Assignment: Creating and Scheduling a Job on Databricks using Notebooks - 17th Sep

SaiPrabath Chowdary S

Task 1:

Prepare Your Notebook

```
csv_path = 'file:/Workspace/Shared/assignment17sep/orders.csv'
dbfs_path = 'dbfs:/Filestore/assignment17sep/orders.csv'
dbutils.fs.cp(csv_path, dbfs_path)

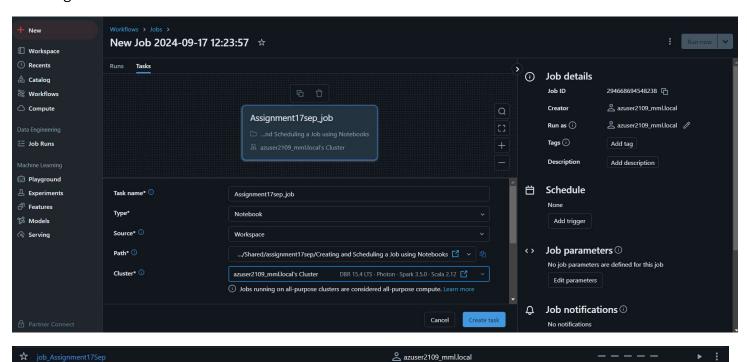
from pyspark.sql.functions import col

# Load CSV data
df = spark.read.format("csv").option("header", True).load(dbfs_path)

# Add TotalAmount column and filter records (I used greater than 2 according to data)
transformed_df = df.withColumn("TotalAmount", col("Quantity") * col("Price")).filter(col("Quantity") > 2)

# Write to Delta table
transformed_df.write.format("delta").mode("overwrite")\
.save("dbfs:/Workspace/Shared/assignment17sep/orders_transformed")
```

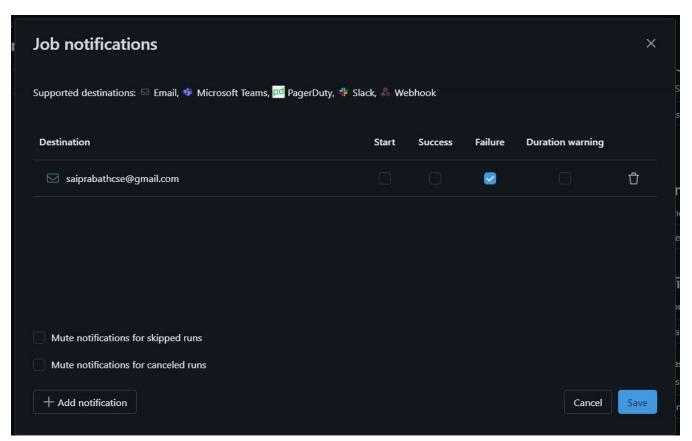
Task 2: Creating a Databricks Job



schedule job every day @8:00:



Task 3: Configure Alerts for Job Failures



Task 5: Run the Job and Monitor Progress



Task 6: Inspect Job Run History

