

Azure Data Factory

Azure console

The screenshot shows the Azure portal interface for a Data Factory (V2) resource named 'querysupportADF'. The top navigation bar includes the Microsoft Azure logo, an 'Upgrade' button, a search bar, and user information. The left sidebar contains a navigation menu with options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings, Getting started, Monitoring, Automation, and Help. The main content area displays the 'Overview' tab for the Data Factory, showing details such as Resource group (295b), Status (Succeeded), Location (East US), Subscription (Azure subscription 1), and Subscription ID (a0a9051b-1d4e-425c-b822-4a6776ad3e69). Below this, there is a large 'Azure Data Factory Studio' logo and a 'Launch studio' button. At the bottom, there are four tiles for 'Quick Starts', 'Tutorials', 'Template Gallery', and 'Training Modules'.

The screenshot shows the Azure Data Factory Studio interface. The top navigation bar includes the Microsoft Azure logo, the Data Factory name 'querysupportADF', a search bar, and user information. The left sidebar contains a navigation menu with options like Factory Resources, Pipelines, Datasets, Data flows, and Power Query. The main content area displays the 'get_files' pipeline, which is currently in the 'Validate' state. The pipeline diagram shows a 'Get Metadata' activity (List s3 files) connected to a 'ForEach' activity (IterateFiles). Inside the 'ForEach' activity, there is a 'getLatestFile' activity connected to a 'compareFile' activity. The bottom of the interface shows tabs for 'Parameters', 'Variables', 'Settings', and 'Output'.

Microsoft Azure | Data Factory | querysupportADF

Search factory and documentation

malipratik297@gmail.com
DEFAULT DIRECTORY

Microsoft recently announced the public preview of Microsoft Fabric, a brand new and exciting way to build cloud-first data analytics. Click [here](#) to get started with Fabric Data Factory!

» Data Factory | Validate all | Publish all | Preview experience | Off

Factory Resources

Filter resources by name

- Pipelines 1
 - get_files
- Change Data Capture (preview) 0
- Datasets 2
- Data flows 0
- Power Query 0

Activities

Search activities

- Move and transform
- Synapse
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & conditionals
- Machine Learning
- Power Query

get_files > iterateFiles

Validate | Debug | Add trigger

Get Metadata

get_latest_file

If Condition

compare_time

True

(X) Get latest modified...

(X) Get latest file name

run_dtb_ntb_k

+

False

+


Edit linked service

 Azure Databricks [Learn more](#) 

Name *

AzureDatabricks1

Description

Connect via integration runtime * 

AutoResolveIntegrationRuntime

Account selection method *

☐ From Azure subscription ☒ Enter manually

Databrick Workspace URL * 

https://adb-1988582850967560.0.azuredatabricks.net

Authentication type *

Access Token

Access token

Azure Key Vault

Access token * 

.....

Select cluster

Save

Cancel

 Test connection


Edit linked service

 Amazon S3 [Learn more](#) 

Name *

AmazonS31

Description

Connect via integration runtime * 

AutoResolveIntegrationRuntime

Authentication type

Access key

Access key ID

Azure Key Vault

Access key ID *

AKIAX5INISKDOPAPOLPE

Secret access key

Azure Key Vault

Secret access key *

.....

Service URL 

https://s3.amazonaws.com

Save

Cancel



Test connection

Azure Databricks

Microsoft Azure

databricks

Search data, notebooks, recents, and more...

295b

P

New

Workspace

Recents

Catalog

Workflows

Compute

SQL

SQL Editor

Queries

Dashboards

Alerts

Query History

SQL Warehouses

Data Engineering

Job Runs

Data Ingestion

Delta Live Tables

Compute >

Pratik Mali's Cluster

More

Start

Edit

Configuration

Notebooks (0)

Libraries

Event log

Spark UI

Driver logs

Metrics

Apps

Spark compute UI - Master

Policy

Personal Compute

Access mode

Single user access

Single user

Pratik Mali

Performance

Databricks Runtime Version

14.3 LTS ML (includes Apache Spark 3.5.0, Scala 2.12)

Use Photon Acceleration

Node type

Standard_DS3_v2

14 GB Memory, 4 Cores

Terminate after 60 minutes of inactivity

Summary

1 Driver

14 GB Memory, 4 Cores

Runtime

14.3.x-cpu-ml-scala2.12

Unity Catalog

Standard_DS3_v2

0.75 DBU/h

Compute >

Pratik Mali's Cluster

More

Start

Edit

Configuration

Notebooks (0)

Libraries

Event log

Spark UI

Driver logs

Metrics

Apps

Spark compute UI - Master

Jobs

Stages

Storage

Environment

Executors

SQL / DataFrame

Spark Jobs (?)

User: root

Total Uptime:

Scheduling Mode: FAIR

Completed Jobs: 15

Event Timeline

Completed Jobs (15)

Page: 1

1 Pages. Jump to 1. Show 100 items in a page. Go

Job Id (Job Group)	Description	Submitted	Duration	Stages: Succeeded/Total
14 (7555624075288637712_9222410458427602438_015410834923416381f08da2e8828d22)	df = spark.read.format("csv").opt... wrapper at /root/.ipykernel/6187/command-3077646539820250-3882870502:6	2024/05/10 06:26:56	2 s	1/1
13 (7555624075288637712_9222410458427602438_015410834923416381f08da2e8828d22)	df = spark.read.format("csv").opt... load at NativeMethodAccessorImpl.java:0	2024/05/10 06:26:55	0.5 s	1/1
12	df = spark.read.format("csv").opt...	2024/05/10	0.2 s	1/1