Document: Create Data Catalog

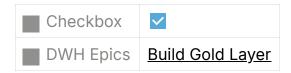| Checkbox	☑
DWH Epics	<u>Build Gold Layer</u>

1. gold.dim_customers (Customer Dimension)

- **Purpose**: Stores customer attributes for analysis (e.g., demographics, location).

- **Columns:**

Column Name	Data Type	Description
customer_key	INT	Surrogate key that uniquely identifies each customer record in the dimension table (system-generated unique ID).
customer_id	INT	Unique numerical identifier assigned to each customer.
customer_number	NVARCHAR(50)	Alphanumeric identifier representing the customer, used for tracking and referencing
first_name	NVARCHAR(50)	Customer's first name.
last_name	NVARCHAR(50)	Customer's last name.
country	NVARCHAR(50)	Customer's country (e.g 'Australia').
marital_status	NVARCHAR(50)	Customer's Marital Status (e.g 'Married', 'Single', etc.)
gender	NVARCHAR(50)	Customer's Gender ('Male', 'Female', 'n/a')
birthdate	DATE	Customer's date of birth in format 'YYYY-MM-DD'
create_date	DATE	When the customer record was created.

2. gold.dim_products (Product Dimension)

- **Purpose**: It contains product details (e.g., name, category, cost).

- **Columns:**

Column Name	Data Type	Description
product_key	INT	Surrogate key that uniquely identifies each product record in the product dimension table (system-generated unique ID).
product_id	INT	A unique identifier assigned to the product for internal tracking and referencing.
product_number	NVARCHAR(50)	A structured alphanumeric cod representing each product, used for categorization and tracking
product_name	NVARCHAR(50)	Descriptive name of product, including key details such as type, color and size
category_id	NVARCHAR(50)	A unique Identifier for the product's category, linking it to its high-level classification
category	NVARCHAR(50)	A broad classification of the product category (e.g. Bikes, components) to group related items
subcategory	NVARCHAR(50)	A more specific classification of the product within the category, i.e. product type.
maintenance	NVARCHAR(50)	It indicates whether the product requires maintenance.
cost	INT	This is the cost of the product, measured in monetary units.
product_line	NVARCHAR(50)	The specific Product line or series to which the product belongs (e.g. Roads, Mountains etc.).
start_date	DATE	The date when the product was introduced for sale or use.

3. gold.fact_sales (Sales Fact Table)

- **Purpose**: Records sales transactions with links to dimensions (customers, products).
- **Columns:**

Column Name	Data Type	Description
order_number	NVARCHAR(50)	A unique alphanumeric identifier for each sales order.
product_key	INT	The surrogate key linking the order to dim_products .

Column Name	Data Type	Description
customer_id	INT	The surrogate key linking the order to dim_customers .
order_date	DATE	When the order was placed.
shipping_date	DATE	When the product was shipped.
due_date	DATE	The date when the order payment was due
sales_amount	INT	The total monetary value from the sale of the line item in whole currency units (e.g. 2, 19).
quantity	INT	The number of units of the product ordered for the line item (e.g. 1).
price	INT	The price per unit for the line item in whole currency units (e.g. 25).