

DATA ENGINEERING 1

DATA WAREHOUSES

LASZLO SALLO

OCTOBER 2023

THE IDEAL WORLD OF DATA PERSISTENCY

WHAT WE WISH



Infinite storage

Immediate data write

Immediate data read

REALITY



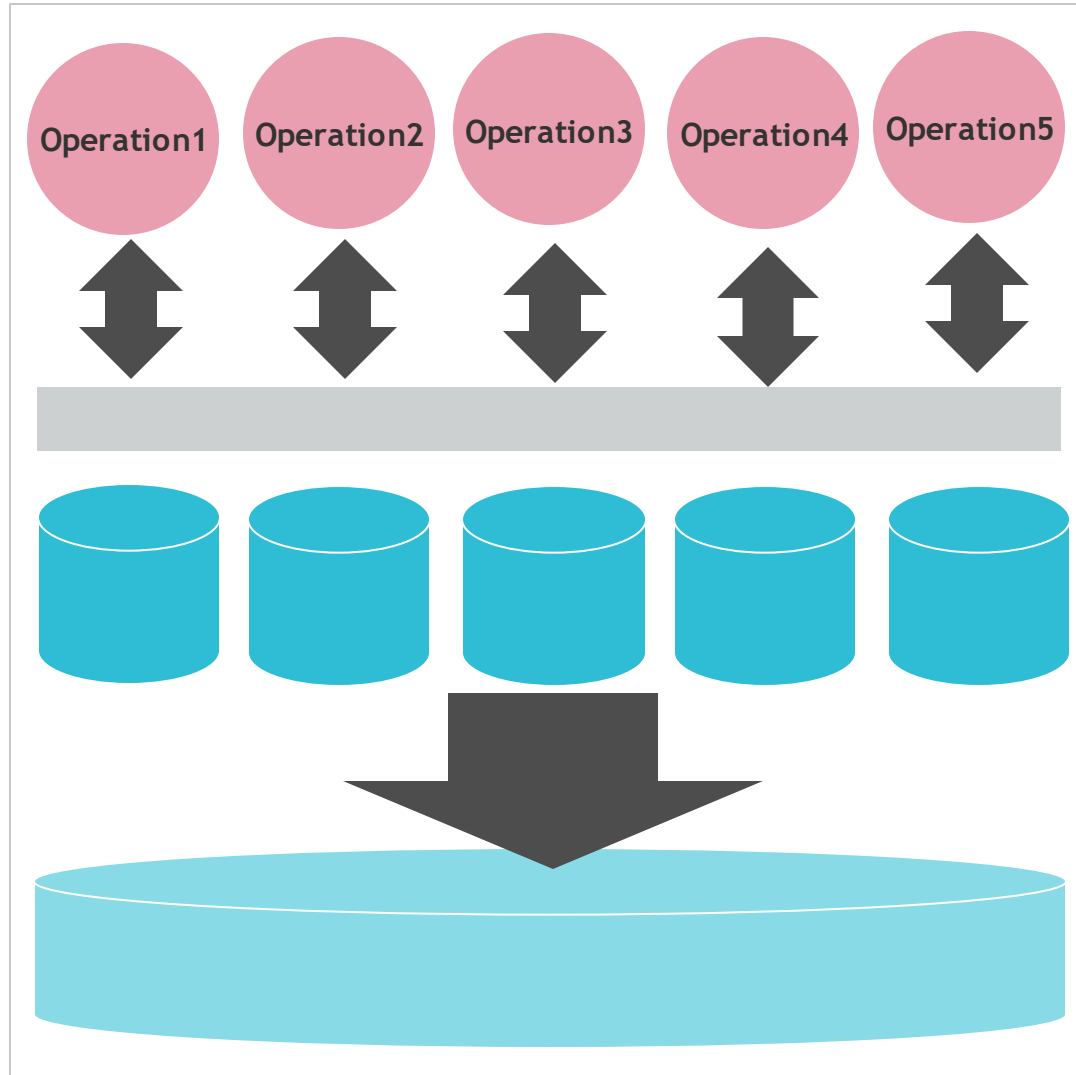
**Read and write depends
on your data volume.**

**Read and write go against
each other.**

PERFORMANCE OR VOLUME ?

WRITE OR READ ?

