**Why Does Olist Use a Constellation Schema?**

The **Olist e-commerce platform** handles a **complex and multi-faceted dataset** involving orders, payments, product details, customer reviews, and delivery logistics. This kind of rich, interconnected data makes the **Constellation Schema** a practical and scalable choice for its data warehouse.

**What Is a Constellation Schema?**

A Constellation Schema, also called a Galaxy Schema, is a data warehouse schema that supports multiple fact tables sharing common dimension tables.

Key Characteristics:

* Multiple Fact Tables: Each represents a major business process or transaction type (e.g., sales, payments).
* Shared Dimension Tables: Common reference tables like date, customer, or product that link to multiple facts.
* Flexible Design: Ideal for large businesses with interrelated datasets across different functions.

In this model design:

* Fact Tables: FCT\_ORDER\_ITEMS, FCT\_PAYMENTS, FCT\_REVIEWS
* Shared Dimensions: DIM\_ORDERS, DIM\_DATE, DIM\_CUSTOMERS, LKP\_STATUS\_DESC

**📊 DIMENSION TABLES**

DIM\_CUSTOMERS

Contains customer-related attributes.

* pk\_customer\_sid: Surrogate key for the customer (Primary Key)
* customer\_unique\_id: 1 customer\_unique\_id can have multiple pk\_customer\_sid. the column customer\_unique\_id serves as a **logical identifier** for a customer across multiple transactions, while pk\_customer\_sid acts as a **surrogate key** specific to each occurrence of a customer in the system.

DIM\_DATE

Centralized calendar dimension, used for multiple date references across the fact tables.

* pk\_date\_sid: Surrogate key for date (Primary Key)
* Includes attributes such as year, month, day\_name, and is\_weekday
* Facilitates time-based analysis like trends and seasonality

**📦 FACT TABLES**

DIM\_ORDERS (Fact Table with degenerate dimensions)

Captures the complete lifecycle and financial summary of orders.

* pk\_order\_sid: Primary Key
* Linked to customers, date dimensions, and status lookup tables
* Derived metrics:
  + total\_payment,
  + order\_amt,
  + freight\_amt,
  + balance\_amt,
  + payment status

FCT\_ORDER\_ITEMS

Line-item details per order; each row represents a product in an order.

* Composite primary key: pk\_order\_id + pk\_order\_item\_id
* References product metadata like category and dimensions
* Includes price, freight\_value, and seller information

FCT\_PAYMENTS

Payment records associated with orders.

* Composite primary key: pk\_order\_id + pk\_payment\_sequential
* Supports multiple payment entries per order
* Includes payment\_type, installments, payment\_value

FCT\_REVIEWS

Captures customer feedback and review activity.

* Composite primary key: pk\_order\_id + pk\_review\_id
* A single order can have multiple reviews, and a single review may be associated with multiple orders.
* Includes review scores, messages, and timestamps

**📘LOOKUP TABLE**

LKP\_STATUS\_DESC

The LKP\_STATUS\_DESC table serves as a **lookup or reference table** that enriches the raw order\_status and payment\_status values with **contextual meanings**, **business classifications**, and **summary metrics**. It provides standardized explanations and categorizations that help analysts and business users better understand and interpret the current state of an order in the Olist e-commerce platform.

This table acts as a **bridge between technical status codes and business-level interpretations**, enhancing reporting, debugging, and business intelligence activities.

**📊 Key Columns and Their Roles**

| **Column** | **Purpose** |
| --- | --- |
| order\_status | The lifecycle status of an order (e.g., *Delivered*, *Canceled*) |
| payment\_status | The status of the associated payment (e.g., *Completed*, *In Progress*) |
| record\_count | The number of records/orders with this exact status combination |
| profit\_lost | A high-level classification of the financial outcome (*Profit*, *Lost*, *In Progress*, *Unknown*) |
| status\_description | Human-readable explanation of the status combination for clarity |

Use Cases Enabled by This Model

* Sales analysis: Product pricing, freight costs, and category distribution
* Customer behavior tracking: Repeat orders, payment methods, satisfaction (via reviews)
* Operational efficiency: Delivery timelines, shipping constraints, geographic segmentation
* Financial KPIs: Revenue, profit/loss mapping (via LKP\_STATUS\_DESC), and installment tracking