**Lab Exercise-2 Using Metaflow with MinIO to Transform CSV Data**

Here's an updated version of your Metaflow project that includes a transformation step for the CSV file data. In this example, we will modify the Age column by adding 5 years to each person's age before uploading the modified DataFrame to the MinIO bucket.

**Updated Directory Structure**

The directory structure remains the same:

metaflow\_minio\_project/

│

├── data/

│ └── sample\_data.csv # This will be created by the flow

│

├── output/

│ └── downloaded\_data.csv # This will be created by the flow

│

├── minio\_metaflow.py # The main Metaflow script

└── requirements.txt # Dependencies

**Updated Code for minio\_metaflow.py**

Here's the complete code with the transformation step added:

from metaflow import FlowSpec, step

import pandas as pd

from minio import Minio

from minio.error import S3Error

import os

class MinIOFlow(FlowSpec):

@step

def start(self):

"""Generate sample data and save to CSV."""

# Create a simple DataFrame

data = {'Name': ['Alice', 'Bob', 'Charlie'],

'Age': [24, 30, 22]}

self.df = pd.DataFrame(data)

# Save to CSV in the data directory

self.data\_path = 'data/sample\_data.csv'

self.df.to\_csv(self.data\_path, index=False)

print(f"Original data saved to {self.data\_path}")

self.next(self.transform\_data)

@step

def transform\_data(self):

"""Transform the data by adding 5 years to the Age."""

print("Transforming data...")

self.df = pd.read\_csv(self.data\_path)

# Add 5 years to each person's age

self.df['Age'] = self.df['Age'] + 5

# Save the transformed DataFrame back to CSV

self.transformed\_data\_path = 'data/transformed\_data.csv'

self.df.to\_csv(self.transformed\_data\_path, index=False)

print(f"Transformed data saved to {self.transformed\_data\_path}")

self.next(self.upload\_to\_minio)

@step

def upload\_to\_minio(self):

"""Upload the transformed CSV file to MinIO."""

# Instantiate MinIO client directly in this step

minio\_client = Minio('localhost:9000',

access\_key='minioadmin',

secret\_key='minioadmin',

secure=False)

# Create bucket if it doesn't exist

try:

if not minio\_client.bucket\_exists('mybucket'):

minio\_client.make\_bucket('mybucket')

except S3Error as e:

print("Error occurred: ", e)

# Upload the transformed file

try:

minio\_client.fput\_object('mybucket', 'transformed\_data.csv', self.transformed\_data\_path)

print("Transformed file uploaded successfully.")

except S3Error as e:

print("Error occurred: ", e)

self.next(self.retrieve\_from\_minio)

@step

def retrieve\_from\_minio(self):

"""Retrieve the transformed CSV file from MinIO."""

# Instantiate MinIO client again in this step

minio\_client = Minio('localhost:9000',

access\_key='minioadmin',

secret\_key='minioadmin',

secure=False)

# Download the file to output directory

self.output\_path = 'output/downloaded\_transformed\_data.csv'

minio\_client.fget\_object('mybucket', 'transformed\_data.csv', self.output\_path)

print("Transformed file downloaded successfully.")

# Read the downloaded file

self.downloaded\_df = pd.read\_csv(self.output\_path)

print("Downloaded Transformed DataFrame:")

print(self.downloaded\_df)

self.next(self.end)

@step

def end(self):

"""End of the flow."""

print("Flow finished!")

if \_\_name\_\_ == '\_\_main\_\_':

MinIOFlow()

**Key Changes Made**

1. **Added a New Step**:
   * The transform\_data step was added to apply a transformation to the DataFrame. Specifically, it adds 5 years to each age in the Age column.
2. **Modified File Names**:
   * The transformed data is saved as transformed\_data.csv in the data/ directory and is uploaded to MinIO as transformed\_data.csv.
3. **Updated Output Path**:
   * The downloaded file will be saved as downloaded\_transformed\_data.csv in the output/ directory.

**Running the Flow Again**

After making these changes, follow these steps to run the flow:

1. **Ensure MinIO is Running**.
2. **Run the Flow**:

python minio\_metaflow.py run

**Checking Results**

After running the flow, check the following:

* The data/sample\_data.csv file will contain the original data.
* The data/transformed\_data.csv file will contain the transformed data (with ages increased by 5).
* The output/downloaded\_transformed\_data.csv file will contain the same transformed data downloaded from the MinIO bucket.
* You can also access the MinIO console at http://localhost:9001 to verify the file upload.

This code now demonstrates data transformation within the Metaflow workflow while still integrating with MinIO for file storage.