Concepts in Data Architecture

paradigms that make BI/DS work at an enterprise level

Presented by: Abdulrahman AlQallaf

Presentation date: 2021-Apr-24

Agenda

- Introduction
- One Developer
- Data Warehouse
- Data Lake
- Data Mesh
- Final Remark
- Questions?
- References

Introduction – about me

About me:

- **Education:**
  - CS → MBA → MCS (in progress)
- **Career:**
  - 7 years in the banking sector (NBK → Boubyan)
  - Work mainly with data, in simple terms trying to figure out how to use it to move things forward

Disclaimer:

- I'm representing myself today and not where I work.
- All the information that I am presenting here today is available in the public domain.

Introduction – motivation and setting expectations

Motivation:

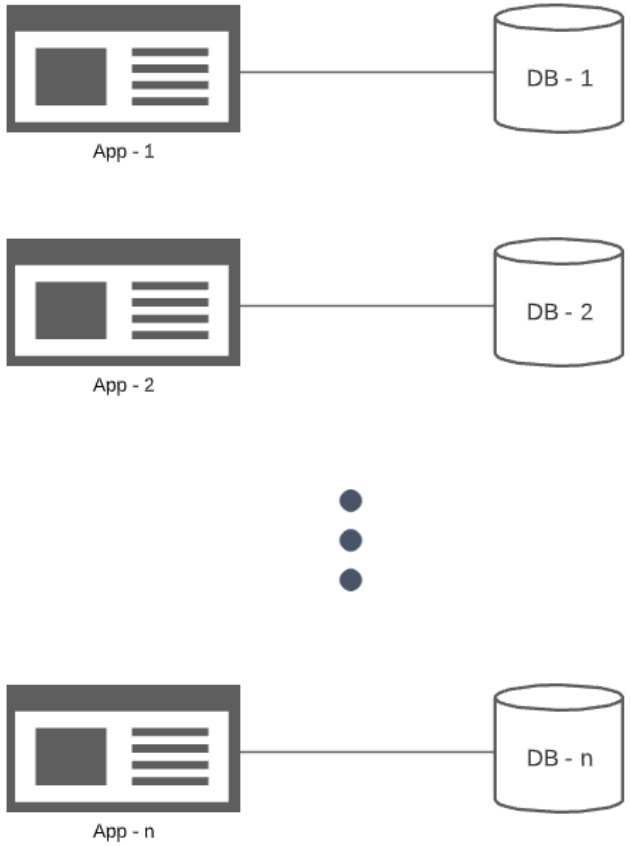
- I've seen a lot of talks recently about DS / DL / AI.
- But not really about what makes them work on an enterprise level (i.e., the stuff behind the scenes).
- Also, even if you want to work as a DS for a company, around 50% of your job will be on data preparation (50% is an optimist percentage, in reality expect more than 80%, easily).
- Therefore, it helps to know the data infrastructure that you will be working with for you to become effective in your job.
- Basically, this talk is what I wish someone would've given me before starting my career in data.

Setting expectations:

- This talk will be on high level concepts (i.e., no terminal, no code).
- The focus is not even on specific technologies, I will try to speak in simple terms. The point is getting the novice up to speed as quickly as possible (and know which terms to search for if more information is needed).

One Developer

Data Warehouse – inception



Data Warehouse – definition

What is a **Data Warehouse** (DW)?

- a repository of current and historical data …
- … of potential interest to employees throughout the organization …
- … with the main purpose of supporting decision making.

