

DATA226HW5

Github link: <https://github.com/lea2105/data226hw>

Please see the GitHub dag as the process of full refresh using SQL and the code.

Q Search

DEV

ADHOC

ANALYTICS

CURATION

INFORMATION\_SCHEMA

PUBLIC

RAW

Tables

COUNTRY\_CAPITAL

LAB1\_STOCK\_DATA

PROB\_HST\_TBL

STOCK\_DATA

STOCK\_DATA\_FORECAST

TEMP\_COUNTRY\_CAPITAL

TEMP\_STOCK\_DATA

Views

Stages

Streams

Tasks

DEV1

SNOWFLAKE

SNOWFLAKE\_SAMPLE\_DATA

DEV / RAW / STOCK\_DATA

Table ACCOUNTADMIN 22 hours ago 90 3.5KB

Table Details Columns Data Preview Copy History Lineage

COMPUTE\_WH 90 Rows Updated just now

|    | SYMBOL | OPEN   | CLOSE  | HIGH   | LOW    | VOLUME   | DATE       |
|----|--------|--------|--------|--------|--------|----------|------------|
| 1  | AAPL   | 237.71 | 235.93 | 240.07 | 234.68 | 53798062 | 2025-03-04 |
| 2  | AAPL   | 241.79 | 238.03 | 244.03 | 236.11 | 47183985 | 2025-03-03 |
| 3  | AAPL   | 236.95 | 241.84 | 242.09 | 230.20 | 56833360 | 2025-02-28 |
| 4  | AAPL   | 239.41 | 237.30 | 242.46 | 237.06 | 41153639 | 2025-02-27 |
| 5  | AAPL   | 244.33 | 240.36 | 244.98 | 239.13 | 44433564 | 2025-02-26 |
| 6  | AAPL   | 248.00 | 247.04 | 250.00 | 244.91 | 48013272 | 2025-02-25 |
| 7  | AAPL   | 244.93 | 247.10 | 248.86 | 244.42 | 51326396 | 2025-02-24 |
| 8  | AAPL   | 245.95 | 245.55 | 248.69 | 245.22 | 53197431 | 2025-02-21 |
| 9  | AAPL   | 244.94 | 245.83 | 246.78 | 244.29 | 32316907 | 2025-02-20 |
| 10 | AAPL   | 244.66 | 244.87 | 246.01 | 243.16 | 32204215 | 2025-02-19 |
| 11 | AAPL   | 244.15 | 244.47 | 245.18 | 241.84 | 48822491 | 2025-02-18 |
| 12 | AAPL   | 241.25 | 244.60 | 245.55 | 240.99 | 40896227 | 2025-02-14 |
| 13 | AAPL   | 236.91 | 241.53 | 242.34 | 235.57 | 53614054 | 2025-02-13 |
| 14 | AAPL   | 231.20 | 236.87 | 236.96 | 230.68 | 45243292 | 2025-02-12 |
| 15 | AAPL   | 228.20 | 232.62 | 235.23 | 228.13 | 53718362 | 2025-02-11 |
| 16 | AAPL   | 229.57 | 227.65 | 230.59 | 227.20 | 33115645 | 2025-02-10 |
| 17 | AAPL   | 232.60 | 227.63 | 234.00 | 227.26 | 39707224 | 2025-02-07 |
| 18 | AAPL   | 231.29 | 233.22 | 233.80 | 230.43 | 29925349 | 2025-02-06 |

