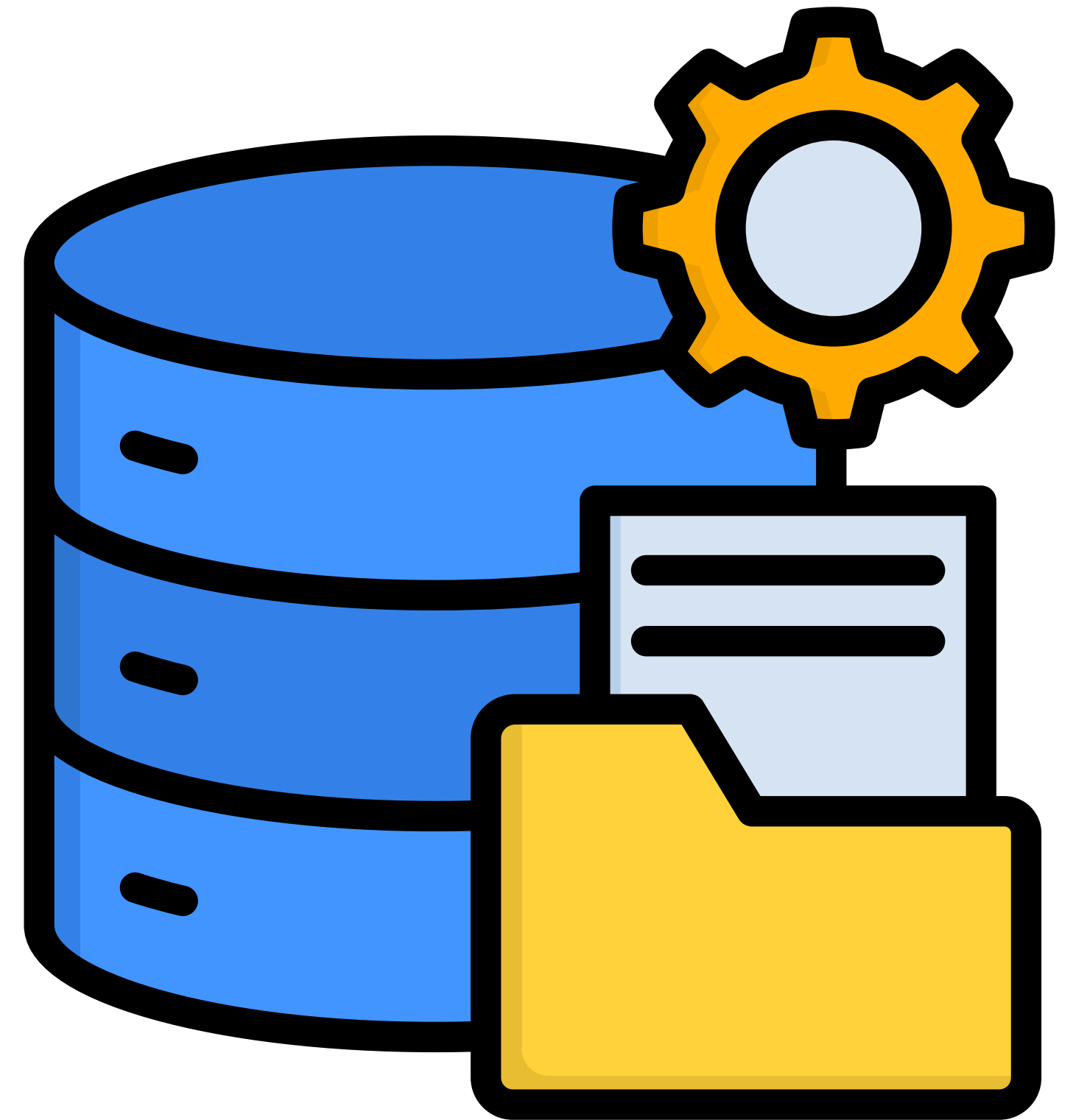


PROJECT DATA WAREHOUSE

Extract - Transform - Load
(ETL)



INTRODUCTION

This project aims to build a Data Warehouse by applying the ETL (Extract, Transform, Load) process to integrate CRM and ERP data. The workflow ensures that raw data is captured, cleaned, standardized, and finally structured into a star schema for effective analysis and reporting.

