Business Template

SUPPLY CHAIN DATA WAREHOUSE

Supply Chain Analytics Platform

CONTENTS

1 BUSINESS DESCRIPTION

1.1 BUSINESS BACKGROUND

Supply Chain Analytics Platform business operates in the global supply chain management and e-commerce analytics domain, focusing on comprehensive end-to-end supply chain optimization. The organization manages complex multi-channel retail operations spanning across various product categories, customer segments, and geographical markets.

The business model encompasses:

- **Multi-channel retail operations** across diverse product categories including consumer electronics, clothing, sports equipment, and household goods
- **Global market presence** with customers spanning multiple countries and regions
- **Complex logistics network** involving multiple warehouses, carriers, and delivery modes
- **Customer-centric approach** with segmented service offerings based on customer profiles and purchasing behavior
- **Data-driven decision making** leveraging big data analytics for supply chain optimization

The organization handles substantial transaction volumes (1 million+ records) across two primary operational systems, requiring sophisticated data management and analytics capabilities to maintain competitive advantage in the rapidly evolving e-commerce landscape.

1.2 PROBLEMS BECAUSE OF POOR DATA MANAGEMENT

The organization faces several critical challenges due to fragmented data management across multiple operational systems:

Data Silos and Integration Issues:

- Order Management System (OMS) and Logistics Management System (LMS) operate independently, creating data inconsistencies
- Lack of unified customer view across sales and shipping operations

Operational Inefficiencies:

- Manual data reconciliation processes consuming significant resources
- Delayed reporting and analytics due to complex data extraction procedures
- Inability to perform real-time cross-system analysis
- Incomplete visibility into end-to-end supply chain performance

Decision-Making Limitations:

- Fragmented insights preventing comprehensive business intelligence
- Inability to correlate sales performance with logistics efficiency
- Limited capacity for predictive analytics and forecasting
- Delayed identification of supply chain bottlenecks and customer satisfaction issues

Compliance and Data Quality Concerns:

- Inconsistent data quality standards across systems
- Difficulty in maintaining data lineage and audit trails
- Challenges in ensuring data privacy and security across multiple platforms
- Compliance reporting complications due to scattered data sources

1.3 BENEFITS FROM IMPLEMENTING A DATA WAREHOUSE

Unified Data Architecture:

- **Single Source of Truth:** Consolidation of OMS and LMS data into a centralized repository ensuring data consistency and accuracy
- **Improved Data Quality:** Implementation of standardized data cleansing, validation, and transformation processes
- **Enhanced Data Governance:** Centralized data management with proper access controls, security measures, and compliance frameworks

Operational Excellence:

- **Real-time Analytics:** Enable near real-time reporting and dashboard capabilities for operational monitoring

Strategic Business Value:

- **Supply Chain Optimization:** Comprehensive analytics enabling identification of bottlenecks, optimization opportunities, and cost reduction strategies
- **Predictive Capabilities:** Advanced analytics foundation for demand forecasting, inventory optimization, and risk management
- **Scalability:** Robust architecture supporting business growth and additional data source integration

1.4 DATASETS DESCRIPTION

1.4.1 **Dataset Overview**

Original Dataset: DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS

- **Total Records:** 1,000,000 (expanded from original dataset)
- **Comprehensive Coverage:** End-to-end supply chain data including orders, customers, products, shipping, and delivery information

System Division Strategy: The original dataset has been strategically divided into two operational systems to reflect real-world enterprise architecture.

OMS Dataset (Order Management System) The OMS dataset focuses on transactional and commercial aspects of order processing. It captures critical sales metrics including transaction amounts, customer segments, product pricing, and order completion status. This dataset serves as the primary source for revenue analysis, customer behavior tracking, and sales performance monitoring across different product categories and customer segments.

LMS Dataset (Logistics Management System) The LMS dataset concentrates on fulfillment and delivery operations. It tracks shipment logistics from warehouse to customer destination, including shipping modes, delivery performance, carrier management, and on-time delivery metrics. This dataset enables supply chain optimization, delivery performance analysis, and logistics cost management.

In a table bellow represented structure of a created datasets for **OMS** and **Logistic Management LMS** driven from extended dataset. In a form of example record for OrderID: 77202_V0 .

Color definition represent what level of transformation has been performed on each record to deliver a dataset.

Preserved from original Transformed/Mapped

New Business field Calculated Field

Attribute Category	Original Dataset	OMS Dataset	LMS Dataset
TRANSACTION KEYS			
Primary Key	Order Id: 77202_V0	OrderID: 77202_V0	ShipmentID: 77202_V0
System Key	N/A	TransactionSK: 1	TransactionSK: 600001
Source System	N/A	SourceSystem: OMS	SourceSystem:LMS
CUSTOMER DATA			
Customer ID	Customer Id: 20755	CustomerID: 20755	CustomerID: 20755
Customer Name	Cally Holloway	Cally Holloway	Not Included
Customer Email, password	XXXX	cally.holloway@mail.com	Not Included

Attribute Category	Original Dataset	OMS Dataset	LMS Dataset
Customer Gender	Not Available	CustomerGender: F	Not Included
Customer Segment	Consumer	Consumer	Not Included
PRODUCT DATA			
Product ID	Product Name: Smart watch	ProductID: Smart watch	ProductID: Smart watch
Product Category	Sporting Goods	Sporting Goods	Not Included
Department	Fitness	Fitness	Not Included
Product Brand	Not Available	ProductBrand: SportSync	Not Included
Product Status	ProductStatus	ProductStatus: Active	Not Included
FINANCIAL DATA			
Sales Amount	Sales: 327.75	SalesAmount: 327.75	Not Included
Order Total	Order Item Total: 314.6400146	OrderTotal: 314.6400146	Not Included
Unit Price	Order Item Product Price: 327.75	UnitPrice: 327.75	Not Included
Unit Cost	Not Available	UnitCost: 235.55	Not Included
Total Cost	Not Available	TotalCost: 235.55	Not Included
Profit	Order Profit Per Order: 92.20	*not in data set but can be Calculated: 92.20	Not Included
Sales per customer	314.6400146	Not Included	Not Included
LOGISTICS DATA			
Quantity	Order Item Quantity: 1	Quantity: 1	ShippedQuantity: 1

Attribute Category	Original Dataset	OMS Dataset	LMS Dataset
Shipping Mode	Standard Class	Not Included	ShippingMode: Standard
Delivery Status	Advance shipping	Not Included	DeliveryStatus: Delivered
Shipping Cost	Not Available	Not Included	ShippingCost: 6.01
Late delivery risk	0	Not Included	Not Included
GEOGRAPHIC DATA			
Customer Location	Caguas, Puerto Rico, PR, 725	Not Included	DestinationCity: Caguas
			DestinationState: PR
			DestinationCountry: Puerto Rico
latitude/ longitude	18.251453/-66.037055	Not Included	Not Included
Order Region	Asia, Java	MarketSegment: Pacific Asia	Not Included
TEMPORAL DATA			
Order Date	2024-06-08	OrderDate: 2024-06-08	OrderDate: 2024-06-08
		OrderYear:2024	
		OrderMonth: 6	
		OrderQuarter:2	
		OrderStatus:Completed	
		OrderDayOfWeek: 2	
		OrderWeekOfYear: 25	
Ship Date	2024-06-13	Not Included	ShipDate: 2024-06-13
Delivery Date	Not Available	Not Included	DeliveryDate: 2024-06-21
PERFORMANCE METRICS			

