Data Warehouse Design

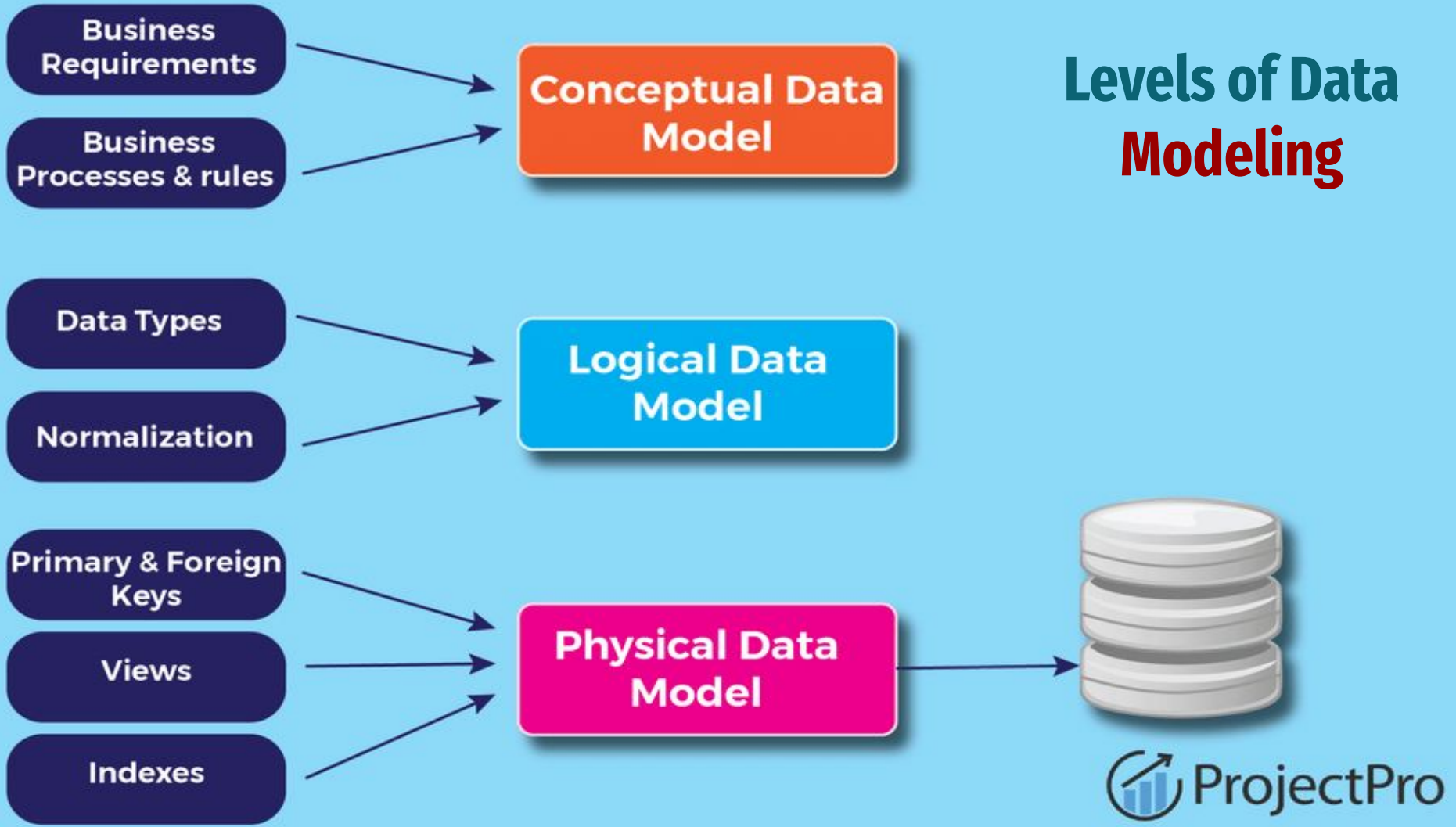
Day 2

ProjectPro

Agenda

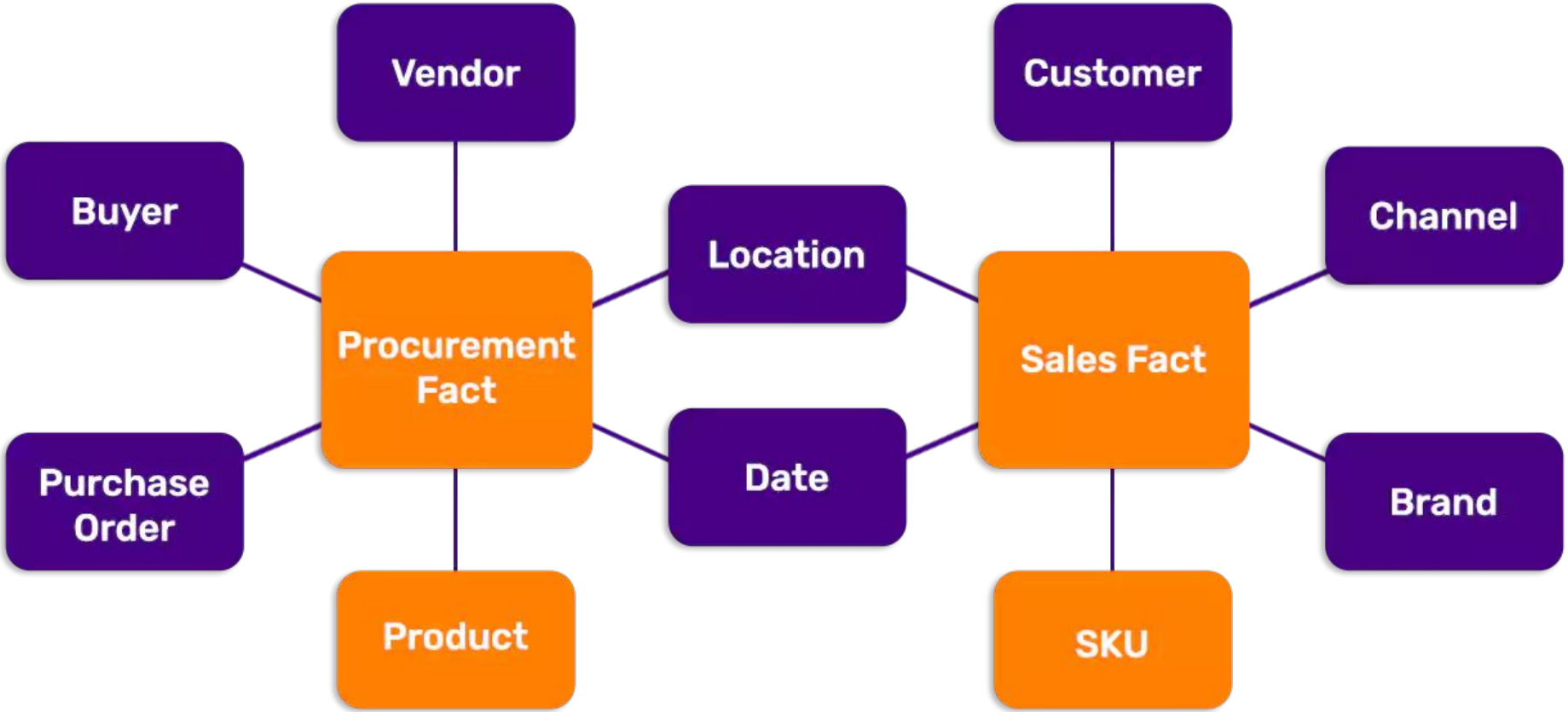
- Levels of Data Modeling
- Types Of Dimensions
- Types of Fact Tables in Data Warehousing
- Types of Measures in Fact Tables “Facts”
- Q&A

Levels of Data Modeling



Types of Dimensions

Conformed Dimensions



Conformed Dimensions

Definition: A dimension shared across multiple fact tables or data marts with the same meaning and values.

Dim_Date

date_key	date	month	year
20250101	2025-01-01	Jan	2025
20250102	2025-01-02	Jan	2025

Fact_Sales (date_key) → Dim_Date
Fact_Purchases (date_key) → Dim_Date

Degenerate Dimension

Definition: A dimension key stored directly in the fact table without a separate dimension table.

Fact_Sales

invoice_no	POS_ID	amount
INV001	101	500
INV002	102	300

Fast Changing Dimension

Definition: A dimension attribute that changes very frequently, often stored in a mini-dimension

Dim_Email

email_id	email_address	effective_date	expiry_date
1	old@mail.com	2025-01-01	2025-02-01
2	new@mail.com	2025-02-01	NULL

Heterogeneous Dimension

Definition: A dimension table that stores multiple different entity types

Dim_Party

party_id | party_type | name
---
1 | Customer | Ahmed
2 | Supplier | East Co.

