Business Intelligence and Data Warehousing (ANL408)

- By Sabarish Nair

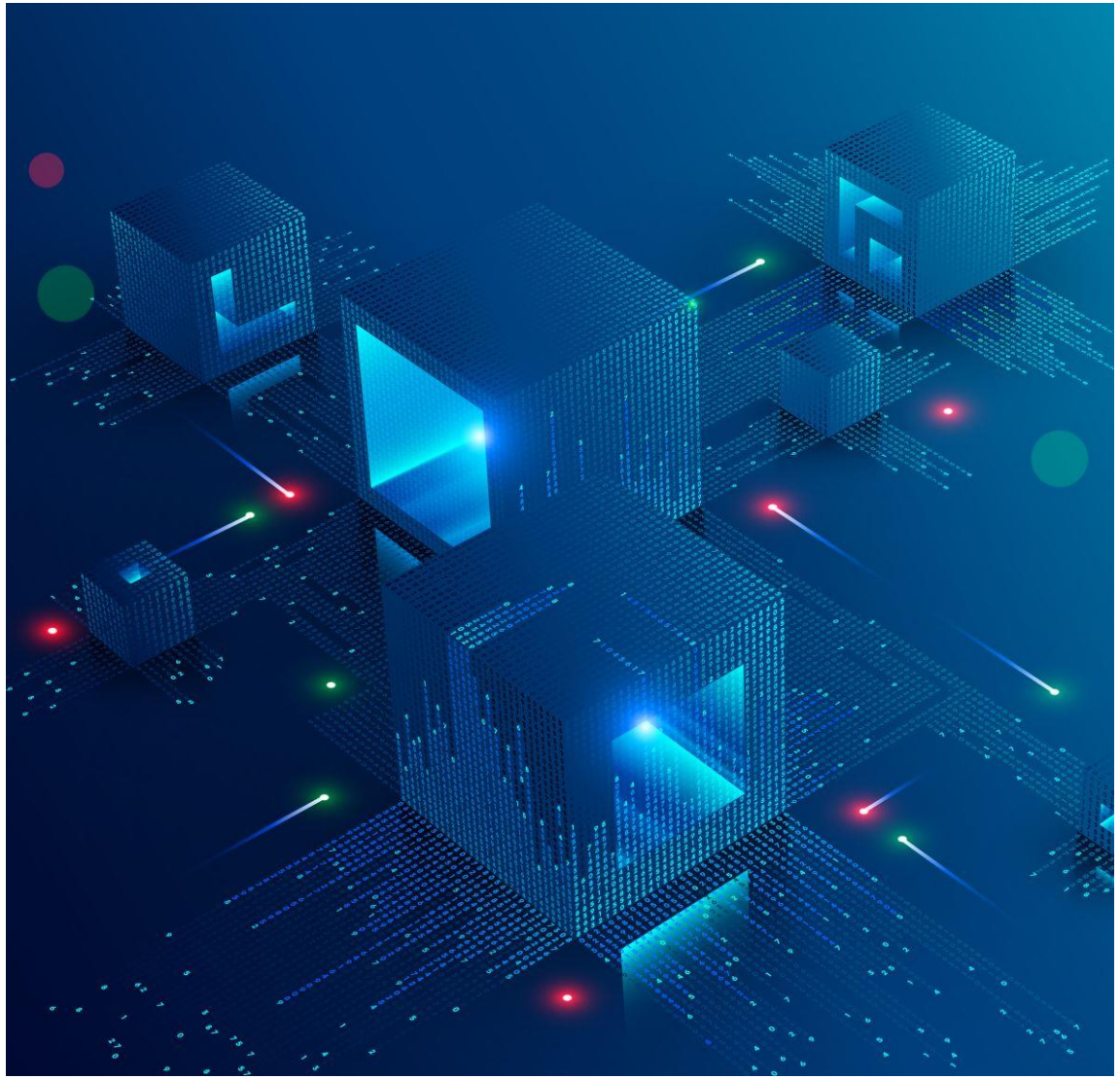
Recap from last week....

