

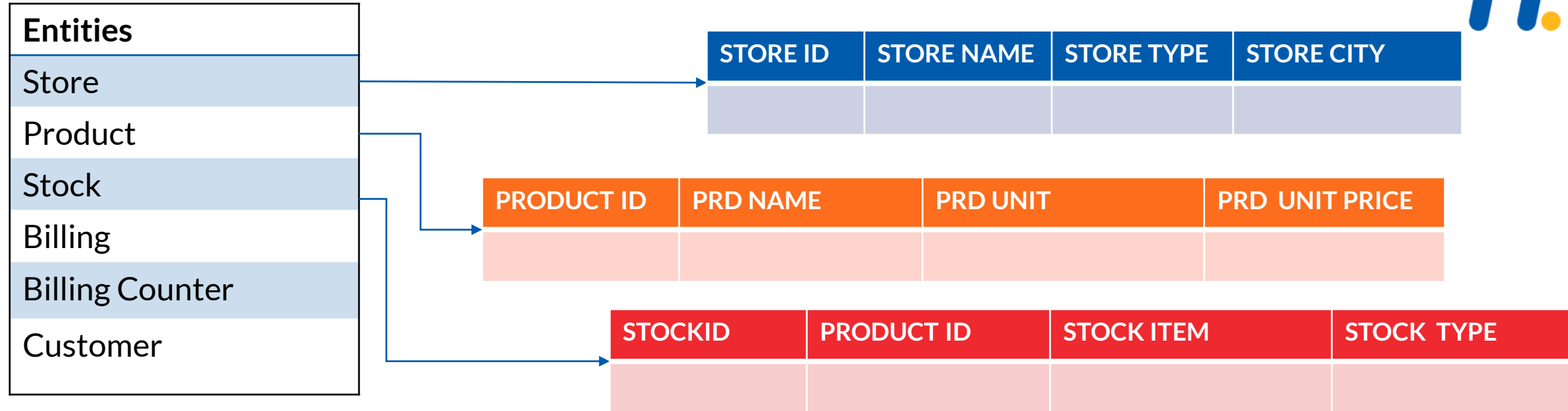


Data Warehouse

An Introduction

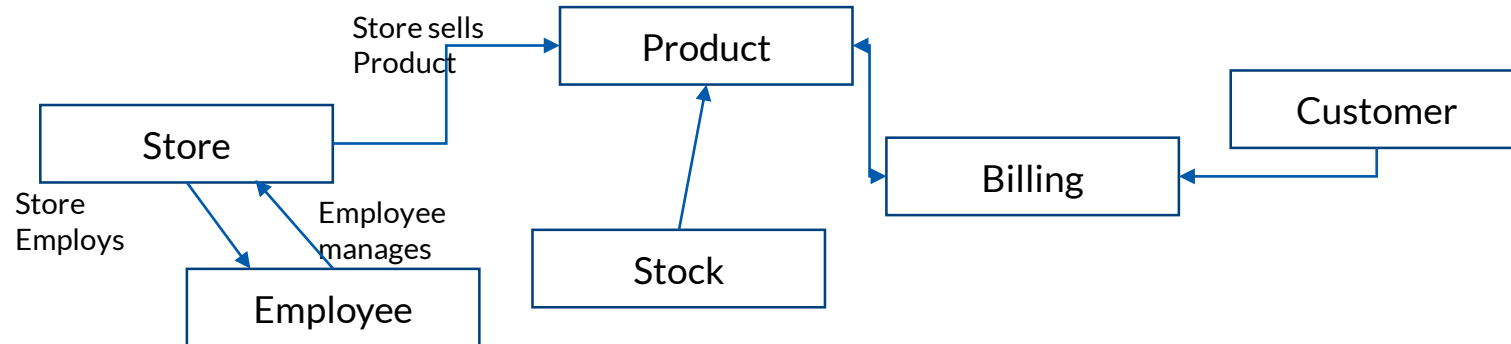


Transaction Systems - OLTP



STORE - S001

RAVI BUYS 10 PACKETS OF LAYS CHIPS



What is an Analytical System



Need to improve sales of my products in Chennai, let me first learn about my sales & customers

