

Data Dictionary: Ekart Performance Data

This document provides a detailed catalog of all data assets within the Ekart logistics data pipeline, covering the Bronze, Silver, and Gold layers.

Bronze Layer

Schema: bronze

The Bronze layer contains raw, unprocessed data ingested directly from the source. All columns are stored as TEXT to prevent load failures and preserve the original state of the data, including errors and inconsistencies.

Table: bronze."Customers"

Stores raw customer registration information.

Column Name	Data Type	Description	Example Raw Data
customer_id	TEXT	Unique identifier for the customer.	'101'
customer_name	TEXT	Full name, with inconsistent casing and whitespace.	' ARJUN KUMAR ', 'arjun kumar'
email	TEXT	Email address, may be invalid or contain spaces.	'arjun.kumar 101@example.com', 'invalid-email'
delivery_address	TEXT	Primary shipping address, may be missing.	'42, Main St, Chennai', 'N/A'
signup_date	TEXT	Registration date in various string formats.	'2024-08-25'

Table: bronze."Orders"

Stores raw data for each customer purchase.

Column Name	Data Type	Description	Example Raw Data
order_id	TEXT	Unique identifier for the order.	'5001'
customer_id	TEXT	Foreign key to Customers, may be NULL.	'101', 'NULL'
order_date	TEXT	Date of purchase in multiple formats.	'2025-01-15', '08/25/2025'
order_total	TEXT	Order value as a string with currency symbols.	'₹1450.75', '\$980.00'

Table: bronze."Shipments"

Stores raw data for the fulfillment of each order.

Column Name	Data Type	Description	Example Raw Data
shipment_id	TEXT	Unique identifier for the shipment.	'7001'
order_id	TEXT	Foreign key to Orders.	'5001'
driver_id	TEXT	Foreign key to Drivers, may be a placeholder.	'201', '-1'
vehicle_id	TEXT	Foreign key to Vehicles.	'55'
dispatch_date	TEXT	Timestamp string, may be logically incorrect.	'2025-01-15 10:00:00'
delivery_date	TEXT	Timestamp string, may be before dispatch_date.	'2025-01-14 18:00:00'

status	TEXT	Delivery status with inconsistent values.	'Delivered', 'delivered', 'D'
--------	------	---	-------------------------------

Table: bronze."Drivers"

Stores a raw log of driver activity, containing many duplicates.

Column Name	Data Type	Description	Example Raw Data
driver_id	TEXT	Identifier for the driver (not unique in this table).	'201'
driver_name	TEXT	Full name of the driver.	'Sai Kumar'
contact_number	TEXT	Phone number in various formats.	'+91 9876543210', '(987) 654-3210'

Table: bronze."Vehicles"

Stores a raw log of vehicle usage, containing many duplicates.

Column Name	Data Type	Description	Example Raw Data
vehicle_id	TEXT	Identifier for the vehicle (not unique).	'55'
license_plate	TEXT	License plate, may have extra whitespace.	' TN 07 AX 1234 '
vehicle_type	TEXT	Vehicle type with inconsistent naming.	'Van', 'van', 'Cargo Van'

Silver Layer

Schema: silver

The Silver layer is the single source of truth. Data has been cleaned, standardized, de-duplicated, and validated. Tables have enforced data types and logical integrity.

Table: silver."Customers"

A clean, unique master list of all valid customers.

Column Name	Data Type	Description	Transformations Applied
customer_id	INTEGER	Primary Key. Unique customer identifier.	Cast from TEXT.
customer_name	VARCHAR	Name with whitespace trimmed and proper-cased.	INITCAP(TRIM())
email	VARCHAR	Cleaned, lowercase email with no spaces.	LOWER(TRIM(REPLACE()))
delivery_address	VARCHAR	The customer's shipping address.	Passed through.
signup_date	DATE	Standardized registration date.	TO_DATE()

Table: silver."Orders"

A clean table of all valid orders placed by known customers.

Column Name	Data Type	Description	Transformations Applied
order_id	INTEGER	Primary Key. Unique order identifier.	Cast from TEXT.
customer_id	INTEGER	Foreign Key to silver."Customers".	Cast from TEXT, validated via JOIN.

order_date	DATE	Standardized order date.	COALESCE(TO_DATE())
order_total	DECIMAL	Order value as a numeric type.	Parsed and cast from TEXT.

