Stock Data Load using Python and Snowflake

Shweta Ajay Shinde

Masters in Data Analytics, San Jose State University

Data 226: Data warehousing

Instructor: Keeyong Han

24th Sept 2024

Stock Data Analysis

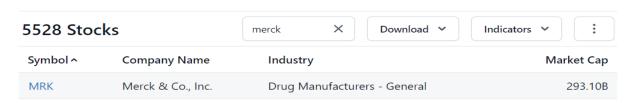
Please refer to the attached Google Colab notebook side by side for more detailed code and analysis.



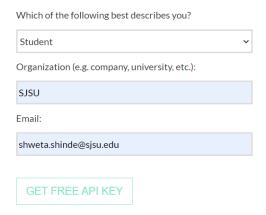
1.(+1) Pick up a stock symbol and get your own API key from Alpha Vantage.

Stock Symbol: MRK :- Merck(Pharmaceutical Company)

All Stock Symbols



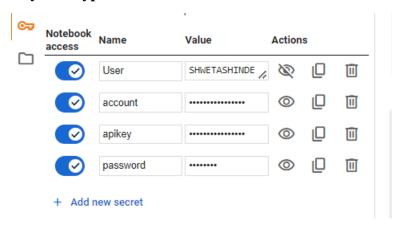
API key from Alpha Vantage.



Welcome to Alpha Vantage! Here is your API key: EP to Look 40th. LOY. Please record this API key at a safe place for future data access.

2.(+1) Secure your Snowflake credentials and Alpha Vantage API key (don't expose them in the code) using the secret in Google Colab

Keys Encrypted:-



3.(+2) Read the last 90 days of the price info via the API (refer to the relevant code snippetLinks to an external site. and you need to add "date").

Python Code

```
# Get Stock dataset from API
from google.colab import userdata
import requests
from datetime import datetime, timedelta
def return_last_90d_price(symbol):
         - return the last 90 days of the stock prices of symbol as a list of json strings
        vantage_api_key = userdata.get('apikey')
        url =
       f'https://www.alphavantage.co/query?function=TIME_SERIES_DAILY&symbol={symb
       ol}&apikey={vantage_api_key}'
        r = requests.get(url)
        data = r.json()
        symbol_value = data["Meta Data"]["2. Symbol"]
        # Get today's date and the date 90 days ago
        today = datetime.now().date()
        start_date = today - timedelta(days=90)
        results = [] # empyt list for now to hold the 90 days of stock info (open, high, low,
       close, volume)
```

```
for d in data.get("Time Series (Daily)", {}): # here d is a date: "YYYY-MM-DD"
      stock_info = data["Time Series (Daily)"][d]
      date = datetime.strptime(d, "%Y-%m-%d").date()
      if start_date <= date <= today: # Filter for the last 90 days
         stock info["date"] = d
         stock_info["symbol"] = symbol
         results.append(stock_info)
   return results
price list = return last 90d price("MRK")
price_list
[{'1. open': '113.6200',
 '2. high': '113.9900',
 '3. low': '112.9400',
 '4. close': '113.0800',
 '5. volume': '11125840',
 'date': '2024-09-26',
 'symbol': 'MRK'},
```

Screenshot:-

```
√ [16] # get Stock dataset from API
          from google.colab import userdata
          import requests
          from datetime import datetime, timedelta
          def return_last_90d_price(symbol):
             - return the last 90 days of the stock prices of symbol as a list of json strings
             vantage_api_key = userdata.get('apikey')
            url = f https://www.alphavantage.co/query?function=TIME_SERIES_DAILY&symbol={symbol}&apikey={vantage_api_key}
             r = requests.get(url)
             data = r.json()
            symbol_value = data["Meta Data"]["2. Symbol"]
# Get today's date and the date 90 days ago
            today = datetime.now().date()
start_date = today - timedelta(days=90)
             results = [] # empyt list for now to hold the 90 days of stock info (open, high, low, close, volume)
             for d in data.get("Time Series (Daily)", {}): # here d is a date: "YYYY-MM-DD"
    stock_info = data["Time Series (Daily)"][d]
                     date = datetime.strptime(d, "%Y-%m-%d").date()
                     if start_date <= date <= today: # Filter for the last 90 days
    stock_info["date"] = d
    stock_info["symbol"] = symbol</pre>
                          results.append(stock_info)
             return results

// [13] price_list = return_last_90d_price("MRK")
price_list
   'date': '2024-07-11',
'symbol': 'MRK'},
{'1. open': '125.8800',
'2. high': '127.4800',
'3. low': '125.8300',
'4. close': '127.3000'
'5. volume': '4588745'
'date': '2024-07-10'
             'date': '2024-07-10',
```

