

Technical Specifications for Data Engineer

Prepared by: Abhilash Mishra, Senior Data Analyst

Objective:

The purpose of this document is to provide clear technical specifications for developing and integrating a portion of the data model that addresses key business requirements. These specifications are based on an in-depth analysis of customer and order data, ensuring that the data structure supports efficient querying and reporting.

Table Structure:

The new data model will integrate multiple data sources to create a comprehensive view of customer transactions. The following table outlines the key columns and their attributes that need to be implemented:

Column Name	Source Table	Description	Data Type
Customer_ID	Customer	Unique identifier for the customer	Integer
First	Customer	Customer's first name	String
Last	Customer	Customer's last name	String
Age	Customer	Customer's age	Integer
Country	Customer	Customer's country	String
Order_ID	Order	Unique identifier for the order	Integer
Item	Order	Name of the item purchased	String
Amount	Order	Total amount spent on the order	Decimal
Shipping_ID	Shipping	Unique identifier for the shipping	Integer
Status	Shipping	Shipping status (Delivered/Pending)	String

Data Integrity:

- Primary Key Consistency:** Ensure that the Customer_ID is consistently applied across all related tables, and that it serves as the primary key for joining these tables.
 - Foreign Key Mapping:** Validate that Order_ID and Shipping_ID correctly link back to their respective Customer_ID records.
 - Data Anomalies:** Address any discrepancies such as missing order information or mismatches between order and shipping data.
-

Business Logic:

- **Order Items:** The Item field must remain in the Orders table to provide detailed insights into product purchases. Despite discussions about separating this data, it will remain in place to avoid disrupting existing reports.
 - **Order Amount:** The Amount field should accurately reflect the total cost of the order, inclusive of taxes and discounts.
-

Performance Optimization:

- **Query Optimization:** The table should be designed with efficient querying in mind, particularly for common business reports like customer demographics and purchasing trends.
-

Data Validation:

- **Null Values:** Ensure there are no null values in any primary or foreign key fields.
 - **Range Checks:** Validate that numeric fields such as Amount and Age fall within expected ranges.
 - **Data Consistency:** Maintain consistency across all tables before merging them into the final joined table.
-

Priority: High

This table structure will serve as the foundation for our analytical work, including the reporting and insights required by the business. These include the total amounts spent by customers, shipping status analysis, and popular products by customer demographics.

Note: This specification serves as the blueprint for developing and optimizing our data models to meet the organization's analytical and operational needs. Your role as a Data Engineer involves implementing this structure while adhering to the outlined performance and data integrity guidelines.