

## Assignment: Creating and Scheduling a Job on Databricks using Notebooks – 17<sup>th</sup> Sep

SaiPrabath Chowdary S

### Task 1:

#### Prepare Your Notebook

```
1
Python

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

New

Workspace

Recents

Catalog

Workflows

Compute

Data Engineering

Job Runs

Machine Learning

Playground

Experiments

Features

Models

Serving

Partner Connect

Workflows > Jobs >

New Job 2024-09-17 12:23:57 ☆

Runs

Tasks

Assignment17sep\_job

...nd Scheduling a Job using Notebooks

azuser2109\_mmllocal's Cluster

Task name\*

Assignment17sep\_job

Type\*

Notebook

Source\*

Workspace

Path\*

.../Shared/assignment17sep/Creating and Scheduling a Job using Notebooks

Cluster\*

azuser2109\_mmllocal's Cluster DBR 15.4 LTS · Photon · Spark 3.5.0 · Scala 2.12

Jobs running on all-purpose clusters are considered all-purpose compute. Learn more

Cancel

Create task

Job details

Job ID 294668694548238

Creator azuser2109\_mmllocal

Run as azuser2109\_mmllocal

Tags Add tag

Description Add description

Schedule

None

Add trigger

Job parameters

No job parameters are defined for this job

Edit parameters

Job notifications

No notifications

☆ job\_Assignment17Sep

azuser2109\_mmllocal

— — — — — ▶ ⋮

schedule job every day @8:00:

## Schedule

Trigger Status

☒ Active

☐ Paused

Trigger type

Scheduled

Schedule type

Simple

Advanced

Schedule ⓘ

Every

Day

at

08

:

00

(UTC+00:00) UTC

☐ Show cron syntax

Cancel

Save

### Task 3: Configure Alerts for Job Failures

## Job notifications

Supported destinations: Email, Microsoft Teams, PagerDuty, Slack, Webhook

Destination	Start	Success	Failure	Duration warning	
saiprabathcse@gmail.com	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	

☐ Mute notifications for skipped runs

☐ Mute notifications for canceled runs

+ Add notification

Cancel

Save

Task 5: Run the Job and Monitor Progress

Workflows > Jobs >

job\_Assignment17Sep ☆

⋮ Run now ▾

azuser2109\_mml.local

100% 100% 100% 100%

Task 6: Inspect Job Run History

Runs Tasks

Runs

Start date 📅 < Previous Next >

Run total duration

Sep 16

39s

19s

data\_loading\_job1

Tasks

Go to the latest successful run

Cancel runs ▾

Start time	Run ID	Launched	Duration	Spark	Status	Error code	Run paramet...	
Sep 17, 2024, 12...	33026787...	Manually	40s	<a href="#">Spark UI / Logs / Metrics</a>	Succe...			⋮

📘 Job details

Job ID

884137515844904 📋

Creator

 azuser2109\_mml.local

Run as ⓘ

 azuser2109\_mml.local ✎

Tags ⓘ

Add tag

Description

Add description

🔗 Git

Not configured

Add Git settings

📅 Schedule

None

Add trigger

🖥️ Compute

● azuser2109\_mml.local's Cluster