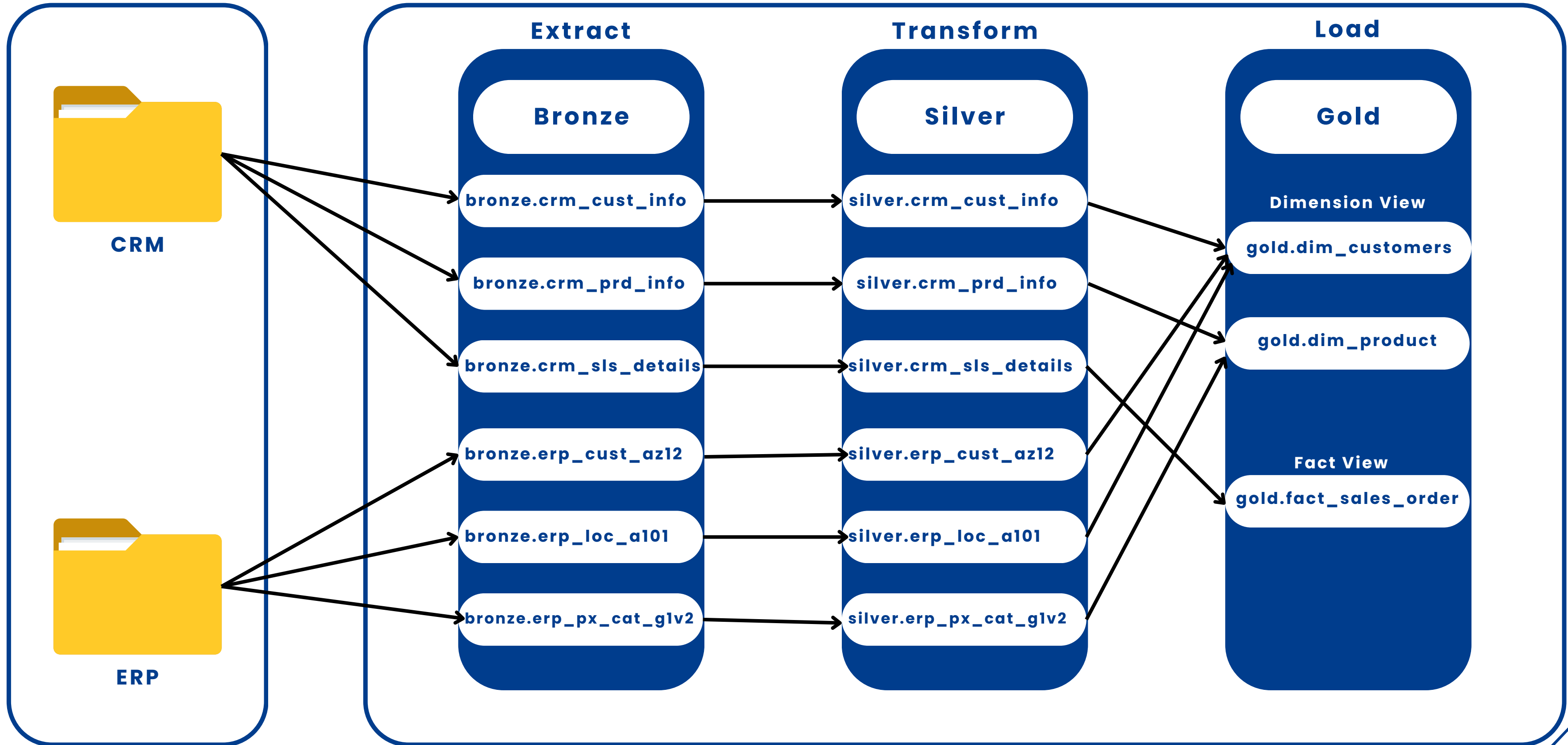
- Objective: Build a Data Warehouse using ETL to unify multiple data sources.
- ETL Workflow:
 - Bronze (Extract) → Load raw CSV files into tables using DDL/DML & Bulk Insert.
 - Silver (Transform) → Perform data cleaning, validation, and standardization.
 - Gold (Load) → Build Dimension Views & Fact Views (Star Schema) for analytics.
- Data Source : <https://www.datawithbaraa.com/sql-introduction/advanced-sql-project/>





Source

Database & Schema



BRONZE LAYER (EXTRACT STAGE)

The Bronze Layer acts as the raw data storage in the Data Warehouse. At this stage, data is directly extracted from multiple CSV files without applying any transformation or cleaning. It ensures that the original source data is captured as-is, providing a reliable foundation before further processing in the Silver layer.