Attribute Category	Original Dataset	OMS Dataset	LMS Dataset
Shipping Days (Scheduled)	4 days	Not Included	Not Included
Shipping Days (Real)	5 days	Not Included	OrderToShipDays: 5
Delivery Days	Not Available	Not Included	DeliveryDays: 8
On-Time Delivery	Not Available	Not Included	OnTimeDelivery: False
OPERATIONAL DATA			
Payment Method	Type:Online	PaymentMethod: PayPal	Not Included
Benefit per order	91.25	Not Included	Not Included
Sales Rep	Not Available	SalesRepID: 22	Not Included
Warehouse	Not Available	Not Included	WarehouseID: WH_005
Carrier	Not Available	Not Included	CarrierID: USPS

1.4.1.1 Key Relationship Insights

1. **OrderID (OMS) = ShipmentID (LMS) = Order Id (Original)** → Primary linking key between systems
2. **CustomerID** is consistent across all three datasets → Reliable customer tracking
3. **ProductID/Product Name** enables product-level analysis across systems
4. **TransactionSK** is system-specific (OMS: 1-600K, LMS: 600K+) → Not a natural business key

Conncections	Transformations	Added Business Values	Calculated
OrderId=ShipmentID	'Standart Class' → 'Standart'	Profitability $92.5 gross	OrderDate elements
CustomerID preserved	'Advance shipping' → 'Delivered'	Cost tracking $235.55	DeliveryDates
ProductID preserved	Real delivery dates (8 ans 3 original)	Performance check (OnTime,WarehouseID,carrierID)	Shipping cost
Date comsistency maintained			Unit cost / totalcost

	OrderToShipDays 5 (calculated)	System context : OMS & LMS view 'CustomerYearOfBirth'Different age distributions by customer segment 'ProductBrand' Generated,Product category determines brand family	Product Status CustomerGender generated based on mapping common male/female names matching + statistical distribution ProductStatus' # ✅ NEW - Business logic: Order date + price determines product lifecycle status

UnitCost Analysis:

Definition: The cost to acquire/produce one unit of a product

Business Purpose: Calculate gross margin per item (**UnitPrice** - **UnitCost**)

Sample Data: UnitCost = 235.55, UnitPrice = 327.75

Gross Margin = $327.75 - $235.55 = $92.20 per unit (28.1% margin)

TotalCost Calculation:

Definition: Total cost for all units in the order line

Business Purpose: Calculate total COGS for the transaction

Formula: UnitCost × Quantity

Sample Verification: $235.55 × 1 = 235.55 ✓ (matches TotalCost)

Profit Calculations:

Order Profit = SalesAmount - TotalCost

Sample: $327.75 - $235.55 = $92.20

Profit Margin = (**Order Profit** / SalesAmount) × 100

Sample: ($92.20 / $327.75) × 100 = 28.1%

1.4.1.2 Key Differences Between Datasets

Data Volume Distribution:

- **OMS (60%):** Higher transaction volume reflecting comprehensive order processing

- **LMS (40%):** Lower volume as represents shipment consolidation (multiple order lines may ship together)

Functional Focus:

- **OMS:** Revenue and profit analysis with detailed cost breakdowns
- **LMS:** Operational efficiency and customer satisfaction tracking

Granularity Levels:

- **OMS:** Order line item level with financial details
- **LMS:** Shipment level with logistics performance metrics

1.4.2 Dimensional Entity Identification

1.4.2.1 OMS Dataset - Identified Dimensions

1. Customer Dimension

- Primary Attributes: CustomerID, Customer Segment (Consumer, Corporate) , CustomerFirstName,CustomerLastName,CustomerGender,CustomerYearOfBirth,Customer Email
- Business Purpose: Customer segmentation, behavior analysis, and targeted marketing
- Granularity: Individual customer level
- Related Measures: Sales per customer, order frequency

2. Product Dimension

- Primary Attributes: ProductID, Product Category, Department, Unit Price ,ProductName,ProductBrand,ProductStatus
- Business Purpose: Product performance analysis, category management, pricing strategies
- Granularity: Individual product level
- Related Measures: Sales amount, profit margins, quantity sold

3. Time Dimension

- Primary Attributes: OrderDate, OrderYear, OrderMonth, OrderQuarter, OrderDayOfWeek, OrderWeekOfYear
- Business Purpose: Temporal analysis, seasonality patterns, trending
- Granularity: Daily level with hierarchical rollups
- Related Measures: All financial and operational metrics across time periods

4. Sales Representative Dimension

- Primary Attributes: SalesRepID

- Business Purpose: Sales performance tracking, commission calculations, territory management
- Granularity: Individual sales representative level
- Related Measures: Sales amount, order count, customer acquisition

5. Order Status Dimension

- Primary Attributes: OrderStatus (Completed, Pending, Cancelled)
- Business Purpose: Order lifecycle tracking, completion rate analysis
- Granularity: Status category level
- Related Measures: Order completion rates, processing times
-

6. Payment Method Dimension

- Primary Attributes: PaymentMethod (PayPal, Credit Card, Debit Card)
- Business Purpose: Payment preference analysis, transaction cost optimization
- Granularity: Payment type level
- Related Measures: Transaction volume by payment type, processing costs

1.4.2.2 LMS Dataset - Identified Dimensions

1. Geographic Dimension

- Primary Attributes: DestinationCity, DestinationState, DestinationCountry
- Business Purpose: Geographic performance analysis, regional optimization, shipping cost analysis
- Granularity: City level with state/country hierarchies
- Related Measures: Delivery performance, shipping costs, regional volumes

2. Shipping Mode Dimension

- Primary Attributes: ShippingMode (Standard, Express, Overnight)
- Business Purpose: Service level analysis, cost optimization, customer preference tracking
- Granularity: Shipping service level
- Related Measures: Shipping costs, delivery times, customer satisfaction

3. Delivery Status Dimension

- Primary Attributes: DeliveryStatus (Delivered, In Transit, Failed, Returned)
- Business Purpose: Delivery performance monitoring, issue identification
- Granularity: Delivery outcome level
- Related Measures: Delivery success rates, return rates

4. Warehouse Dimension

- Primary Attributes: WarehouseID, Warehouse Location
- Business Purpose: Warehouse performance analysis, capacity planning, cost allocation
- Granularity: Individual warehouse level
- Related Measures: Shipping volumes, processing times, operational costs

5. Carrier Dimension

- Primary Attributes: CarrierID (UPS, USPS, FedEx)
- Business Purpose: Carrier performance comparison, contract optimization, service quality tracking
- Granularity: Individual carrier level
- Related Measures: Delivery performance, shipping costs, reliability metrics

6. Time Dimension (Logistics)

- Primary Attributes: OrderDate, ShipDate, DeliveryDate
- Business Purpose: Logistics timeline analysis, performance trending
- Granularity: Daily level with calculated intervals
- Related Measures: Order-to-ship days, delivery days, on-time performance

