

# Data

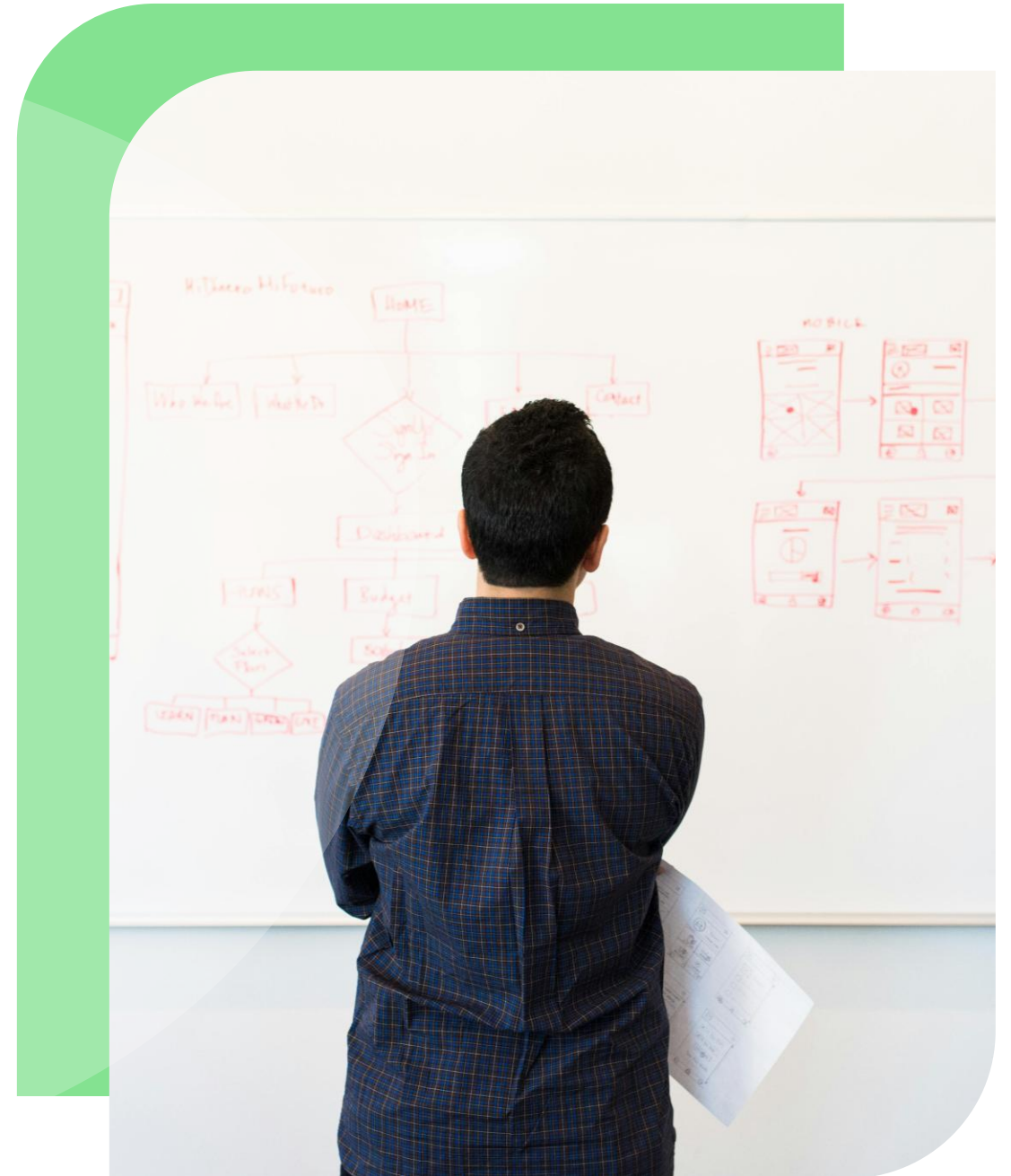
## Warehouse & Analytics

# Portfolio

Let's Connect



Elfrid Hasman



# Project Overview

**Goal :** To demonstrate the ability to build a robust data warehouse (DW) using the Bronze, Silver, Gold architecture.

## Business Challenge

The raw data from various sources (CRM and ERP) is separate, inconsistent, and unstructured, making it difficult to use for reliable business analysis.

**Data issues:** Inconsistent formats ('F' vs. 'Female'), missing or invalid values (NULL), duplicate records, and inconsistent column naming.