- Purpose: Store raw, unmodified data from different sources as the single source of truth.
- Process:
 - Data sources include CRM tables (cust_info, prd_info, sales_details) and ERP tables (cust_az12, loc_a101, px_cat_glv2).
 - Extraction is handled through a stored procedure (bronze.load_bronze).
 - Target tables are truncated first to avoid duplication.
 - Data is loaded via BULK INSERT with CSV delimiters.
 - Execution time for each load is measured and printed for monitoring.
 - Error handling is implemented using a TRY...CATCH block for logging issues.
- Key Point: The Bronze layer ensures that raw, untouched data is preserved, serving as the baseline before quality checks and transformations in the Silver stage.



BRONZE TABLE

File

Edit

View

Git

Project

Tools

Extensions

Window

Help

Search

Solution1

File

Edit

View

Query

Git

Project

Tools

Extensions

Window

SILVER LAYER (TRANSFORM STAGE)

The Silver Layer is where raw Bronze data is transformed into clean, standardized, and business-ready data. This stage ensures data quality, consistency, and compliance with business rules before feeding into the Gold (analytics) layer.

- Purpose: Convert raw data into trusted, structured information by applying validation, cleaning, and standardization.
- Process:
 - a. Data Quality Check (DQC)
 - Detect duplicates, nulls, invalid dates, negative values, and broken relationships.
 - Check standardization for gender, marital status, country codes, and categories.
 - Validate business rules such as $\text{Sales} = \text{Quantity} \times \text{Price}$.
 - b. Transformation with silver.load_silver Procedure
 - CRM Data: Remove duplicates (keep latest), trim extra spaces, decode marital status & gender, validate product cost, product line, and effective dates.
 - Sales Data: Convert integer fields to proper DATE, recalculate sales if inconsistent, and fix invalid or missing prices.
 - ERP Data: Standardize customer IDs, gender values, and country codes, remove invalid birthdates (future dates), normalize location and category attributes.
- Key Point: The Silver Layer guarantees data accuracy, standardization, and compliance with business rules — making it a trusted input for the Gold stage reporting and analytics.



