

## Copilot

### Enhanced Data Engineering Demo Project Learning Flow

#### Data Collection (Using Pandas)

- **Fetch Data:** Access the NSE India Historical Index Data to retrieve 1-year data on Nifty sectoral indices in CSV format. This data provides a comprehensive view of the market trends and sectoral performance.
- **File Management:** Consolidate all the downloaded files in a single location for easy access and processing.
- **File Identification:** Append a unique symbol to each file. This step ensures that each file can be easily identified and retrieved when needed.
- **Local Storage:** Save the CSV files locally. This step not only ensures data availability but also reduces the dependency on the internet for data access.

#### ETL Process (Using AWS Lambda)

- **Column Rearrangement:** Swap the position of columns as per the requirements. This step enhances data readability and accessibility.
- **Column Renaming:** Modify the column names for better understanding and consistency across all files.
- **Data Type Verification:** Check the data types of each column and modify them as needed. This step ensures data integrity and accuracy.

#### Data Warehousing (Using Amazon Redshift)

- **Table Creation:** Create a table in Redshift, applying distribution and sort keys for optimized performance.
- **Primary Key Assignment:** Assign 'date' as the primary key for the table. This step ensures data uniqueness and facilitates efficient data retrieval.
- **Data Loading:** Import the CSV data from AWS S3 into the Redshift table. This step centralizes data storage and enables large-scale data processing.
- **Data Querying:** Perform queries as required to extract insights from the data. This step transforms raw data into meaningful information.

#### Data Visualization (Using Tableau)

- **Connection Setup:** Establish a connection between Redshift and Tableau. This step enables seamless data transfer between the two platforms.
- **Data Model Configuration:** Set up the data model in Tableau. This step structures the data for effective visualization.
- **Insight Generation:** Use Tableau's visualization tools to generate insights. This step transforms data into visually appealing and easily understandable formats.

This enhanced flow provides a detailed roadmap for your data engineering project, covering all stages from data collection to insight generation. It incorporates various tools and technologies, making it a comprehensive and robust learning experience. Remember to replace placeholders with actual values when implementing this project. Best of luck with your demo project!