

Reflection

In this lab, I built a simple data pipeline using Snowflake and Snowpark. The goal was to take a CSV file containing Apple balance sheet data, load it into Snowflake, clean the data, and run SQL queries for analysis.

First, I read the CSV file into a pandas DataFrame inside a Snowflake Notebook. This allowed me to inspect the data before loading it into Snowflake. Then, using Snowpark, I created a session and wrote the DataFrame into a Snowflake table.

One issue I encountered was that the column names contained spaces and special characters. Because Snowflake treats quoted column names as case-sensitive, this caused SQL errors. To fix this, I cleaned the column names by converting them to uppercase and replacing spaces with underscores. After that, I saved the cleaned version as a new table.

Finally, I ran SQL queries to analyze financial metrics such as Net Debt, Total Debt, and Working Capital. This showed how raw CSV data can be transformed into structured data inside Snowflake and used for analysis.

Overall, this lab helped me understand how data ingestion, cleaning, and SQL analysis work together in a cloud data platform. It also helped me better understand Snowflake's handling of identifiers and how Snowpark integrates Python with SQL.