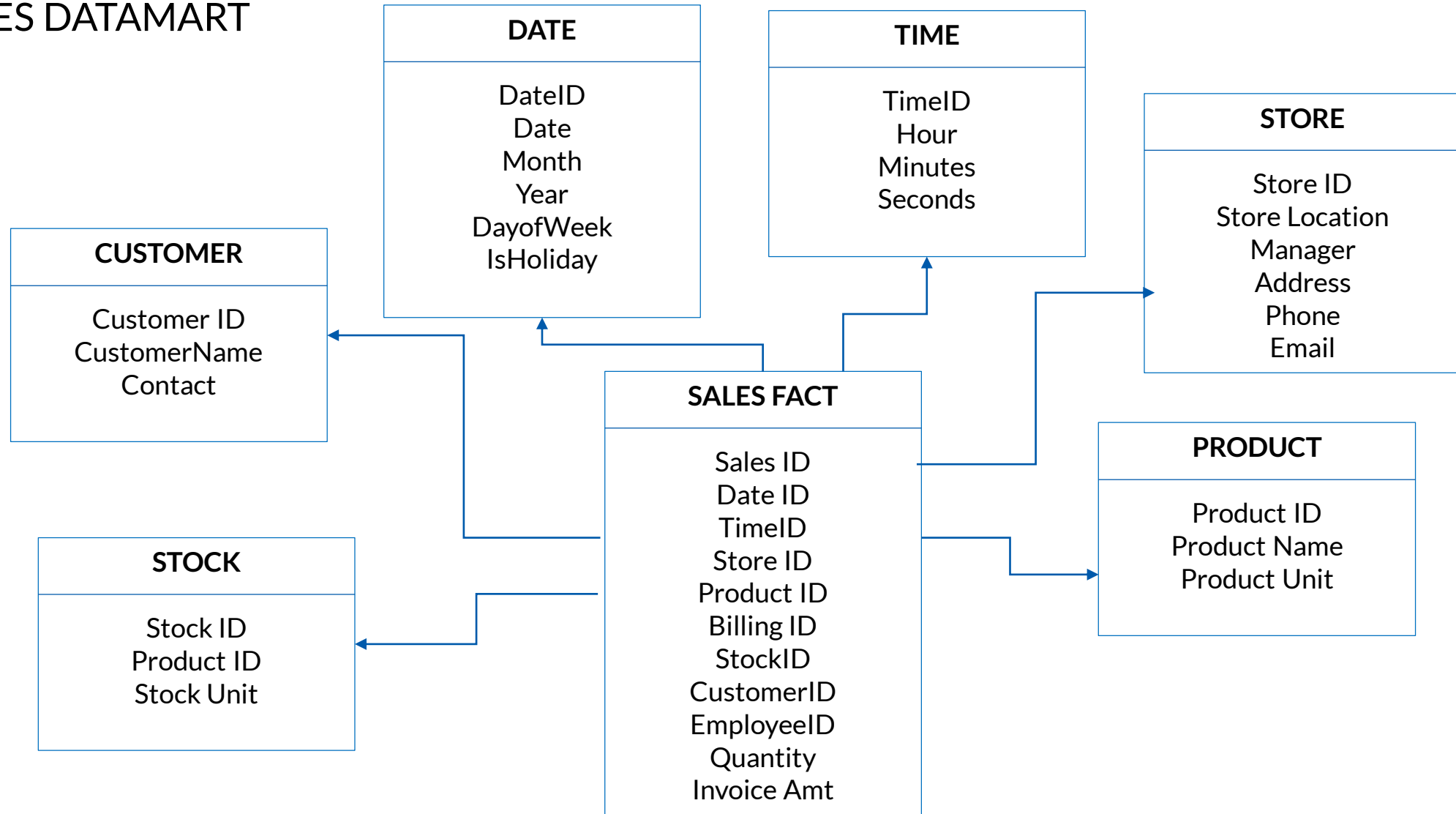
I would like to know



1. TOTAL SALES IN CHENNAI REGION
2. TOTAL SALES OF LAYS CHIPS IN EACH STORE IN CHENNAI REGION
3. HOW MANY COLLEGE GOERS BOUGHT LAYS CHIPS?
4. AGE GROUP WISE, ANALYSIS OF QUANTITY OF LAYS CHIPS PURCHASED?
5. TIME OF DAY WISE - LAYS CHIPS CUSTOMER DEMOGRAPHICS
6. WHAT ARE THE OTHER PRODUCTS BOUGHT ALONG WITH LAYS CHIPS?
7. CUSTOMER TOTAL BILL VS QUANTITY OF LAYS CHIPS



SALES DATAMART



What is a Datawarehouse?



DWH is a database design that enables business analytics.

- Managing Historic Records
- Analyzing the data to gain a better understanding of the business and to improve the business

A copy of transaction data specifically structured for query and analysis.

Ralph Kimball

A data warehouse is a subject-oriented, integrated, time-variant, non-volatile collection of data that is used primarily in organizational decision making

