# **Introduction**

This document outlines the functional requirements for the data pipeline involving SnowSQL for data ingestion, Snowflake for data engineering, and PowerBI for data visualization. The system aims to handle data loading, transformation, and visualization from January 2023 to January 2024.

# **Pre-requisites**

1. Snowflake Training

* Snowflake badges

|  |
| --- |
| <https://learn.snowflake.com/en/courses/uni-essdww101/> |
| <https://learn.snowflake.com/en/courses/uni-ess-dlkw/> |
| <https://learn.snowflake.com/en/courses/uni-ess-dngw/> |
| <https://learn.snowflake.com/en/courses/uni-lvlup-101/> |
| <https://learn.snowflake.com/en/courses/uni-lvlup-104/> |
| <https://learn.snowflake.com/en/courses/uni-lvlup-106/> |
| <https://learn.snowflake.com/en/courses/uni-lvlup-107/> |
| <https://learn.snowflake.com/en/courses/uni-lvlup-108/> |
| <https://learn.snowflake.com/en/courses/uni-lvlup-109/> |
| <https://learn.snowflake.com/en/courses/uni-lvlup-201/> |
| <https://learn.snowflake.com/en/courses/uni-lvlup-202/> |

1. PowerBI Training

[PowerBI Training](https://www.youtube.com/watch?v=5X5LWcLtkzg&t=4507s&pp=ygU5SGFuZHMtT24gUG93ZXIgQkkgVHV0b3JpYWw6IEJlZ2lubmVyIHRvIFBybyBbRnVsbCBDb3Vyc2Vd)

# **Data Ingestion**

1. **Data Ingestion (Bulk)**

* **Purpose**: To ingest data into Snowflake from local to snowflake.
* **Activity**: Perform bulk data loading using SnowSQL.
* **Scope of Data**: From January 2023 to December 2023.

1. **Chart**



# **Transformations**

1. **Source\_Table**

* **Description**: Initial staging area for raw data ingestion.
* **Data Types**: All fields loaded as VARCHAR.

1. **Staging\_Table**

* **Description**: Transformation stage for data type conversions.
* **Data Types**:
  + **DATE\_BOOKED**: VARCHAR
  + **ORIGIN**: VARCHAR
  + **DESTINATION**: VARCHAR
  + **ORDER\_REF**: VARCHAR
  + **TICKET\_NO**: VARCHAR
  + **SEATNO**: NUMBER
  + **DATE\_REDEEMED**: VARCHAR
  + **EMAIL**: VARCHAR
  + **MOBILENO**: VARCHAR
  + **FARE**: NUMBER
  + **CONVENIENCE\_FEE**: NUMBER
  + **DISCOUNT**: VARCHAR
  + **DEPARTURE\_DATE**: VARCHAR
  + **DEPARTURE\_TIME**: TIME
  + **BUS\_TYPE**: VARCHAR
  + **NUMBER\_OF\_VOUCHERS\_BOOKED**: VARCHAR
  + **BOOK\_DT**: DATE
  + **BOOK\_TM**: TIME
  + **DEPARTURE\_DAY**: VARCHAR
  + **BOOKING\_INFO**: VARCHAR
  + **VOUCHER**: VARCHAR
  + **ROUTE**: VARCHAR
  + **TRIP\_ID**: VARCHAR
  + **TRIP\_ID\_DAY**: VARCHAR
  + **REDEEMED\_FLAG**: VARCHAR

1. **Analytical\_Business\_Table**

* **Description**: Further transformation stage with calculations or formulas applied.
* **New Fields**:
  + **BOOK\_DT** – Date from Date\_Booked (MM/DD/YY)
  + **BOOK\_TM** – Time from Date\_Booked (HH24.MI.SS)
  + **DEPARTURE\_DAY** – First 3 letters of the day in Departure\_Date
  + **BOOKING\_INFO** – If Order\_Ref contains “OFF”. Booking is Offline, else online.
  + **VOUCHER** = Use the **first number** in NUMBER\_OF\_VOUCHERS\_BOOKED
    - **1** - Individual
    - **2** - Pair
    - **3 or more** - Group
  + **ROUTE** - Concatenate ‘Origin’ ‘to’ ‘Destination’ (Alabang to Santa Rosa)
  + **TRIP\_ID** - Concatenate the below. Expected format is PAS-BAG-SAT-21
    - **Origin** – first 3 letters
    - **Destination** - first 3 letters
    - **Day** - first 3 letters
    - **Departure Time** - 24 hr format & first 2 numbers
  + **TRIP\_ID\_DAY** – Concatenate the below. Expected format is PAS-BAG-SAT.
    - **Origin** – first 3 letters
    - **Destination** - first 3 letters
    - **Day** - first 3 letters
  + **REDEEMED\_FLAG** - If Date\_Redeemed =is Unredeemed then N else Y
* **Data Types**: Same in Staging\_Table

1. **Chart**

****

# **Operationalization**

1. **Data Ingestion (Daily)**

* **Purpose**: To ingest data into Snowflake from local to snowflake.
* **Activity**: Perform **daily** data loading using SnowSQL and Windows Task Scheduler.
* **Scope of Data**: January 2024

1. **Automated Transformation Inside Snowflake**

* **Purpose**: Automate transformation of data once Snowflake ingests the data (**daily**).
* **Activity**: Create a Task in Snowflake that appends the ingested data (daily) in the Source\_Table and converts the data types in the Staging\_Table and perform business requirements in the Analytical\_Business\_Table.
* **Scope of Data:** January 2024

1. **Chart**

**A diagram of snowflakes

Description automatically generated**

# **Visualization and Insights Generation**

Resources should be able to present visualization and insights generated from the data given.

1. **Visualization**

* **Tool**: Use PowerBI Desktop for visualization.
* **Data**: Use the transformed final data (Analytical Base Table).
* **Requirements**: 3 Dashboards with a minimum of 5 elements and maximum of 8. Each dashboard requires KPI that needs to be part of the report.
  + Executive Summary Dashboard
    - Required KPI’s:
      * Total Sales
      * Total Tickets Sold
      * Monthly Sales
  + Trips Dashboard
    - Required KPI’s:
      * Top 5 Routes
      * Bottom 5 Routes
  + Sales Dashboard
    - Required KPI’s:
      * Online vs Offline Ticket Sales
      * Sales per hour of departure
      * Sales per day

1. **Insights Generation**

* Resource should be able to provide insights generated from the different visualizations.
* Problems that need to be solved:
  + The transport company wants to improve their sales. They need to come up with new campaign strategies to bump up sales.
* Insights can be of the following:
  + General insights
  + Marketing/campaign suggestions