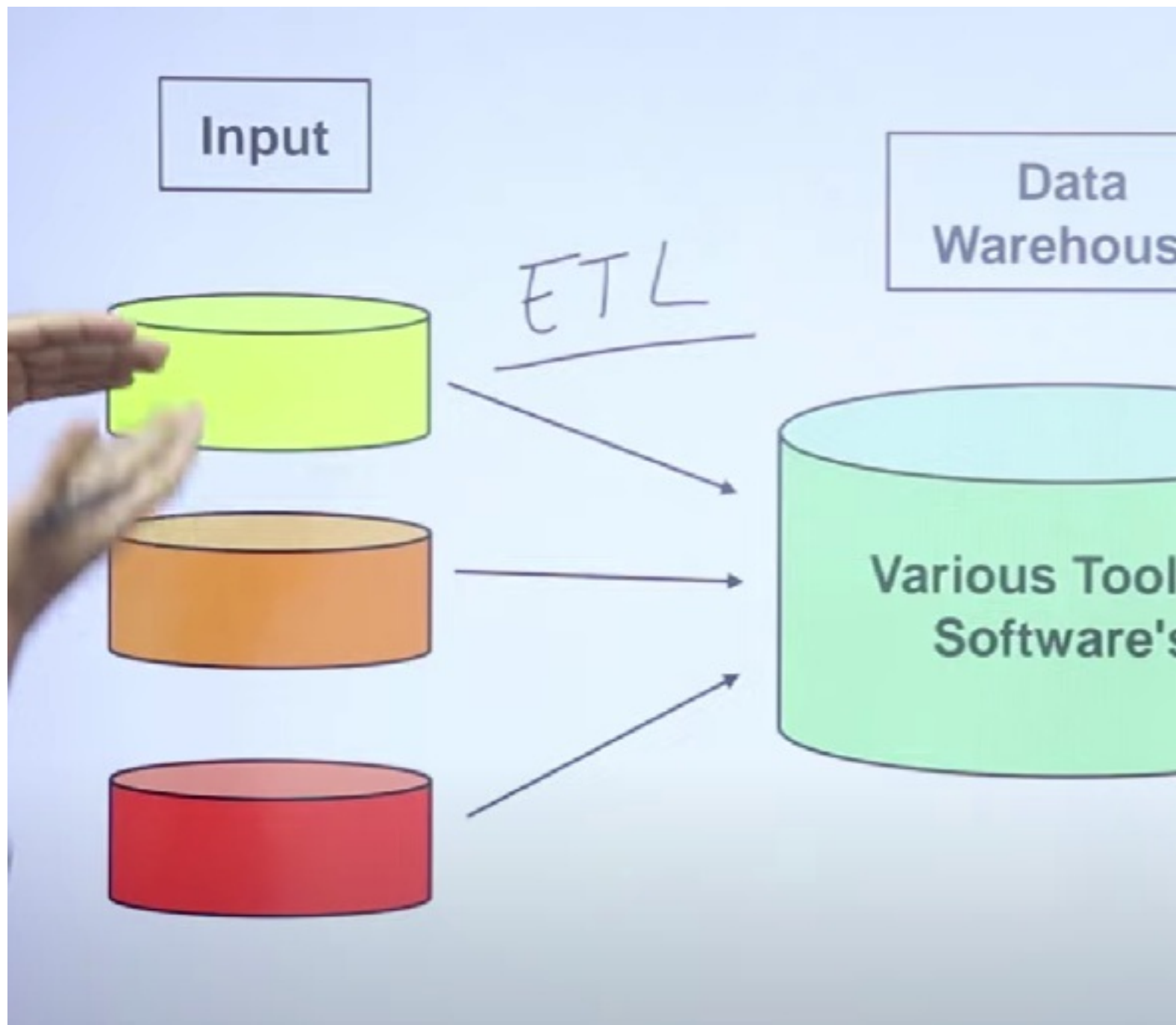


Data warehouse : multiple locations se data is collected and stored here

ETL : extract transform load



Tools like Microsoft ssis , Apache spark do ETL

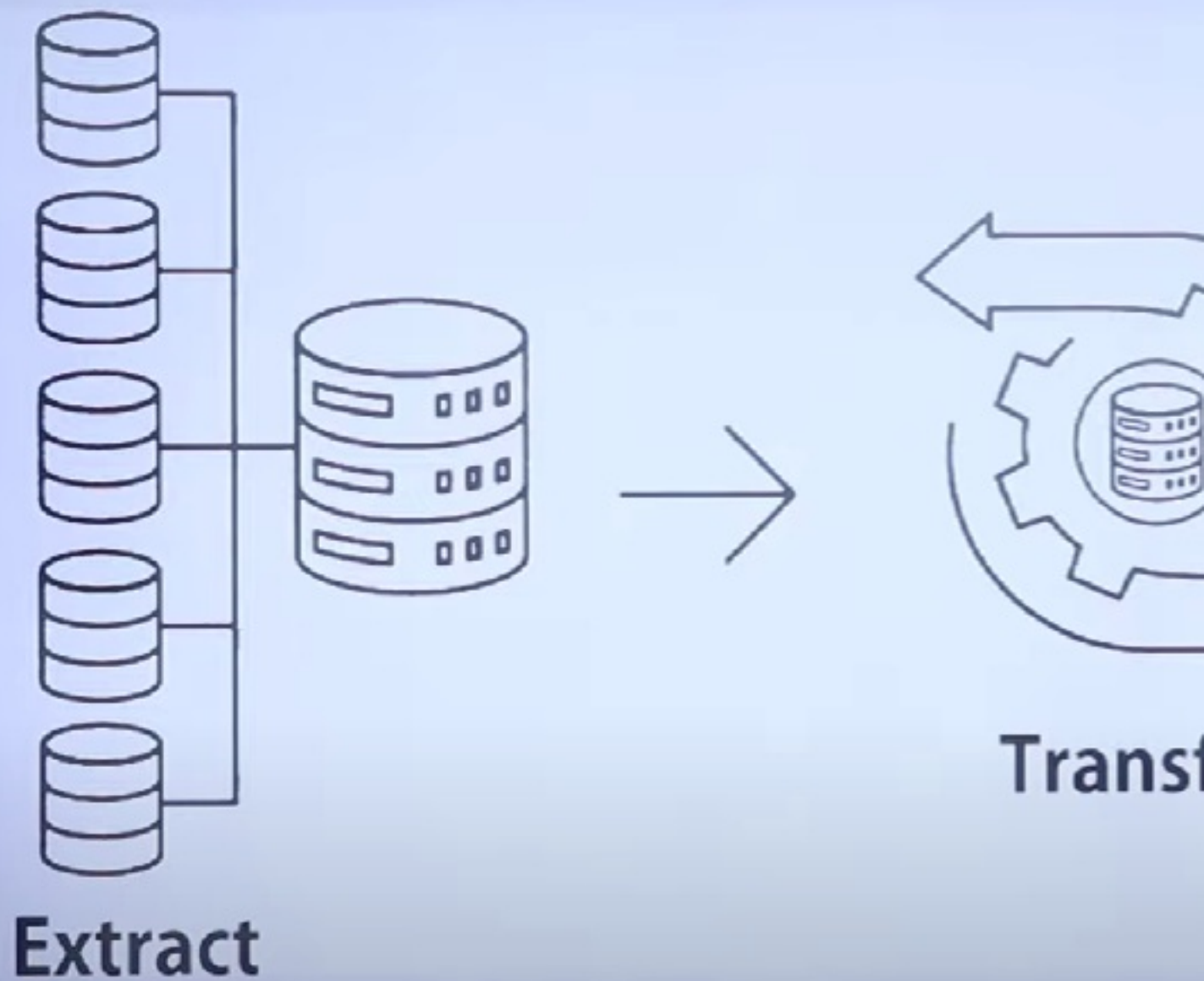
Exraction : eg : multiple reliance stores generate bills , reliance extracts data from all these sources so as to make future predictions

Extracting data from different sources to integrate them at one place. Removing all the errors like duplicacy , inconsistency

Transform : missing values , null values , duplicate values , error removal and replacements, formatting , cleaning, joins , filtering , aggregation, converting data to a proper model

Here we transform data into star or snowflake model

Load : loading data in warehouses



Data models : star schema and snowflake schema

These tell how data is stored structurally in the warehouse

Star Schema

- A star schema is a type of data model used in data warehousing to represent data in a star-like structure.
- The fact table in a star schema contains quantitative measures.
- The dimension tables in a star schema contain descriptive attributes of the measures in the fact table.



