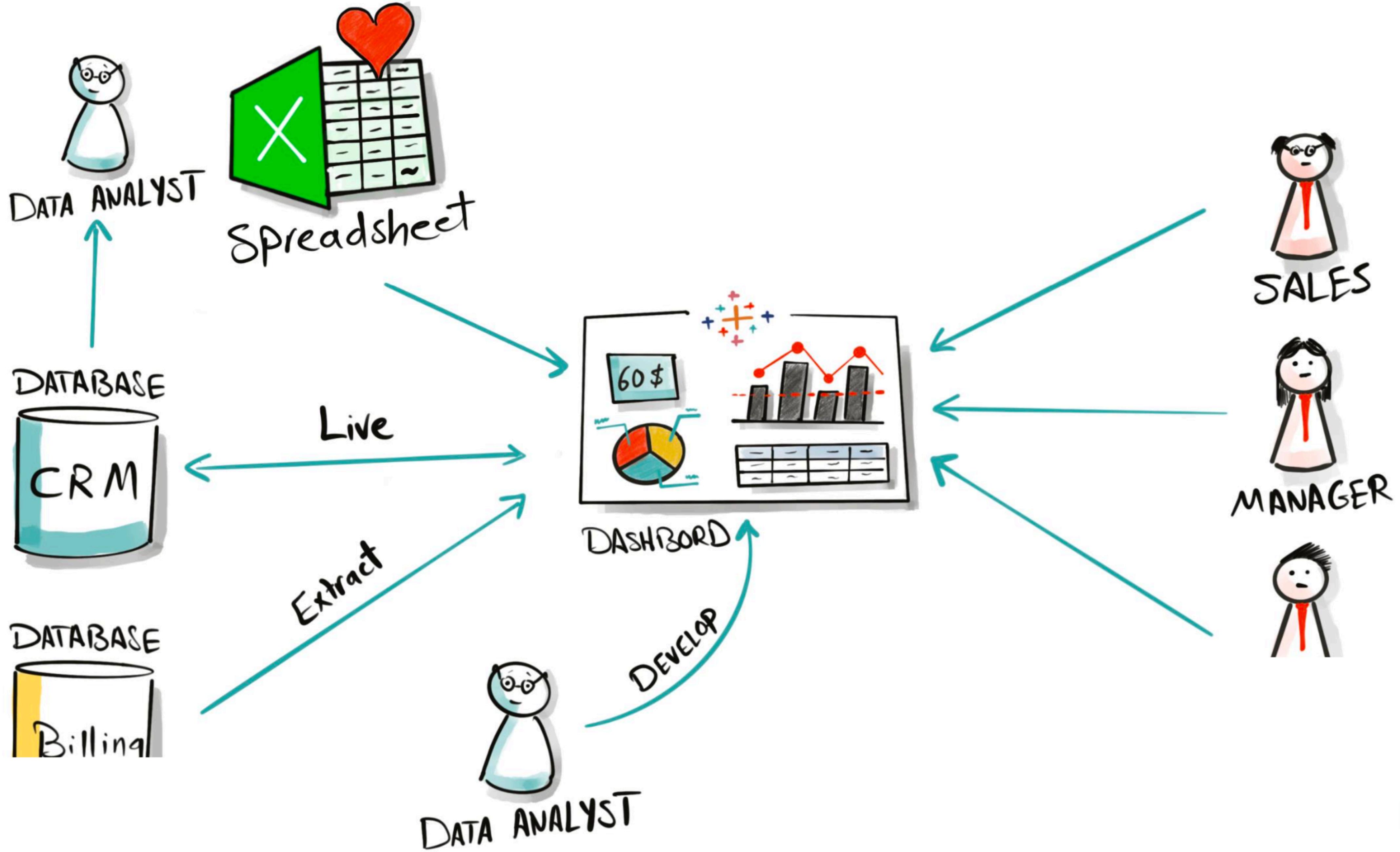
MANAGER



FINANCE



# BI with Tableau & Excel



# Benefits of Tableau vs Excel



Automation ✓



Security ✓



BIG DATA ✓



Interactive ✓

