HLD for Teradata to Snowflake Migration

|  |  |
| --- | --- |
| Author: | Shajahan Hussain |
| Classification: | Internal |
| Date: | September 2025 |
| Approved by: | <approver> |

INTELLECTUAL PROPERTY RIGHTS

CONFIDENTIAL & PROPRIETARY

This document is confidential and is furnished under license and may be used only in accordance with the terms of that license and with the inclusion of the copyright notice set forth below.

COPYRIGHTS

Copyright © 2020 Elait Ltd. All rights reserved.

Reproduction, adaptation, or translation without prior written permission is prohibited, except as allowed under copyright law or license from Elait Ltd.

LICENSING

This document is licensed for sole use on the Teradata to Snowflake migration POC.

**Version Control**

| **Author** | **Date** | **Version No.** | **Details** |
| --- | --- | --- | --- |
| Shajahan Hussain | 15/09/2025 | 0.1 | Introduction, Architecture, workflow and tools & technologies |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Related documents**

|  |  |
| --- | --- |
| **Title** | **Author** |
|  |  |

**Distribution List**

|  |  |  |
| --- | --- | --- |
| **Name** | **Email Address** | **Action (\*)** |
| Sreeni Rao | Sreeni.rao@elait.com | Review |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

(\*) Action types: Approve, Inform and Review

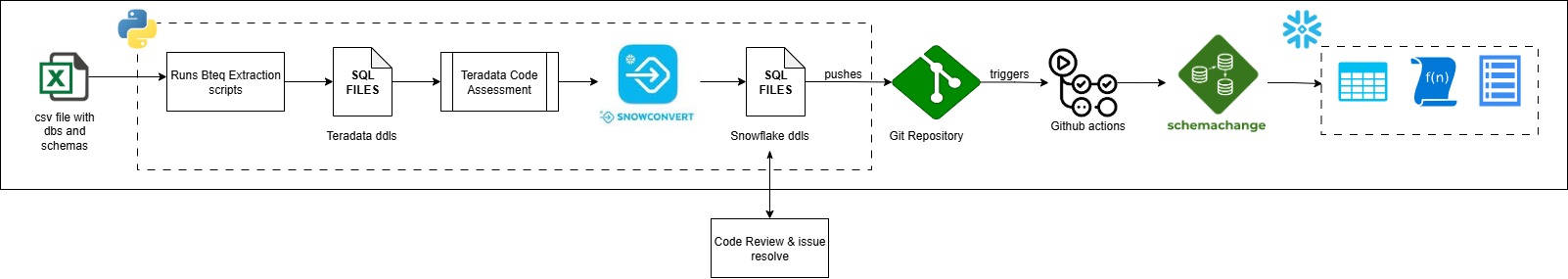
# Introduction

This document outlines the **high-level design (HLD)** for the Teradata to Snowflake migration project.  
It describes the migration of **database objects** (tables, views, functions, procedures) and **data**.  
The process leverages automation for **DDL extraction, conversion, and deployment**.  
It also includes **data migration pipelines** for exporting from Teradata and loading into Snowflake.

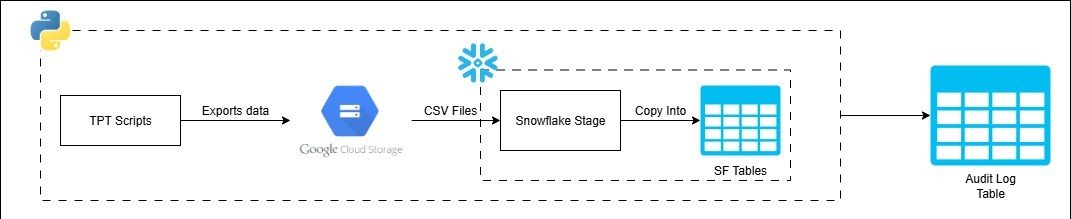
# Architecture Overview

The migration consists of two main components:  
1. DDL Migration – Extract, convert, and deploy Teradata objects to Snowflake.  
2. Data Migration – Export data from Teradata, stage in GCS, and load into Snowflake.