SILVER TABLE

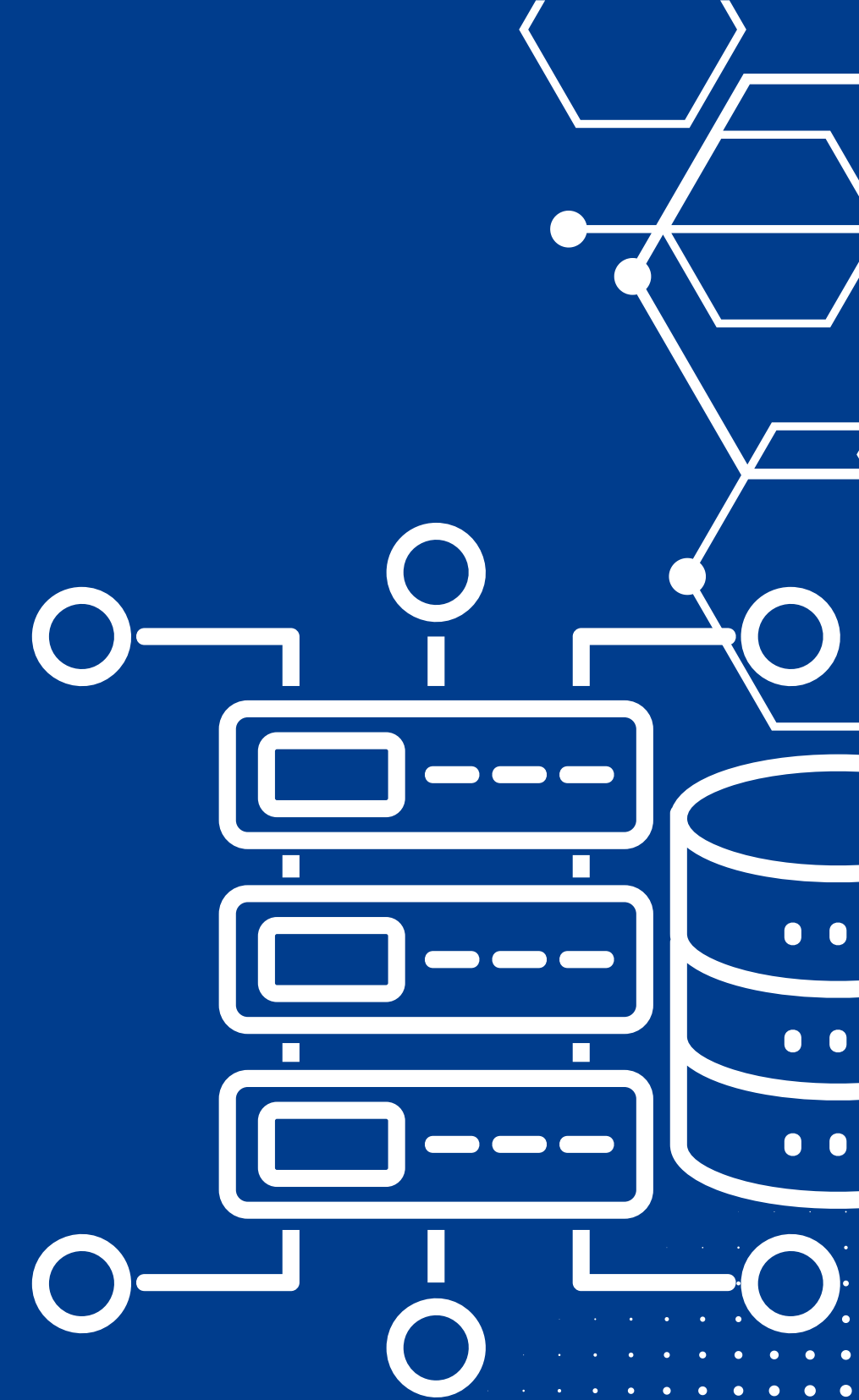
SQLQuery1.sql...0\acer (70))*											
Results											
	cst_id	cst_key	cst_firstname	cst_lastname	cst_marital_status	cst_gndr	cst_create_date	dwh_create_date			
1	11000	AW00011000	Jon	Yang	Married	Male	2025-10-06	2025-09-04 12:20:41.3766667			
2	11001	AW00011001	Eugene	Huang	Married	Male	2025-10-06	2025-09-04 12:20:41.3766667			
3	11002	AW00011002	Ruben	Torres	Married	Male	2025-10-06	2025-09-04 12:20:41.3766667			
4	11003	AW00011003	Christy	Zhu	n/a	Female	2025-10-06	2025-09-04 12:20:41.3766667			
5	11004	AW00011004	Elizabeth	Johnson	n/a	Female	2025-10-06	2025-09-04 12:20:41.3766667			
	prd_id	cat_id	prd_key	prd_nm	prd_cost	prd_line	prd_start_dt	prd_end_dt	dwh_create_date		
1	478	AC_BC	BC-M005	Mountain Bottle Cage	4	Mountain	2013-07-01	NULL	2025-09-04 12:20:41.5200000		
2	479	AC_BC	BC-R205	Road Bottle Cage	3	Road	2013-07-01	NULL	2025-09-04 12:20:41.5200000		
3	477	AC_BC	WB-H0...	Water Bottle - 30 oz.	2	Other Sales	2013-07-01	NULL	2025-09-04 12:20:41.5200000		
4	483	AC_BR	RA-H123	Hitch Rack - 4-Bike	45	Other Sales	2013-07-01	NULL	2025-09-04 12:20:41.5200000		
5	486	AC_BS	ST-1401	All-Purpose Bike St...	59	Mountain	2013-07-01	NULL	2025-09-04 12:20:41.5200000		
	sls_ord_num	sls_prd_key	sls_cust_id	sls_order_dt	sls_ship_dt	sls_due_dt	sls_sales	sls_quantity	sls_price	dwh_create_date	
1	SO43697	BK-R93R-62	21768	2010-12-29	2011-01-05	2011-01-10	3578	1	3578	2025-09-04 12:20:41.5333333	
2	SO43698	BK-M82S-44	28389	2010-12-29	2011-01-05	2011-01-10	3400	1	3400	2025-09-04 12:20:41.5333333	
3	SO43699	BK-M82S-44	25863	2010-12-29	2011-01-05	2011-01-10	3400	1	3400	2025-09-04 12:20:41.5333333	
4	SO43700	BK-R50B-62	14501	2010-12-29	2011-01-05	2011-01-10	699	1	699	2025-09-04 12:20:41.5333333	
5	SO43701	BK-M82S-44	11003	2010-12-29	2011-01-05	2011-01-10	3400	1	3400	2025-09-04 12:20:41.5333333	
	cid	bdate	gen	dwh_create_date							
1	AW00011000	1971-10-06	Male	2025-09-04 12:20:42.1400000							
2	AW00011001	1976-05-10	Male	2025-09-04 12:20:42.1400000							
3	AW00011002	1971-02-09	Male	2025-09-04 12:20:42.1400000							
4	AW00011003	1973-08-14	Fe...	2025-09-04 12:20:42.1400000							
5	AW00011004	1979-08-05	Fe...	2025-09-04 12:20:42.1400000							