SEPARATION OF DATA LAYERS



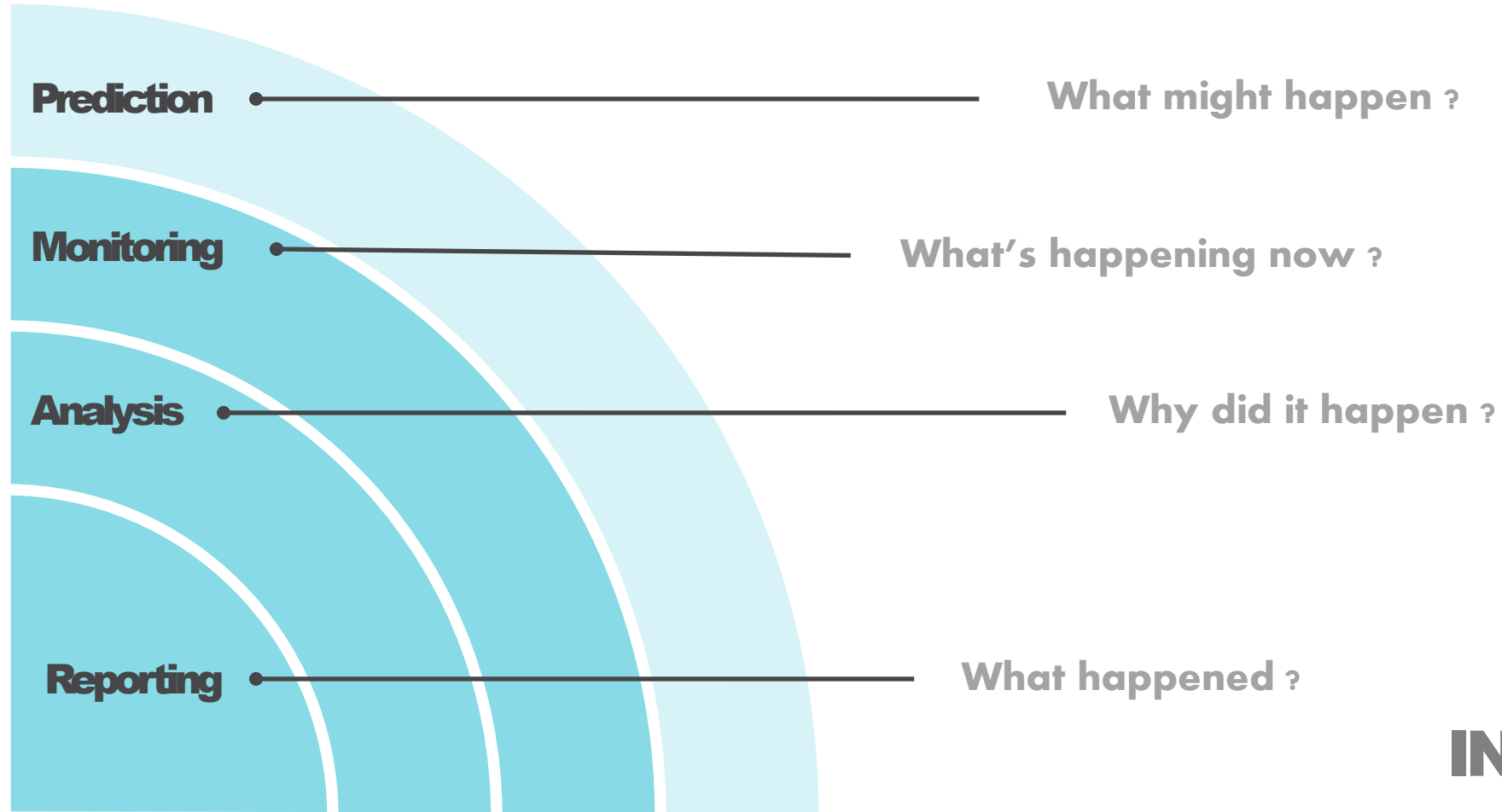
← **Operational Data Layer**

← **Analytical Data Layer**

DATA LAYERS

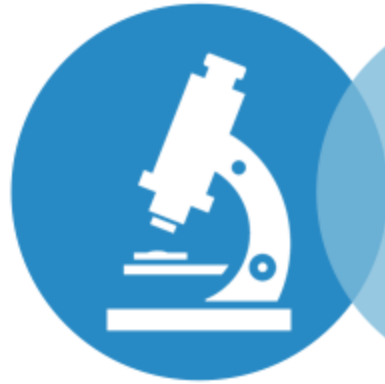
	OPERATIONAL	ANALYTICAL
Data age	60-90 days	3-10 years
Immutability	Can update / delete	No update / rare delete
Users	Operation and administration	Analysts and Stakeholders
Speed	Performance sensitive	Not performance critical
Optimized for	Many users	Few users
	Simple queries	Complex queries
	Real time	Delayed
	Write optimized	Read optimized

TYPES OF ANALYTICS



**BUSINESS
INTELLIGENCE (BI)**

TYPES OF ANALYTICS



Descriptive
Explains what
happened.



Diagnostic
Explains why it
happened.