1. DDL Migration Flow



1. Data Migration Flow



# DDL Migration

## DDL Extraction

* + Python script triggers BTEQ scripts to extract Teradata DDLs.
  + DDLs are pulled for all objects (tables, views, functions, procedures).
  + Extracted files are stored in a structured directory for further processing.

## DDL Conversion

* + Teradata DDLs are processed through **SnowConvert** for automated translation.
  + Conversion ensures compatibility with Snowflake SQL syntax and features.
  + Generates a new set of Snowflake-ready .sql files.

## DDL Refinement

* + Converted DDLs are analysed for unsupported syntax or incompatibilities.
  + **SnowConvert AI assistant** helps in identifying and correcting issues.
  + All corrections are handled manually by developers.
  + Ensures that final DDLs are fully aligned with Snowflake standards before deployment

## Push to Git

* + Refined SQL files are committed and pushed into the central Git repository.
  + Version control enables collaboration and rollback if needed.
  + Each commit is associated with a branch or merge request for controlled integration.

## GitHub Actions (Schemachange Deployment)

* + Merge request approval triggers **GitHub Actions workflow**.
  + Workflow executes **schemachange** to deploy Snowflake objects.
  + Deployment includes tables, views, functions, and stored procedures.
  + Logs and deployment reports are generated for audit and traceability.

# Data Migration

## Data Export

* + Python script orchestrates **TPT scripts** to export data from Teradata.
  + Data is extracted in parallel for performance optimization.
  + Exported data is stored as CSV files in **Google Cloud Storage (GCS)**.

## Copy into Snowflake

* + Snowflake external stage points to the GCS bucket.
  + COPY INTO command loads CSV data into Snowflake tables.
  + Load is validated, and audit log tables capture record counts.
  + Ensures consistency between source and target datasets.

## Audit table Structure

The following SQL statement defines the structure of the audit log table:

*create or replace TABLE ADVENTUREWORKSDW.PUBLIC.AUDIT\_LOG (*

*ID NUMBER(38,0),*

*DATABASE\_NAME VARCHAR(16777216),*

*SCHEMA\_NAME VARCHAR(16777216),*

*TABLE\_NAME VARCHAR(16777216),*

*TD\_ROW\_COUNT NUMBER(38,0),*

*SF\_ROW\_COUNT NUMBER(38,0)*

*);*

# Tools & Technologies

## Teradata

* BTEQ – used for extracting DDLs.
* TPT – used for exporting data in parallel for high performance.

## Conversion

* SnowConvert – automatically converts Teradata SQL to Snowflake-compatible SQL.
* SnowConvert AI – assists in identifying conversion issues and suggesting resolutions.

## Version Control & CI/CD

* Git – manages version control of converted DDL files.
* GitHub Actions – automates workflows for deployment.

## Deployment

* Schemachange – manages and deploys Snowflake objects (tables, views, functions, procedures).

## Cloud Storage

* Google Cloud Storage (GCS) – used as staging area for Teradata data exports.

## Target Data Warehouse

* Snowflake – cloud data warehouse where converted objects and migrated data are deployed.

## Automation Scripts

* Python Script #1 – automates DDL extraction from Teradata.
* Python Script #2 – orchestrates TPT data export and manages Snowflake data loading.

# RAIDs

This section describes any risks, assumptions, issues, and dependencies where these might impact delivery timings, functionality or have a reliance on external factors (especially software, licenses, other deliveries, or teams).

## Risks

| **ID** | **Description & Impact** | **Severity** | **Likelihood** | **Owner** | **Mitigating Actions** |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

## Assumptions

| **ID** | **Description & Impact** | **Confidence** | **Owner** | **Mitigating Actions** |
| --- | --- | --- | --- | --- |
| A1 |  |  |  |  |
| A2 |  |  |  |  |

## Issues

| **ID** | **Description & Impact** | **Severity** | **Likelihood** | **Owner** | **Mitigating Actions** |
| --- | --- | --- | --- | --- | --- |
| I1 | None |  |  |  |  |

## Dependencies

| **ID** | **Description & Impact** | **Confidence** | **Owner** | **Mitigating Actions** |
| --- | --- | --- | --- | --- |
| D1 |  |  |  |  |
| D2 |  |  |  |  |