SQLQuery1.sql...0\acer (70))*											
Results											
	cid	cntry	dwh_create_date								
1	AW00011000	Australia	2025-09-04 12:20:42.2866667								
2	AW00011001	Australia	2025-09-04 12:20:42.2866667								
3	AW00011002	Australia	2025-09-04 12:20:42.2866667								
4	AW00011003	Australia	2025-09-04 12:20:42.2866667								
5	AW00011004	Australia	2025-09-04 12:20:42.2866667								
	id	cat	subcat	maintance	dwh_create_date						
1	AC_BR	Accessories	Bike Racks	Yes	2025-09-04 12:20:42.4466667						
2	AC_BS	Accessories	Bike Stands	No	2025-09-04 12:20:42.4466667						
3	AC_BC	Accessories	Bottles and Cages	No	2025-09-04 12:20:42.4466667						
4	AC_CL	Accessories	Cleaners	Yes	2025-09-04 12:20:42.4466667						
5	AC_FE	Accessories	Fenders	No	2025-09-04 12:20:42.4466667						

GOLD LAYER (ANALYTICS STAGE)

The Gold Layer is the analytics-ready stage of the Data Warehouse. At this point, cleaned and standardized data from the Silver layer is modeled into a star schema using dimension and fact views, enabling reporting, dashboards, and business insights.

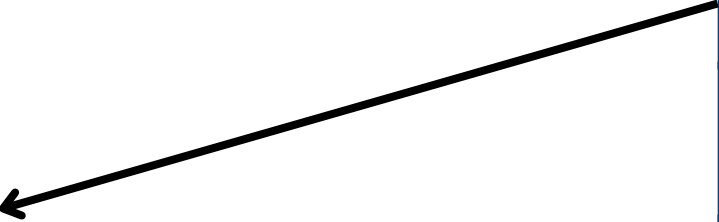
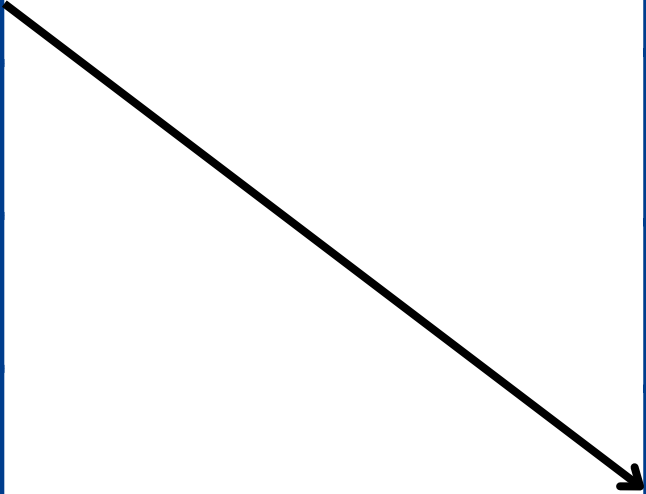
- Purpose: Deliver business-ready data for analytics by integrating dimensions and facts in a star schema.
- Process:
 - a. Dimension Views
 - `dim_customers`: Combines CRM & ERP customer data, enriched with gender, marital status, and country.
 - `dim_product`: Consolidates product details with ERP categories, filtering only active products.
 - b. Fact Views
 - `fact_sales_order`: Central fact table linking sales transactions with customer and product dimensions.
 - Provides key measures such as Sales Amount, Quantity, Unit Price, and dates (Order, Ship, Due).
- Key Point: The Gold Layer transforms structured data into a star schema (facts & dimensions), powering dashboards, KPIs, and advanced business reporting.