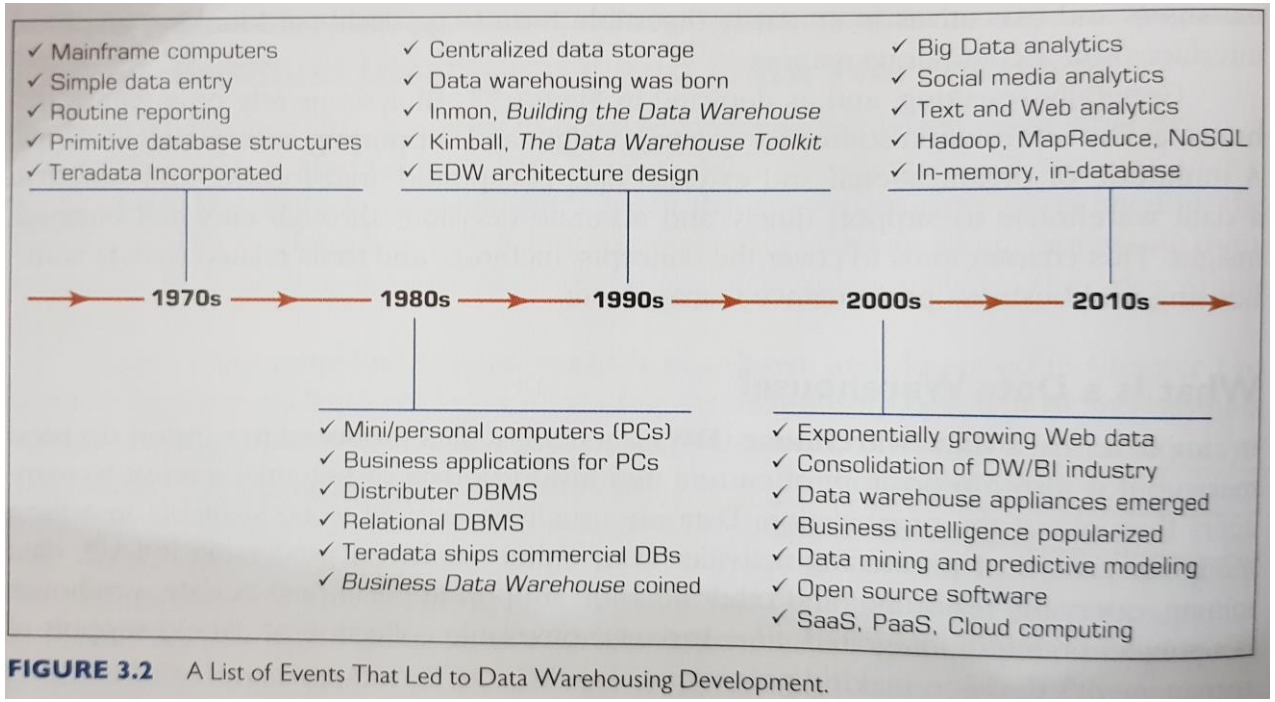


FIGURE 3.2 A List of Events That Led to Data Warehousing Development.

Data Warehouse – support for BI

Modern BI Mantra: Employees need the…

- right information …
- … at the right time
- … and in the right place.

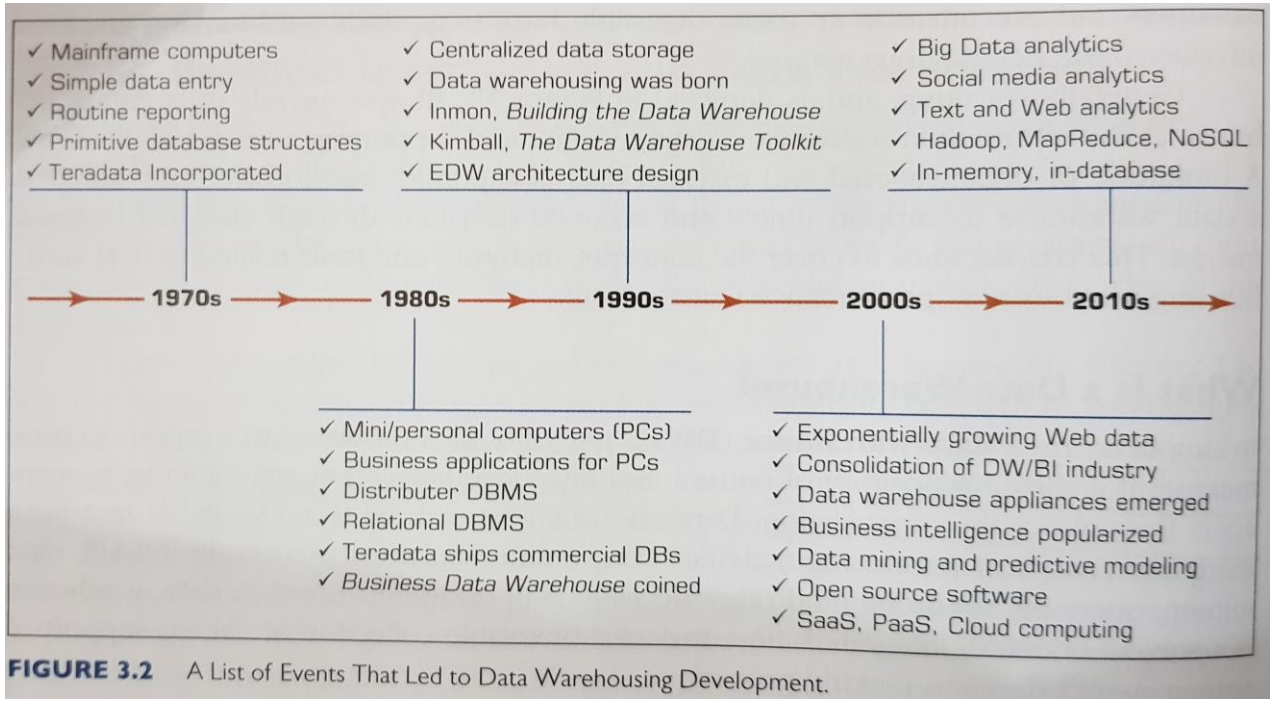


FIGURE 3.2 A List of Events That Led to Data Warehousing Development.