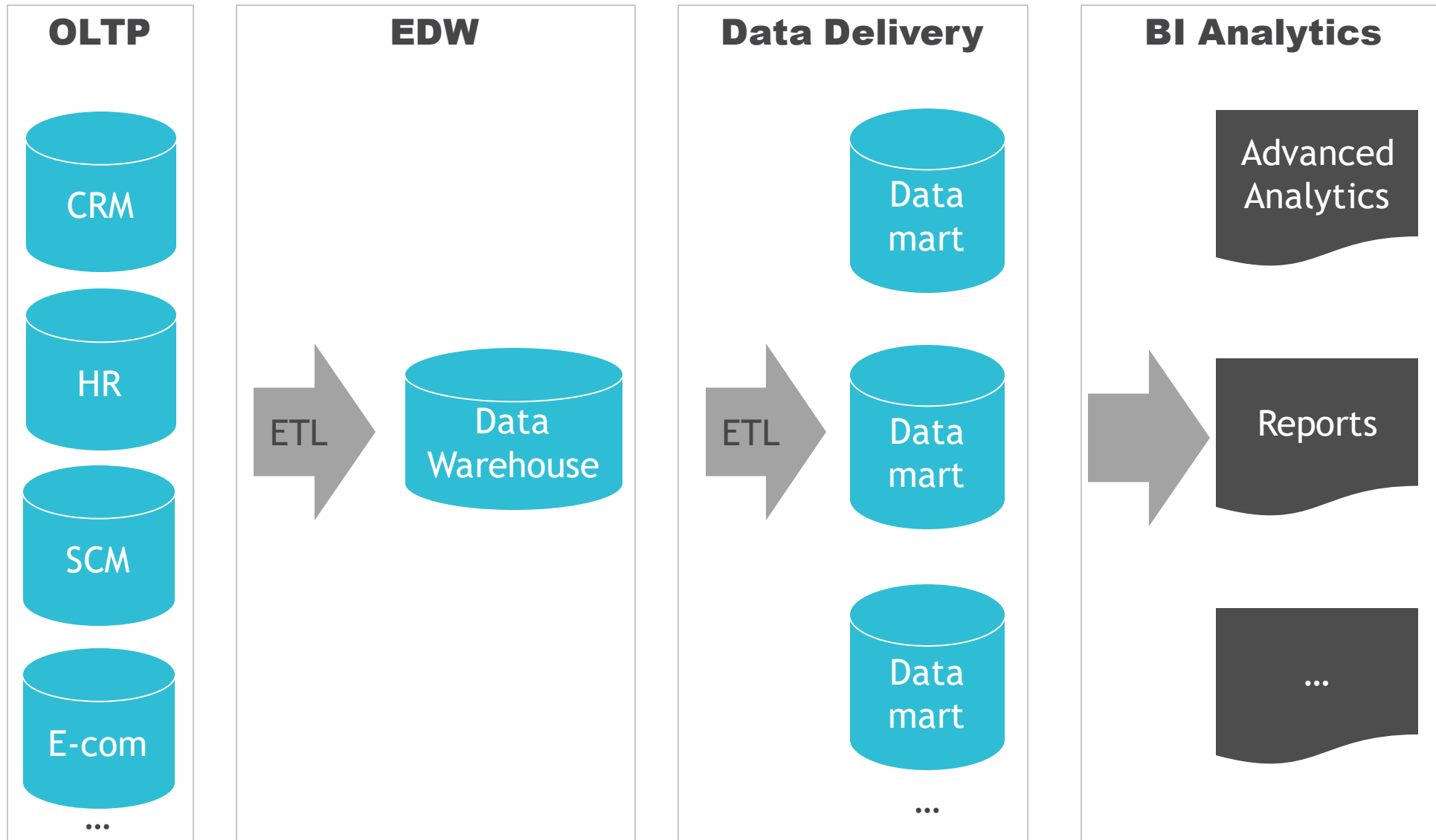


Predictive
Forecasts what
might happen.



Prescriptive
Recommends an
action based on
the forecast.