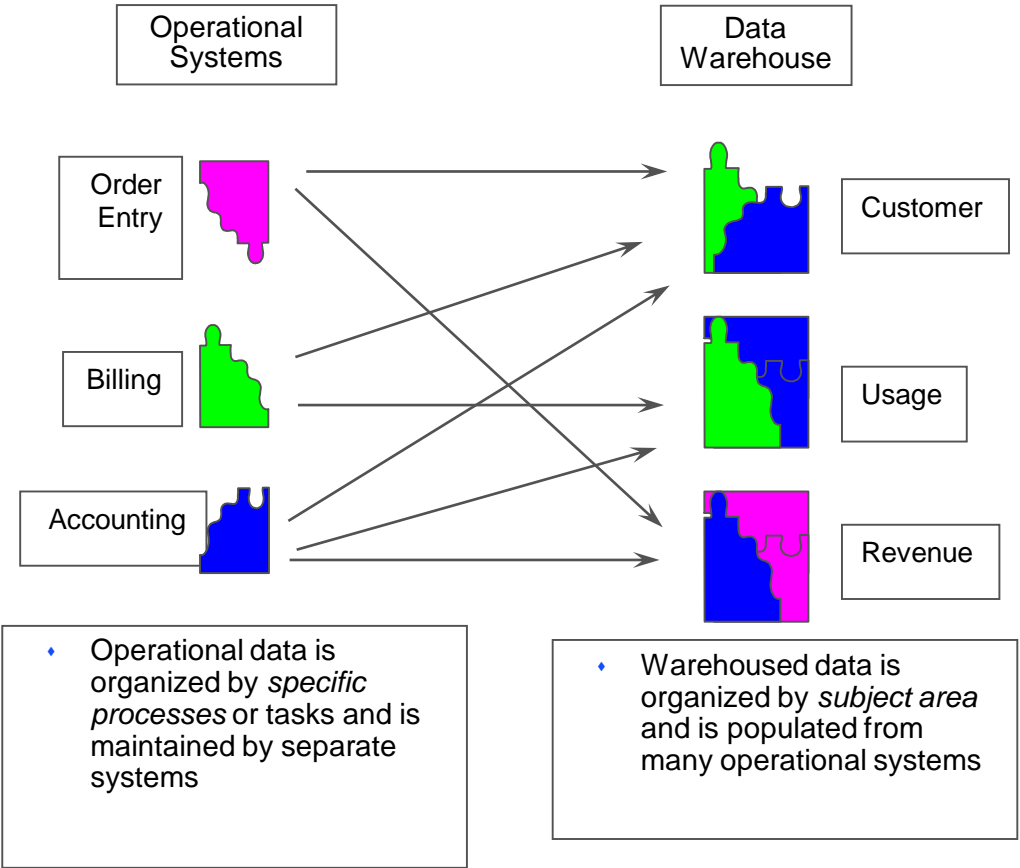
W.H. Inmon



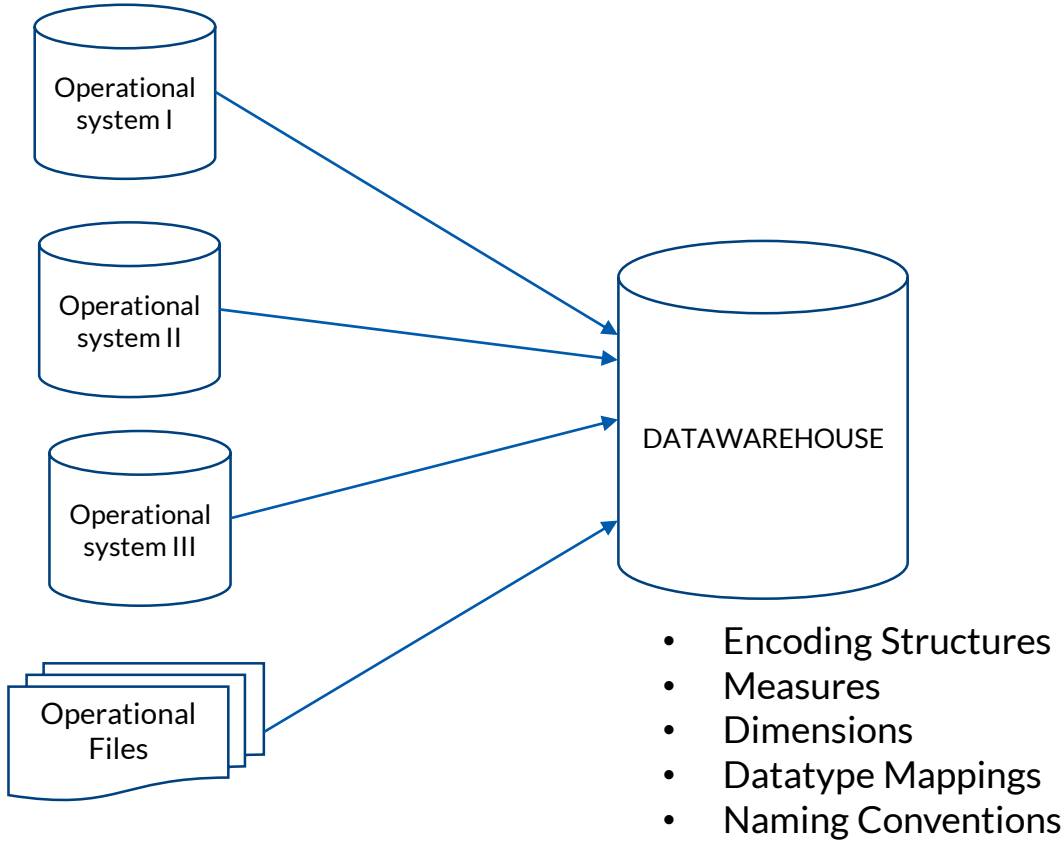
DWH ATTRIBUTES?