Data Warehouse – benefits

- data synergy.
- data consistency.
- faster reporting/analytics/DS.
- data timeliness.
- relieve processing from production systems.

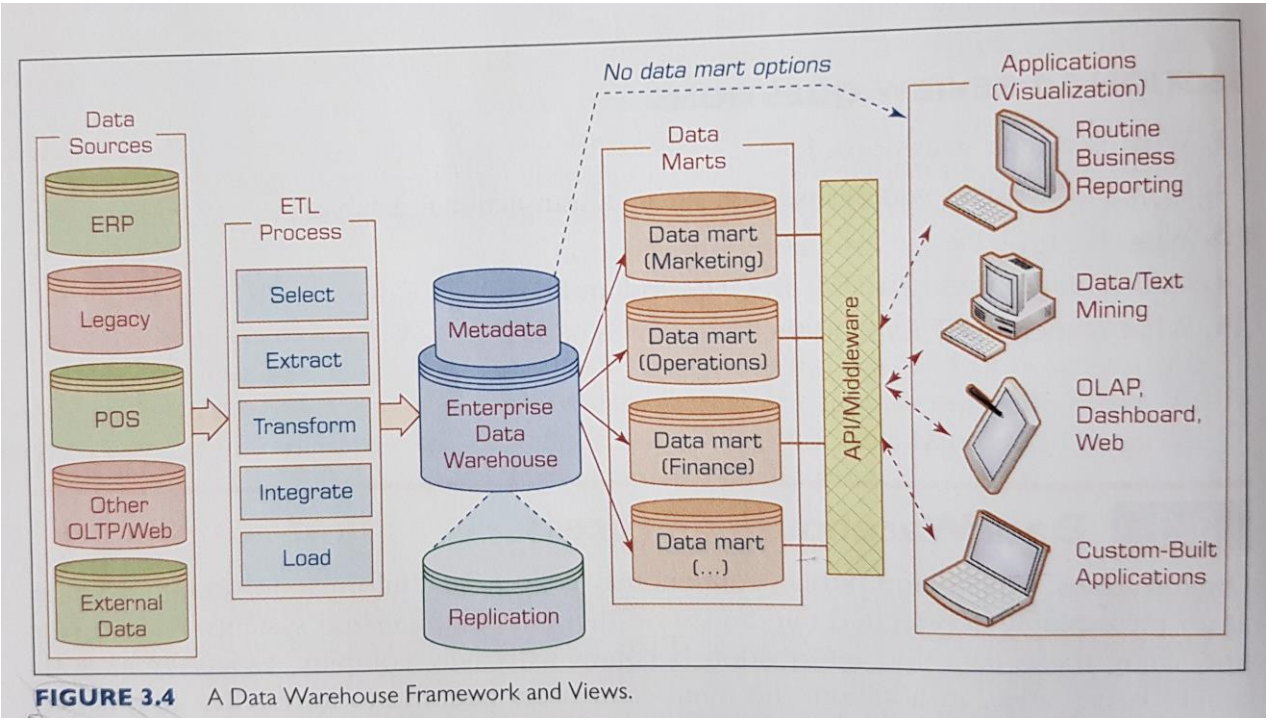
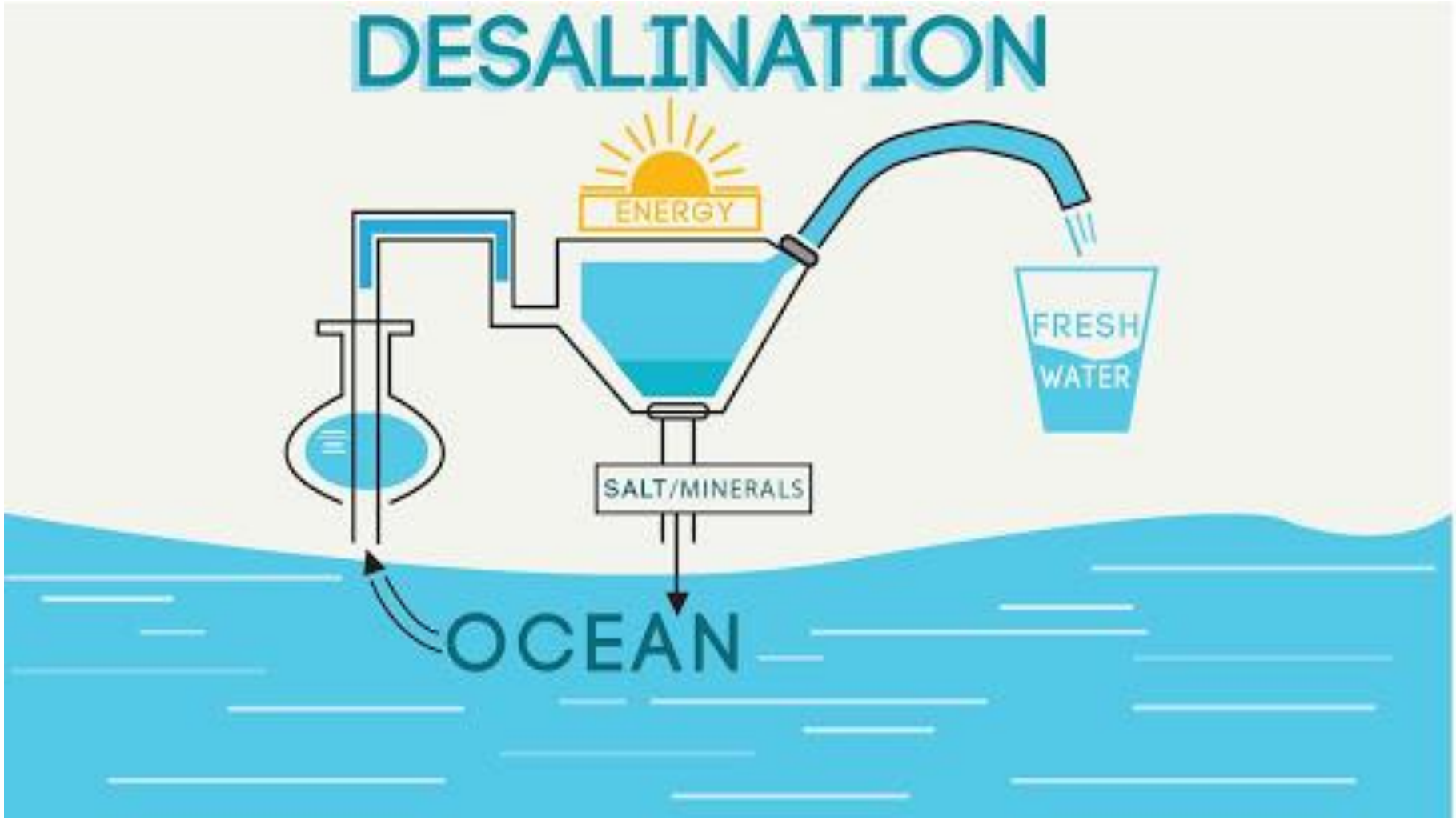