1.4.3 Sales and Cost Measures Definition

1.4.3.1 Sales Measures

Primary Sales Metrics:

- **SalesAmount**: Individual transaction value
  - Aggregation: SUM, AVG across dimensions
  - Business Use: Revenue tracking, performance analysis
- **OrderTotal**: Complete order value including discounts
  - Aggregation: SUM, COUNT for order analysis
  - Business Use: Order size analysis, discount impact assessment
- **Quantity**: Units sold per transaction
  - Aggregation: SUM for volume analysis
  - Business Use: Demand planning, inventory management

Derived Sales Metrics:

- **Average Order Value (AOV)**: OrderTotal / Order Count
- **Sales per Customer**: Total Sales / Unique Customers
- **Revenue per Product Category**: Sales grouped by product category

1.4.3.2 Cost Measures

Primary Cost Metrics:

- **UnitCost**: Cost of goods sold per unit
    - Aggregation: Weighted average by quantity
    - Business Use: Margin analysis, pricing decisions
- **TotalCost**: Complete cost of goods for transaction
    - Aggregation: SUM for total cost analysis
    - Business Use: Profitability assessment, cost control
- **ShippingCost**: Logistics expense per shipment
    - Aggregation: SUM, AVG across shipping dimensions
    - Business Use: Logistics cost optimization, carrier comparison

Derived Cost Metrics:

- **Gross Profit**: SalesAmount - TotalCost
- **Profit Margin**: (SalesAmount - TotalCost) / SalesAmount
- **Total Shipping Cost**: Sum of all shipping expenses
- **Cost per Unit Shipped**: TotalCost / Quantity

1.4.3.3 Performance Measures

Operational Efficiency:

- **OrderToShipDays**: Time from order to shipment
- **DeliveryDays**: Time from shipment to delivery
- **OnTimeDelivery**: Boolean metric for delivery performance
- **Delivery Success Rate**: Percentage of successful deliveries

Quality Metrics:

- **Customer Satisfaction Score**: Based on delivery performance
- **Order Fulfillment Rate**: Percentage of orders completed successfully

1.4.4 **Sum-up**

While both datasets share common identifiers (TransactionSK, CustomerID, ProductID), they serve complementary purposes: **OMS** provides the commercial foundation for understanding what was sold and to whom, while **LMS** delivers operational insights into how orders are fulfilled and delivered. The **OMS** dataset emphasizes financial metrics and customer relationships, whereas **LMS** prioritizes operational efficiency and delivery performance. Together, they provide end-to-end visibility from order placement through final delivery.

1.5 GRAIN / DIM / FACT

1.5.1 4-Step Dimensional Modeling Process

1.5.1.1 Step 1: Select Business Process

Selected Business Process: **Integrated Supply Chain Management**

This business process encompasses the complete order-to-delivery lifecycle, combining:

- **Order Management (OMS)**: Customer orders, sales transactions, payment processing
- **Logistics Management (LMS)**: Shipping, delivery, warehouse operations

1.5.1.2 Step 2: Determine the Grain

Selected Grain: **One row per order line item per shipment**

Grain Definition: Each fact table row represents:

- A specific product (ProductID)
- From a specific order (OrderID)
- Shipped in a specific shipment (ShipmentID)
- To a specific customer (CustomerID)
- On specific dates (OrderDate, ShipDate, DeliveryDate)

Why This Grain Was Selected:

1. **Business Requirement Alignment**: Supports both sales analysis (order-level) and logistics analysis (shipment-level) in a unified model
2. **Data Relationship**: OrderID serves as the natural bridge between OMS and LMS datasets
3. **Analytical Flexibility**: Enables analysis at multiple levels - order, shipment, product, customer, time periods
4. **Operational Insight**: Allows tracking of complete order lifecycle from placement to delivery
5. **Performance Measurement**: Supports end-to-end KPIs like order-to-delivery time, fulfillment rates, and customer satisfaction

1.5.1.3 Step 3: Identify Dimensions

Based on the datasets, the following dimensions are identified.

SCD Type Classifications for Dimensions:

Type 2 (Track Historical Chleanges):

- **DIM_PRODUCTS_SCD**: Product catalog with categories and departments

Type 1 (Overwrite Changes):

- **DIM_SALES_REPRESENTATIVES**: Sales rep information

- **DIM_CUSTOMERS**: Customer information and segmentation
- **DIM_GEOGRAPHIES**: Destination locations
- **DIM_WAREHOUSES**: Warehouse facilities
- **DIM_CARRIERS**: Shipping carriers

Type 0 (Static Lookups):

- **DIM_ORDER_STATUSES**: Order lifecycle statuses
- **DIM_PAYMENT_METHODS**: Payment processing methods
- **DIM_SHIPPING_MODES**: Shipping service levels
- **DIM_DELIVERY_STATUSES**: Delivery outcome tracking
- **DIM_TIME_DD**: Daily time dimension for all date attributes

1.5.1.3.1 Dimensions From OMS Dataset:

- **DIM_CUSTOMERS**: Customer information and segmentation

Column Name	Description	Data Type
CUSTOMER_ID	Primary key for customer dimension surrogate key	BIGINT NOT NULL
CUSTOMER_SRC_ID	Source system customer identifier	BIGINT NOT NULL
CUSTOMER_FIRST_NAME	Customer first name	VARCHAR(50) NOT NULL
CUSTOMER_LAST_NAME	Customer last name	VARCHAR(50) NOT NULL
CUSTOMER_GENDER	Customer gender identifier F/M	VARCHAR(1) NOT NULL
CUSTOMER_YEAR_OF_BIRTH	Customer year of birth	INT NOT NULL
CUSTOMER_EMAIL	Customer email name+surname+@mail.com	VARCHAR(100) NOT NULL
CUSTOMER_SEGMENT	Customer business segment classification	VARCHAR(50) NOT NULL
TA_INSERT_DT	Technical insert timestamp	TIMESTAMP NOT NULL
TA_UPDATE_DT	Technical update timestamp	TIMESTAMP NOT NULL

- Example with filled data:

Column Name	Record 1	Record 2
CUSTOMER_ID	1001	1002
CUSTOMER_SRC_ID	20755	19492
CUSTOMER_FIRST_NAME	Cally	Irene

Column Name	Record 1	Record 2
CUSTOMER_LAST_NAME	Holloway	Luna
CUSTOMER_GENDER	F	F
CUSTOMER_YEAR_OF_BIRTH	1987	1993
CUSTOMER_EMAIL	cally.holloway@mail.com	irene.luna@mail.com
CUSTOMER_SEGMENT	Consumer	Consumer
TA_INSERT_DT	2024-06-22 08:30:15	2024-06-22 08:30:15
TA_UPDATE_DT	2024-06-22 08:30:15	2024-06-22 08:30:15