Subject Oriented



Integrated

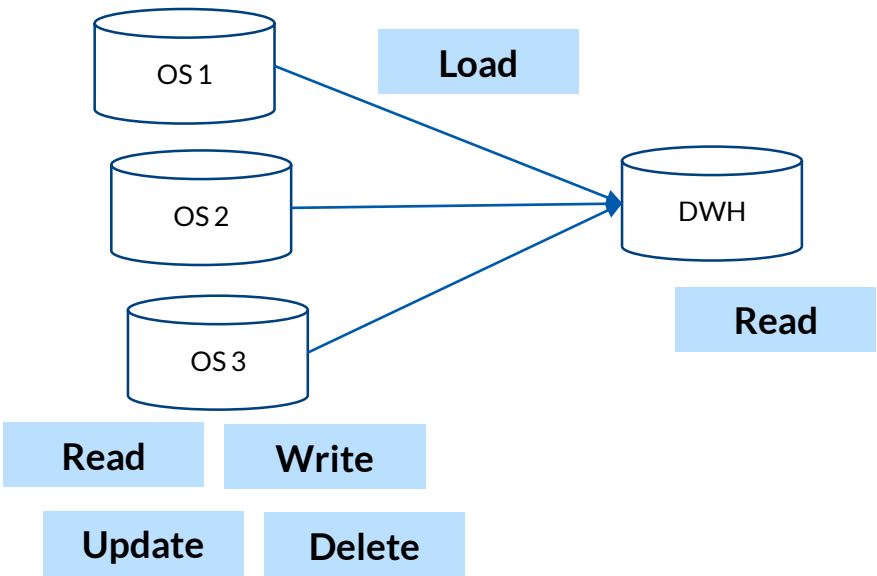


DWH ATTRIBUTES?



Non Volatile

Data from operational systems are moved into a DWH in specific intervals
Records in DWH once loaded are not usually updated
Data in DWH are not deleted