Q Search

DEV

ADHOC

ANALYTICS

CURATION

INFORMATION\_SCHEMA

PUBLIC

RAW

Tables

COUNTRY\_CAPITAL

LAB1\_STOCK\_DATA

PROB\_HST\_TBL

STOCK\_DATA

STOCK\_DATA\_FORECAST

TEMP\_COUNTRY\_CAPITAL

TEMP\_STOCK\_DATA

Views

Stages

Streams

Tasks

DEV1

SNOWFLAKE

SNOWFLAKE\_SAMPLE\_DATA

DEV / RAW / TEMP\_STOCK\_DATA

Table ACCOUNTADMIN 22 hours ago 90 3.5KB

Table Details Columns Data Preview Copy History Lineage

COMPUTE\_WH 90 Rows Updated just now

|    | SYMBOL | OPEN   | CLOSE  | HIGH   | LOW    | VOLUME   | DATE       |
|----|--------|--------|--------|--------|--------|----------|------------|
| 1  | AAPL   | 237.71 | 235.93 | 240.07 | 234.68 | 53798062 | 2025-03-04 |
| 2  | AAPL   | 241.79 | 238.03 | 244.03 | 236.11 | 47183985 | 2025-03-03 |
| 3  | AAPL   | 236.95 | 241.84 | 242.09 | 230.20 | 56833360 | 2025-02-28 |
| 4  | AAPL   | 239.41 | 237.30 | 242.46 | 237.06 | 41153639 | 2025-02-27 |
| 5  | AAPL   | 244.33 | 240.36 | 244.98 | 239.13 | 44433564 | 2025-02-26 |
| 6  | AAPL   | 248.00 | 247.04 | 250.00 | 244.91 | 48013272 | 2025-02-25 |
| 7  | AAPL   | 244.93 | 247.10 | 248.86 | 244.42 | 51326396 | 2025-02-24 |
| 8  | AAPL   | 245.95 | 245.55 | 248.69 | 245.22 | 53197431 | 2025-02-21 |
| 9  | AAPL   | 244.94 | 245.83 | 246.78 | 244.29 | 32316907 | 2025-02-20 |
| 10 | AAPL   | 244.66 | 244.87 | 246.01 | 243.16 | 32204215 | 2025-02-19 |
| 11 | AAPL   | 244.15 | 244.47 | 245.18 | 241.84 | 48822491 | 2025-02-18 |
| 12 | AAPL   | 241.25 | 244.60 | 245.55 | 240.99 | 40896227 | 2025-02-14 |
| 13 | AAPL   | 236.91 | 241.53 | 242.34 | 235.57 | 53614054 | 2025-02-13 |
| 14 | AAPL   | 231.20 | 236.87 | 236.96 | 230.68 | 45243292 | 2025-02-12 |
| 15 | AAPL   | 228.20 | 232.62 | 235.23 | 228.13 | 53718362 | 2025-02-11 |
| 16 | AAPL   | 229.57 | 227.65 | 230.59 | 227.20 | 33115645 | 2025-02-10 |
| 17 | AAPL   | 232.60 | 227.63 | 234.00 | 227.26 | 39707224 | 2025-02-07 |
| 18 | AAPL   | 231.29 | 233.22 | 233.80 | 230.43 | 29925349 | 2025-02-06 |

Task scheduled

```
@task
def extract(url):
    response = requests.get(url)
    response.raise_for_status() # This will raise an error for a failed request
    return response.json()
```

```
@task
def transform(data):
    results = []
    for d in data["Time Series (Daily)"]:
        stock_data = data["Time Series (Daily)"][d]
        stock_data["date"] = d
        results.append(stock_data)

        if len(results) >= 90:
            break

    return results
```

@task

```
def load_stock_data(con, records):
    staging_table = "dev.raw.temp_stock_data"
    target_table = "dev.raw.stock_data"

    try:
        con.execute("BEGIN;")
        con.execute(f"""
            CREATE OR REPLACE TABLE {target_table} (
                symbol VARCHAR,
                open DECIMAL(10, 2),
                close DECIMAL(10, 2),
                high DECIMAL(10, 2),
                low DECIMAL(10, 2),
                volume BIGINT,
                date DATE,
                PRIMARY KEY (symbol, date)
            );
        """)

        con.execute(f"""
            CREATE OR REPLACE TABLE {staging_table} (
                symbol VARCHAR,
                open DECIMAL(10, 2),
                close DECIMAL(10, 2),
                high DECIMAL(10, 2),
                low DECIMAL(10, 2),
                volume BIGINT,
                date DATE,
                PRIMARY KEY (symbol, date)
            );
        """)

        con.execute(f'DELETE FROM {target_table}')
        con.execute(f'DELETE FROM {staging_table}')

        # Insert the incoming stock data into the staging table
        for r in records:
            open_price = r['1. open']
```

```

# Insert the incoming stock data into the staging table
for r in records:
    open_price = r['1. open']
    close_price = r['4. close']
    high_price = r['2. high']
    low_price = r['3. low']
    volume = r['5. volume']
    date = r['date']

    sql = f"""
        INSERT INTO {staging_table} (symbol, open, close, high, low, volume, date)
        VALUES ('AAPL', {open_price}, {close_price}, {high_price}, {low_price}, {volume}, '{date}');
    """
    con.execute(sql)

# Perform the MERGE operation to upsert data from the staging table into the target table
upsert_sql = f"""
    MERGE INTO {target_table} AS target
    USING {staging_table} AS stage
    ON target.symbol = stage.symbol AND target.date = stage.date
    WHEN MATCHED THEN
        UPDATE SET
            target.open = stage.open,
            target.close = stage.close,
            target.high = stage.high,
            target.low = stage.low,
            target.volume = stage.volume
    WHEN NOT MATCHED THEN
        INSERT (symbol, open, close, high, low, volume, date)
        VALUES (stage.symbol, stage.open, stage.close, stage.high, stage.low, stage.volume, stage.date);
"""
con.execute(upsert_sql)

con.execute("COMMIT;")

except Exception as e:
    con.execute("ROLLBACK;")
    print(f"Error: {e}")
    raise e

