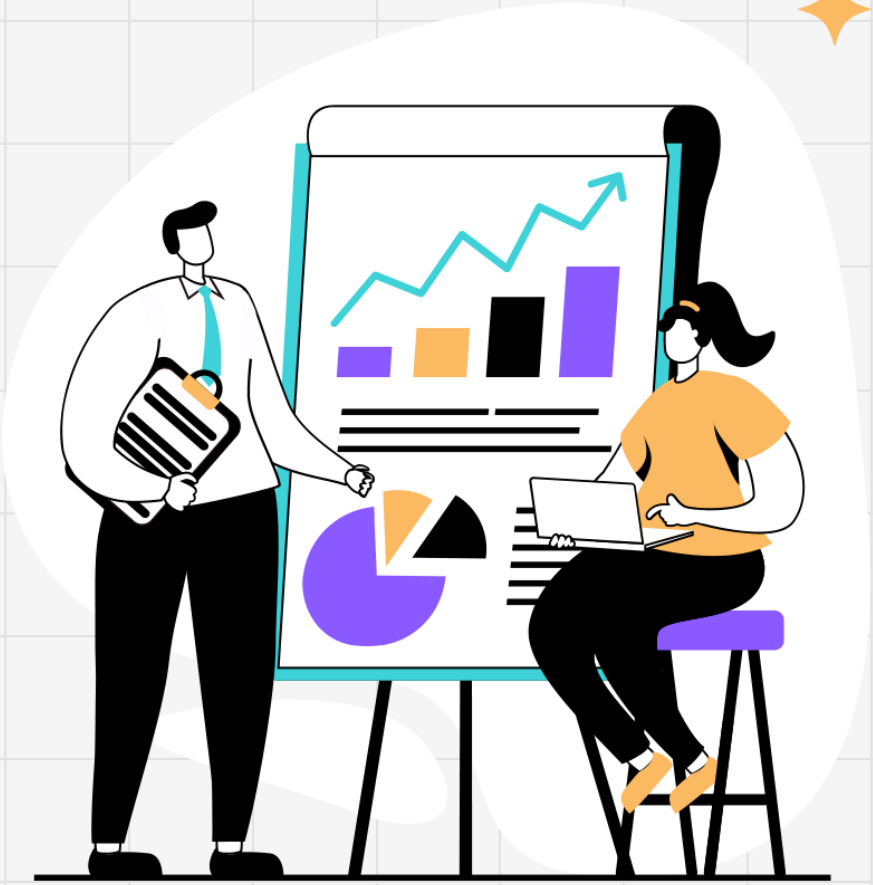
Binisha Maharjan

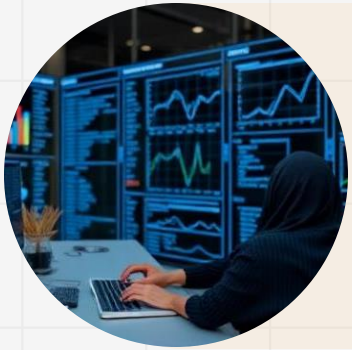
ETL SYSTEM FOR SALES, CUSTOMER,PRODUCT & INVENTORY MANAGEMENT



INTRODUCTION

- ETL system has been designed to automate the daily processing of Sales, Customer, Product and inventory data .
- Workflow ensures that the raw data from flat files is cleaned ,transformed, validated, and loaded into a structured data warehouse.

PROBLEM STATEMENT



Current State

Business lacks automated system for daily data ingestion, processing, and historical tracking of key operational data.

Need

ETL solution required to automate daily workflow for accurate reporting, analytics on sales trends, inventory tracking, and product performance.

OBJECTIVES

Develop ETL Solution

Implement SSIS to automate daily ingestion and processing of Customer, Product, and Sales data feeds.

Ensure Data Quality

Implement deduplication for Sales data and use configurable file paths for reliable data processing.

Implement Historical Tracking

Maintain full historical tracking for Customer and Product changes using SCD Type 2 logic.

Manage Inventory

Maintain daily Inventory table with correct Beginning On Hand and End On Hand calculations.