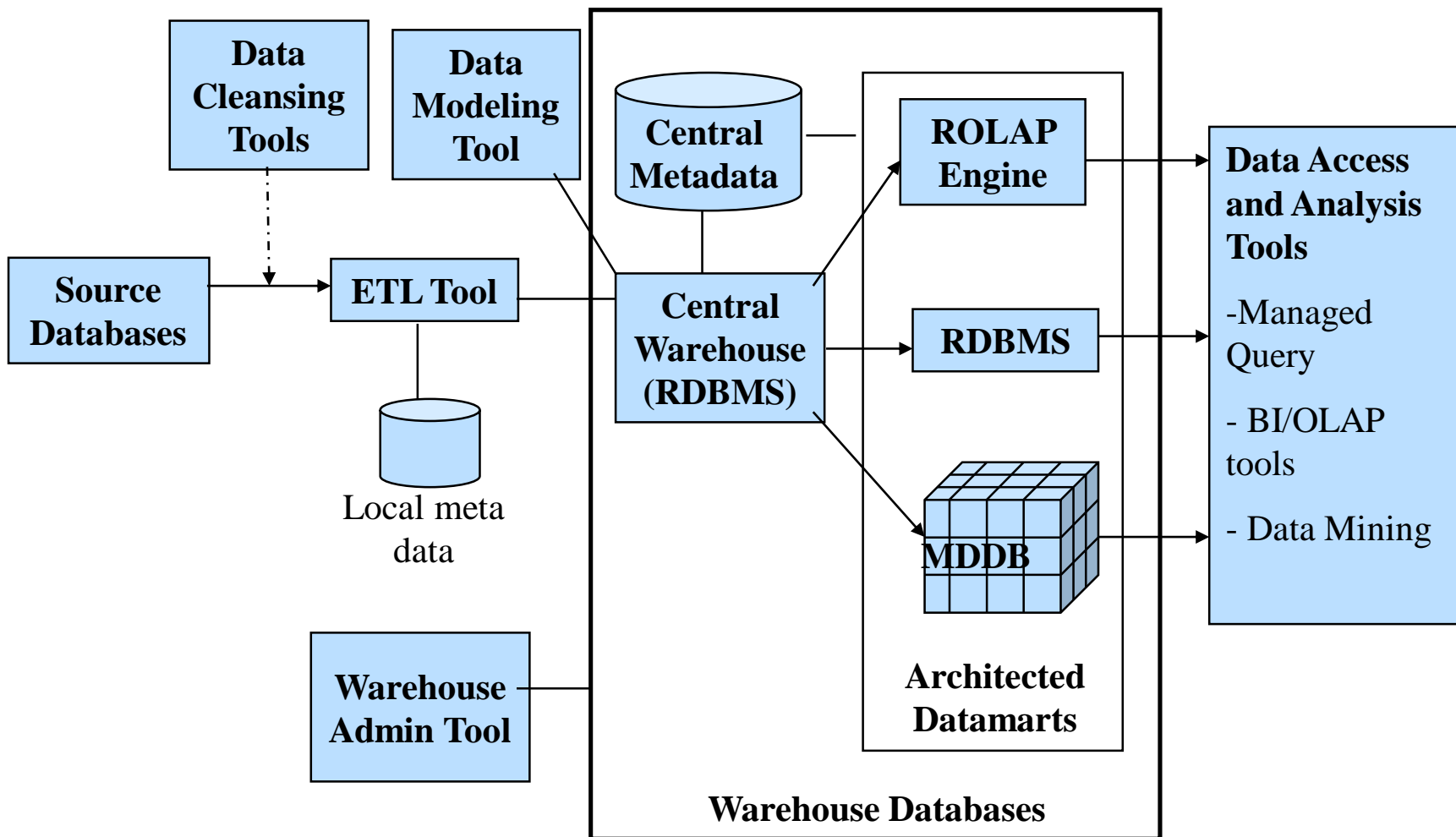


Time Variant

DWH always stores an element of time and is nothing but a series of snapshots, taken as of some moment in time.

Operational	DWH
Time horizon 60 – 90 days	Time horizon – 5-10 years
Definitions or values change with time For example: When a product name changes, these systems are updated with the new name	Definitions or values are held for a point in time. For example: When a product name changes, DWH holds the old name till the date it was changes and the new date from when it is changed
Data can be updated	Once a snapshot is made, records cannot be updated