Fact table stores quantitative info and dimension tables use its description

Product Dimension

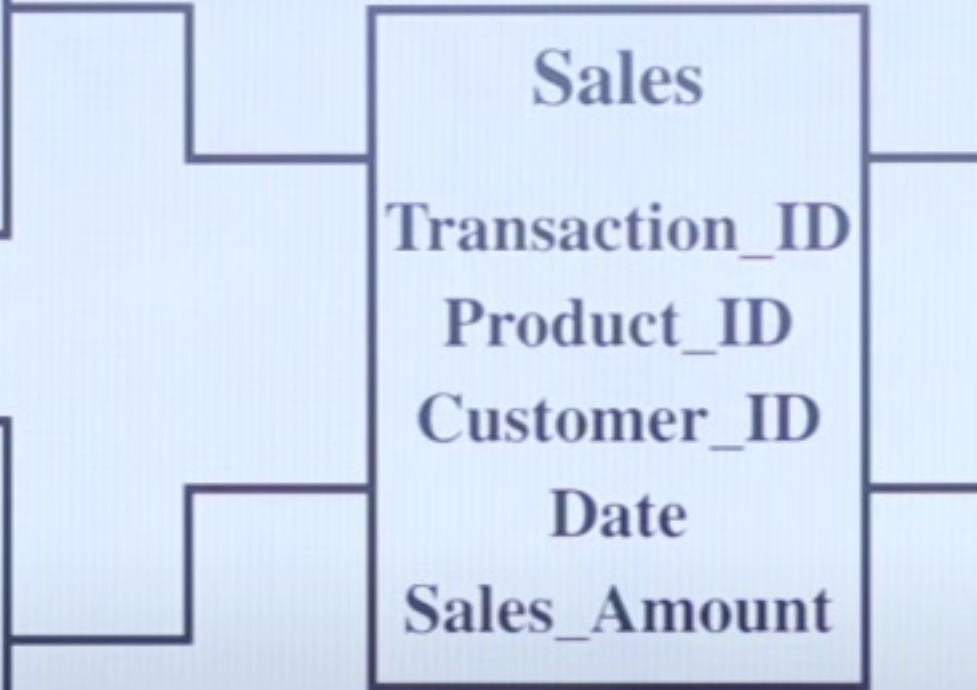
Product_ID
Product_name
Category
Brand

Employee Dimension

Emp_ID
Emp_name
Mobile_no.
Department

Sales

Transaction_ID
Product_ID
Customer_ID
Date
Sales_Amount



Transaction ID	Date	Product ID	Customer ID
101	2023-07-21	P1	C1
102	2023-07-21	P2	C2
103	2023-07-22	P3	C1
...

Product ID	Product Name	Cate
P1	Laptop	Elect
P2	Smart Watch	Elect
...

Customer ID	Customer Name	Loca
C1	Varun	Char
C2	Ravi	Delh

Advantages of the Star Sch

- Query Performance
- Simplicity
- Flexibility
- Scalability



*Star schema is also known as star join

* star schema is type of data modelling technique to represent data

*star schema is multi dimensional data modelling to organize data

*star schema contains only one fact table or multiples dimension table

*star schema can be applied data warehouse, database, data mart or other tool for organized data *all the dimension are connected to fact table through the foreign key

*primary key of each dimension table is part of the fact table

Snowflake Schema : normalized star schema

Star:

Product Dimension

Product_ID
Product_name
Category
Brand

Employee Dimension

Employee_ID
Employee_name
Mobile_no
Department

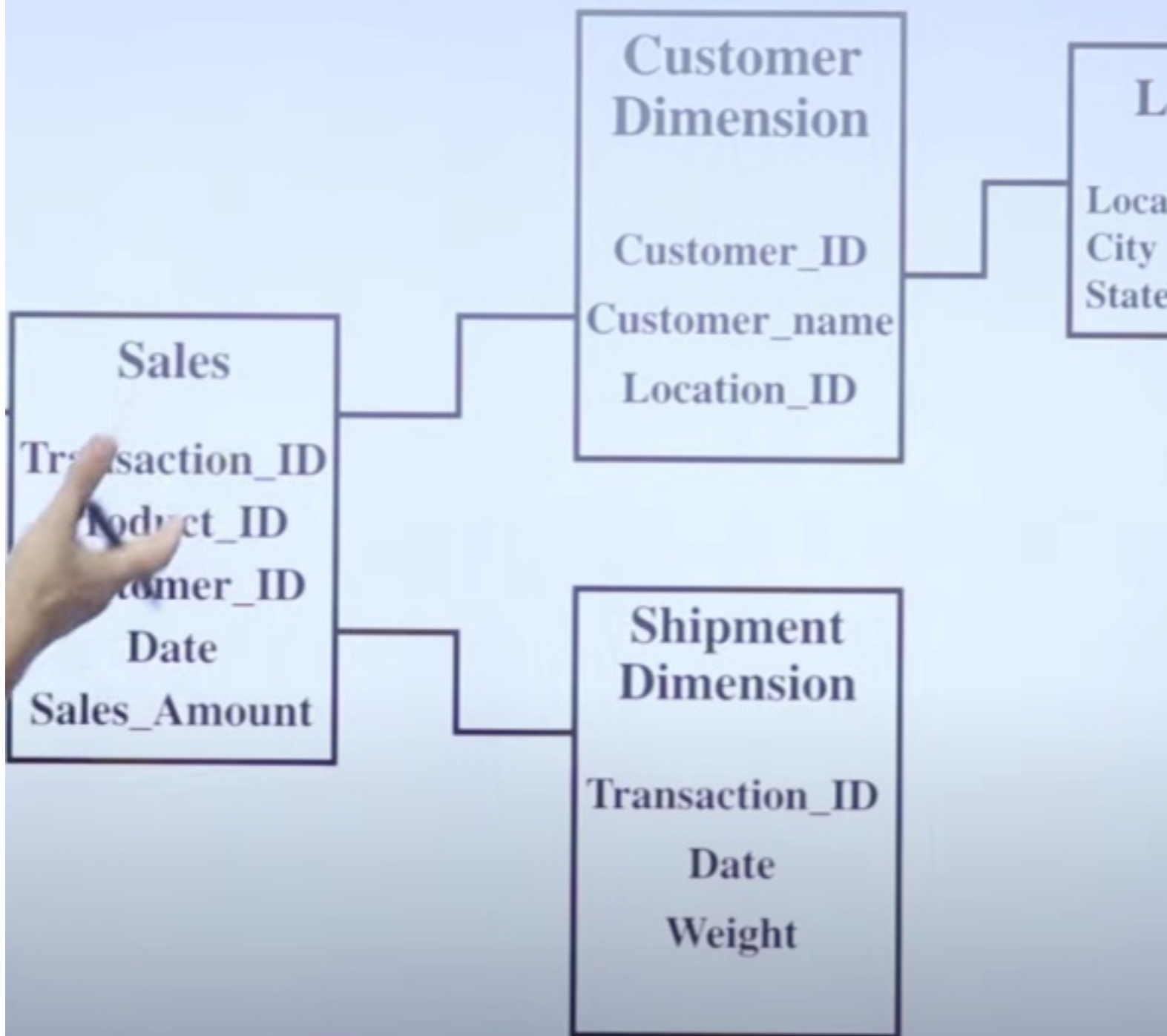
Sales

Transaction_ID
Product_ID
Customer_ID
Date
Sales_Amount

C

T

snowflake:





Customer ID	Customer Name	Location_ID	City
C1	Varun	L1	Jaipur
C2	Ravi	L2	Lucy
C3	Amrit	L1	Jaipur
C4	Nitin	L1	Jaipur
C5	Jacky	L2	Lucy
	Navi	L1	Jaipur

Suppose the city jaipur's name is changed to jayapur then we will have to update many rows

Location_ID	City
L1	Jaipur
L2	<u>Lucknow</u>
L1	<u>Jaipur</u>
L1	<u>Jaipur</u>
L2	<u>Lucknow</u>
L1	<u>Jaipur</u>

So on normalizing ::

:

Customer_ID	Customer_ Name	Location_ID	City	State
C1	Varun	L1	Jaipur	Rajast
C2	Ravi	L2	Lucknow	UP
C3	Amrit	L1	Jaipur	Rajast
C4	Nitin	L1	Jaipur	Rajast
C5	ckv		Lucknow	UP
C6		L1	Jaipur	Rajast

J/K

Customer_ ID	Customer Name	Location_ ID
C1	Varun	L1
C2	Ravi	L2
C3	Amrit	L1
C4	Nitin	L1
C5	Jacky	L2
C6	Navi	L1

<u>Location_ID</u>	City	State
L1	Jaipur	Rajasthan
L2	Lucknow	UP



But major disadvantage : more joins !! more complexity !! slower than star low query performance.