FIGURE 3.4 A Data Warehouse Framework and Views.

Data Lake – simplified explanation



Data Lake – a hybrid approach

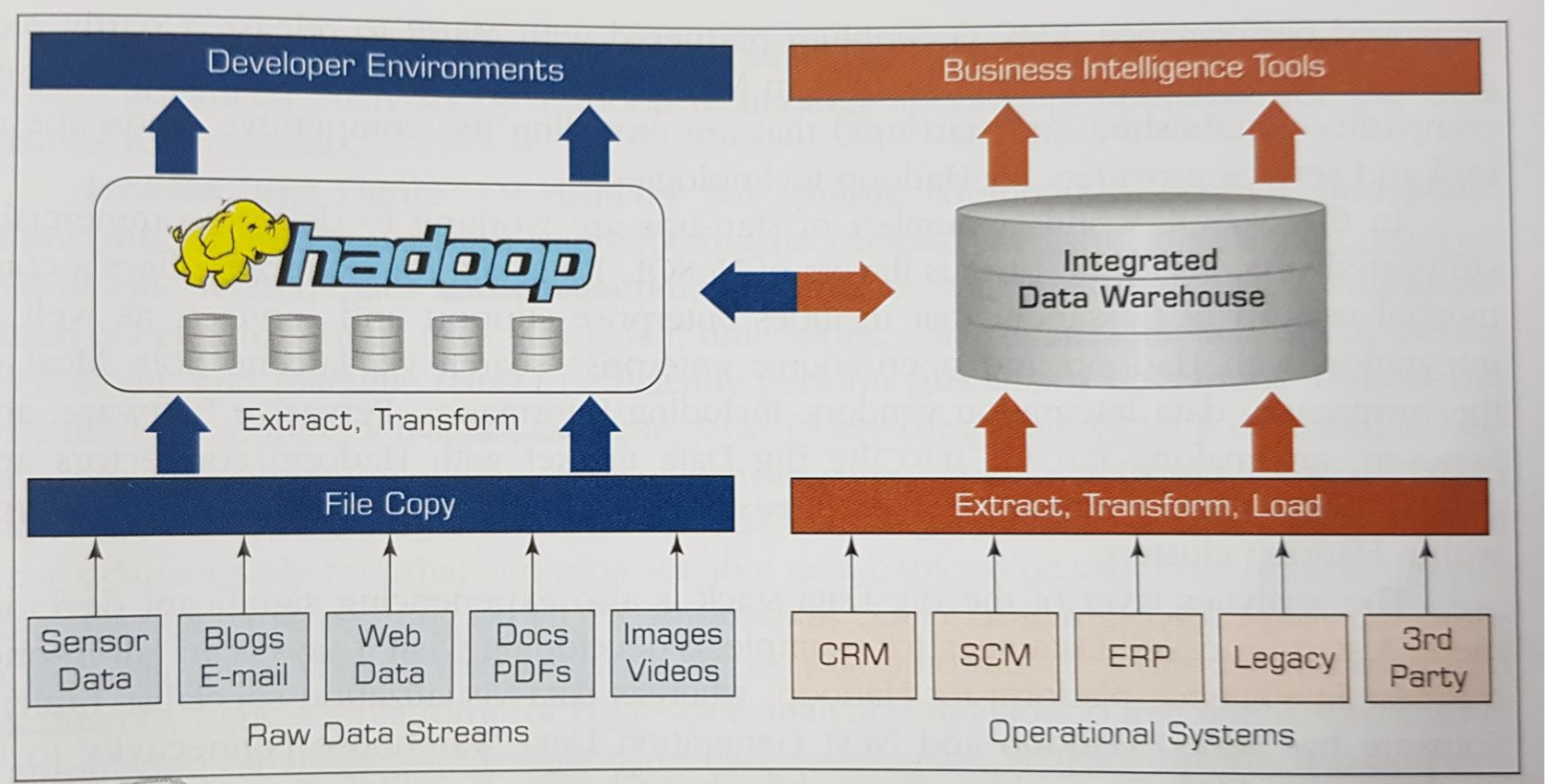


FIGURE 7.8 Coexistence of Hadoop and Data Warehouses. *Source:* Teradata Corp.

Data Lake – key differences

TABLE 3.6 A Simple Comparison between a Data Warehouse and a Data Lake

Dimension	Data Warehouse	Data Lake
The nature of data	Structured, processed	Any data in raw/native format
Processing	Schema-on-write (SQL)	Schema-on-read (NoSQL)
Retrieval speed	Very fast	Slow
Cost	Expensive for large data volumes	Designed for low-cost storage
Agility	Less agile, fixed configuration	Highly agile, flexible configuration
Novelty/newness	Not new/matured	Very new/maturing
Security	Well-secured	Not yet well-secured

Data Mesh – the next paradigm shift?