Junk Dimension

Definition: Combines miscellaneous flags and codes into one dimension table

Dim_Junk

junk_id	is_returned	payment_method_flag
1	Y	C
2	N	P

Multi-Valued Dimension

Definition: Entity can have multiple values for the same attribute.

Dim_Customer_Phone

customer_id	phone_number
1	0101111121111
1	0102222222

Outrigger Dimension Dimension

Definition: A dimension linked to another dimension instead of directly to the fact table

Dim_Customer

customer_id	name	region_id_Fk
1	Ali	10

Dim_Region

region_id	region_name
10	Cairo

Role Playing Dimension

Definition: The same dimension table is used multiple times in the same fact table for different roles

Fact_Sales

order_date_key | ship_date_key | amount

20250101 | 20250105 | 500

Assume we have a Dim Date.

Shrunken Dimension

Definition: A reduced version of a dimension, usually for a smaller data mart

Dim_Date_Shrunk

date_key	date	year
20230101	2023-01-01	2023

Assume we have a Dim Date.

Slowly Changing Dimension (SCD0,1,2,3,...)

Definition:A dimension that changes slowly.

Dim_Date_Shrunk (SDC2)

customer_id	name	address	start_date	end_date
1	Ali	Old Street	2020-01-01	2023-01-01
1	Ali	New Street	2023-01-01	NULL

SDC 0: no change
SDC1 : Updated
SDC2 : Historical
SDC 3 : Add Column
SDC 4 : Add table

Snowflaked Dimension

Definition: A normalized dimension broken into multiple related tables

Dim_Customer

```
customer_id | name  | city_id
-----------------------------
1           | Ali   | 101
```

Dim_City

```
city_id | city_name | country_id
-------------------------------
101     | Cairo     | 20
```

Dim_Country

```
country_id | country_name
-------------------------
20         | Egypt
```

Swappable Dimension

Definition: A dimension that can be replaced with another similar one without affecting the facts

Dim_Currency (Source A)

```
currency_id | currency_name
--------------------------
1           | USD
```

Dim_Currency (Source B)

```
currency_id | currency_name
--------------------------
1           | USD
```

Types of Fact Tables in Data Warehousing

Transactional Fact

Definition: Records each business transaction at the most detailed (granular) level

Fact_Sales

date_key	customer_id	product_id	quantity	amount
20250101	101	201	2	500
20250101	102	202	1	300

Periodic Snapshot Fact

Definition: Captures aggregated measures at regular time intervals (daily, monthly, etc.)

Fact_Daily_Sales

date_key	total_orders	total_amount
20250101	150	75000
20250102	120	65000

Accumulating Snapshot Fact

Definition: Tracks the progress of a process over time, updating as milestones are reached.

Fact_Order_Process

order_id	order_date	ship_date	delivery_date	status
1001	20250101	20250103	20250105	Delivered
1002	20250102	20250104	NULL	Shipped