OLTP : transactions ko database mei save krta hai

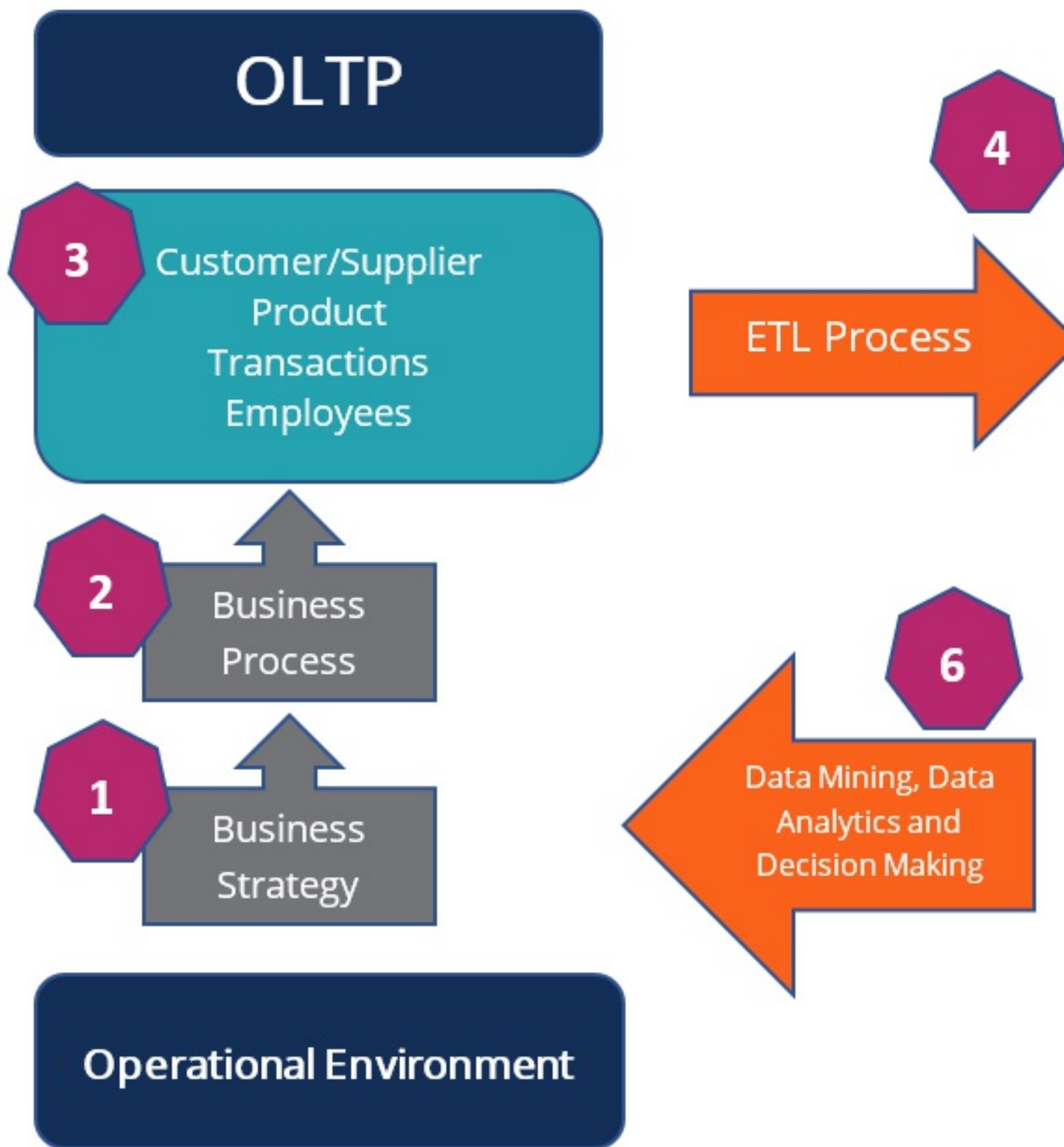
OLTP (Online Transaction

- Retail and E-commerce
- Banking and Finance
- Travel and Hospitality
- Healthcare
- Telecommunications
- Education
- etc.....

Gpay, paytm , amazon are examples of oltp

OLTP Architecture

- **Database Server**
- **Application Server**
- **User Interfaces**



OLTP (Online Transaction Processing)

- Current data
- Day-to-day transactional operations
- Normalized data structures
- Simple Queries
- Used by Front Line Employs, Managers
- Require fast response times
- Data in OLTP systems is updated in real time.
- Oracle, MySQL, SQL Server, DB2