- ELT Basics
- Extract
- Extracting Types
- Full Load
- Incremental Load
- Transform
- Consolidate Data
- Reshape Data
- Different Transformations
- UPDATE/INSERT/DELETE
- Vendor ETL Tools
- Choosing ETL Tools
- ETL Tools Evaluation Matrix
- ETL vs ELT

Online Analytical Processing (OLAP)

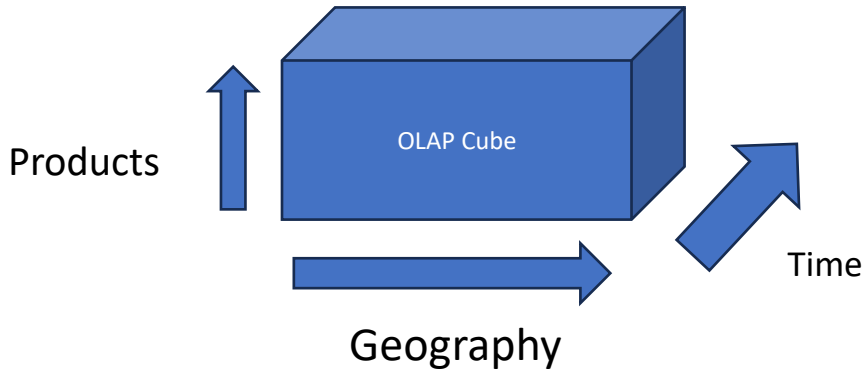
- *Technology used in Business Intelligence (BI) to enable interactive analysis of large volumes of data from multiple perspectives.*

- *Category of technology that enables users to gain insight into their data in a fast, interactive and easy to use manner.*



OLAP features

- Multidimensional Information viewing capabilities – Browse and Navigation (Slice and Dice)
- Calculation Intensive Capabilities
- Time Intelligence – Time Series Analysis



OLAP Architecture

- MOLAP
- ROLAP
- HOLAP
- DOLAP

MOLAP

- Multidimensional OLAP
- Data stored in multidimensional cubes or arrays
- Data is pre-aggregated and stored in a proprietary format
- Provide fast query performance for multidimensional analysis due to pre-computed aggregations
- Example: Microsoft Analysis Services

ROLAP

- Relational OLAP.
- Stores data in relational databases, typically using a star or snowflake schema.
- Extended DBMS to store and manage warehouse data.
- Can leverage functionalities inherent in the relational database.
- Example: IBM Cognos

HOLAP

- Hybrid OLAP
- MOLAP + ROLAP
- MOLAP = Store summary data in Multidimensional format
- ROLAP = Store detailed data in relational format
- Example : Microsoft SQL Server Analysis Services

DOLAP

- Desktop OLAP
- Lightweight OLAP tools designed for individual users or small workgroups.
- They typically run on desktop computers and do not require a dedicated server.
- Low-cost tools.
- Extract Relational data into local (or server) Multidimensional cubes
- Example : Tableau Desktop

Practical

Data cleansing

Define fact and dimension tables

```sql
-- Step 1: Dump all data from staging to temporary tables
CREATE TABLE temp_tblproduct AS
SELECT * FROM "Staging"."tbl_ProductsData";

-- Verify of the data exists in the temporary table
SELECT * FROM temp_tblproduct ORDER BY customer_id;
```

Query 2: Result

Data Output | Messages | Notifications

	sales_id integer	date_sales date	product_id integer	product_name character varying	category character varying	price numeric	customer_id integer	customer_name character varying	city character varying	country character varying
1	1	2022-01-01	1	Laptop	Electronics	1200	1	John Doe	New York	USA
2	2	2022-01-02	2	Smartphone	Electronics	800	2	Jane Smith	London	UK
3	3	2022-01-03	3	Speaker	Electronics	[null]	3	[null]	Paris	France
4	4	2022-01-04	4	TV	Electronics	1500	[null]	[null]	Berlin	Germany
5	5	2022-01-05	5	Tablet	Electronics	-100	5	Michael Lee	Tokyo	Japan
6	6	2022-01-06	[null]	Mouse	Electronics	25	6	[null]	Sydney	Australia
7	7	2022-01-07	7	Camera	Electronics	750	7	Emily Wang	Rome	Italy
8	8	2022-01-08	8	Headphones	Electronics	[null]	[null]	Madrid	Spain	[null]
9	9	2022-01-09	9	Keyboard	Electronics	30	9	Alex Johnson	Madrid	Spain