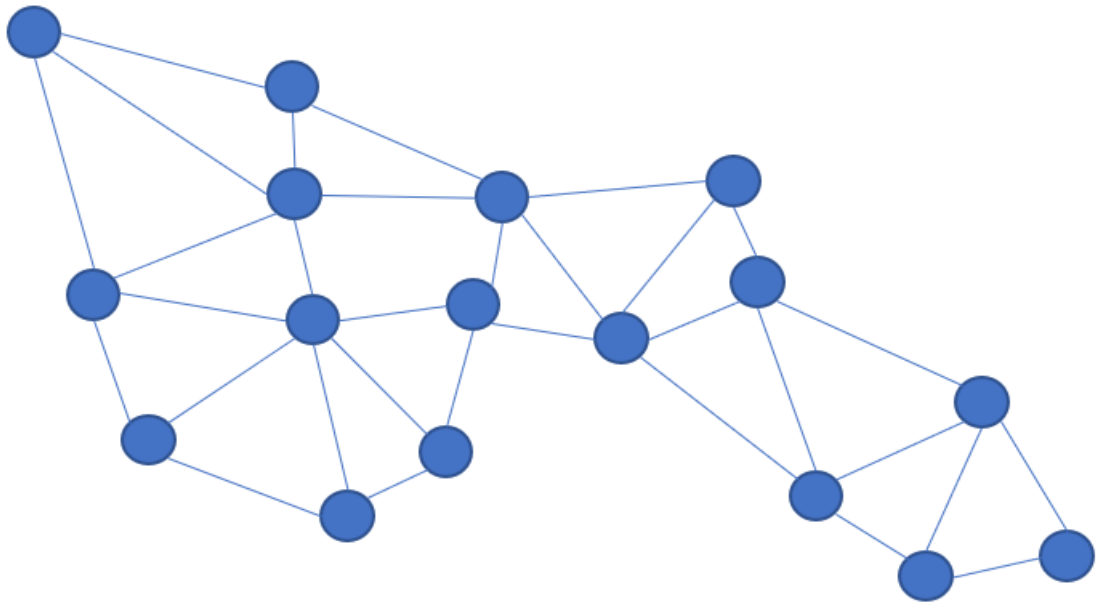
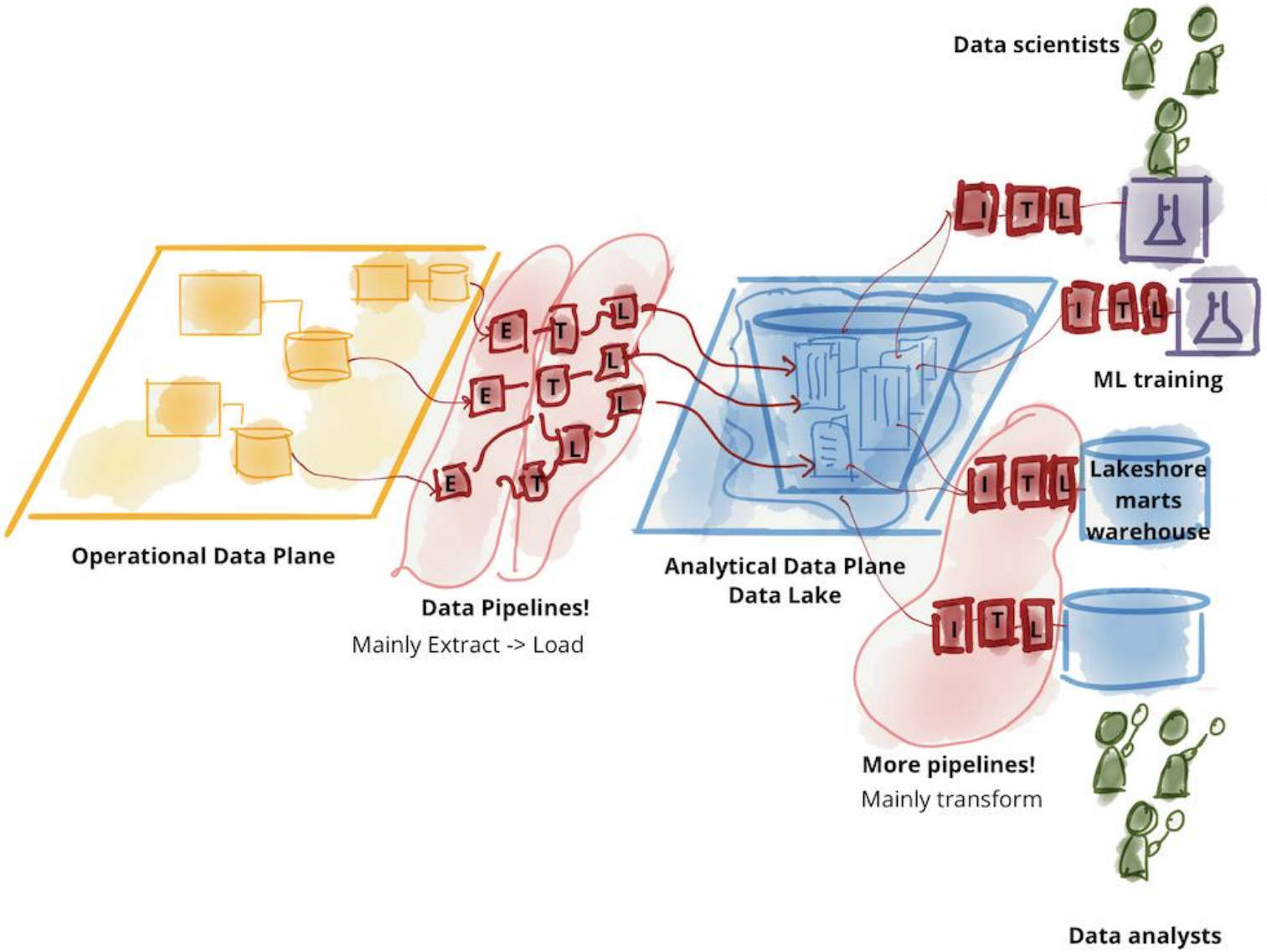