Table: silver."Shipments"

A clean table of all valid shipments with logical integrity.

Column Name	Data Type	Description	Transformations Applied
shipment_id	INTEGER	Primary Key. Unique shipment identifier.	Cast from TEXT.
order_id	INTEGER	Foreign Key to silver."Orders".	Cast from TEXT, validated via JOIN.
driver_id	INTEGER	Foreign Key to silver."Drivers".	Cast from TEXT, validated via JOIN.
vehicle_id	INTEGER	Foreign Key to silver."Vehicles".	Cast from TEXT, validated via JOIN.
dispatch_date	TIMESTAMP	Standardized dispatch timestamp.	Cast from TEXT.
delivery_date	TIMESTAMP	Standardized delivery timestamp.	Cast from TEXT.
status	VARCHAR	Standardized delivery status.	Normalized via CASE statement.

Table: silver."Drivers"

A de-duplicated master list of all unique drivers.

Column Name	Data Type	Description	Transformations Applied
-------------	-----------	-------------	-------------------------

driver_id	INTEGER	Primary Key. Unique driver identifier.	De-duplicated, cast from TEXT.
driver_name	VARCHAR	Cleaned driver name.	TRIM()
driver_display_name	VARCHAR	A unique name for charts (Name + ID).	Concatenated from name and ID.
contact_number	VARCHAR	Cleaned phone number with only digits.	REGEXP_REPLACE()

Table: silver."Vehicles"

A de-duplicated master list of all unique vehicles.

Column Name	Data Type	Description	Transformations Applied
vehicle_id	INTEGER	Primary Key. Unique vehicle identifier.	De-duplicated, cast from TEXT.
license_plate	VARCHAR	Cleaned license plate.	TRIM()
vehicle_type	VARCHAR	Standardized vehicle type.	Normalized via CASE statement.

Gold Layer

Schema: gold

The Gold layer contains tables optimized for business intelligence and analytics. These tables are often aggregated or denormalized to make querying fast and simple.

Table: gold."Monthly_Driver_Performance"

An aggregate table summarizing driver performance metrics for each month.

Column Name	Data Type	Description
driver_id	INTEGER	The unique ID for the driver.
driver_display_name	VARCHAR	The unique, chart-friendly name of the driver.
performance_year	DOUBLE	The year of the performance summary.
performance_month	DOUBLE	The month of the performance summary.
total_shipments	BIGINT	The total number of shipments assigned to the driver that month.
on_time_shipments	BIGINT	The count of shipments delivered within 72 hours.
avg_delivery_hours	DOUBLE	The driver's average delivery time in hours for that month.

Table: gold."Vehicle_Utilization_Summary"

An aggregate table summarizing vehicle usage by type for each month.

Column Name	Data Type	Description
vehicle_type	VARCHAR	The standardized type of vehicle.
usage_year	DOUBLE	The year of the usage summary.
usage_month	DOUBLE	The month of the usage summary.
total_shipments	BIGINT	The total number of shipments handled by this vehicle type.

avg_delivery_hours	DOUBLE	The average delivery time in hours for this vehicle type.
--------------------	--------	---

Table: gold."Full_Shipment_Details"

A wide, denormalized table joining all silver tables. Optimized for BI dashboarding.

Column Name	Data Type	Description
shipment_id	INTEGER	The unique ID for the shipment.
shipment_status	VARCHAR	The clean status of the shipment.
dispatch_date	TIMESTAMP	When the package left the warehouse.
delivery_date	TIMESTAMP	When the package was delivered.
delivery_hours	DOUBLE	Total delivery duration in hours.
order_id	INTEGER	The original customer order ID.
order_date	DATE	The original customer order date.
order_total	DECIMAL	The monetary value of the order.
customer_id	INTEGER	The unique ID of the customer.
customer_name	VARCHAR	The name of the customer.
delivery_address	VARCHAR	The full delivery address.
customer_city	VARCHAR	The city extracted from the address.

driver_id	INTEGER	The ID of the assigned driver.
driver_display_name	VARCHAR	The unique, chart-friendly name of the driver.
vehicle_id	INTEGER	The ID of the vehicle used.
license_plate	VARCHAR	The license plate of the vehicle.
vehicle_type	VARCHAR	The type of vehicle used.