- **DIM_PRODUCTS_SCD**: Product catalog with categories and departments

Column Name	Description	Data Type
PRODUCT_ID	Primary key for product dimension, surrogate key	BIGINT NOT NULL
PRODUCT_SRC_ID	Source system product identifier	BIGINT NOT NULL
PRODUCT_NAME	Name from souse catalogue	VARCHAR(200) NOT NULL
PRODUCT_STATUS	Order date + price determines product lifecycle status	VARCHAR(100) NOT NULL
PRODUCT_BRAND	Product category determines brand family	VARCHAR(100) NOT NULL
PRODUCT_CATEGORY	Product category classification	VARCHAR(100) NOT NULL
DEPARTMENT	Department managing the product	VARCHAR(100) NOT NULL
START_DT	SCD Type 2 start date	DATE NOT NULL
END_DT	SCD Type 2 end date	DATE NOT NULL
TA_INSERT_DT	Technical insert timestamp	TIMESTAMP NOT NULL
TA_UPDATE_DT	Technical update timestamp	TIMESTAMP NOT NULL

- Example with filled data:

Column Name	Record 1
PRODUCT_ID	2001
PRODUCT_SRC_ID	1360
PRODUCT_NAME	Smart watch

Column Name	Record 1
PRODUCT_STATUS	Active
PRODUCT_BRAND	SportSync
PRODUCT_CATEGORY	Sporting Goods
DEPARTMENT	Fitness
START_DT	2023-01-01
END_DT	2023-01-01
TA_INSERT_DT	2024-06-22 08:30:15
TA_UPDATE_DT	2024-06-22 08:30:15

- **DIM_SALES_REPRESENTATIVES**: Sales rep information

Column Name	Description	Data Type
SALES_REP_ID	Primary key for sales representative dimension, surrogate key	BIGINT NOT NULL
SALES_REP_SRC_ID	Source system sales rep identifier	BIGINT NOT NULL
TA_INSERT_DT	Technical insert timestamp	TIMESTAMP NOT NULL
TA_UPDATE_DT	Technical update timestamp	TIMESTAMP

- Example with filled data:

SALES_REP_ID	SALES_REP_SRC_ID	TA_INSERT_DT	TA_UPDATE_DT
30001	49	2024-06-23 08:30:15	2024-06-25 08:30:15
30002	40	2024-06-23 08:30:15	2025-06-18 08:30:15
30003	33	2024-06-23 08:30:15	2025-06-20 08:30:15

- **DIM_ORDER_STATUSES**: Order lifecycle statuses

Column Name	Description	Data Type
ORDER_STATUS_ID	Primary key for order status dimension, surrogate key	BIGINT NOT NULL

Column Name	Description	Data Type
ORDER_STATUS	Order processing status	VARCHAR(50) NOT NULL
TA_INSERT_DT	Technical insert timestamp	TIMESTAMP NOT NULL
TA_UPDATE_DT	Technical update timestamp	TIMESTAMP NOT NULL

- Example with filled data:

ORDER_STATUS_ID	ORDER_STATUS	TA_INSERT_DT	TA_UPDATE_DT
40001	Completed	2024-06-23 08:30:15	2024-06-25 08:30:15
40002	Pending	2024-06-23 08:30:15	2025-06-18 08:30:15
40003	Cancelled	2024-06-23 08:30:15	2025-06-20 08:30:15

- **DIM_PAYMENT_METHODS**: Payment processing methods

Column Name	Description	Data Type
PAYMENT_METHOD_ID	Primary key for payment method dimension, surrogate key	BIGINT NOT NULL
PAYMENT_METHOD	Payment processing method	VARCHAR(50) NOT NULL
TA_INSERT_DT	Technical insert timestamp	TIMESTAMP NOT NULL
TA_UPDATE_DT	Technical update timestamp	TIMESTAMP NOT NULL

- Example with filled data:

PAYMENT_METHOD_ID	PAYMENT_METHOD	TA_INSERT_DT	TA_UPDATE_DT
50001	Debit Card	2024-06-23 08:30:15	2024-06-25 08:30:15
50002	Bank Transfer	2024-06-23 08:30:15	2025-06-18 08:30:15
50003	Credit Card	2024-06-23 08:30:15	2025-06-20 08:30:15

1.5.1.3.2 Dimensions From LMS Dataset:

- **DIM_GEOGRAPHIES**: Destination locations (City/State/Country hierarchy)

Column Name	Description	Data Type
GEOGRAPHY_ID	Primary key for geography dimension, surrogate key	BIGINT NOT NULL
DESTINATION_CITY	Destination city name	VARCHAR(100) NOT NULL
DESTINATION_STATE	Destination state/province	VARCHAR(100) NOT NULL
DESTINATION_COUNTRY	Destination country name	VARCHAR(100) NOT NULL
TA_INSERT_DT	Technical insert timestamp	TIMESTAMP NOT NULL
TA_UPDATE_DT	Technical update timestamp	TIMESTAMP NOT NULL

- Example with filled data:

GEOGRAPHY_ID	DESTINATION_CITY	DESTINATION_STATE	DESTINATION_COUNTRY
60001	Caguas	PR	Puerto Rico
60002	San Jose	CA	EE. UU.

- **DIM_SHIPPING_MODES**: Shipping service levels

Column Name	Description	Data Type
SHIPPING_MODE_ID	Primary key for shipping mode dimension, surrogate key	BIGINT NOT NULL
SHIPPING_MODE	Shipping service level	VARCHAR(50) NOT NULL
TA_INSERT_DT	Technical insert timestamp	TIMESTAMP NOT NULL
TA_UPDATE_DT	Technical update timestamp	TIMESTAMP NOT NULL

- Example with filled data:
-

SHIPPING_MODE_ID	SHIPPING_MODE	TA_INSERT_DT	TA_UPDATE_DT
70001	Standard	2024-06-23 08:30:15	2024-06-25 08:30:15
70002	Express	2024-06-23 08:30:15	2025-06-18 08:30:15
70003	Overnight	2024-06-23 08:30:15	2025-06-20 08:30:15

- **DIM_DELIVERY_STATUSES**: Delivery outcome tracking

Column Name	Description	Data Type
DELIVERY_STATUS_ID	Primary key for delivery status dimension, surrogate key	BIGINT NOT NULL
DELIVERY_STATUS	Final delivery outcome	VARCHAR(50) NOT NULL
TA_INSERT_DT	Technical insert timestamp	TIMESTAMP NOT NULL
TA_UPDATE_DT	Technical update timestamp	TIMESTAMP NOT NULL

- Example with filled data:

DELIVERY_STATUS_ID	DELIVERY_STATUS	TA_INSERT_DT	TA_UPDATE_DT
80001	Delivered	2024-06-23 08:30:15	2024-06-25 08:30:15
80002	Delayed	2024-06-23 08:30:15	2025-06-18 08:30:15
80003	In Transit	2024-06-23 08:30:15	2025-06-20 08:30:15
80004	Failed	2024-06-23 08:30:15	2025-06-20 08:30:15

- **DIM_WAREHOUSES**: Warehouse facilities

Column Name	Description	Data Type
WAREHOUSE_ID	Primary key for warehouse dimension, surrogate key	BIGINT NOT NULL
WAREHOUSE_SRC_ID	Source system warehouse identifier	VARCHAR(50) NOT NULL
TA_INSERT_DT	Technical insert timestamp	TIMESTAMP NOT NULL
TA_UPDATE_DT	Technical update timestamp	TIMESTAMP NOT NULL

- Example with filled data:

