

# Assignment- JOB Workflow

Submitted By-

Subrat Shukla, DE-1

Hands-on 1: Creating a job run with the sample load tutorial given.

The screenshot shows the Databricks Jobs page for a workflow named 'Example job'. The 'Runs' tab is active, displaying a bar chart of run durations. The 'Job details' panel on the right shows the Job ID (47203570833377), Creator (azuser2374\_mmllocal), and Run as (azuser2374\_mmllocal). The 'Schedule' section is set to 'None'.

Start time	Run ID	Launch...	Duration	Spark	Status	Error code	Run para...

Output:

The screenshot shows the Databricks Jobs page for a specific run of the 'Ingestion-Extract' job. The run is marked as 'Succeeded'. The 'Output' section displays the code executed and the resulting table of data sets.

```
# Let's see what example data sets exist
display(dbutils.fs.ls('/databricks-datasets/'))
```

id	path	name	size
1	dbfs/databricks-datasets/COVID/	COVID/	
2	dbfs/databricks-datasets/README.md	README.md	9
3	dbfs/databricks-datasets/Rdatasets/	Rdatasets/	
4	dbfs/databricks-datasets/SPARK_README.md	SPARK_README.md	33
5	dbfs/databricks-datasets/adult/	adult/	
6	dbfs/databricks-datasets/airlines/	airlines/	
7	dbfs/databricks-datasets/amazon/	amazon/	
8	dbfs/databricks-datasets/asa/	asa/	
9	dbfs/databricks-datasets/atlas_higgs/	atlas_higgs/	
10	dbfs/databricks-datasets/bikeSharing/	bikeSharing/	

## Hands-on 2: Creating a job run with the existing notebook.

The screenshot shows the Databricks Jobs page for a job named 'Job\_with\_existing\_notebook'. The 'Runs' tab is active, displaying a table with columns: Start time, Run ID, Launch..., Duration, Spark, Status, Error code, and Run para... The table shows a single run from Dec 02, 2024, with Run ID 129881..., launched manually, with a duration of 10s, using Spark UI / Logs / Metrics, and a status of 'Suc...'. To the right, the 'Job details' panel shows Job ID 1006653838785363, Creator azuser2374\_mml.local, Run as azuser2374\_mml.local, and a description 'Add description'. The 'Schedule' panel shows 'None' and an 'Add trigger' button.

## Output:

The screenshot shows the 'Transform run' details for a job run. The 'Output' tab is active, displaying the code and the resulting data. The code is as follows:

```
from pyspark import SparkContext
from pyspark.sql import SparkSession

sc = SparkContext.getOrCreate()
spark = SparkSession.builder.appName('pyspark first program').getOrCreate()
spark.table("default.export")
data = spark.read.format("delta").load("dbfs:/user/hive/warehouse/export")
data.show()
```

The output shows a DataFrame with 11 rows and 6 columns. The data is as follows:

	id	long	firstName	string	...	6 more fields
3	Quyen	Marlen	Dome	F	1970-10-11 04:00:00	957-57-8246   53417
4	Coralie	Antonina	Marshal	F	1990-04-11 04:00:00	963-39-4885   94727
5	Terrie	Wava	Bonar	F	1980-01-16 05:00:00	964-49-8051   79908
6	Chassidy	Concepcion	Bourthouloume	F	1990-11-24 05:00:00	954-59-9172   64652
7	Geri	Tambra	Mosby	F	1970-12-19 05:00:00	968-16-4020   38195
8	Patria	Nancy	Arstall	F	1985-01-02 05:00:00	984-76-3770   102053
9	Terese	Alfredia	Tocque	F	1967-11-17 05:00:00	967-48-7309   91294
10	Wava	Lyndsey	Jeandon	F	1963-12-30 05:00:00	997-82-2946   56521
11	Sophie	Emerita	Hearn	F	1979-09-17 04:00:00	977-66-4483   90920

The 'Task run' panel on the right shows details for the task run, including Job ID 1006653838785363, Job run ID 129881547026086, Task run ID 11611633844732, Run as azuser2374\_mml.local, Launched Manually, Started 12/02/2024, 10:49:42 AM, Ended 12/02/2024, 10:49:51 AM, Duration 9s, Queue duration -, and Status Succeeded.