```

## Variables

| List Variable   |                     |                                 |             |                 |
|---|---------------------|---------------------------------|-------------|-----------------|
| Search -  |                     |                                 |             |                 |
| <div> <div>+</div> <div>Actions</div> <div>←</div> </div> |                     |                                 |             | Record Count: 7 |
| <input type="checkbox"/>                                  | Key                 | Val                             | Description | Is Encrypted    |
| <input type="checkbox"/>                                  | api_key             | *****                           |             | False           |
| <input type="checkbox"/>                                  | country_capital_url | https://s3-geospatial.s3.us-... |             | False           |
| <input type="checkbox"/>                                  | snowflake_account   | ANJOSHA-AK15174                 |             | False           |
| <input type="checkbox"/>                                  | snowflake_password  | *****                           |             | False           |
| <input type="checkbox"/>                                  | snowflake_userid    | leali                           |             | False           |
| <input type="checkbox"/>                                  | symbol              | 'AAPL'                          |             | False           |
| <input type="checkbox"/>                                  | url                 | https://www.alphavantage.c...   |             | False           |

Connection details

Edit Connection

|                    |  |
|--------------------|--|
| Connection Id *    | snowflake_con  |
| Connection Type *  | Snowflake<br><small>Connection Type missing? Make sure you've installed the corresponding Airflow Provider Package.</small>                |
| Description        |  |
| Schema             | RAW  |
| Login              | LEALI  |
| Password           | snowflake password   |
|                    | <pre>{<br/>  "account": "ANJOSHA-AK15174",<br/>  "database": "DEV",<br/>  "insecure_mode": false,<br/>  "role": "ACCOUNTADMIN"<br/>}</pre> |
| Account            | ANJOSHA-AK15174  |
| Warehouse          | snowflake warehouse name   |
| Database           | DEV  |
| Region             | snowflake hosted region  |
| Role               | ACCOUNTADMIN   |
| Private key (Path) | Path of snowflake private key (PEM Format)   |

Airflow web ui, homepage & log

DAGs

All 6Active 0Paused 0

Running 1Failed 1

Filter DAGs by tag

Search DAGs

Auto-refresh

| DAG  | Owner   | Runs | Schedule   | Last Run             | Next Run             | Recent Tasks | Actions     | Links       |
|--|---------|------|------------|----------------------|----------------------|--------------|-------------|-------------|
| <div>CountryCapital</div> <div>ETL</div>           | airflow | 4    | 0 2 * * *  | 2025-03-04, 02:00:00 | 2025-03-05, 02:00:00 | 5            | <div></div> | <div></div> |
| <div>CountryCapital_v2</div> <div>ETL</div>        | airflow | 2    | 30 2 * * * | 2025-03-05, 05:28:55 | 2025-03-05, 02:30:00 | 3            | <div></div> | <div></div> |
| <div>HelloWorld</div> <div>example</div>           | keeyong | 5    | 0 2 * * *  | 2025-03-04, 02:00:00 | 2025-03-05, 02:00:00 | 5            | <div></div> | <div></div> |
| <div>Lab1Task</div> <div>ELTpredict</div>          | airflow | 165  | 10 * * * * | 2025-03-06, 02:10:00 | 2025-03-06, 06:10:00 | 212          | <div></div> | <div></div> |
| <div>stock_data_pipeline</div> <div>ETLstock</div> | airflow | 1    | None       | 2025-03-05, 08:07:52 |                      | 3            | <div></div> | <div></div> |
| <div>TrainPredict</div> <div>ELTML</div>           | airflow | 1    | 30 2 * * * | 2025-03-05, 02:30:00 | 2025-03-06, 02:30:00 | 1            | <div></div> | <div></div> |