len(price_list)

Number of records:-

Connected to Snowflake successfully!

- 4(+1) Create or replace a table with a primary key under raw_data schema to capture the info from the API.
 - 1. It should have **date**, open, high, low, close, volume and symbol fields

First, we need to connect to DB from python

```
Python Code:
#Configure
                                                                           #Step 2
                                                                           #Configure
   user id = userdata.get('User')
                                                                           user id = userdata.get('User')
   password = userdata.get('password')
                                                                           password = userdata.get('password')
   account = userdata.get('account')
                                                                           account = userdata.get('account')
                                                                           import snowflake.connector
   import snowflake.connector
   try:
                                                                           try:
       conn = snowflake.connector.connect(
                                                                              conn =
          user=user_id,
                                                                           snowflake.connector.connect(
          password=password,
                                                                                user=user_id,
          account=account,
                                                                                password=password,
                                                                                account=account,
          database='dev',
                                                                                database='dev',
       print(conn)
                                                                              print(conn)
       print("Connected to Snowflake successfully!")
                                                                              print("Connected to Snowflake
   except Exception as e:
                                                                           successfully!")
       print("Error connecting to Snowflake:", e)
                                                                           except Exception as e:
                                                                              print("Error connecting to
       raise
                                                                           Snowflake:", e)
                                                                              raise
   <snowflake.connector.connection.SnowflakeConnection object at 0x7b52e1812200>
```

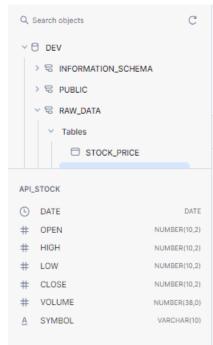
Python query to create table in snowflake

```
def create_table_api(conn):
  target_table = "dev.raw_data.stock_price"
  try:
    with conn.cursor() as cursor:
      cursor.execute(f"""
         CREATE OR REPLACE TABLE {target_table} (
            date DATE PRIMARY KEY NOT NULL,
            open DECIMAL(10, 2) NOT NULL,
            high DECIMAL(10, 2) NOT NULL,
            low DECIMAL(10, 2) NOT NULL,
            close DECIMAL(10, 2) NOT NULL,
            volume BIGINT NOT NULL,
            symbol VARCHAR(10) NOT NULL
           ("""
    # Commit the changes
    conn.commit()
    print(f"Table '{target_table}' created successfully.")
  except Exception as e:
    print(e)
    raise e
create_table_api(conn)
```

Screenshot:

def create_table_api(conn): target_table = "dev.raw_data.stock_price" with conn.cursor() as cursor: cursor.execute(f""" CREATE OR REPLACE TABLE {target_table} (date DATE PRIMARY KEY NOT NULL, open DECIMAL(10, 2) NOT NULL, high DECIMAL(10, 2) NOT NULL, low DECIMAL(10, 2) NOT NULL, close DECIMAL(10, 2) NOT NULL, volume BIGINT NOT NULL. symbol VARCHAR(10) NOT NULL # Commit the changes conn.commit() print(f"Table '{target_table}' created successfully.") except Exception as e: print(e) raise e create_table_api(conn)

Table Created in snowflake:



5.(+2) Populate the table with the records from step 3 using INSERT SQL (refer to the relevant code snippetLinks to an external site. as a starting point)

Python Code:

Table 'dev.raw_data.stock_price' created successfully.

```
def load_records(table, results):
             target_table = "dev.raw_data.stock_price"
                       with conn.cursor() as cursor:
                         for r in results:
                             open = r["1. open"]
                             high = r["2. high"]
                             low = r["3. low"]
                              close = r["4. close"]
                              volume = r["5. volume"]
                              date=r['date']
                               symbol=r['symbol']
                               insert sql = f"INSERT INTO {table} (date, open, high, low,
close, volume, symbol) VALUES ('{date}',{open}, {high}, {low}, {close}, {volume},
'{symbol}')"
                               cursor.execute(insert_sql)
                               conn.commit()
                          print(f"Recordes inserted successfully in table '{target_table}' ")
                      except Exception as e:
                      print(e)
```

raise e

load_records(dev.raw_data.stock_price_list)

Screenshot:

```
#Step 4
      #Load the stock data set in the table created in snowflake.
      def load_records(table, results):
         to complete this, first create a cursor object via
              with conn.cursor() as cursor:
                for r in results:
                  or r in results:

open = r["1. open"]

high = r["2. high"]

low = r["3. low"]

close = r["4. close"]

volume = r["5. volume"]

date=r['date']
                   symbol=r['symbol']
insert_sql = f"INSERT INTO {table} (date, open, high, low, close, volume, symbol) VALUES ('{date}', {open}, {high}, {low}, {close}, {volume}, '{symbol}')"
                   cursor.execute(insert_sql)
               # Commit the changes
                  conn.commit()
              print(f"Recordes inserted successfully in table '{table}' ")
        except Exception as e:
                raise e
[ ] load_records("dev.raw_data.stock_price", price_list)
Recordes inserted successfully in table 'dev.raw_data.stock_price'
```

Snowflake Screenshot



- 6.(+4) Steps 4 and 5 need to be done together
 - 1. Use try/except along with SQL transaction (use the code hereLinks to an external site. as reference)

For this we can use 2 methods

- Incremental Load: An incremental load transfers only new or changed data from a source to a target database. This method saves time and resources by updating only what's necessary. It's commonly used for regular updates to keep the target database current.
 - # Load data in existing table.
 - #Step1:- Create or replace table and load the data in Staging table.
 - #Step2:- Using the staging table insert the records if not exist in target table or update if any data changed for those records.

Python code:-

```
def create load incremental(conn, records):
  staging_table = "dev.raw_data.stock_price_stage"
  target_table = "dev.raw_data.stock_price"
  try:
   with conn.cursor() as cursor:
     cursor.execute(f"""
        CREATE TABLE IF NOT EXISTS {target_table} (
          date DATE PRIMARY KEY NOT NULL,
          open DECIMAL(10, 2) NOT NULL,
          high DECIMAL(10, 2) NOT NULL,
          low DECIMAL(10, 2) NOT NULL,
          close DECIMAL(10, 2) NOT NULL,
          volume BIGINT NOT NULL,
          symbol VARCHAR(10) NOT NULL
     ## Create or replace the staging table
     cursor.execute(f"""
       CREATE OR REPLACE TABLE {staging_table} (
          date DATE PRIMARY KEY NOT NULL,
          open DECIMAL(10, 2) NOT NULL,
          high DECIMAL(10, 2) NOT NULL,
          low DECIMAL(10, 2) NOT NULL,
          close DECIMAL(10, 2) NOT NULL,
          volume BIGINT NOT NULL,
          symbol VARCHAR(10) NOT NULL
              );
     # Insert records into the staging table
     for r in records:
       open = r["1. open"]
       high = r["2. high"]
       low = r["3. low"]
       close = r["4. close"]
       volume = r["5. volume"]
       date=r['date']
       symbol=r['symbol']
       insert_sql = f"INSERT INTO {staging_table} (date, open, high, low, close,
volume, symbol) VALUES ('{date}',{open}, {high}, {low}, {close}, {volume},
'{symbol}')"
       cursor.execute(insert_sql) # Execute within the with block
   conn.commit()
```

