



PRESENTATION

Data Modeling for eCommerce Data
Warehouse

Dina Mostafa

1/15/2025

PROJECT OVERVIEW

Objective:

Build a data warehouse model to serve business queries for an eCommerce platform.

Main Phases

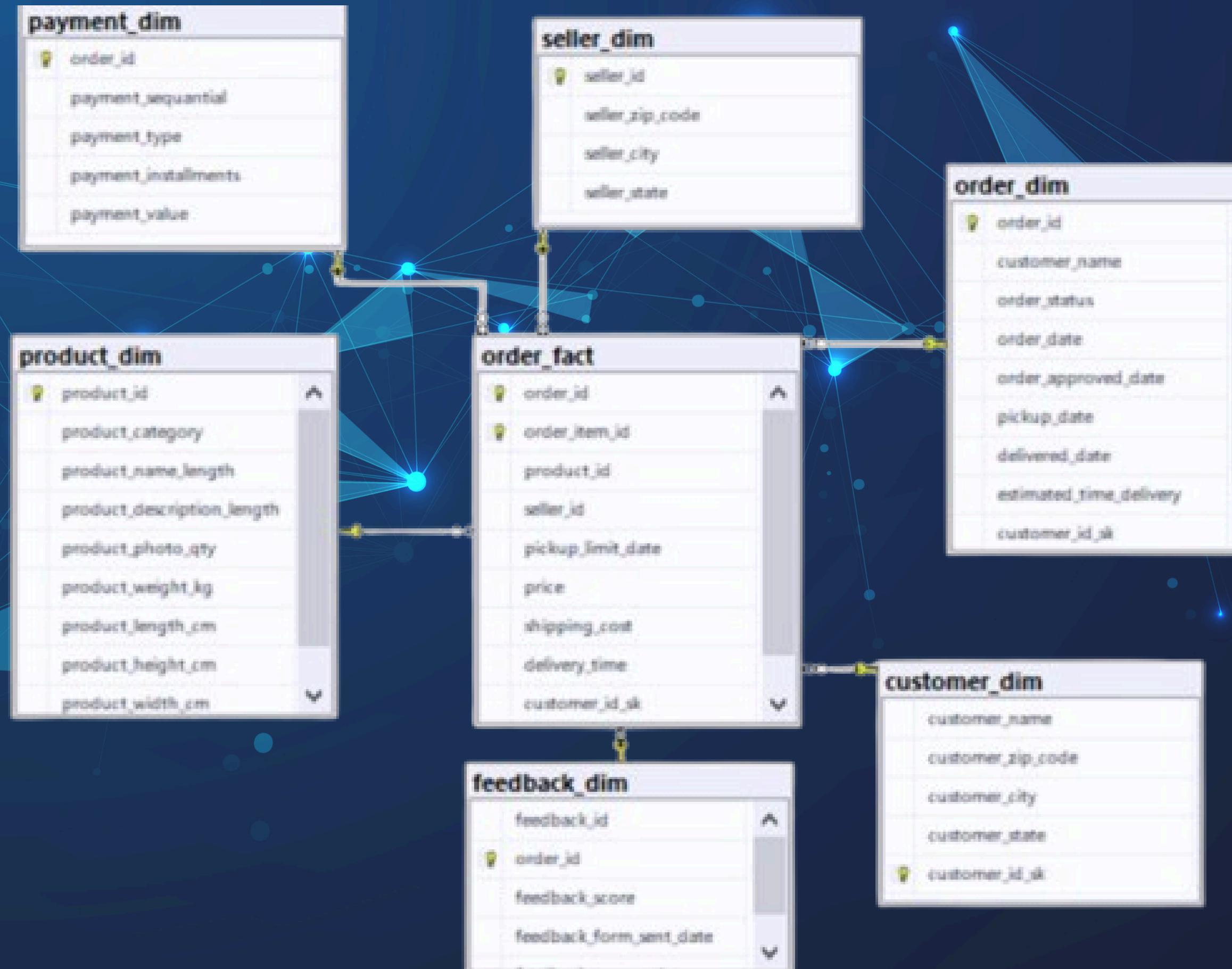
- Data Warehouse Modeling
- ETL Pipeline Development
- Analytical Query Execution

DATA SOURCE:

7 TABLES

- User Dataset
- Seller Dataset
- Product Dataset
- Order Dataset
- Payment Dataset
- Order Item Dataset
- Feedback Dataset

DATA MODELING:



ETL PROCESS

Python Pipeline

- Reading Data from Excel Files
- Data Cleaning and Transformation
- Loading Data into the Data Warehouse

EXTRACT

Data Extracted from 7 exel files, then has been converted to pandas data frame

TRANSFORM

- Missing Values: Left as it is as it has meaning.
- Date Columns : Converting to datetime .
- PK problem : Add a surrogate key to *customer_dim*.
- *Data Model* : Merge *order fact* to all dimensions.
- *Final Fact* : Extract only the columns of the fact table with the new values on MERGED data frame.

LOAD

Data Storage:

- Loaded into a Data Warehouse schema consisting of:
 - Fact Tables: Order Fact Table
 - Dimension Tables: Customer, Order, Product, Seller, feedback, and Payment Dimensions
- Loading Method: Batch loading using pandas DataFrame to SQL export



A complex network graph is displayed against a dark blue background. The graph consists of numerous glowing cyan and white nodes (dots) connected by thin, translucent cyan lines forming a web of triangles. The nodes are more concentrated in the center and become sparser towards the edges. Some nodes are significantly brighter than others, creating a sense of depth and highlighting certain connections.

THANK YOU