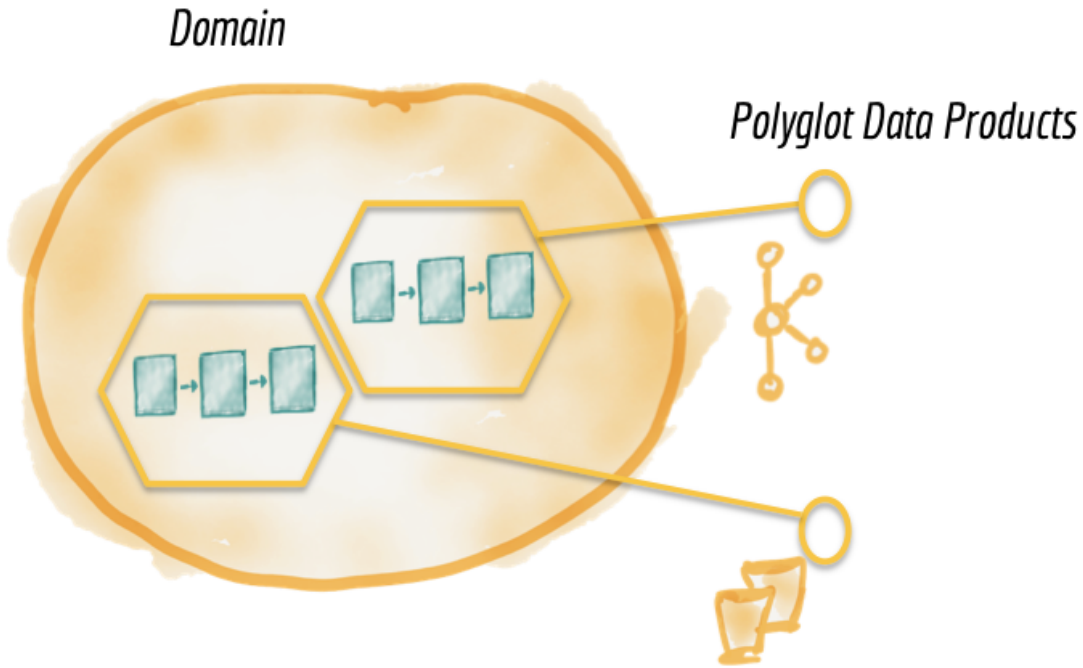