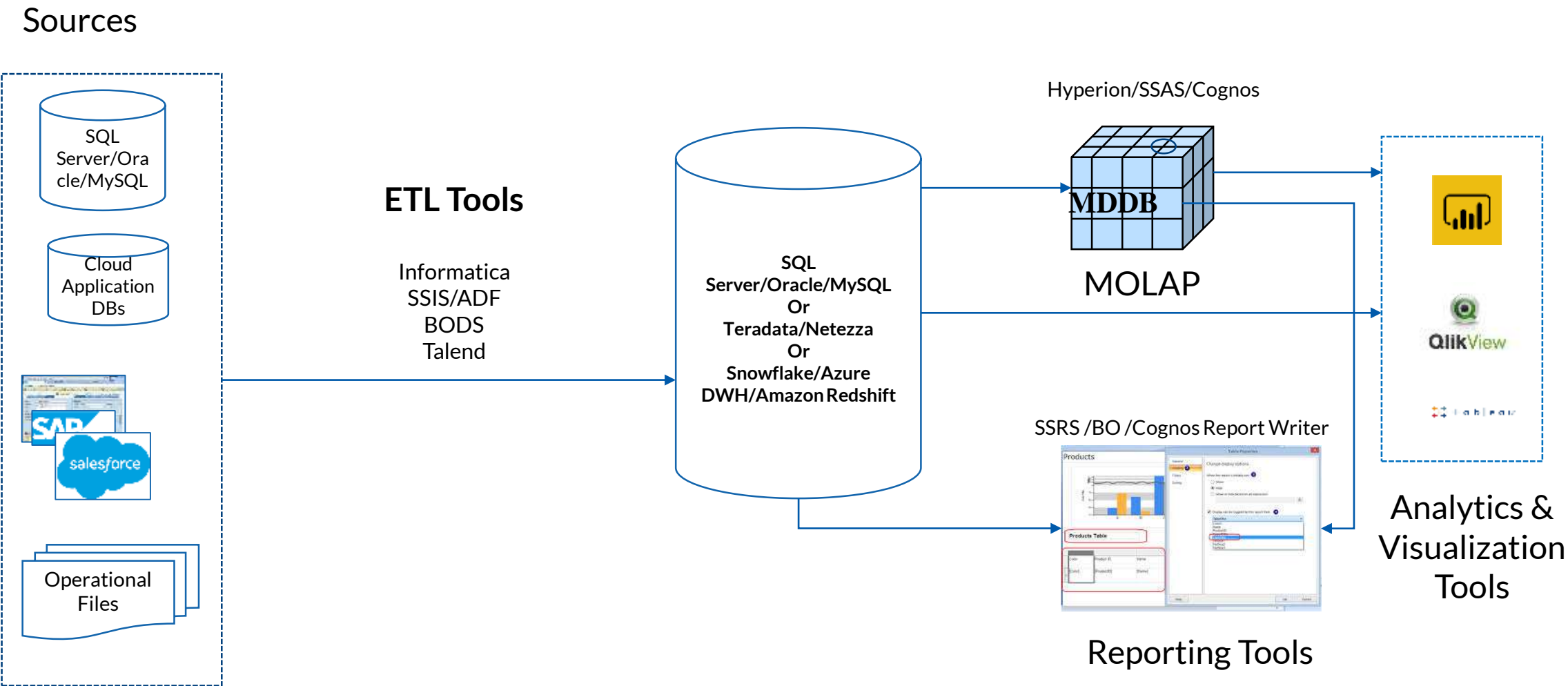
DWH Architecture



Components

- Source System(s)
- Extraction, Transformation, Loading (ETL) tool
- Data Warehouse maintenance and administration tools
- Data modeling tool or interface to external data models
- Target data warehouse databases
- End-user data access and analysis tools (BI/OLAP/DSS)

DWH Architecture

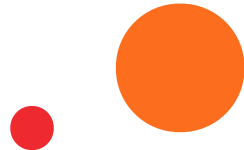


ODS – Operational Data Store



The purpose of the Operation Data Store (ODS) is to integrate corporate data from different heterogeneous data sources in order to facilitate real time or near real time operational reporting.

- Often their structure is similar to that of source systems
- During integration, data cleansing, de-duplication and standardization are performed
- Usually maintains data in the lowest granularity level
- ODS is not optimized to store and analyze huge set of historical data





OLAP – Online Analytical Processing

- OLAP means Online Analytical processing
- Enables decision making for future actions (DSS)
- OLAP ranges from basic navigation and browsing to calculations to serious analysis such as a time series analysis or complex modeling
- Drill Up & Down
- Slice & Dice
- Multi-dimensional Analysis

Different types of OLAP

- ROLAP (Relational OLAP)
- MOLAP (Multi-dimensional OLAP)
- HOLAP (Hybrid OLAP)

OLAP – Online Analytical Processing



ROLAP – Relational OLAP

- Data stored in the relational database
- Each action of slicing and dicing is equivalent to adding a "WHERE" clause in the SQL statement.

Advantages:

- Can handle large amounts of data
- Can leverage the strengths of a relational database
- More common and easy to initiate

Disadvantages:

- Performance can be limited
- Limited by the capabilities of SQL when compared to MDX

MOLAP – Multi-dimensional OLAP

- Data stored in a multidimensional cube
- Storage is not in the relational database, but in proprietary formats.