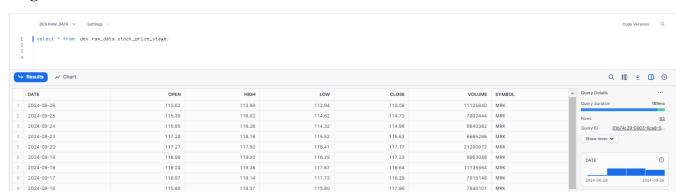
```
# Perform UPSERT
    upsert_sql = f"""
       MERGE INTO {target_table} AS target
       USING {staging_table} AS stage
       ON target.date = stage.date
       WHEN MATCHED THEN
         UPDATE SET
            target.date = stage.date,
            target.open = stage.open,
            target.high = stage.high,
            target.low = stage.low,
            target.close = stage.close,
            target.volume = stage.volume,
            target.symbol = stage.symbol
       WHEN NOT MATCHED THEN
         INSERT (date, open, high, low, close, volume, symbol)
         VALUES (stage.date, stage.open, stage.high, stage.low, stage.close,
stage.volume, stage.symbol)
    with conn.cursor() as cursor: # Open a new cursor for the upsert operation
      cursor.execute(upsert_sql)
    #Commit the change
    conn.commit()
    print(f"Stage Table {staging_table}, Target table create '{target_table}', Data loaded
successfully in both the tables using Incremental Load ")
  except Exception as e:
    print(e)
    raise e
create_load_incremental (conn, price_list)
```

Python code

```
#Incremental Load
                                                                                                                                                                                                         symbol=r['symbol']
\ensuremath{\sharp} Creating the table and and loading the data again
                                                                                                                                                                                                          insert_sql = f"INSERT INTO [staging_table] (date, open, high, low, close, volume, symbol) VALUES ('{date}', [open], {high}, {low}, {close}, {volume},
#Step1:- Load the data in Staging table.
                                                                                                                                                                                                         cursor.execute(insert_sql) # Execute within the with block
#5tep2:- Using the stageing table insert the records if not exist or update if any data changed for that records.
def create_load_incremental(conn, records):
    staging_table = "dev.raw_data.stock_price_stage"
                                                                                                                                                                                                    # perform UPSERT
    target table = "dev.raw data.stock price"
                                                                                                                                                                                                   upsert_sql = f***
                                                                                                                                                                                                       MERGE INTO {target_table} AS target
       with conn.cursor() as cursor:
                                                                                                                                                                                                       USING {staging_table} AS stage
          cursor.execute(f"""
                                                                                                                                                                                                       ON target.date = stage.date
               CREATE TABLE IF NOT EXISTS {target_table} (
                 date DATE PRIMARY KEY NOT NULL,
                                                                                                                                                                                                       WHEN MATCHED THEN
                                                                                                                                                                                                           UPDATE SET
                  open DECIMAL(10, 2) NOT NULL,
                  high DECIMAL(10, 2) NOT NULL,
                                                                                                                                                                                                              target.date = stage.date,
                  low DECIMAL(10, 2) NOT NULL,
                                                                                                                                                                                                              target.open = stage.open,
                  close DECIMAL(10, 2) NOT NULL,
                                                                                                                                                                                                              target.high = stage.high,
                   volume BIGINI NOT NULL.
                                                                                                                                                                                                              target.low = stage.low,
                   symbol VARCHAR(10) NOT NULL
                                                                                                                                                                                                              target.close = stage.close,
                                                                                                                                                                                                              target.volume = stage.volume,
                                                                                                                                                                                                              target.symbol = stage.symbol
          ## Create or replace the staging table
                                                                                                                                                                                                        WHEN NOT MATCHED THEN
          cursor.execute(f"""
                                                                                                                                                                                                           INSERT (date, open, high, low, close, volume, symbol)
             CREATE OR REPLACE TABLE {staging_table} (
                                                                                                                                                                                                           VALUES (stage.date, stage.open, stage.high, stage.low, stage.close, stage.volume, stage.symbol)
                  date DATE PRIMARY KEY NOT NULL,
                  open DECIMAL(10, 2) NOT NULL,
                  high DECIMAL(10, 2) NOT NULL.
                                                                                                                                                                                                   with conn.cursor() as cursor: # Open a new cursor for the upsert operation
                   low DECIMAL(10, 2) NOT NULL,
                                                                                                                                                                                                      cursor.execute(upsert_sql)
                  close DECIMAL(10, 2) NOT NULL,
                                                                                                                                                                                                   #Commit the change
                   volume BIGINT NOT NULL,
                                                                                                                                                                                                  conn.commit()
                   symbol VARCHAR(10) NOT NULL
                                                                                                                                                                                                  print(f"Stage Table {staging table}, Target table create '{target_table}', Data loaded successfully in both the tables using Incremental Load ")
                                                                                                                                                                                                except Exception as e:
                                                                                                                                                                                                   print(e)
                                                                                                                                                                                                    raise e
          # Insert records into the staging table
          for r in records:
              open = r["1. open"]
              high = r["2. high"]
              low = r["3, low"]
              close = r["4. close"]
                                                                                                                                                                                        create load incremental (conn, price list)
              volume = r["5. volume"]
              date=r['date']
                                                                                                                                                                                        🛨 Stage Table dev.raw_data.stock_price_stage, Target table create 'dev.raw_data.stock_price', Data loaded successfully in both the tables using Incremental Load
              symbol=r['symbol']
              insert_sql = f"INSERT INTO {staging_table} (date, open, high, low, close, volume, symbol) VALUES ('{date}', (open), {high}, {low}, {close}, {volume}, '{symbol}')"
```

