# Storage Verification and Orchestration Report

## Section 1: Proof of Data Ingestion

To verify successful data flow across the ETL pipeline, row counts were recorded for each stage:

**Verification query-1:**

* From: **GitHub**

To: **Bronze**

**A screenshot of a computer

AI-generated content may be incorrect.**

A screenshot of a computer

AI-generated content may be incorrect.

Querying File from Databricks Notebook: **bronze\_silver**

A screenshot of a computer

AI-generated content may be incorrect.

Total Records: 253680

A screenshot of a computer

AI-generated content may be incorrect.

In addition, the presence of files was verified in the respective directories for /mnt/diabetes/bronze/, /mnt/diabetes/silver/, and /mnt/diabetes/gold/.

**Verification query-2:**

* From: **Bronze**

To: **Silver**

Writing Data to Silver Container

**A screenshot of a computer

AI-generated content may be incorrect.**

Data is written Successfully in Silver container of Azure Data Lake gen2 storage

A screenshot of a computer

AI-generated content may be incorrect.

Verifying through a query from Azure DataBricks

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

Total Records in Silver: **229193**

**Verification query-3:**

* From: **Silver**

To: **Gold**

Notebook: **silver\_to\_gold**  
A screenshot of a computer code

AI-generated content may be incorrect.

The Data is written to gold in delta format with AgeGroup as partition column

Verifying data at Gold Layer

A screenshot of a computer

AI-generated content may be incorrect.

Verifying through a query from Azure DataBricks Notebook (gold\_to\_dashboard)

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

## Section 2: ETL Orchestration in Azure Data Factory

Azure Data Factory (ADF) was used to orchestrate the complete ETL flow. The pipeline consists of four steps:  
1. Copy raw data to Bronze  
2. Transform Bronze ➜ Silver using `bronze\_to\_silver` notebook  
3. Transform Silver ➜ Gold using `silver\_to\_gold` notebook  
4. Generate visual insights using `gold\_to\_dashboard` notebook  
  
Each step is linked and scheduled sequentially in ADF. The screenshot below shows successful execution.

A screenshot of a computer

AI-generated content may be incorrect.

## Section 3: Summary

All datasets were successfully processed and written to their respective layers. The orchestration via ADF ensures repeatability, scalability, and automation of the entire pipeline from ingestion to dashboard delivery.