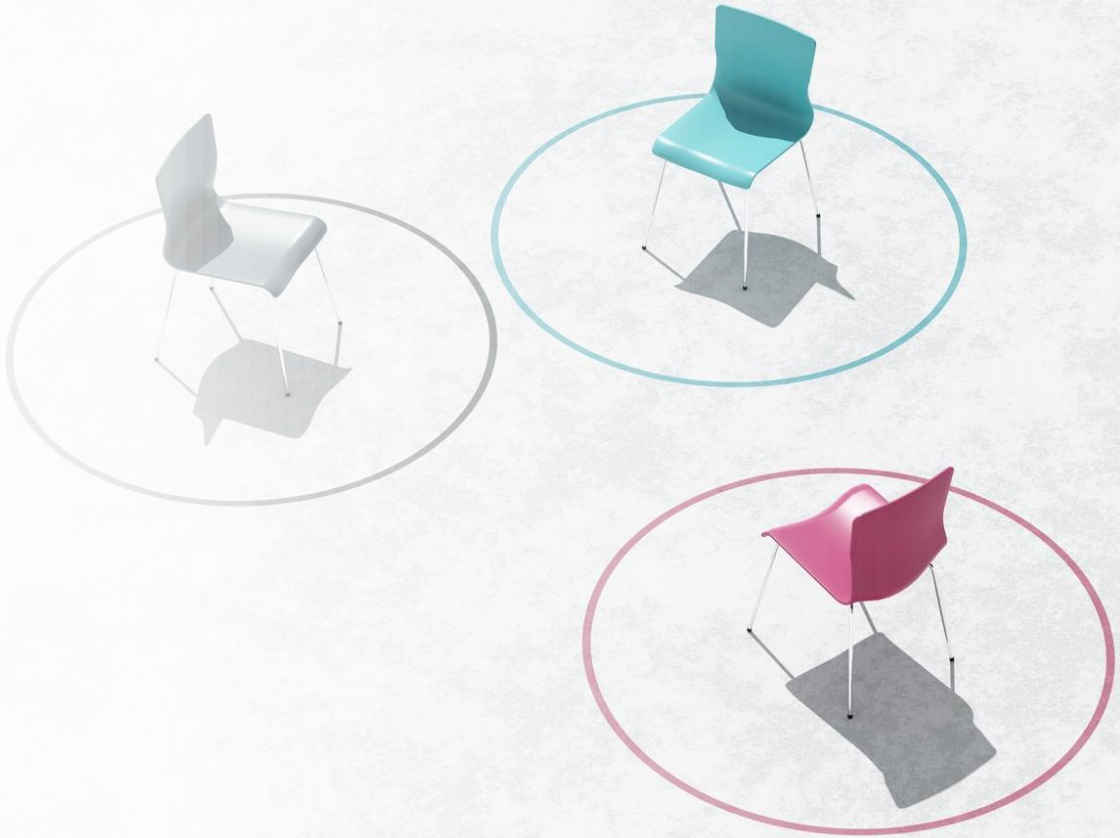
Create a Temporary Table

UPDATE Query

UPDATE <table_name>

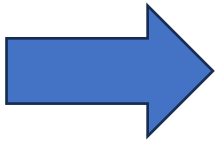
SET <column_name> = <value>

WHERE <column_name> = <value>



Example: Update Query

Emp_ID	Emp_Name	Emp_City
1	Sab	Dublin
2	Viv	Cork



Emp_ID	Emp_Name	Emp_City
1	Sab	Limerick
2	Viv	Cork

```
UPDATE tbl_Employee
SET Emp_City = 'Limerick'
WHERE Emp_ID = 1
```