BI ARCHITECTURE





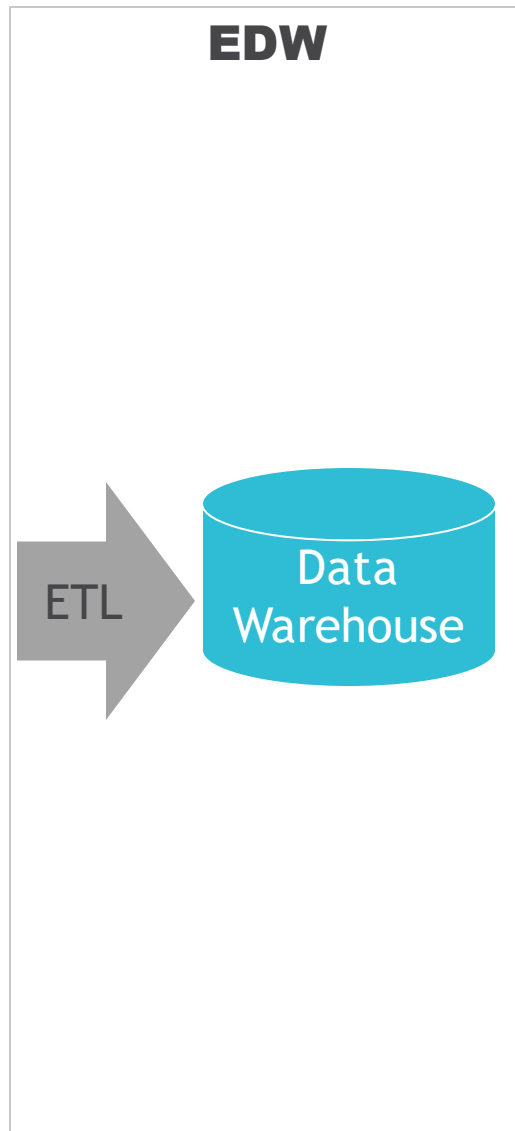
OLTP

Online Transaction Processing Systems

Operational data

Write optimized

Unsuitable for analytics



EDW

Enterprise Data Warehouse

Data from OLTP

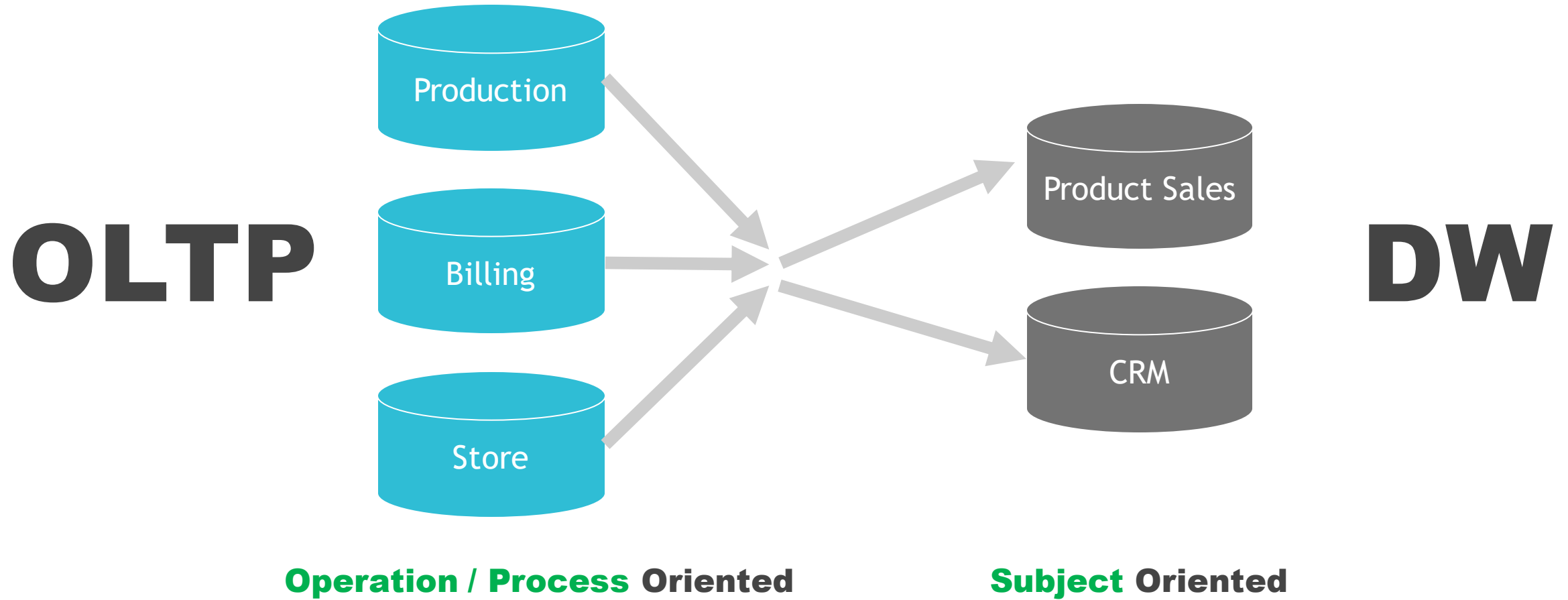
Historical data

Single source of truth

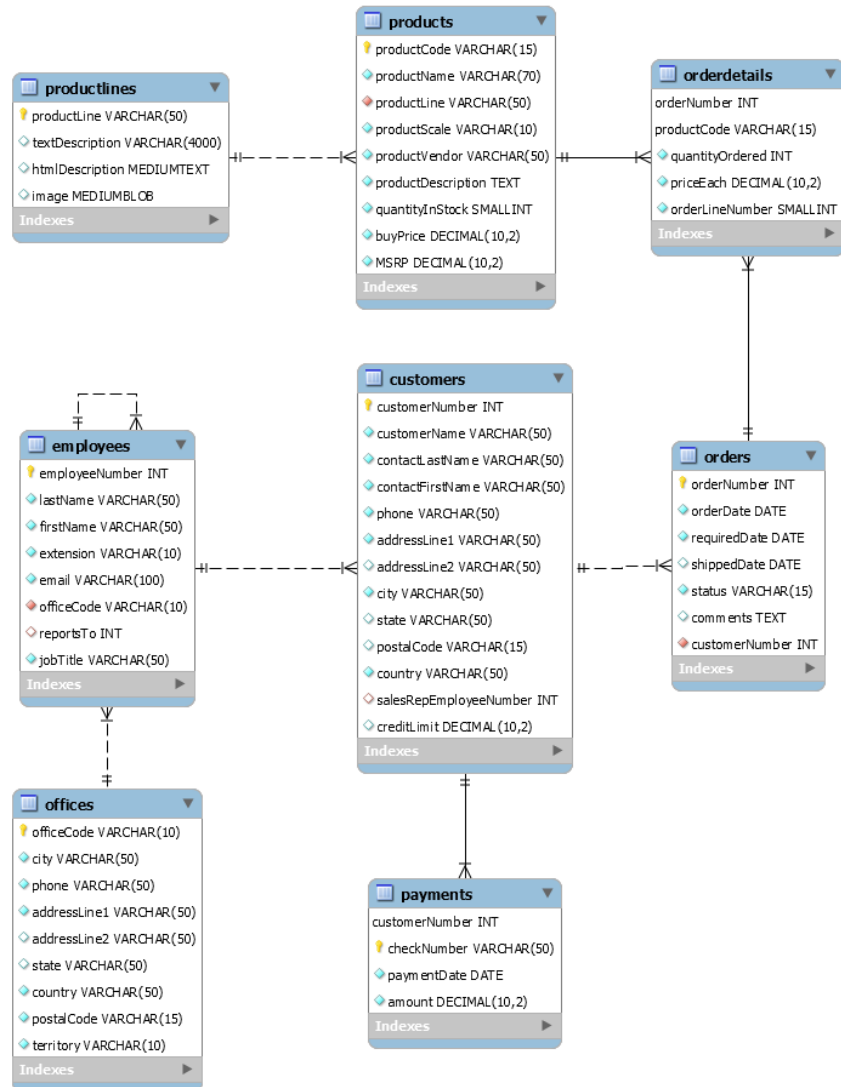
Read optimized