TOOLS AND TECHNOLOGY

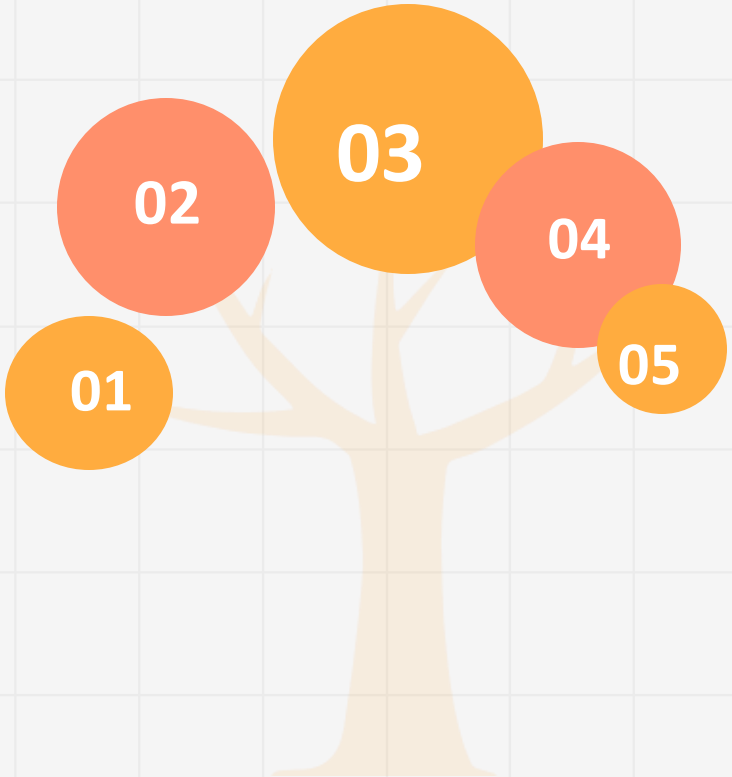
01 SSMS 

02 Visual Studio 

03 Copilot 

04 Excel 

IMPLEMENTATION PLAN



Source to Destination

Flat File Source transforms to OLE DB Destination through SSIS components.

Incremental load

To append the data from the SaleDate greater than the last_run_date

SCD Implementation

Custom logic tracks changes with StartDate, EndDate, and IsActive flags for historical data.

Product Sales Metrics

Calculate total_qty_sold for each product cumulatively across all transactions

Currency Conversion

SQL Functions convert USD to NPR .

