**Lab Exercise 3 -dbt model using dbt core for Snowflake**

Here's the complete code for creating a **source database** (raw\_db) with a customers table and inserting records into it. We'll also create a **target database** (analytics\_db) for dbt to store the transformed data.



**Prerequisites**

1. **Install dbt**: Ensure dbt is installed. You can install it using:

pip install dbt-core

pip install dbt-snowflake

1. **Set Up dbt Project**: Create a new dbt project if you don’t already have one.

dbt init my\_snowflake\_project

**1. Set Up Source Database and Schema in Snowflake**

First, log in to Snowflake and run these SQL statements to create the **source database** (raw\_db), the **schema** (raw\_data), and the **table** (customers).

-- Create source database

CREATE DATABASE IF NOT EXISTS raw\_db;

-- Use the source database

USE DATABASE raw\_db;

-- Create schema in source database

CREATE SCHEMA IF NOT EXISTS raw\_data;

-- Create the source table in raw\_db.raw\_data schema

CREATE OR REPLACE TABLE raw\_data.customers (

customer\_id INT,

first\_name STRING,

last\_name STRING,

email STRING,

created\_at TIMESTAMP

);

-- Insert sample data into raw\_db.raw\_data.customers

INSERT INTO raw\_data.customers VALUES

(1, 'Alice', 'Smith', 'alice@example.com', '2024-01-01 10:00:00'),

(2, 'Bob', 'Johnson', 'bob@example.com', '2024-02-01 11:30:00'),

(3, 'Charlie', 'Williams', 'charlie@example.com', '2024-03-01 14:15:00');

This will create and populate raw\_db.raw\_data.customers, which will be our **source table** for dbt.

This will be the **target database and schema** where dbt will create its transformed models.

**2. Create the Model in dbt**

Create a model that pulls data from the source table in raw\_db and applies a simple transformation.

**customer\_summary.sql**

In ***models/customer\_summary.sql***, create a basic transformation model:

-- models/customer\_summary.sql

{{ config(

    materialized = 'table'

) }}

WITH base AS (

    SELECT

        customer\_id,

        first\_name,

        last\_name,

        email,

        created\_at

    FROM raw\_db.raw\_data.customers

)

SELECT

    customer\_id,

    first\_name,

    last\_name,

    email,

    created\_at,

    YEAR(created\_at) AS signup\_year

FROM base

WHERE created\_at >= '2024-01-01'

This query selects data from the raw\_db.raw\_data.customers table, adds a signup\_year column, and filters the records based on the created\_at date.

**3. Run dbt to Create the Model**

After setting up your dbt project and the above files, run the model with:

dbt run --select customer\_summary

**4. Verify the Results in Snowflake**

Check if the transformed model customer\_summary has been created in the **target database and schema** analytics\_db.transformed\_data:

SELECT \* FROM analytics.hks\_schema.customer\_summary;

**Summary of Database Setup**

* **Source Database**: raw\_db
  + **Source Schema**: raw\_data
  + **Source Table**: customers