Figure 1 Mesh network

Data Mesh


Data scientists
I T L
I T L
ML training
Lakeshore marts warehouse
I T L
I T L
Data analysts
E T L
E T L
E T L
Operational Data Plane
Data Pipelines!
Mainly Extract -> Load
Analytical Data Plane
Data Lake
More pipelines!
Mainly transform

Data Mesh



DISCOVERABLE

ADDRESSABLE

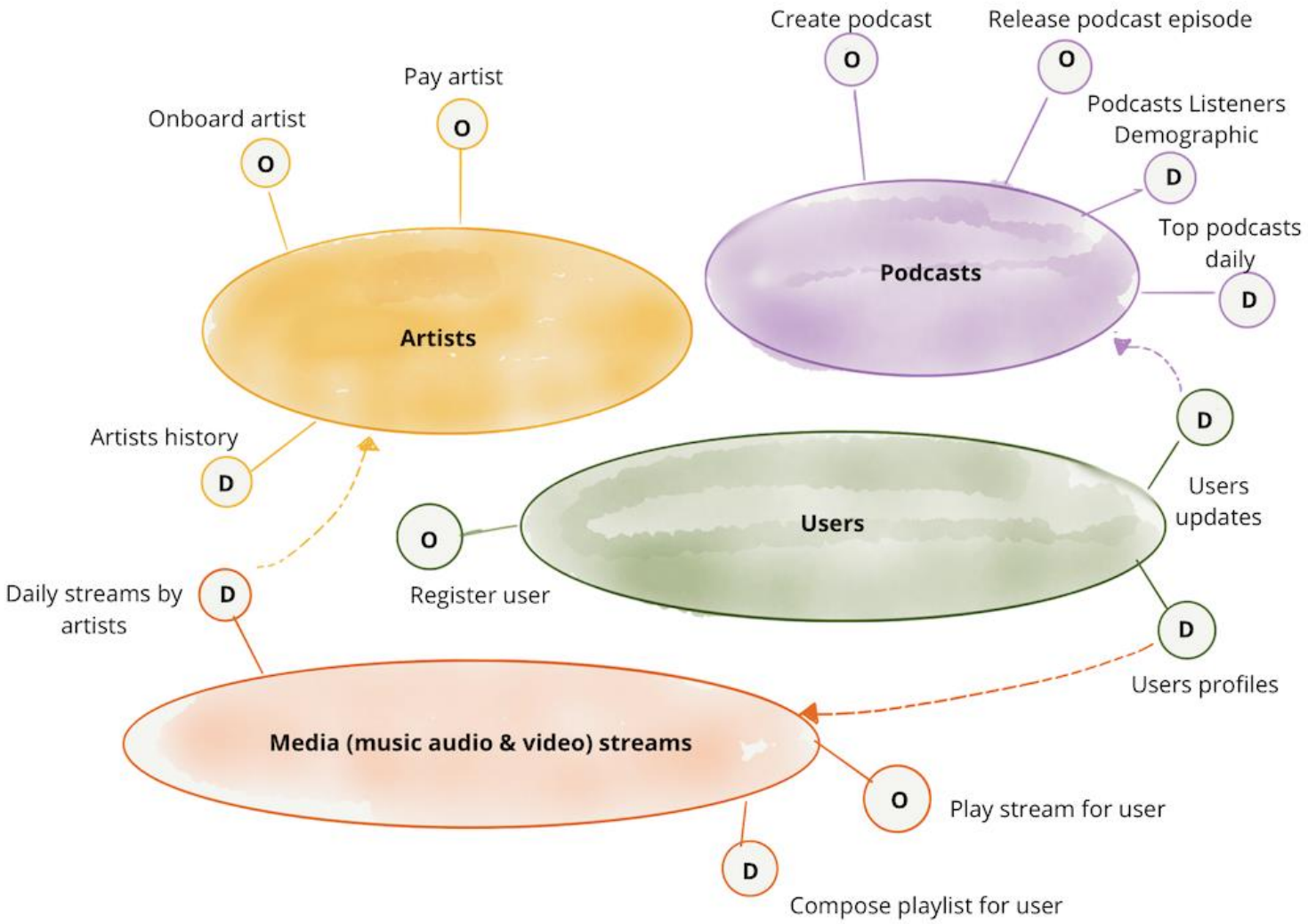
TRUSTWORTHY
(DEFINED & MONITORED SLOs)