# Creating storage accounts:

Home >

Storage accounts

Techademy Learning Solutions Private Limited

+ Create

↶ Restore

⚙️ Manage view

↻ Refresh

📄 Export to CSV

🔗 Open query

🏷️ Assign tags

🗑️ Delete

Filter for any field...

Subscription equals all

Resource group equals all

Location equals all

+ Add filter

Showing 1 to 2 of 2 records.

No grouping

List view

<input type="checkbox"/>	Name ↑↓	Type ↑↓	Kind ↑↓	Resource group ↑↓	Location ↑↓	Subscription ↑↓	
<input type="checkbox"/>	datalakestoredb	Storage account	StorageV2	rg-azuser2374_mmllocal-CTJmW	Central India	MML Learners	...
<input type="checkbox"/>	datastoresindb	Storage account	StorageV2	rg-azuser2374_mmllocal-CTJmW	East US	MML Learners	...

## Hands-on 3: Creating a datalake storage account.

Home > Storage accounts >

datalakestoredb

Storage account

Search

Overview

Activity log

Tags

Diagnose and solve problems

Access Control (IAM)

Data migration

Events

Storage browser

Partner solutions

Data storage

Security + networking

Data management

Settings

Monitoring

Upload

Open in Explorer

Delete

Move

Refresh

Open in mobile

CLI / PS

Feedback

Essentials

Resource group (...): rg-azuser2374\_mmllocal-CTJmW

Location: centralindia

Primary/Secondary ...: Primary: Central India, Secondary: South India

Subscription (move): MML Learners

Subscription ID: 2a3c6418-97b9-4d96-a24b-2c2d7633d375

Disk state: Primary: Available, Secondary: Available

Tags (edit): Add tags

Performance: Standard

Replication: Read-access geo-redundant storage (RA-GRS)

Account kind: StorageV2 (general purpose v2)

Provisioning state: Succeeded

Created: 12/2/2024, 3:08:15 PM

Properties

Monitoring

Capabilities (5)

Recommendations (0)

Tutorials

Tools + SDKs

Data Lake Storage

Hierarchical namespace: Enabled

Default access tier: Hot

Blob anonymous access: Disabled

Blob soft delete: Enabled (7 days)

Security

Require secure transfer for REST API operations: Enabled

Storage account key access: Enabled

Minimum TLS version: Version 1.2

Infrastructure encryption: Disabled

## Hands-on 4: Creating a datastore storage account.

Home > Storage accounts >

datastoresindb

Storage account

Search

Overview

Activity log

Tags

Diagnose and solve problems

Access Control (IAM)

Data migration

Storage browser

Storage Mover

Partner solutions

Data storage

Security + networking

Data management

Settings

Monitoring

Upload

Open in Explorer

Delete

Move

Refresh

Open in mobile

CLI / PS

Feedback

Essentials

Resource group (...): rg-azuser2374\_mmllocal-CTJmW

Location: eastus

Subscription (move): MML Learners

Subscription ID: 2a3c6418-97b9-4d96-a24b-2c2d7633d375

Disk state: Available

Tags (edit): Add tags

Performance: Standard

Replication: Locally-redundant storage (LRS)

Account kind: StorageV2 (general purpose v2)

Provisioning state: Succeeded

Created: 12/2/2024, 3:05:04 PM

Properties

Monitoring

Capabilities (0)

Recommendations (0)

Tutorials

Tools + SDKs

Blob service

Hierarchical namespace: Disabled

Default access tier: Hot

Blob anonymous access: Disabled

Blob soft delete: Enabled (7 days)

Container soft delete: Enabled (7 days)

Security

Require secure transfer for REST API operations: Enabled

Storage account key access: Enabled

Minimum TLS version: Version 1.2

Infrastructure encryption: Disabled