Relevant data for analytics

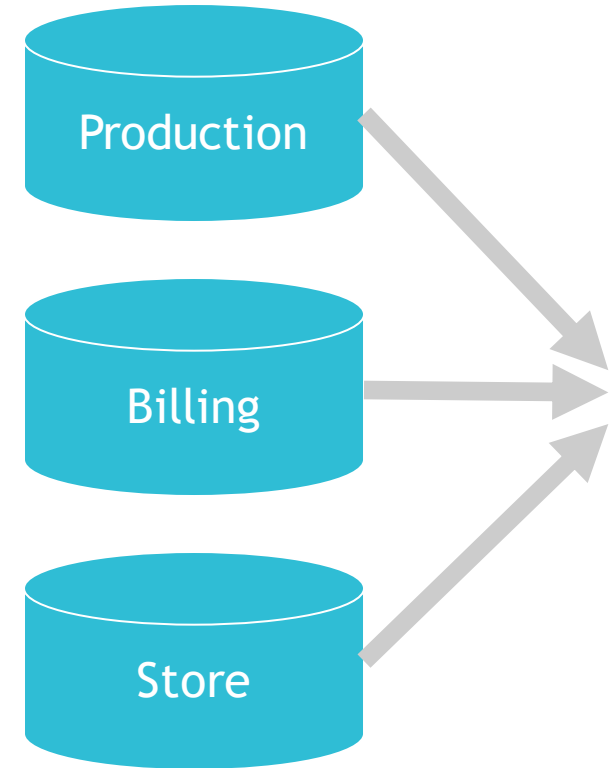
SUBJECT ORIENTATION



MAPPING TO OUR SAMPLE DB

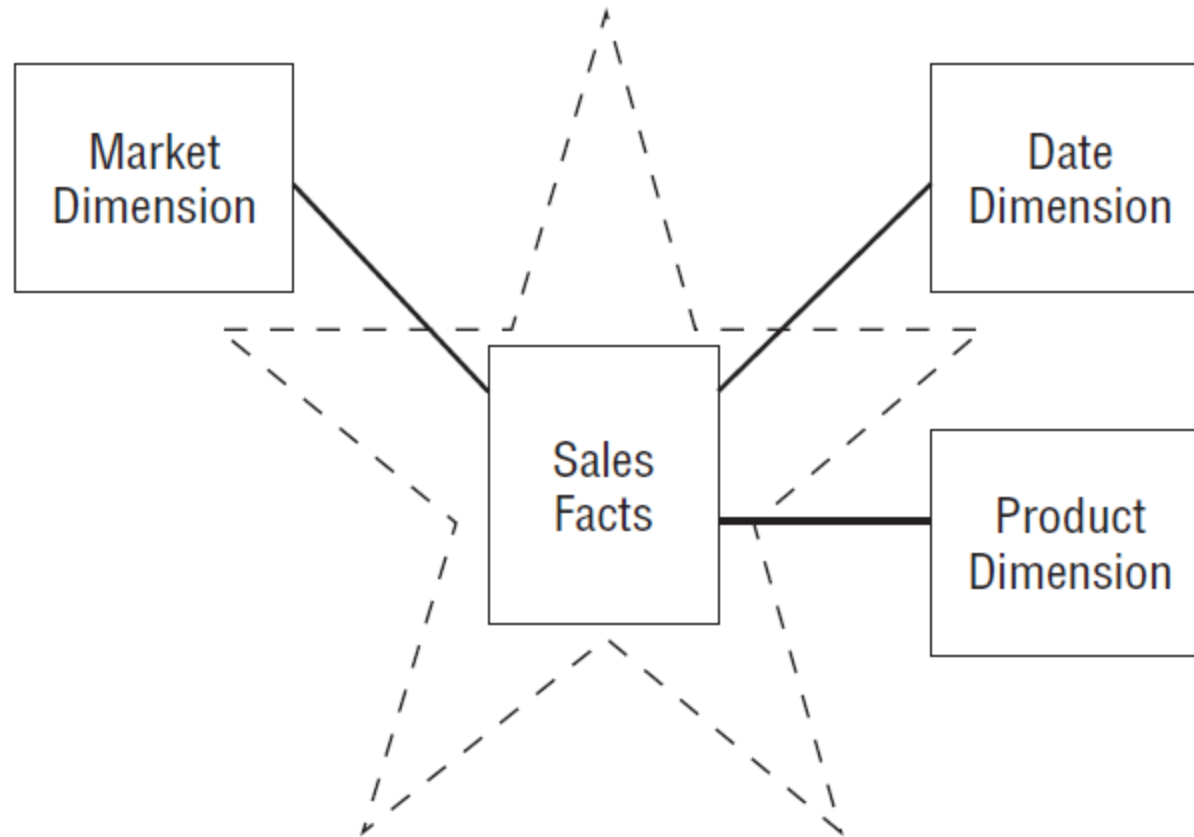


OLTP



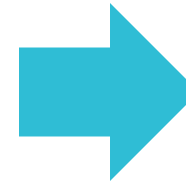
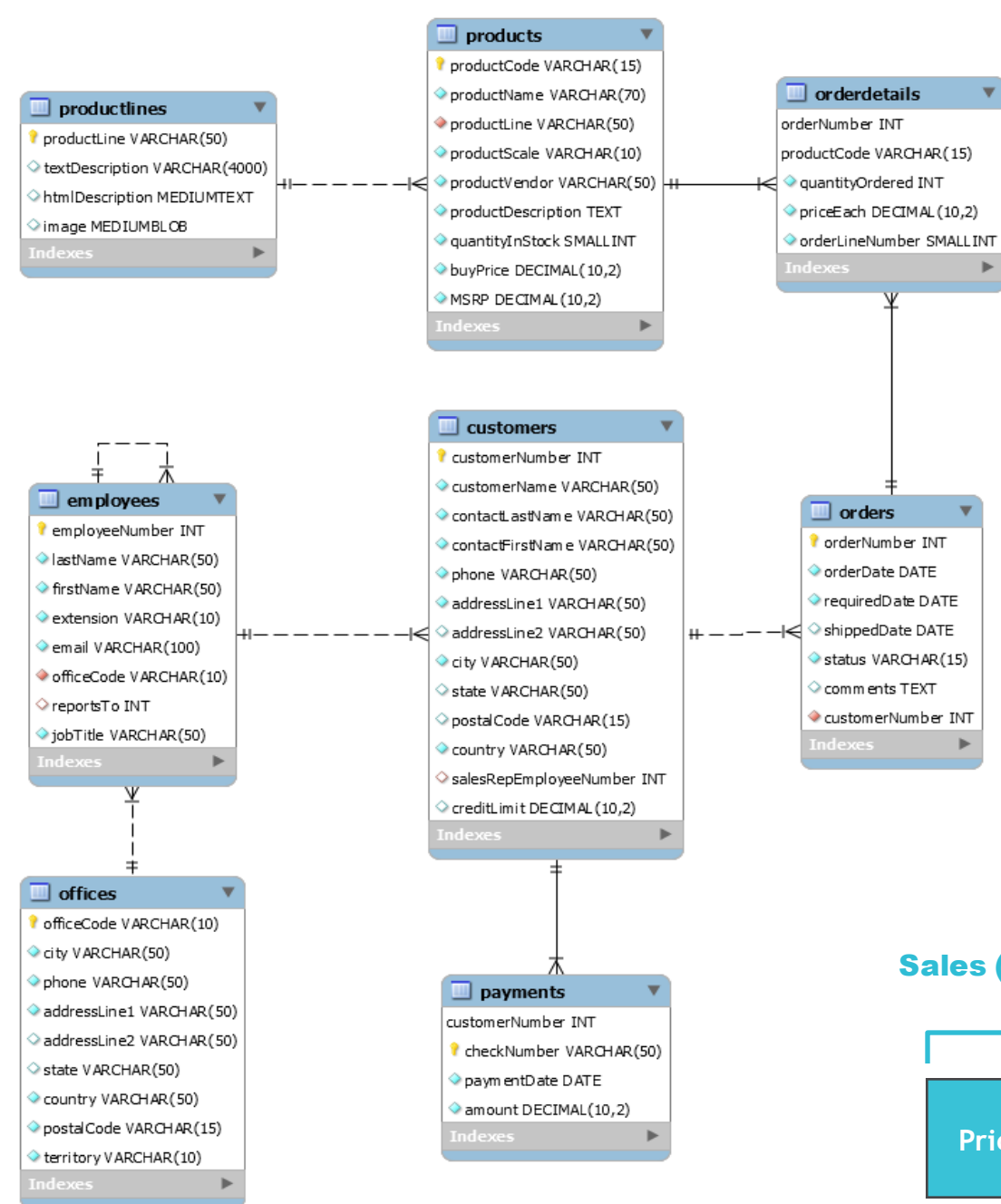
Operation / Process Oriented

PRODUCT SALES - STAR SCHEMA



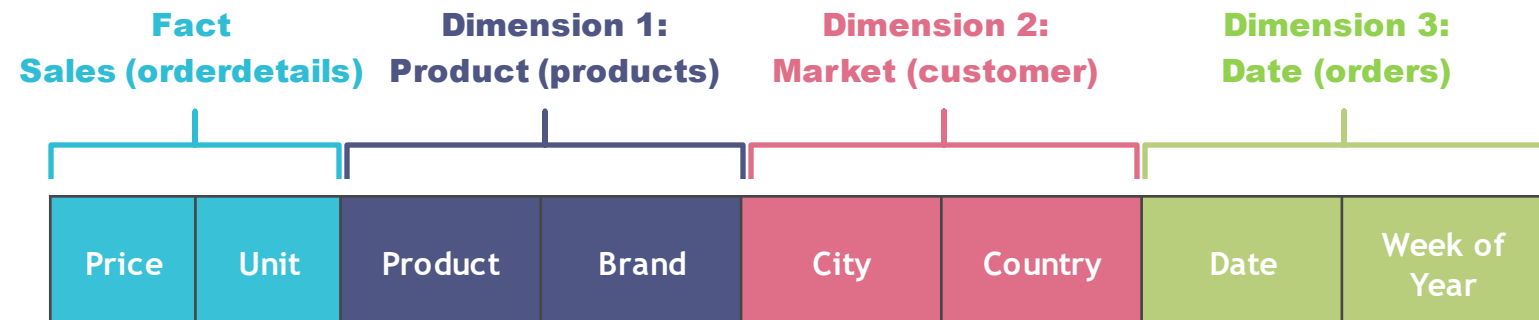
PRODUCT SALES

Fact Sales		Dimension 1: Product		Dimension 2: Market		Dimension 3: Date	
Price	Unit	Product	Brand	City	Country	Date	Week of Year
10.000	1	Chair	Ikea	Budapest	HU	2019.09.20	38
500	1	Flower	Zip	London	UK	2019.09.20	38
1.000.000	3	Spaceship	Virgin	Budapest	HU	2019.09.21	38
2.000	10	Weblogic	Oracle	New York	USA	2019.09.21	38