WAREHOUSE_ID	WAREHOUSE_SRC_ID	TA_INSERT_DT	TA_UPDATE_DT
90001	WH_003	2024-06-23 08:30:15	2024-06-25 08:30:15
90002	WH_001	2024-06-23 08:30:15	2025-06-18 08:30:15
90003	WH_004	2024-06-23 08:30:15	2025-06-20 08:30:15

- **DIM_CARRIERS**: Shipping carriers

Column Name	Description	Data Type
CARRIER_ID	Primary key for carrier dimension	BIGINT NOT NULL
CARRIER_NAME	Shipping carrier company name	VARCHAR(100) NOT NULL
TA_INSERT_DT	Technical insert timestamp	TIMESTAMP NOT NULL
TA_UPDATE_DT	Technical update timestamp	TIMESTAMP NOT NULL

- Example with filled data:

CARRIER_ID	CARRIER_NAME	TA_INSERT_DT	TA_UPDATE_DT
95001	UPS	2024-06-23 08:30:15	2024-06-25 08:30:15
95002	USPS	2024-06-23 08:30:15	2025-06-18 08:30:15
95003	FedEx	2024-06-23 08:30:15	2025-06-20 08:30:15

1.5.1.3.3 Shared Dimensions:

- **DIM_TIME_DD**: Daily time dimension for all date attributes

I have choose creation of SK to using FULL_DATE as PK for DIM_TIME_DD as **performance issues propagation .** Date joins are slower than integer joins and Database optdsimizers work better with integer keys

Column Name	Description	Data Type
TIME_ID	Primary key for time dimension - YYYYMMDD format, surrogate key	BIGINT NOT NULL
FULL_DATE	Complete date value	DATE NOT NULL
YEAR	Calendar year	INTEGER NOT NULL
MONTH	Calendar month number	INTEGER NOT NULL
QUARTER	Calendar quarter	INTEGER NOT NULL
DAY_OF_WEEK	Day of week number	INTEGER NOT NULL
WEEK_OF_YEAR	Week number in year	INTEGER NOT NULL
IS_WEEKEND	Weekend flag	BOOLEAN NOT NULL

Column Name	Description	Data Type
MONTH_NAME	Month name	VARCHAR(20) NOT NULL
DAY_NAME	Day name	VARCHAR(20) NOT NULL
TA_INSERT_DT	Technical insert timestamp	TIMESTAMP NOT NULL
TA_UPDATE_DT	Technical update timestamp	TIMESTAMP NOT NULL

- Example with filled data:

TIME_ID	FULL_DATE	YEAR	MONTH	QUARTER	DAY_OF_WEEK	WEEK_OF_YEAR
20240622	2024-06-22	2024	6	2	5	25
20240831	2024-08-31	2024	8	3	5	35
20231110	2023-11-10	2023	11	4	4	45

1.5.1.4 Step 4: Identify Facts

Fact Table: **FCT_SUPPLY_CHAIN_DD**

Measures (KPIs):

- **Sales Measures**: SalesAmount, Quantity, OrderTotal, UnitPrice, UnitCost, TotalCost
- **Logistics Measures**: ShippedQuantity, ShippingCost, DeliveryDays, OrderToShipDays
- **Performance Measures**: OnTimeDelivery (Boolean flag)

Degenerate Dimensions:

- OrderID, ShipmentID (stored as attributes in fact table for operational tracking)

Foreign Keys: All IDs are surrogate keys pointing to respective dimension tables

In Kimball methodology suggested to have pk for FACT table as composit key .

Primary Key = All Foreign Keys + Degenerate Dimensions (needed for uniqueness)

I have choose **hybrid approach Surogat PK** applying **Unique constraint on business key** combination for faster ETL jobs and error recovery as it easy to identify specific rows.

SUPPLY_CHAIN_ID PK "NOT NULL, SEQUENCE" *-- For operations*

UNIQUE (CUSTOMER_ID, PRODUCT_ID, ORDER_DATE_ID, ORDER_SRC_ID, SHIPMENT_SRC_ID)

Column Name	Description	Data Type
SUPPLY_CHAIN_ID	Primary key for supply chain fact table, surrogate key	BIGINT NOT NULL
CUSTOMER_ID	Foreign key to customer dimension	BIGINT NOT NULL
PRODUCT_ID	Foreign key to product dimension	BIGINT NOT NULL
SALES_REP_ID	Foreign key to sales representative dimension	BIGINT NOT NULL
ORDER_STATUS_ID	Foreign key to order status dimension	BIGINT NOT NULL
PAYMENT_METHOD_ID	Foreign key to payment method dimension	BIGINT NOT NULL
GEOGRAPHY_ID	Foreign key to geography dimension	BIGINT NOT NULL
SHIPPING_MODE_ID	Foreign key to shipping mode dimension	BIGINT NOT NULL
DELIVERY_STATUS_ID	Foreign key to delivery status dimension	BIGINT NOT NULL
WAREHOUSE_ID	Foreign key to warehouse dimension	BIGINT NOT NULL
CARRIER_ID	Foreign key to carrier dimension	BIGINT NOT NULL
ORDER_DATE_ID	Foreign key to time dimension for order date	BIGINT NOT NULL
SHIP_DATE_ID	Foreign key to time dimension for ship date	BIGINT NOT NULL
DELIVERY_DATE_ID	Foreign key to time dimension for delivery date	BIGINT NOT NULL
ORDER_SRC_ID	Source system order identifier	VARCHAR NOT NULL
SHIPMENT_SRC_ID	Source system shipment identifier	VARCHAR NOT NULL
TRANSACTION_SRC_ID	Source system transaction identifier	BIGINT NOT NULL
SALES_AMOUNT	Revenue amount from the sale	DECIMAL(15,2) NULL
QUANTITY	Number of units ordered	INTEGER NULL
ORDER_TOTAL	Total order value	DECIMAL(15,2) NULL
UNIT_PRICE	Price per unit	DECIMAL(15,2) NULL
UNIT_COST	Cost per unit	DECIMAL(15,2) NULL
TOTAL_COST	Total cost of goods sold	DECIMAL(15,2) NULL
SHIPPED_QUANTITY	Number of units shipped	INTEGER NULL

Column Name	Description	Data Type
SHIPPING_COST	Cost of shipping	DECIMAL(15,2) NULL
DELIVERY_DAYS	Number of days for delivery	INTEGER NULL
ORDER_TO_SHIP_DAYS	Days from order to shipment	INTEGER NULL
ON_TIME_DELIVERY	Whether delivery was on time	BOOLEAN NULL
TA_INSERT_DT	Technical insert timestamp	TIMESTAMP NOT NULL
TA_UPDATE_DT	Technical update timestamp	TIMESTAMP NOT NULL

Example with filled data:

Column Name	Record 1	Record 2	Record 3
SUPPLY_CHAIN_ID	1	2	3
CUSTOMER_ID	10001	10002	10003
PRODUCT_ID	20001	20001	20001
SALES_REP_ID	30001	30002	30003
ORDER_STATUS_ID	40001	40001	40001
PAYMENT_METHOD_ID	50001	50002	50001
GEOGRAPHY_ID	60001	60001	60002
SHIPPING_MODE_ID	70001	70001	70001
DELIVERY_STATUS_ID	80001	80002	80001
WAREHOUSE_ID	90001	90002	90003
CARRIER_ID	95001	95001	95002
ORDER_DATE_ID	20240622	20240831	20231110
SHIP_DATE_ID	20240714	20240916	20231207
DELIVERY_DATE_ID	20240809	20241007	20240109
ORDER_SRC_ID	77202_V0	75939_V0	75938_V0
SHIPMENT_SRC_ID	77202_V0	75939_V0	75938_V0
TRANSACTION_SRC_ID	1	2	3
SALES_AMOUNT	327.75	327.75	327.75

Column Name	Record 1	Record 2	Record 3
QUANTITY	1	1	1
ORDER_TOTAL	314.64	311.36	309.72
UNIT_PRICE	327.75	327.75	327.75
UNIT_COST	238.94	236.85	231.71
TOTAL_COST	238.94	236.85	231.71
SHIPPED_QUANTITY	1	1	1
SHIPPING_COST	15.34	14.33	18.47
DELIVERY_DAYS	26	21	33
ORDER_TO_SHIP_DAYS	22	16	27
ON_TIME_DELIVERY	FALSE	FALSE	FALSE
TA_INSERT_DT	2024-06-22 08:30:15.123	2024-08-31 09:15:42.456	2023-11-10 14:22:33.789
TA_UPDATE_DT	2024-06-22 08:30:15.123	2024-09-01 10:22:18.234	2023-11-10 14:22:33.789

2 BUSINESS LAYER 3NF

The 3NF Business Layer serves as the normalized foundation for our supply chain data warehouse, transforming denormalized dimensional data into a clean, normalized structure that supports enterprise data architecture.

2.1 3NF DECOMPOSITION

2.1.1 Dimension Decomposition

The flat dimension tables were broken down into normalized entities:

`DIM_PRODUCTS_SCD` (contained product, category, brand, department) →

- `CE_PRODUCTS_SCD` (just product info)
- `CE_BRANDS` (brand details)
- `CE_CATEGORIES` (category info)
- `CE_DEPARTMENTS` (department info)
- `CE_PRODUCT_STATUSES` (status lookup) ‘

The product hierarchy follows a 4-level normalized structure:
Department → Category → Brand → Product

`DIM_GEOGRAPHIES` (flat city/state/country) →

- `CE_GEOGRAPHIES` (geography points)
- `CE_CITIES`, `CE_STATES`, `CE_COUNTRIES`, `CE_REGIONS` (proper hierarchy)

The geography hierarchy follows a 4-level normalized structure:
Region → Country → State → City → Geography

2.1.2 SCD Type 2 Implementation

From our Dimentions definition in a part #1.5.1.3.Step 3: Identify Dimensions|outline were we defined SCD of types 0-2 we get out entity SCD type 2:

CE_PRODUCTS_SCD
Primary Key: (PRODUCT_ID + START_DT)
Tracks: Product category and department changes over time

We added Composit PK(PRODUCT_ID + START_DT), and as well as souce triplets clmns(PRODUCT_SRC_ID, SOURCE_SYSTEM, SOURCE_ENTITY)and flag IS_ACTIVE ('Y'/'N')

2.1.3 Fact Table Decomposition

Entity Identification:

1. ORDERS (Order-level information)
2. ORDER_LINES (Product-line information)
3. SHIPMENTS (Shipping information)
4. SHIPMENT_LINES (Product-shipment intersection)
5. DELIVERIES (Delivery outcomes)
6. TRANSACTIONS (Financial records)

Result: 6 separate entity tables (CE_ORDERS,CE_ORDER_LINES,CE_SHIPMENTS,CE_SHIPMENT_LINES,CE_DELIVERIES,CE_TRANSACTIONS)

Eliminated: Repeating groups and mixed business concepts

2.1.4 Elimination of Redundancy

Denormalization → Normalization

- **FCT table**: One row = one complete supply chain transaction
- **3NF tables**: Separated by business entities and their relationships

Composite Keys → Surrogate Keys

- **FCT**: Used dimensional foreign keys
- **3NF**: Each table has its own `<NAME>_ID` primary key

Date Dimensions → Date Columns

- **FCT**: `ORDER_DATE_ID`, `SHIP_DATE_ID`, `DELIVERY_DATE_ID`
- **3NF**: `ORDER_DATE`, `SHIP_DATE`, `DELIVERY_DATE`

Calculated Fields → Source Fields

- **FCT**: Pre-calculated KPIs like `TOTAL_COST`, `ORDER_TO_SHIP_DAYS`
- **3NF**: Store base values, calculate derived metrics in ETL to dimensional layer

Throe all entities implemented:

Source Triplets pattern:

```
SOURCE_SYSTEM VARCHAR -- 'SA_OMS', 'SA_LMS', 'MANUAL'
SOURCE_ENTITY VARCHAR -- 'ORDERS_TABLE', 'CUSTOMER_MASTER'
<NAME>_SRC_ID VARCHAR -- Natural business key from source
```

Technicall Audit fields for SCD 2:

```
TA_INSERT_DT DATE -- Record creation timestamp
TA_UPDATE_DT DATE -- Last modification timestamp
```

Default Rows:

Every table should includes -1 default row for "Not Applicable"
Dates in range *'1990-01-01'*, *'9999-12-31'*

2.1.5 Business Flow Modeling

CUSTOMER MANAGEMENT:

CE_CUSTOMERS

Customers are managed through a slowly changing dimension (SCD type 1) structure. Each customer can place multiple orders throughout their relationship with the company.

PRODUCT CATALOG HIERARCHY:

CE_DEPARTMENTS → CE_CATEGORIES → CE_BRANDS → CE_PRODUCTS_SCD

↑

CE_PRODUCT_STATUSES

The product hierarchy follows a structured approach: Departments organize broad product categories, which contain specific brands that operate within those categories. Each brand manufactures multiple products, and product statuses define current availability (active, discontinued, seasonal, etc.).

GEOGRAPHY HIERARCHY:

CE_REGIONS → CE_COUNTRIES → CE_STATES → CE_CITIES → CE_GEOGRAPHIES

↓

CE_SALES_REPRESENTATIVES

The geographic structure follows a standard administrative hierarchy: Regions contain multiple countries, countries contain states or provinces, states contain cities, and cities define specific geographic locations used throughout the system for addresses, territories, and logistics planning.

TERRITORY & LOGISTICS SETUP:

CE_WAREHOUSES ←→ CE_GEOGRAPHIES CE_WAREHOUSES

CE_CARRIERS

CE_SHIPPING_MODES

CE_DELIVERY_STATUSES

Geographic territories define operational boundaries for both sales representatives and warehouse coverage areas. This enables territory-based performance analysis and optimal logistics planning based on geographic proximity.

ORDER MANAGEMENT LOOKUPS:

CE_ORDER_STATUSES

CE_PAYMENT_METHODS

The sales workflow begins when sales representatives manage customer orders. Orders are tracked through various statuses (pending, confirmed, processing, cancelled) and processed using different payment methods (credit card, bank transfer, cash). Each order contains multiple product lines with specific quantities and pricing.