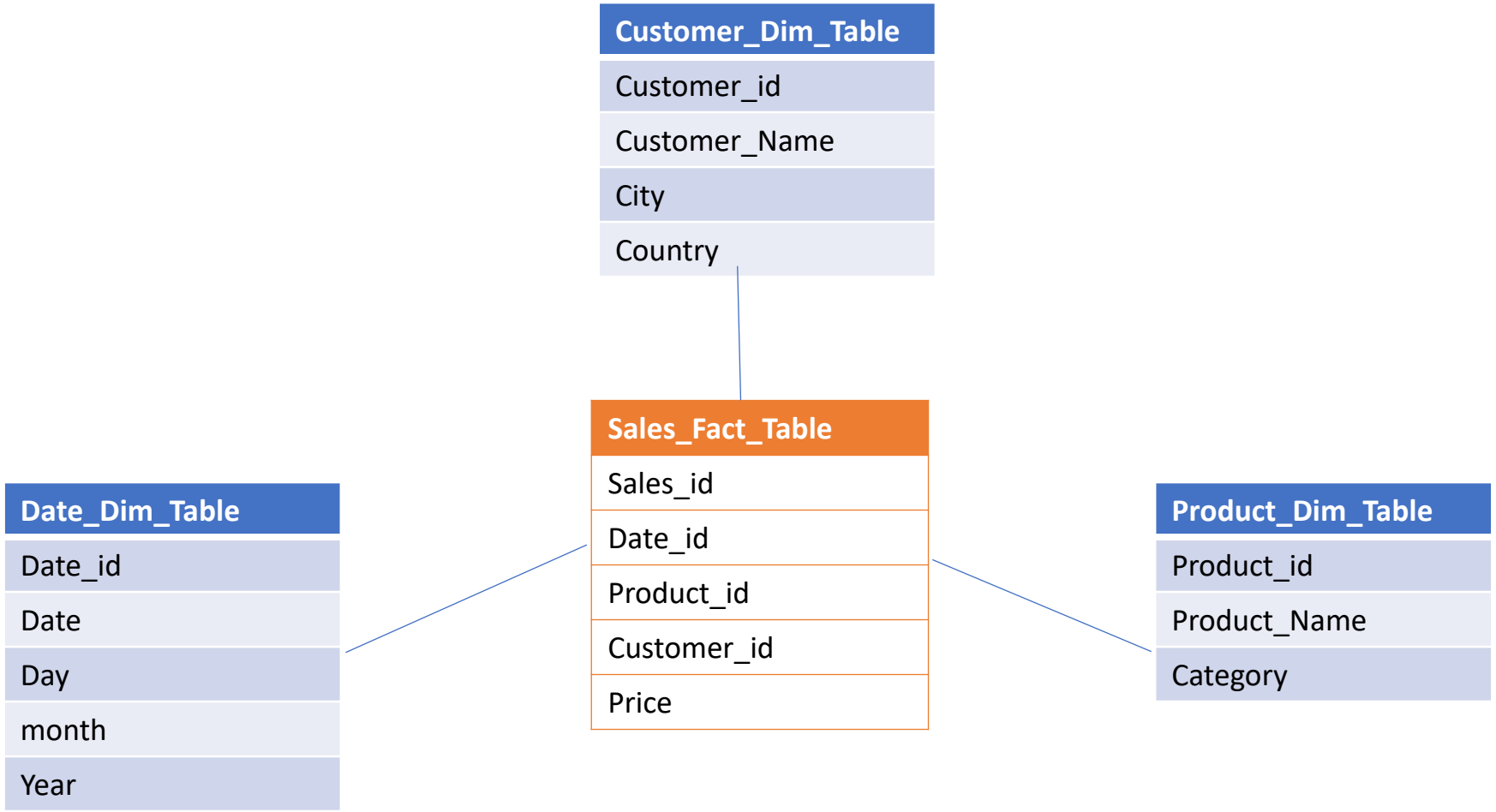
Cleansing the textual data

```sql
--UPDATE THE product name to UNKNOWN where it is NULL
UPDATE temp_tblproduct
SET product_name = 'UNKNOWN'
WHERE product_name IS NULL OR UPPER(product_name) = 'NULL';

--UPDATE THE customer_name to UNKNOWN where it is NULL
UPDATE temp_tblproduct
SET customer_name = 'UNKNOWN'
WHERE customer_name IS NULL OR UPPER(customer_name) = 'NULL';

--UPDATE THE country to UNKNOWN where it is NULL
UPDATE temp_tblproduct
SET country = 'UNKNOWN'
WHERE country IS NULL OR UPPER(country) = 'NULL';
```

Cleansing the numerical data

```
-- Correcting negative price values by taking absolute values
UPDATE temp_tblproduct
SET price = ABS(price)
WHERE price < 0;

-- Set price as 0 where it is NULL
UPDATE temp_tblproduct
SET price = 0
WHERE price IS NULL;
```

Identify the fact and dimension tables




Thank you