DAG: stock\_data\_pipeline

03/06/202505:47:23 AM

All Run TypesAll Run States

Clear Filters

Auto-refresh25

Press **shift** + **/** for Shortcuts

deferredfailedqueuedremovedrestartingrunningcheduledshutdownskippedsuccessup\_for\_rescheduleup\_for\_retryupstream\_failedno\_status

stock\_data\_pipeline / 2025-03-05, 08:07:52 UTC / transform

Clear taskMark state as...Filter DAG by task

DetailsGraphGanttCodeEvent LogLogsXComTask Duration

All LevelsAll File SourcesWrapDownloadSee More

48f7359eac4c  
Found local files:  
\* /opt/airflow/logs/dag\_id=stock\_data\_pipeline/run\_id=manual\_2025-03-05T08:07:52.317884+00:00/task\_id=transform/attempt=1.log  
[2025-03-05, 08:08:00 UTC] (local\_task\_job\_runner.py:123) Pre task execution logs  
[2025-03-05, 08:08:00 UTC] (python.py:240) INFO - Done. Returned value was: [{'1. open': '237.7850', '2. high': '240.0700', '3. low': '234.6000', '4. close': '235.9300', '5. volume': '53798062', 'data': 'AAPL'}]  
[2025-03-05, 08:08:00 UTC] (taskinstance.py:340) Post task execution logs

DAG: stock\_data\_pipeline

03/06/202505:47:23 AM

All Run TypesAll Run States

Clear Filters

Auto-refresh25

Press **shift** + **/** for Shortcuts

deferredfailedqueuedremovedrestartingrunningcheduledshutdownskippedsuccessup\_for\_rescheduleup\_for\_retryupstream\_failedno\_status

stock\_data\_pipeline / 2025-03-05, 08:07:52 UTC / extract

Clear taskMark state as...Filter DAG by task

DetailsGraphGanttCodeEvent LogLogsXComTask Duration

All LevelsAll File SourcesWrapDownloadSee More

48f7359eac4c  
Found local files:  
\* /opt/airflow/logs/dag\_id=stock\_data\_pipeline/run\_id=manual\_2025-03-05T08:07:52.317884+00:00/task\_id=extract/attempt=1.log  
[2025-03-05, 08:07:56 UTC] (local\_task\_job\_runner.py:123) Pre task execution logs  
[2025-03-05, 08:07:57 UTC] (python.py:240) INFO - Done. Returned value was: {'Meta Data': {'1. Information': 'Daily Prices (open, high, low, close) and Volumes', '2. Symbol': 'AAPL', '3. Last Refresh': '2025-03-05, 08:07:57 UTC'}]  
[2025-03-05, 08:07:57 UTC] (taskinstance.py:340) Post task execution logs

DAG: stock\_data\_pipeline

03/06/202505:47:23 AM

All Run TypesAll Run States

Clear Filters

Auto-refresh25

Press **shift** + **/** for Shortcuts

deferredfailedqueuedremovedrestartingrunningcheduledshutdownskippedsuccessup\_for\_rescheduleup\_for\_retryupstream\_failedno\_status

stock\_data\_pipeline / 2025-03-05, 08:07:52 UTC / load\_stock\_data

Clear taskMark state as...Filter DAG by task

DetailsGraphGanttCodeEvent LogLogsXComTask Duration

All LevelsAll File SourcesWrapDownloadSee More

48f7359eac4c  
Found local files:  
\* /opt/airflow/logs/dag\_id=stock\_data\_pipeline/run\_id=manual\_2025-03-05T08:07:52.317884+00:00/task\_id=load\_stock\_data/attempt=1.log  
[2025-03-05, 08:08:02 UTC] (local\_task\_job\_runner.py:123) Pre task execution logs  
[2025-03-05, 08:08:02 UTC] (cursor.py:1156) INFO - Number of results in first chunk: 1  
[2025-03-05, 08:08:02 UTC] (cursor.py:1156) INFO - Number of results in first chunk: 1  
[2025-03-05, 08:08:03 UTC] (cursor.py:1156) INFO - Number of results in first chunk: 1  
[2025-03-05, 08:08:13 UTC] (cursor.py:1156) INFO - Number of results in first chunk: 1  
[2025-03-05, 08:09:13 UTC] (python.py:240) INFO - Done. Returned value was: None  
[2025-03-05, 08:09:13 UTC] (taskinstance.py:340) Post task execution logs