Factless Fact

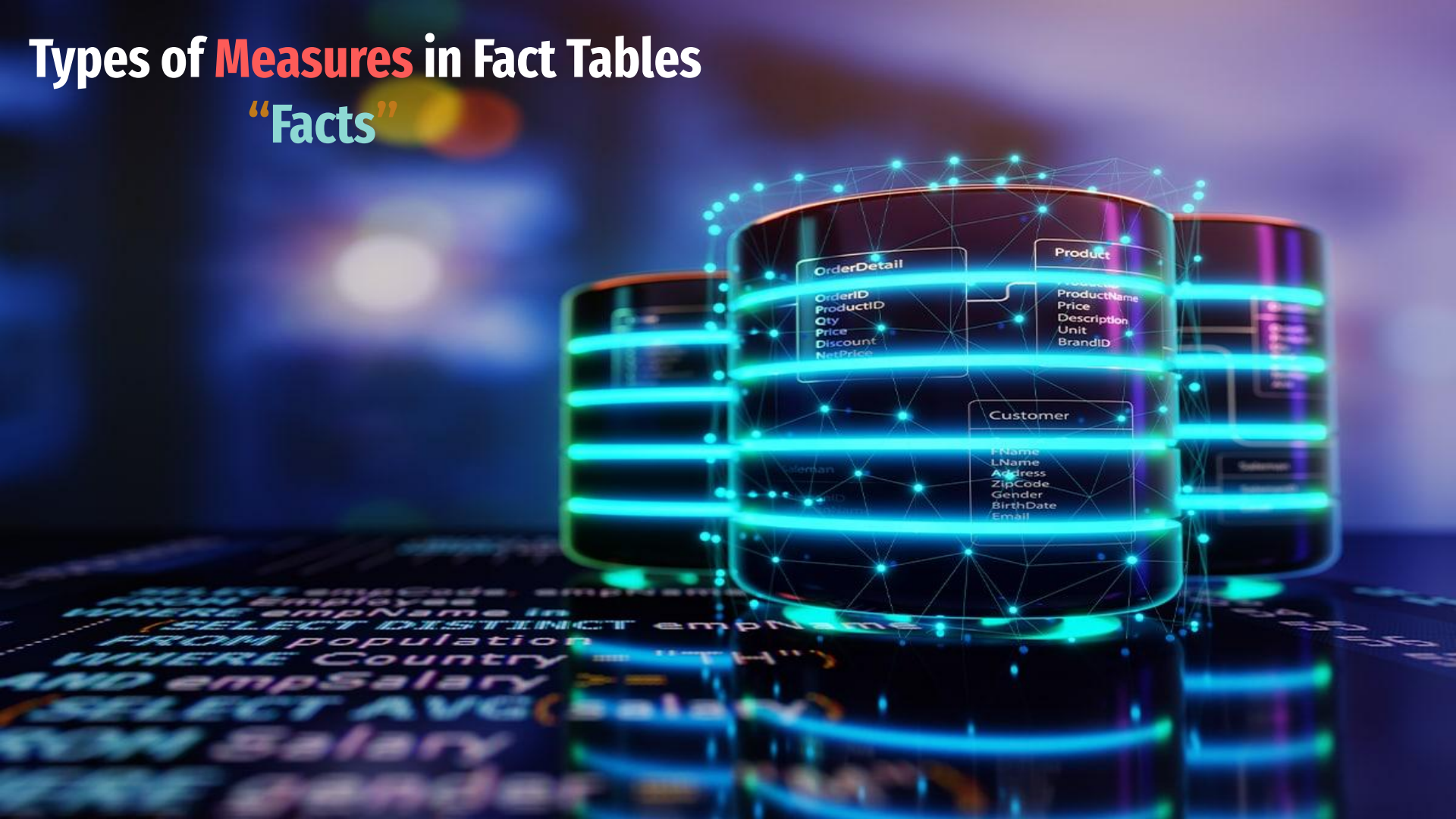
Definition: Contains no numeric measures, only records the occurrence of events or relationships

Fact_Student_Attendance

date_key	student_id	course_id
20250101	201	301
20250101	202	302

Types of Measures in Fact Tables

"Facts"



Additive Facts

Definition: Can be summed across all dimensions.

Example: Sales amount can be added across time, product, and region.

Fact_Sales

date_key	product_id	region_id	amount
20250101	101	1	500
20250101	102	1	300

Semi-Additive Facts

Definition: Can be summed across some dimensions but not all (e.g., across products but not time).

Example: Account balance can be summed across accounts but not over days.

Fact_Account_Balance

```
date_key  | account_id | balance
-----------------------------
20250101  | 201        | 1000
20250102  | 201        | 1200
```

Non-Additive Facts

Definition: Cannot be summed across any dimension; need a different aggregation like average or ratio.

Example: Profit margin percentage.

Fact_Sales

date_key	product_id	margin_pct
20250101	101	25%
20250101	102	30%

Derived Facts

Definition: Calculated from other facts using formulas.

Example: Profit = Sales Amount – Cost.

Fact_Sales

date_key	product_id	amount	cost	profit
20250101	101	500	300	200

Factless Facts

Definition: Facts without numeric measures, only indicating event occurrence.

Example: Student attendance.

Fact_Student_Attendance

date_key	student_id	course_id
20250101	201	301

Textual Facts

Definition: Non-numeric descriptive facts, usually rare in fact tables.

Example: Comments or status text.

Fact_Customer_Feedback

date_key	customer_id	feedback_text
20250101	101	"Fast delivery"


Any
QUESTIONS?