

Project Summary: Automated Sales Reporting Pipeline

1. Project Objective and Overview

The goal of this project was to eliminate a time-consuming and error-prone manual process for reporting sales data, which was stored across a large-volume **Google Sheet** (over 250,000 records).

The solution was a **fully automated, end-to-end data pipeline** designed to extract live data, ensure transactional integrity, and deliver a self-refreshing dashboard in Power BI. This reduced data latency from 2 days to under 10 minutes.

2. Technical Implementation and Flow

This three-tiered pipeline ensures data moves securely and is transformed correctly at each stage.

Component	Technology	Specific Functionality
Data Extraction (E)	Python (Pandas, gspread)	Securely read all 250K+ records from Google Sheets and performed primary data cleansing (e.g., date standardization).
Data Transformation & Load (T/L)	Python (pyodbc) & SQL Staging	Used Python's batch loading (chunking) capabilities to efficiently push data into a temporary SQL Server staging table, preventing system overload.
Data Integrity Logic	SQL Server Express (MERGE)	A dedicated SQL Stored Procedure ran the MERGE statement based on OrderID to perform reliable Upserts (Update if record exists, Insert if new) into the final Sales_Dataset

		table.
Automation Scheduling	Windows Task Scheduler	Configured to trigger the Python script and SQL Stored Procedure automatically, eliminating all manual steps.
Visualization & Reporting	Power BI Desktop/Service	Created a dynamic sales dashboard, linked to the live SQL table via the On-premises Data Gateway for cloud refresh.

3. Key Achievements and Impact

The successful implementation of this pipeline provided immediate, quantifiable benefits:

- **Automation:** Eliminated approximately **4 hours of manual labor per day** previously spent on data extraction, cleaning, and reconciliation.
- **Speed:** Reduced the **data latency from 24-48 hours to less than 10 minutes**, enabling near real-time decision-making for executive teams.
- **Data Reliability:** Achieved **guaranteed data consistency and accuracy** using the SQL MERGE logic, resolving issues with duplicate transactions and missed record updates.
- **Deliverable:** Provided an interactive, automatically refreshing **Power BI dashboard** that visualizes sales trends, profitability, and regional performance.

The project successfully transformed a manual data bottleneck into a reliable, automated business intelligence asset.