gold.dim_customers	
Customer_Key	PK
Customer_ID	
Customer_Number	
Customer_Firstname	
Customer_LastName	
Birthdate	
Gender	
Marital_Status	
Country	
Create_Date	

gold.fact_sales_order	
	Order_Number
FK	Product_Key
FK	Customer_Key
	Order_Date
	Ship_Date
	Due_Date
	Sales
	Quantity
	Price

gold.dim_product	
PK	Product_Key
	Product_id
	Product_Number
	Product_Name
	Category_id
	Category
	Subcategory
	Maintance
	Cost
	Product_Line
	Start_Date



GOLD VIEWS

File Edit View Query Git Project Tools Extensions Window Help Search Solution1

Datawarehouse Execute

SQLQuery1.sql...0\acer (70))*

Results Messages

	Customer_Key	Customer_id	Customer_Number	Customer_FirstName	Customer_LastName	Birthdate	Gender	Marital_Status	Country	Create_Date
1	1	11000	AW00011000	Jon	Yang	1971-10-06	Male	Married	Australia	2025-10-06
2	2	11001	AW00011001	Eugene	Huang	1976-05-10	Male	Married	Australia	2025-10-06
3	3	11002	AW00011002	Ruben	Torres	1971-02-09	Male	Married	Australia	2025-10-06
4	4	11003	AW00011003	Christy	Zhu	1973-08-14	Female	n/a	Australia	2025-10-06
5	5	11004	AW00011004	Elizabeth	Johnson	1979-08-05	Female	n/a	Australia	2025-10-06

	Product_Key	Product_Id	Product_Number	Product_Name	Category_Id	Category	Subcategory	Maintance	Cost	Product_Line	Start_Date
1	1	211	FR-R92R-58	HL Road Frame - Red- 58	CO_RF	Components	Road Frames	Yes	0	Road	2003-07-01
2	2	210	FR-R92B-58	HL Road Frame - Black...	CO_RF	Components	Road Frames	Yes	0	Road	2003-07-01
3	3	291	FR-M94S-44	HL Mountain Frame - Sil...	CO_MF	Components	Mountain Fr...	Yes	707	Mountain	2011-07-01
4	4	274	FR-R72R-44	ML Road Frame - Red- 44	CO_RF	Components	Road Frames	Yes	352	Road	2011-07-01
5	5	275	FR-R72R-48	ML Road Frame - Red- 48	CO_RF	Components	Road Frames	Yes	352	Road	2011-07-01

	Order_Number	Product_Key	Customer_Key	Order_Date	Ship_Date	Due_Date	Sales	Quantity	Price
1	SO43697	31	10769	2010-12-29	2011-01-05	2011-01-10	3578	1	3578
2	SO43698	20	17390	2010-12-29	2011-01-05	2011-01-10	3400	1	3400
3	SO43699	20	14864	2010-12-29	2011-01-05	2011-01-10	3400	1	3400
4	SO43700	60	3502	2010-12-29	2011-01-05	2011-01-10	699	1	699
5	SO43701	20	4	2010-12-29	2011-01-05	2011-01-10	3400	1	3400

CONTACT



tsaqifu15@gmail.com



www.linkedin.com/in/tsaqifu-mu-tashim





THANK YOU