SELF-DESCRIBING

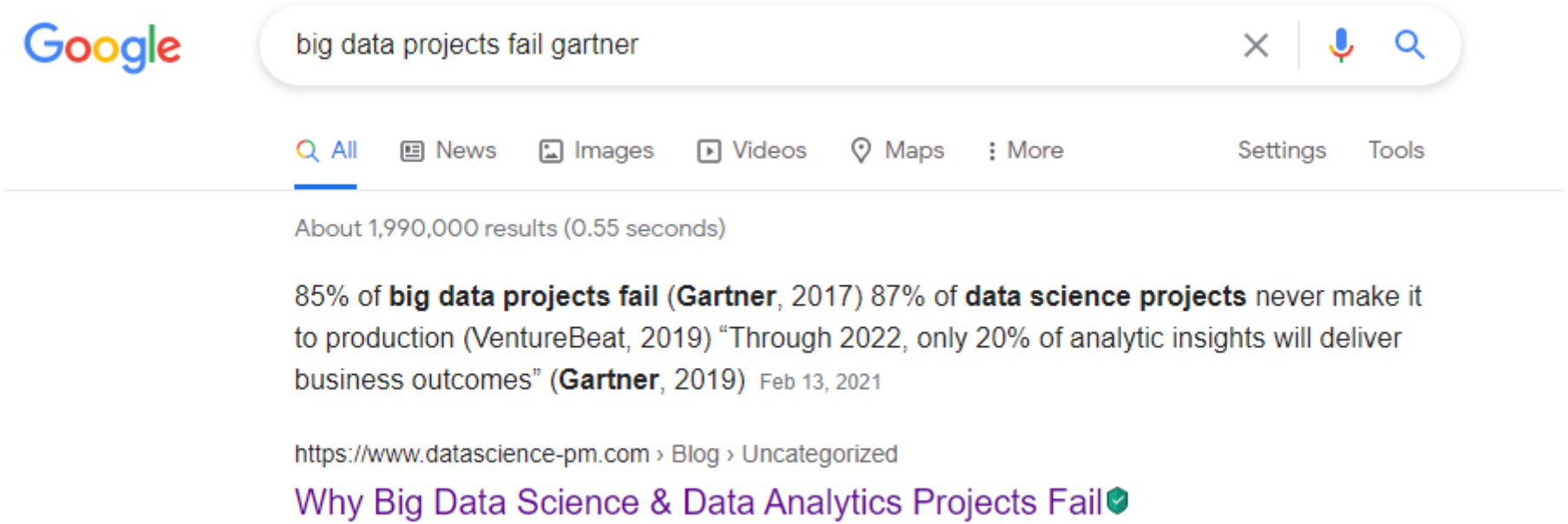
INTER OPERABLE
(GOVERENED BY OPEN STANDARDS)

SECURE

Data Mesh


Create podcast
Release podcast episode
Pay artist
Onboard artist
Podcasts Listeners Demographic
Top podcasts daily
Podcasts
Artists
Artists history
Users updates
Users
Register user
Daily streams by artists
Users profiles
Media (music audio & video) streams
Play stream for user
Compose playlist for user

Final remarks



Questions?

References

- Business Intelligence, Analytics, and Data Science: A Managerial Perspective 4th Edition
- Data Mesh Paradigm Shift in Data Platform Architecture
- Data Mesh Principles and Logical Architecture
- Future of Data Engineering