Advantages:

- Excellent performance
- Can perform complex calculations

Disadvantages:

- Limited in the amount of data it can handle
- High investment

HOLAP – Hybrid OLAP

- HOLAP technologies attempt to combine the advantages of MOLAP and ROLAP
- For summary-type information, HOLAP leverages cube technology for faster performance
- When detail information is needed, HOLAP can "drill through" from the cube into the underlying relational data.

OLTP Vs OLAP



OLTP	OLAP
Orientation of data	
Aligned by application	Aligned by dimension
Different systems hold different types of data	All types of data Integrated into one system
Data is inherently organized by application	Data is organized by defined dimensions of the business
Contains detailed up-to-date data(hundreds of MB to GB)	Historical, Summarized and consolidated data from several operational databases over long periods of time projected hundreds of GB to TB
Relational data model	Multidimensional data cube. Dimensions are hierarchical.
Integration	
Typically not Integrated	Must be integrated
Different key structures	Has a standard key structure
It has different naming conventions	Standard naming conventions
Different file formats	Standard file formats
Data access and manipulation	
Transaction	Bulk Process
Allow inserts, updates, deletes	Selects only
Non redundancy	Redundancy is present
High probability of access	Low when compared to OLTP

OLTP Vs OLAP



OLTP	OLAP
History	
Recent or current data	Historical data
60-90 days	2+ years
Current values only	Historical snapshots of data
No time key	Time key
No time series analysis	Time series analysis
Usage	
It is fairly consistent	Spiked or uneven
Serves those requiring current and detailed information	Serves those requiring historical and summarized information



Thank you

Innovative Services



Passionate Employees

Delighted Customers