DIFFERENT TABLES IN ETL SYSTEM

Stage tables

- stg_customers
- Stg_products
- Stg_Sales

Dimension Tables

- dim_customers
- dim_products

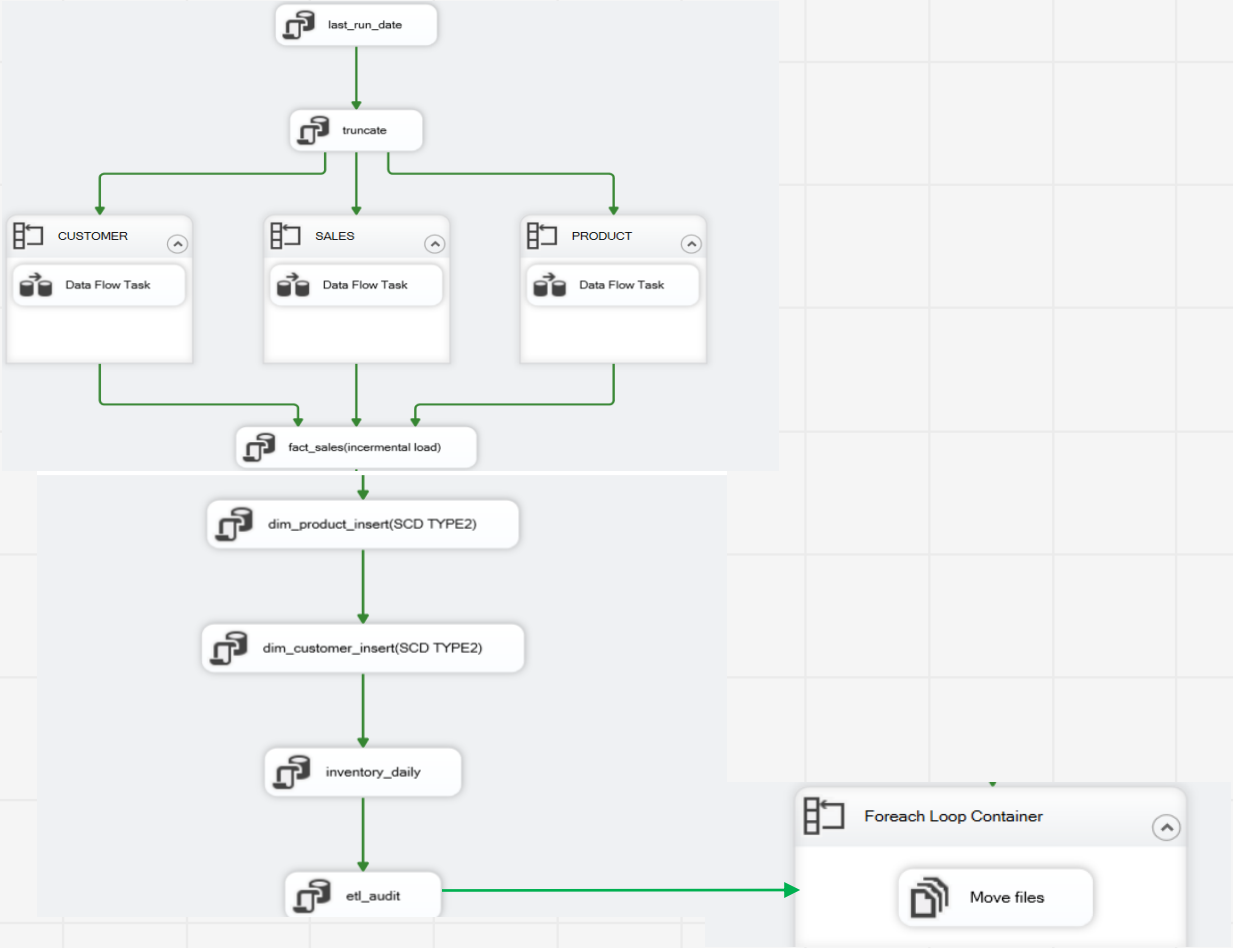
Fact Tables

- Inventory_daily
- Fact_sales

Audit Tables

- Etl_audit

CONTROL FLOW



SCD TYPE 2 IMPLEMENTATION

	customer_sk	customer_id	first_name	last_name	email	city	total_amt_spent	is_active	start_date	end_date	LastUpdatedDateTime
1	1	101	Isha	Shrestha	isha.shrestha@example.com	Lalitpur	NULL	0	2024-01-10 10:15:00.000	2025-12-07 08:02:42.053	2025-12-07 08:02:42.0533333
2	2	102	Sonu	KC	sonu.kc@example.com	Lalitpur	NULL	1	2024-01-11 09:12:00.000	NULL	2025-12-07 08:01:01.2633333
3	3	103	Bhuwan	Adhikari	bhuwan.adhikari@example.com	Jhapa	NULL	0	2024-01-12 11:25:00.000	2025-12-07 08:02:42.053	2025-12-07 08:02:42.0533333
4	4	104	Pragya	Bhattarai	pragya.bhattarai@example.com	Biratnagar	NULL	1	2024-01-13 13:45:00.000	NULL	2025-12-07 08:01:01.2633333
5	5	105	Norgen	Moktan	norgen.moktan@example.com	Chitwan	NULL	1	2024-01-14 15:00:00.000	NULL	2025-12-07 08:01:01.2633333
6	6	106	Anisha	Thapa	anisha.thapa@example.com	Dharan	NULL	1	2024-01-15 12:10:00.000	NULL	2025-12-07 08:01:01.2633333
7	7	107	Bibek	Lama	bibek.lama@example.com	Hetauda	NULL	1	2024-01-16 14:20:00.000	NULL	2025-12-07 08:01:01.2633333
8	8	108	Pooja	Maharjan	pooja.maharjan@example.com	Bhaktapur	NULL	1	2024-01-17 16:30:00.000	NULL	2025-12-07 08:01:01.2633333
9	9	109	Nirajan	Rai	nirajan.rai@example.com	Janakpur	NULL	1	2024-01-18 09:50:00.000	NULL	2025-12-07 08:01:01.2633333
10	10	110	Smriti	Karki	smriti.karki@example.com	Butwal	NULL	1	2024-01-19 08:40:00.000	NULL	2025-12-07 08:01:01.2633333
11	11	101	Isha	Shrestha	isha.shrestha@example.com	Kathma...	NULL	1	2024-01-20 10:15:00.000	NULL	2025-12-07 08:02:42.0566667
12	12	103	Bhuwan	Adhikari	bhuwan.adhikari@example.com	Morang	NULL	1	2024-01-22 11:25:00.000	NULL	2025-12-07 08:02:42.0566667

Historical Data Tracking

Uses StartDate, EndDate, and IsActive flags to maintain complete history of customer and product changes.

Update Process

Sets IsActive=0 and EndDate for old records, inserts new records with IsActive=1 when changes occur.

Record Discontinuation

Records discontinuation by setting IsActive=0 and EndDate to current load date.

