

# Assignment- JOB Workflow

Name: E.R Harish

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

The screenshot shows the Databricks Jobs interface for a job named 'Example job'. The left sidebar contains navigation options like Workspace, Recents, Catalog, Workflows, Compute, Data Engineering, Job Runs, Machine Learning, Playground, Experiments, Features, Models, and Serving. The main panel displays the 'Runs' tab for the job, showing a bar chart of run durations (3m 20s, 1m 40s) and a table of tasks (Ingestion-Extract). The right sidebar shows 'Job details' including Job ID (47203570833377), Creator (azuser2374\_mmllocal), Run as (azuser2374\_mmllocal), Tags (Add tag), Description (Add description), Git (Not configured), and Schedule (None). A 'Run now' button is visible in the top right.

Output:

The screenshot shows the Databricks Jobs interface for a specific job run named 'Ingestion-Extract run' with status 'Succeeded'. The left sidebar is the same as the previous screenshot. The main panel displays the 'Output' tab, showing a code snippet and a table of data sets. The right sidebar shows 'Task run' details including Job ID (47203570833377), Job run ID (318476716474240), Task run ID (996074168934478), Run as (azuser2374\_mmllocal), Launched (Manually), Started (12/02/2024, 10:33:37 AM), Ended (12/02/2024, 10:36:57 AM), Duration (3m 20s), Queue duration (0), and Status (Succeeded).

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 interface. The left sidebar contains navigation options: New, Workspace, Recents, Catalog, Workflows, Compute, Data Engineering, Job Runs, Machine Learning, Playground, Experiments, Features, Models, and Serving. The main area displays the job 'Job\_with\_existing\_notebook'. A 'Runs' tab shows a single run with a duration of 10s and a status of 'Suc...'. A 'Job details' panel on the right shows the Job ID, Creator, Run as, Tags, Description, Git, and Schedule. A 'Run now' button is visible in the top right.

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P HexawareWorkspace

New Workspace Recents Catalog Workflows Compute Data Engineering Job Runs Machine Learning Playground Experiments Features Models Serving

Workflows > Jobs > Job\_with\_existing\_notebook ☆ Send feedback Run now

Runs Tasks

Run total duration 9s 4s Dec 02

Transform

Go to the latest successful run Cancel runs

Start time Run ID Launch... Duration Spark Status Error code Run para... Spark UI / Logs / Metrics

Dec 02, 2024... 129881... Manually 10s Logs / Metrics

Job details

Job ID 1006653838785363

Creator azuser2374\_mml.local

Run as azuser2374\_mml.local

Tags Add tag

Description Add description

Git Not configured Add Git settings

Schedule None Add trigger

## Output:

The screenshot shows the Databricks Jobs interface with the 'Transform run' selected. The 'Output' tab displays the code and the resulting DataFrame. The 'Task run' panel on the right shows details about the run, including the Job ID, Job run ID, Task run ID, Run as, Launched, Started, Ended, Duration, Queue duration, and Status.

Workflows > Jobs > Job\_with\_existing\_notebook > Run 129881547026086 > Transform run Succeeded Edit task Repair run

Output Hide code Export as HTML

2.28 seconds 1

```
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()
```

data: pyspark.sql.dataframe.DataFrame = [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	Marshall	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

Task run Details Metrics

Details

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 dura... -

Status Succeeded

## Hands-on 3: Creating a Storage Account.

The screenshot shows the Microsoft Azure portal interface. The left sidebar displays the 'Storage accounts' section with a list of accounts: 'datalakegen2hexaware' and 'datastoragehexaware'. The main content area shows the details for the 'datastoragehexaware' storage account. The 'Essentials' tab is active, displaying the following information:

Property	Value
Resource group	(move) rg-azuser2365_mml.local-Ta7JA
Location	westus3
Subscription	(move) MML Learners
Subscription ID	2a3c6418-97b9-4d96-a24b-2c2d7633d375
Disk state	Available
Performance	Standard
Replication	Locally-redundant storage (LRS)
Account kind	StorageV2 (general purpose v2)
Provisioning state	Succeeded
Created	12/2/2024, 2:37:40 PM

The bottom of the screen shows a Windows taskbar with the date and time as 15:20 on 02-12-2024.

The screenshot shows the Microsoft Azure portal interface, displaying the 'Properties' tab for the 'datastoragehexaware' storage account. The 'Properties' tab is active, showing the following settings:

Category	Property	Value
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)
	Versioning	Disabled
	Change feed	Disabled
	NFS v3	Disabled
	Allow cross-tenant replication	Disabled
	Storage tasks assignments	None
Security	Require secure transfer for REST API operations	Enabled
	Storage account key access	Enabled
	Minimum TLS version	Version 1.2
Networking	Infrastructure encryption	Disabled
	Allow access from	All networks
	Private endpoint connections	0
	Network routing	Microsoft network routing
File service	Access for trusted Microsoft services	Yes
	Endpoint type	Standard

The bottom of the screen shows a Windows taskbar with the date and time as 15:20 on 02-12-2024.

## Hands-on 4: Creating a Data lake gen2 account.

Home > Storage accounts >

### Storage accounts

Techademy Learning Solutions Private Limited (tec...

+ Create Restore ...

Filter for any field...

Name ↑↓

- datalakegen2hexaware
- datastoragehexaware

Page 1 of 1

#### datalakegen2hexaware

Storage account

Upload Open in Explorer Delete Move Refresh Open in mobile CLI / PS Feedback

JSON View

##### Essentials

Resource group (move)	Performance
rg-azuser2365_mml.local-Ta7JA	Standard
Location	Replication
westus3	Locally-redundant storage (LRS)
Subscription (move)	Account kind
MML Learners	StorageV2 (general purpose v2)
Subscription ID	Provisioning state
2a3c6418-97b9-4d96-a24b-2c2d7633d375	Succeeded
Disk state	Created
Available	12/2/2024, 3:06:00 PM

Tags (edit)  
Add tags

Home > Storage accounts >

### Storage accounts

Techademy Learning Solutions Private Limited (tec...

+ Create Restore ...

Filter for any field...

Name ↑↓

- datalakegen2hexaware
- datastoragehexaware

Page 1 of 1

#### datalakegen2hexaware

Storage account

##### Data Lake Storage

Hierarchical namespace	Enabled
Default access tier	Hot
Blob anonymous access	Disabled
Blob soft delete	Enabled (7 days)
Container soft delete	Enabled (7 days)
Versioning	Disabled
Change feed	Disabled
NFS v3	Disabled
SFTP	Disabled
Storage tasks assignments	None

##### File service

Large file share	Enabled
------------------	---------

##### Security

Require secure transfer for REST API operations	Enabled
Storage account key access	Enabled
Minimum TLS version	Version 1.2
Infrastructure encryption	Disabled

##### Networking

Allow access from	All networks
Private endpoint connections	0
Network routing	Microsoft network routing
Access for trusted Microsoft services	Yes
Endpoint type	Standard