E-Commerce Data Pipeline:

From Ingestion to Insights

How to Turn Data Into Insights From Ingestion?

Description:

This project aims to have a hands-on experience in seeing the full Data Engineering Lifecycle at play. For the purpose of this project we will act as engineers working for an e-commerce company. We will start with Account data in Salesforce. Salesforce is a leading CRM (Customer Relationship Manager) product, meaning it is used by hundreds of thousands of businesses around the world. The odds that we will work with salesforce data in our career is very high. We will then grab that data using an automated ETL tool (Airbyte) and bring that data into Snowflake, a popular cloud-based data warehousing platform. In order for our data to properly load into Snowflake we must create our database, schemas and tables. Once we have created our objects using SQL in Snowflake, we will surface our data into staging tables - clean the data as needed - and copy the data into Prod tables. Now that the data is ready to consume, we are going to create a connection to Snowflake from Tableau Desktop. We will also build a dashboard using the data in snowflake. The goal of this project is to simulate the end to end movement of data from ingestion to serving.

Objectives:

- Gain proficiency in using Snowflake for data warehousing and processing.
- Familiarize ourselves with automated ETL (Extract, Transform, Load) tools.
- Understand the concepts and principles of database objects (schemas, tables, etc)
- Develop skills in data transformation and manipulation.
- Practice data validation and error handling techniques.
- Explore best practices for serving data analytics and insights through Tableau

Technology Requirements:

- Salesforce (CRM Tool)
- Airbyte (ETL)
- Snowflake (Datawarehouse)
- Tableau Desktop (Visualization Tool)

Requirements:

1. **Data Generation**: Generate Account data in Mockaroo and load that data into the Salesforce account.
2. **Data Preparation**: Create and configure the database objects in Snowflake
    a. One Database
    b. Two Schemas (One Stage and One Prod)
    c. 3 Dimension Tables and One Fact Tables
3. **Data Loading**: Utilize Airbyte to extract data from Salesforce and load it into the stage tables in Snowflake.

4. **Data Transformation**: Implement transformation logic to cleanse, filter, aggregate, and enrich the data in the stage tables.
5. **Data Validation**: Validate data transformation and copy the clean data into the Prod tables.
6. **Data Serving**: Establish a connection to the Snowflake database through Tableau Desktop. Develop a Dashboard of our choosing that analyzes the data we have now ingested.
    a. One Dashboard
    b. Includes at least five different visualizations
    c. Properly used containers throughout the dashboard
    d. Ability to filter the entire dashboard by a dimension
    e. Use a parameter to toggle between two dimensions
7. **Documentation**: A final, formal document that includes; our analysis and insights from the data, the process of building the Tableau Dashboard and all SQL scripts used.

Deliverables:

- Part 1: Generating Data and Configuring Your Tools
    - Mockaroo CSV (Data Generation)
    - Screenshots of Salesforce data successfully loaded (Data Generation)
    - SQL script of the creation of the database objects in Snowflake (Data Preparation)
    - Screenshots showing successful data load into Snowflake from Airbyte (Data Loading)
- Part 2: Staging and Storing Data
    - SQL script to clean data and copy data into Production Tables (Data Transformation & Validation)
    - Screenshot showing successful Tableau connection to your DB
- Part 3: Serving Data, Analytics and Insights and Documentation
    - Tableau Dashboard following the above requirements (Data Serving)
    - A final formal document that includes all of our SQL scripts used and written observations from start to finish of the project. Use this as an opportunity to flex your analytics skills and present your insights and findings to me. (Documentation)