CORE BUSINESS PROCESS FLOW:

CE_CUSTOMERS ←→ CE_ORDERS ←→ CE_ORDER_LINES ←→ CE_PRODUCTS_SCD

↑ ↑

CE_SALES_REPRESENTATIVES CE_BRANDS

CE_ORDER_STATUSES ↑

CE_PAYMENT_METHODS CE_CATEGORIES

↓ ↓ ↑

CE_SHIPMENTS ←→ CE_SHIPMENT_LINES

↑ CE_DEPARTMENTS

CE_GEOGRAPHIES

CE_SHIPPING_MODES

CE_WAREHOUSES

CE_CARRIERS

↓

CE_DELIVERIES

↑

CE_DELIVERY_STATUSES

Once orders are confirmed, they generate shipments for fulfillment. The system tracks which specific order line items are included in each shipment through shipment lines, allowing for partial shipments and split deliveries across multiple shipments.

Shipments are orchestrated through multiple **logistics components**: shipping modes define the delivery method (standard, express, overnight), warehouses serve as the origin point, carriers handle transportation, and geographies specify the destination locations.

Each shipment results in **delivery records** that track completion status. Delivery statuses monitor the final mile (delivered, failed delivery, returned to sender, in transit) and capture performance metrics like delivery timeframes and on-time delivery rates.

FINANCIAL PROCESS:

CE_ORDERS → CE_TRANSACTIONS

Orders generate corresponding financial transactions that record the monetary exchange. This separation allows for complex payment scenarios like partial payments, refunds, and installment plans while maintaining order integrity.

3 BUSINESS LAYER DIMENSIONAL MODEL

3.1 COMPLETE METRICS DESCRIPTION

FACT TABLE: FCT_ORDER_LINE_SHIPMENTS_DD

Grain: One row per order line item per shipment
Business Purpose: Comprehensive supply chain analytics from order to delivery

For example used Smart Watch Order Example (Order #77202_V0)
Customer: Cally Holloway (Consumer segment)
Product: Smart Watch (SportSync brand, Sporting Goods category)
Timeline: Ordered 2024-05-11, Shipped 2024-05-31, Delivered 2024-06-25

3.1.1 Sales & Revenue Metrics

UNIT_PRICE_ACT

- **Type**: DECIMAL
- **Business Definition**: The selling price per unit of product on the order line
- **Calculation**: Direct from order line item
- **Aggregation**: Average (for unit analysis), Sum×Quantity (for revenue)
- **Business Use**: Pricing analysis, revenue per unit tracking
- **Example**: `Smart watch (SportSync brand) sold at $327.75 per unit to Cally Holloway`

UNIT_COST_ACT

- **Type**: DECIMAL
- **Business Definition**: The cost per unit of product (COGS - Cost of Goods Sold)
- **Calculation**: Direct from product cost data
- **Aggregation**: Average (for cost analysis), Sum×Quantity (for total cost)
- **Business Use**: Profitability analysis, margin calculations
- **Example**: `Smart watch costs $242.54 to produce/acquire`

LINE_TOTAL_ACT

- **Type**: DECIMAL
- **Business Definition**: Total sales amount for the order line
- **Calculation**: `UNIT_PRICE_ACT × ORDERED_QUANTITY_CNT`
- **Aggregation**: Additive (SUM)
- **Business Use**: Order value analysis, revenue tracking
- **Example**: `1 unit × $327.75 = $327.75 (Order #77202_V0, OrderItemID: 180517)`

ORDERED_QUANTITY_CNT

- **Type**: INTEGER
- **Business Definition**: Number of units ordered by customer for this product line
- **Calculation**: Direct from order line
- **Aggregation**: Additive (SUM)
- **Business Use**: Demand analysis, volume tracking
- **Example**: `Cally Holloway ordered 1 unit of Smart watch on 2024-05-11`

PROFIT_AMOUNT_ACT *(Calculated)*

- **Type**: DECIMAL
- **Business Definition**: Gross profit for the shipped quantity
- **Calculation**: `(UNIT_PRICE_ACT - UNIT_COST_ACT) × SHIPPED_QUANTITY_CNT`
- **Aggregation**: Additive (SUM)
- **Business Use**: Profitability analysis, margin tracking
- **Example**: `($327.75 - $242.54) × 1 = $85.21 profit on Smart watch order`

3.1.2 SHIPPING & LOGISTICS METRICS

SHIPPED_QUANTITY_CNT

- **Type**: INTEGER
- **Business Definition:** Actual number of units shipped in this specific shipment
- **Calculation:** Direct from shipment line
- **Aggregation:** Additive (SUM)
- **Business Use:** Fulfillment analysis, fill rate calculations
- **Example:** 1 unit of Smart watch shipped from WH_005 on 2024-05-31 (same as ordered - 100% fill rate)

SHIPPING_COST_ACT

- **Type:** DECIMAL
- **Business Definition:** Total shipping cost for the entire shipment
- **Calculation:** Direct from shipment record
- **Aggregation:** Additive (SUM) - but may cause double counting
- **Business Use:** Logistics cost analysis
- **Example:** $9.75 to ship Smart watch via UPS Standard to Caguas, PR
-

ALLOCATED_SHIPPING_COST_ACT *(Calculated)*

- **Type**: DECIMAL

- **Business Definition**: Shipping cost allocated to this specific order line within the shipment
- **Calculation**: `SHIPPING_COST_ACT × (Line_Value / Total_Shipment_Value)`
- **Aggregation**: Additive (SUM)
- **Business Use**: True profitability analysis including shipping
- **Example**: $9.75 × ($327.75 / $327.75) = $9.75 (single item shipment gets full allocation)

UNSHIPPED_QUANTITY_CNT *(Calculated)*

- **Type**: DECIMAL
- **Business Definition**: Quantity ordered but not yet shipped
- **Calculation**: `ORDERED_QUANTITY_CNT - SHIPPED_QUANTITY_CNT`
- **Aggregation**: Additive (SUM)
- **Business Use**: Backorder analysis, fulfillment gaps
- **Example**: 1 ordered - 1 shipped = 0 unshipped (perfect fulfillment)

FILL_RATE_PCT *(Calculated)*

- **Type**: DECIMAL
- **Business Definition**: Percentage of ordered quantity that was shipped
- **Calculation**: `(SHIPPED_QUANTITY_CNT / ORDERED_QUANTITY_CNT) × 100`
- **Aggregation**: Weighted Average
- **Business Use**: Fulfillment performance, service level measurement
- **Example**: (1 / 1) × 100 = 100% fill rate for Smart watch order

3.1.3 DELIVERY & PERFORMANCE METRICS

DELIVERY_DAYS_CNT

- **Type**: INTEGER
- **Business Definition**: Number of days from ship date to actual delivery
- **Calculation**: `DELIVERY_DATE - SHIP_DATE`
- **Aggregation**: Average (for performance), Distribution analysis
- **Business Use**: Delivery performance, customer satisfaction
- **Example**: `Smart watch shipped 2024-05-31, delivered 2024-06-25 = 25 days`

ORDER_TO_SHIP_DAYS_CNT

- **Type**: INTEGER
- **Business Definition**: Number of days from order date to ship date