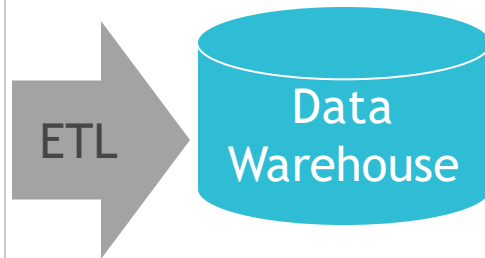


```

SELECT
    orders.orderNumber AS SalesId,
    orderdetails.priceEach AS Price,
    orderdetails.quantityOrdered AS Unit,
    products.productName AS Product,
    products.productLine AS Brand,
    customers.city AS City,
    customers.country AS Country,
    orders.orderDate AS Date,
    WEEK(orders.orderDate) AS WeekOfYear
FROM
    orders
INNER JOIN
    orderdetails USING (orderNumber)
INNER JOIN
    products USING (productCode)
INNER JOIN
    customers USING (customerNumber)
  
```



Data Warehouse



ETL

Extract Transform Load

Link between OLTP and DW

Data enrichment from external source

Tool or code

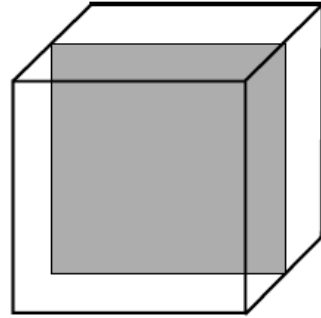
DW = RAW DATA

Still too slow

Different business fields needs different data / aggregation

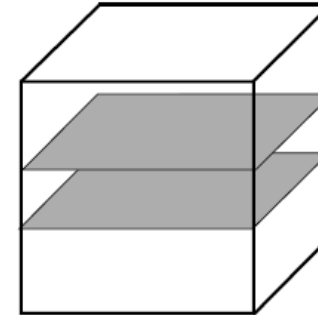
SLICING AND DICING

All Flower sale

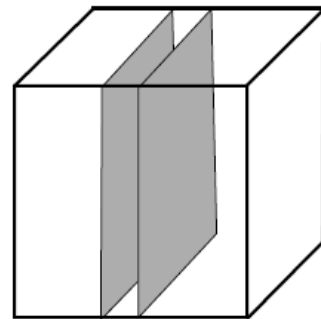


Product manager

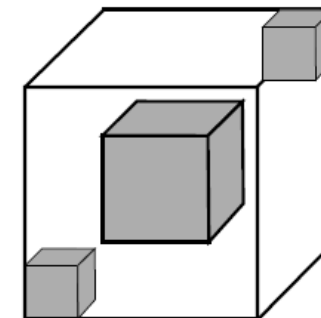
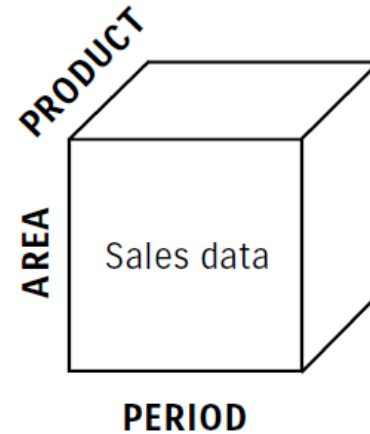
All sale of Budapest and New York