Change Management

Preserves historical data while maintaining current state for accurate reporting and analysis.

INVENTORY LOGIC

	inventory_date	product_id	boh	eoh
1	2025-12-07	1001	100	84
2	2025-12-07	1002	100	89
3	2025-12-07	1003	100	83
4	2025-12-07	1004	100	90
5	2025-12-07	1005	100	85
6	2025-12-07	1006	100	80
7	2025-12-07	1007	100	86
8	2025-12-07	1008	100	75
9	2025-12-07	1009	100	92
10	2025-12-07	1010	100	89
11	2025-12-07	1011	100	89
12	2025-12-07	1012	100	91
13	2025-12-07	1013	100	85
14	2025-12-07	1014	100	80

	inventory_date	product_id	boh	eoh
21	2025-12-08	1001	84	68
22	2025-12-08	1002	89	78
23	2025-12-08	1003	83	66
24	2025-12-08	1004	90	80
25	2025-12-08	1005	85	70
26	2025-12-08	1006	80	60
27	2025-12-08	1007	86	72
28	2025-12-08	1008	75	50
29	2025-12-08	1009	92	84
30	2025-12-08	1010	89	78
31	2025-12-08	1011	89	78
32	2025-12-08	1012	91	82
33	2025-12-08	1013	85	70
34	2025-12-08	1014	80	60

Daily Inventory Calculation

EOH = BOH – Total_quantity_sold
BOH derived from previous day's EOH.

Historical Tracking

System maintains daily inventory records for each product, product, enabling trend analysis and stock level monitoring. monitoring.

INCREMENTAL LOAD PROCESS

	product_sk	sale_id	product_id	customer_id	qty	Price	Total_amt	sale_date	LoadedDateTime
1	1	100	1004	101	3	48	20448	2025-12-07	2025-12-07 08:01:01.1500000
2	2	101	1008	104	2	82	23288	2025-12-07	2025-12-07 08:01:01.1500000
3	3	102	1004	109	1	48	6816	2025-12-07	2025-12-07 08:01:01.1500000
4	4	103	1019	107	1	6	852	2025-12-07	2025-12-07 08:01:01.1500000
5	5	104	1001	102	2	420	119280	2025-12-07	2025-12-07 08:01:01.1500000

	product_sk	sale_id	product_id	customer_id	qty	Price	Total_amt	sale_date	LoadedDateTime
101	101	100	1004	101	3	48	20448	2025-12-08	2025-12-07 08:02:41.9633333
102	102	101	1008	104	2	82	23288	2025-12-08	2025-12-07 08:02:41.9633333
103	103	102	1004	109	1	48	6816	2025-12-08	2025-12-07 08:02:41.9633333
104	104	103	1019	107	1	6	852	2025-12-08	2025-12-07 08:02:41.9633333
105	105	104	1001	102	2	420	119280	2025-12-08	2025-12-07 08:02:41.9633333

It stores all sales events by appending the data of SalesDate > Last_run_date

PACKAGE FAILURE HANDLING

	audit_id	package_name	run_start	run_end	status	error_message
1	1	Package	2025-11-18 00:00:00.000	2025-11-18 00:00:00.000	Success	NULL
2	2	Package	2025-12-07 08:01:01.340	2025-12-07 08:01:01.340	Success	NULL
3	3	Package	2025-12-07 08:02:42.120	2025-12-07 08:02:42.120	Success	NULL
4	4	Package	2025-12-07 08:25:21.133	2025-12-07 08:25:21.133	Fail	Executing the query "select max(run_end) from ...
5	5	Package	2025-12-07 08:25:55.740	2025-12-07 08:25:55.740	Success	NULL
6	6	Package	2025-12-08 11:12:38.523	NULL	Fail	Executing the query "select max(run_end) from ...

1

Robust logging system

Tracks operations, status, and errors across all ETL packages.

2

Event handlers

Uses SSIS Error_Description and Package_name as variable mapping.

3

ETL_Log table

Records package execution details including success/failure status and error messages.

4

Copy Files

Copies files to the error-handler folder if failure in the execution occurs

DATA VALIDATION TESTS

File Handling

Verify successful data loading from configurable path.

Incremental Loading

Ensure new transactions load.

SCD2 Validation

Confirm updates/insertions correctly manage rows in dimension tables.

Data Transformation

Check accuracy of USD to NPR conversion in sales facts.

LEARNINGS

- ETL design pattern in SSIS
- Concept of SCD Type 2
- Concept of Incremental load