- **Calculation**: `SHIP_DATE - ORDER_DATE`
- **Aggregation**: Average (for performance analysis)
- **Business Use**: Order processing efficiency, fulfillment speed
- **Example**: `Cally's order placed 2024-05-11, shipped 2024-05-31 = 20 days processing time`

ON_TIME_DELIVERY_FLAG

- **Type**: DECIMAL (0 or 1)
- **Business Definition**: Indicator if delivery met the planned timeline
- **Calculation**: `IF(DELIVERY_DAYS_CNT <= PLANNED_DELIVERY_DAYS_CNT, 1, 0)`
- **Aggregation**: Average (gives percentage), Sum (gives count)
- **Business Use**: Service level measurement, carrier performance
- **Example**: `Smart watch delivered in 25 days vs 7 planned = 0 (late delivery)`

LATE_DELIVERY_FLAG *(Calculated)*

- **Type**: DECIMAL (0 or 1)
- **Business Definition**: Indicator if delivery was late
- **Calculation**: `IF(DELIVERY_DAYS_CNT > PLANNED_DELIVERY_DAYS_CNT, 1, 0)`
- **Aggregation**: Average (gives percentage), Sum (gives count)
- **Business Use**: Problem identification, carrier evaluation
- **Example**: `Smart watch took 25 days when 7 were planned = 1 (late delivery to Caguas, PR)`

3.1.4 DERIVED FINANCIAL METRICS *(Calculated at Query Time)*

SHIPPED_SALES_AMOUNT_ACT *(Calculated)*

- **Type**: DECIMAL
- **Business Definition**: Revenue for the actual shipped quantity
- **Calculation**: `UNIT_PRICE_ACT × SHIPPED_QUANTITY_CNT`
- **Aggregation**: Additive (SUM)
- **Business Use**: Actual revenue recognition, shipped vs ordered analysis
- **Example**: `$327.75 × 1 shipped = $327.75 revenue recognized for Smart watch`

3.2 KEY PERFORMANCE INDICATORS (KPIS)

Order Fulfillment KPIs

- **Perfect Order Rate**: Orders with 100% fill rate, on-time delivery
- **Average Fill Rate**: Weighted average of FILL_RATE_PCT

- **Order Cycle Time**: ORDER_TO_SHIP_DAYS_CNT + DELIVERY_DAYS_CNT

Financial KPIs

- **Revenue per Order**: Sum of SHIPPED_SALES_AMOUNT_ACT by order
- **Average Order Value**: Revenue / Number of orders
- **Gross Margin**: Sum of PROFIT_AMOUNT_ACT / Sum of SHIPPED_SALES_AMOUNT_ACT

Logistics KPIs

- **On-Time Delivery Rate**: Average of ON_TIME_DELIVERY_FLAG
- **Average Delivery Time**: Average of DELIVERY_DAYS_CNT
- **Shipping Cost as % of Revenue**: ALLOCATED_SHIPPING_COST_ACT / SHIPPED_SALES_AMOUNT_ACT

Operational KPIs

- **Backorder Rate**: UNSHIPPED_QUANTITY_CNT / ORDERED_QUANTITY_CNT
- **Order Processing Speed**: Average ORDER_TO_SHIP_DAYS_CNT
- **Carrier Performance**: ON_TIME_DELIVERY_FLAG by carrier

To implement in query time:

TOTAL_LANDED_COST_ACT *(Calculated)*

- **Type**: DECIMAL
- **Business Definition**: Total cost including product cost and allocated shipping
- **Calculation**: `(UNIT_COST_ACT × SHIPPED_QUANTITY_CNT) + ALLOCATED_SHIPPING_COST_ACT`
- **Aggregation**: Additive (SUM)
- **Business Use**: True cost analysis, margin calculations
- **Example**: `($242.54 × 1) + $9.75 = $252.29 total landed cost`

NET_PROFIT_ACT *(Calculated)*

- **Type**: DECIMAL
- **Business Definition**: Profit after including shipping costs
- **Calculation**: `SHIPPED_SALES_AMOUNT_ACT - TOTAL_LANDED_COST_ACT`
- **Aggregation**: Additive (SUM)
- **Business Use**: True profitability analysis
- **Example**: `$327.75 - $252.29 = $75.46 net profit on Smart watch order`

PROFIT_MARGIN_PCT *(Calculated)*

- **Type**: DECIMAL

- **Business Definition**: Profit margin percentage including all costs
- **Calculation**: `(NET_PROFIT_ACT / SHIPPED_SALES_AMOUNT_ACT) × 100`
- **Aggregation**: Weighted Average
- **Business Use**: Profitability analysis, pricing decisions
- **Example**: `($75.46 / $327.75) × 100 = 23.0% profit margin on Smart watch`

4 LOGICAL SCHEME

High-level architectural overview showing logical data flow from source systems (OMS, LMS) through CSV files, staging areas, 3NF normalized schema, dimensional model, to data marts and end-user tools. Demonstrates the layered approach with clear separation between back-room operations (ETL) and front-room analytics (BI, reporting).

Shows multiple consumption patterns including direct dimensional access and specialized data marts for sales, customer analytics, and supply chain analysis.

5 DATA FLOW

Comprehensive Data Flow Diagram (DFD) - Gane/Sarson Notation showing the complete ETL pipeline from OMS/LMS source systems through staging areas (SA_OMS, SA_LMS), 3NF relational layer (BL_3NF), to dimensional model (BL_DM).

Illustrates 12 distinct processes including data extraction, staging transformations, SCD Type 1/2 implementations, and fact table loading. Contains 25+ data stores with actual table names, key fields, and business metrics for order line shipments.

6 FACT TABLE PARTITIONING STRATEGY

The `FCT_ORDER_LINE_SHIPMENTS_DD` fact table implements **native PostgreSQL range partitioning** with monthly granularity and rolling window management for optimal performance and data lifecycle management.

6.1.1 Partitioning Scheme

- **Partition Key:** `EVENT_DT` (Order Date)
- **Granularity:** Monthly partitions (YYYY-MM format)
- **Naming Convention:** `FCT_ORDER_LINE_SHIPMENTS_DD_YYYYMM`
- **Date Range:** 2023-2025 (36 partitions total)

Rolling Window Strategy

Historical Data (2023-2024):

- Status: **Static partitions** - permanently attached
- Behavior: Never detached, always accessible
- Purpose: Long-term historical reporting and compliance

Current Data (2025+):

- Status: **Active rolling window** - dynamically managed
- Window Size: **3 months** (configurable)
- Behavior: Automatically detach partitions older than 3 months
- Purpose: Optimize query performance on recent data

Data Loading Integration

- **Incremental Loads:** Only affected partitions are processed
- **Delta Detection:** Based on `TA_UPDATE_DT` timestamp comparison
- **Data Quality:** Built-in validation and missing dimension detection
- **Logging:** Comprehensive audit trail with partition-specific metrics

Business Continuity

- **Historical Preservation:** 2023-2024 data always available
- **Flexible Windows:** Rolling window size adjustable based on business needs
- **Disaster Recovery:** Each partition can be backed up/restored independently
- **Compliance:** Audit trail preserved across all partition operations

Implementation Result: 500,392 records successfully distributed across 36 monthly partitions with 0ms query response time for partition-aware queries.