Snowflake Screenshot:-

Stage table



Target Table



2. **Full Load:** A full load is when all the data from a source is moved to a target database, replacing any old data. This process makes sure that the target has the latest information.

#Step 1: We will create or replace table every time we run the load.

#Step 2: Once the table is created, we will insert the records again from initial.

Python Code

```
def create_load_full(conn, records):
     target_table = "dev.raw_data.stock_price"
     try:
       with conn.cursor() as cursor:
        cursor.execute(f"""
           CREATE OR REPLACE TABLE {target table} (
              date DATE PRIMARY KEY NOT NULL,
              open DECIMAL(10, 2) NOT NULL,
              high DECIMAL(10, 2) NOT NULL,
              low DECIMAL(10, 2) NOT NULL,
              close DECIMAL(10, 2) NOT NULL,
              volume BIGINT NOT NULL,
              symbol VARCHAR(10) NOT NULL
                 );
        # Insert records into the staging table
        for r in records:
           open = r["1. open"]
           high = r["2. high"]
           low = r["3. low"]
           close = r["4. close"]
           volume = r["5. volume"]
           date=r['date']
           symbol=r['symbol']
           insert_sql = f"INSERT INTO {target_table} (date, open, high, low, close, volume,
   symbol) VALUES ('{date}', {open}, {high}, {low}, {close}, {volume}, '{symbol}')"
           cursor.execute(insert_sql) # Execute within the with block
       conn.commit()
```

```
print(f"Target table create '{target_table}', Data loaded successfully in both the tables
using full load ")
except Exception as e:
    print(e)
    raise e
```

Screenshot:

create_load_full(conn, price_list)

Snowflake screenshot



7.(+1) Demonstrate your work ensures Idempotency by running it twice in a row and checking the number of records.

Idempotency means that when you perform an action multiple times, the result will be the same as if you only did it once. In case of data, even if we run the load multiple time it should not create duplicate records. If it is a full load, it should be deleted and loaded again. In case of incremental load, it should not be inserted unless it's a new record or and update in the existing one .

For Incremental Load(CREATE TABLE IF NOT EXIST):-





Snowflake Target:



For Full Load(CREATE OR REPLACE TABLE):-

```
[15] create_load_full(conn, price_list)

Target table create 'dev.raw_data.stock_price', Data loaded successfully in both the tables using full load
```

After running the code multiple times in a row.

```
[17] create_load_full(conn, price_list)

Target table create 'dev.raw_data.stock_price', Data loaded successfully in both the tables using full load
```

Snowflake table:



8.(+1) Follow today's demo (you can find relevant slides from today's lecture notes too) and capture your Cloud Composer Environment screen.

Composer Environment Screenshot:-