Area manager



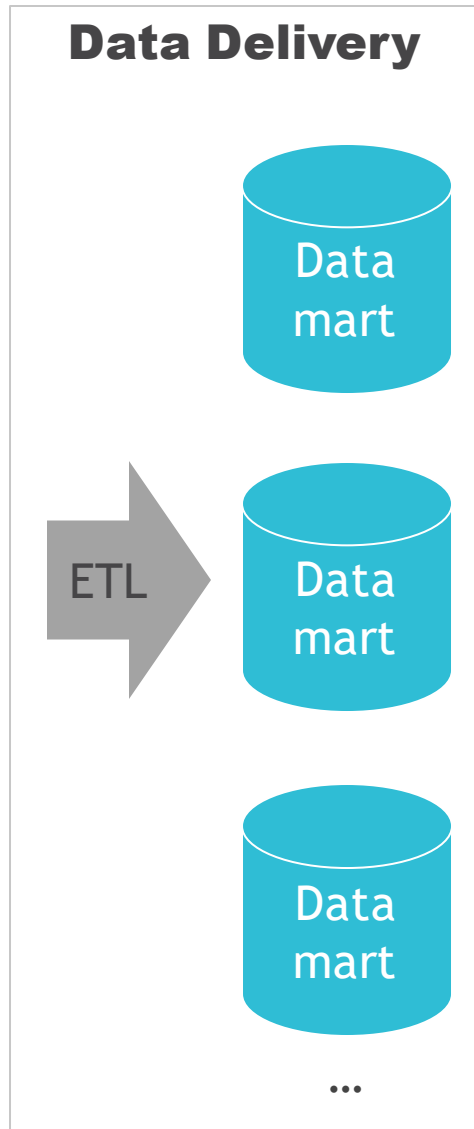
Chief financial
officer



Ad hoc views

Furniture sales of a year in Europe

The current sale vs prev. weeks sales



DATA

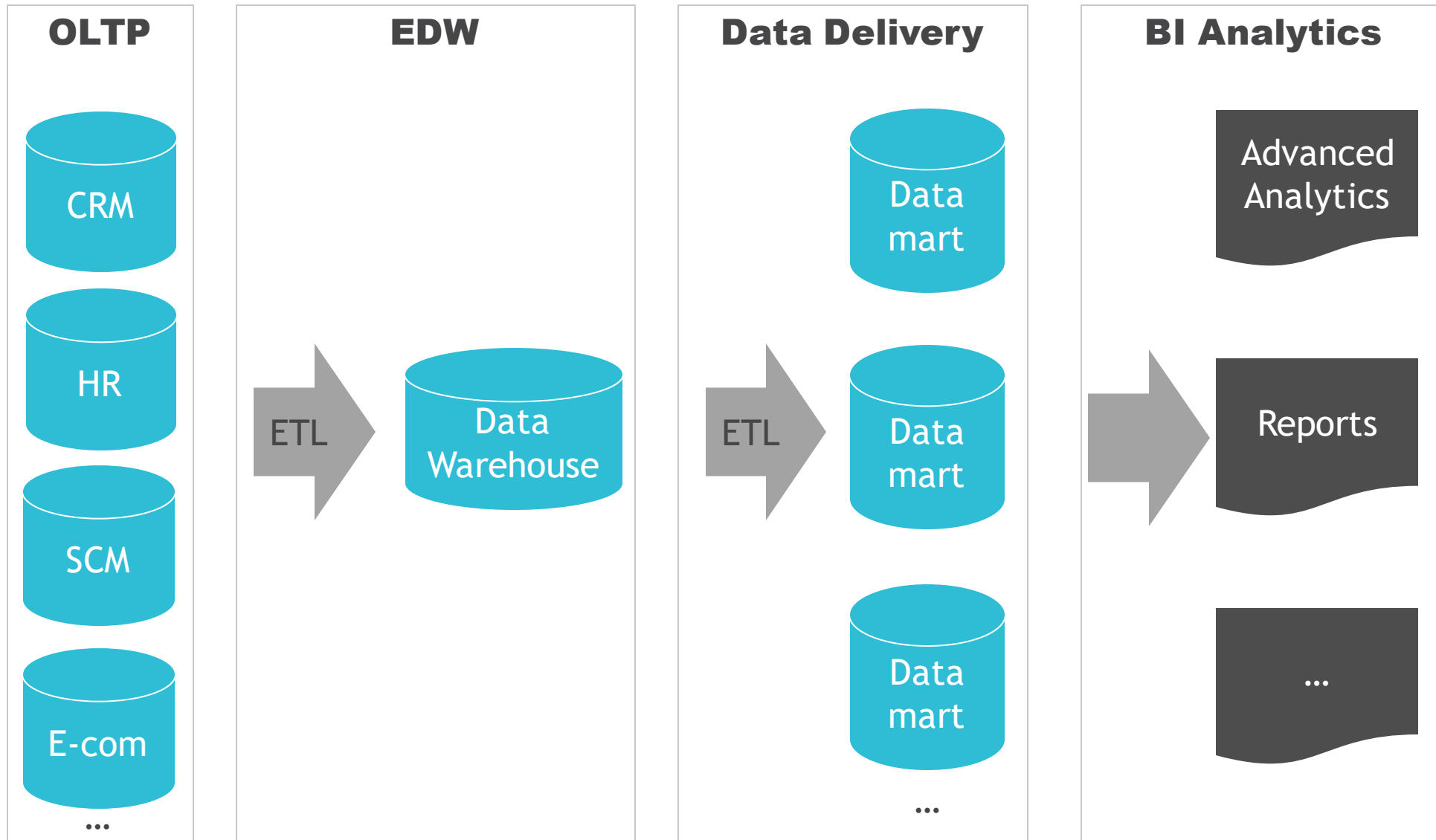
Delivery

Analytics driven data marts

Subset Of DW

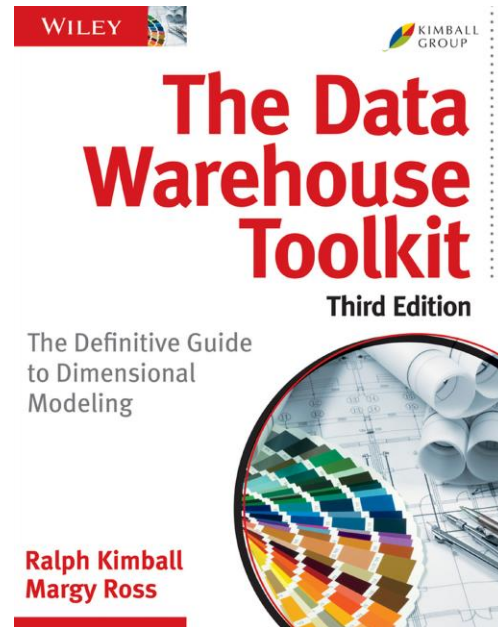
Slice of a subject

BI ARCHITECTURE



Reading recommendation

BOOK:



ARTICLE:

**Data
Warehouse
Design –
Inmon
versus
Kimball**

<http://tdan.com/data-warehouse-design-inmon-versus-kimball/20300>