## Solution & Architecture

Building a Data Warehouse using the medallion architecture to systematically transform data into a clean, unified source of truth.

# Bronze Layer: Raw Data

## Description:

This layer is the "landing zone" for raw data. The data is stored exactly as it came from the source, without any modifications.

## EL (Extract & Load) Process:

Leverages the high-performance `BULK INSERT` command to efficiently load data from CSV files into the database.

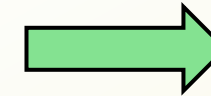
## Code Highlight (`proc_load_bronze.sql`):

Show the `BULK INSERT` syntax for various tables (`crm_cust_info`, `erp_cust_az12`, etc.), emphasizing its speed and efficiency for large datasets.

## Source Data



source\_crm



cust\_info



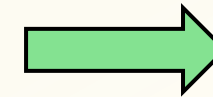
prd\_info



sales\_details



source\_erp



CUST\_AZ12



LOC\_A101



PX\_CAT\_G1V2

# Silver Layer: Trusted & Cleaned Data

This is the "trusted layer" where raw data from the Bronze layer is cleaned, validated, and standardized.

## Data Cleansing

**TRIM()** to remove excess spaces

**CASE WHEN** to normalize values ('S' to 'Single'), and **ISNULL()** to handle nulls.

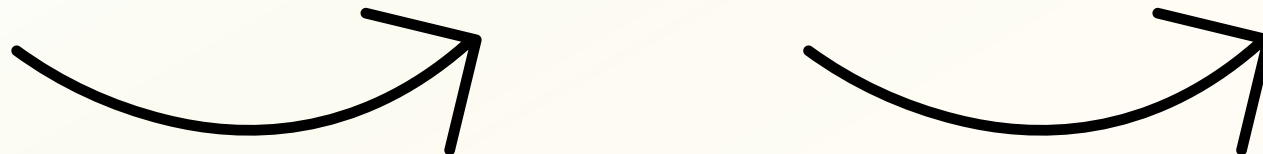
## Data Validation

**LEN()** and **CASE WHEN** to fix invalid dates, and recalculating values like `sls_sales` if invalid.

## Data Enrichment

Using **SUBSTRING()** to derive new values and

**ROW\_NUMBER()** to handle duplicates by selecting the most recent record.



# Gold Layer: Business- Ready Data

This is the "**presentation layer**" where data is modeled into a Star Schema, optimized for fast and intuitive analytics.

## Concepts :

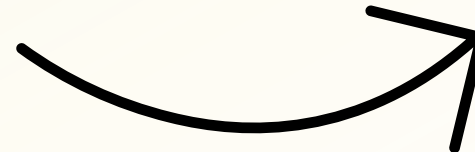
Explaining the difference between Dimension Tables (who, what, where) and Fact Tables (how much, when).

## Building Dimensions

- **gold.dim\_customers** : A Surrogate Key (customer\_key) is created using ROW\_NUMBER() to uniquely identify each customer. Data from CRM and ERP sources is unified using JOIN and COALESCE to handle inconsistencies in fields like gender.
- **gold.dim\_products** : A product\_key is created to link product details, categories, and subcategories from various source tables.

## Building Fact Table

- **gold.fact\_sales** : This view contains key sales metrics (sales\_amount, quantity, price) and foreign keys (customer\_key, product\_key) that connect to the dimension tables.
- This structure allows for slicing and dicing sales data by various customer and product attributes.



# Business Impact

## Project Summary

Successfully transformed fragmented, raw data into a structured and reliable single source of truth.

---

Emphasis on data quality and validation at every stage, from Bronze to Gold.

---

The resulting Star Schema provides an efficient foundation for business intelligence and reporting.

## Business Benefits

01

### Faster & More Accurate Analytics

Business users can directly query clean, structured data, eliminating the need for manual data preparation.

02

**Deeper Insights:** Enables richer analysis, such as:

- Sales trends over time.
- Customer behavior analysis based on demographics.
- Product performance tracking.

02

### Data-Driven Decisions

The reliable data empowers stakeholders to make strategic decisions with confidence.



# Acknowledgements



This project was developed by following the excellent tutorial and instructions provided by [Data with Bara](#).



A big thank you for the detailed and hands-on guidance in building this data warehouse from scratch.



SQL Data Warehouse from Scratch | Full Hands-On  
Data Engineering Project

# Thank You

## Project Details



You can find all project details, including the code and documentation, on [my GitHub repository](#).

## Let's Connect



Elfrid Hasman



---

Elfrid Hasman