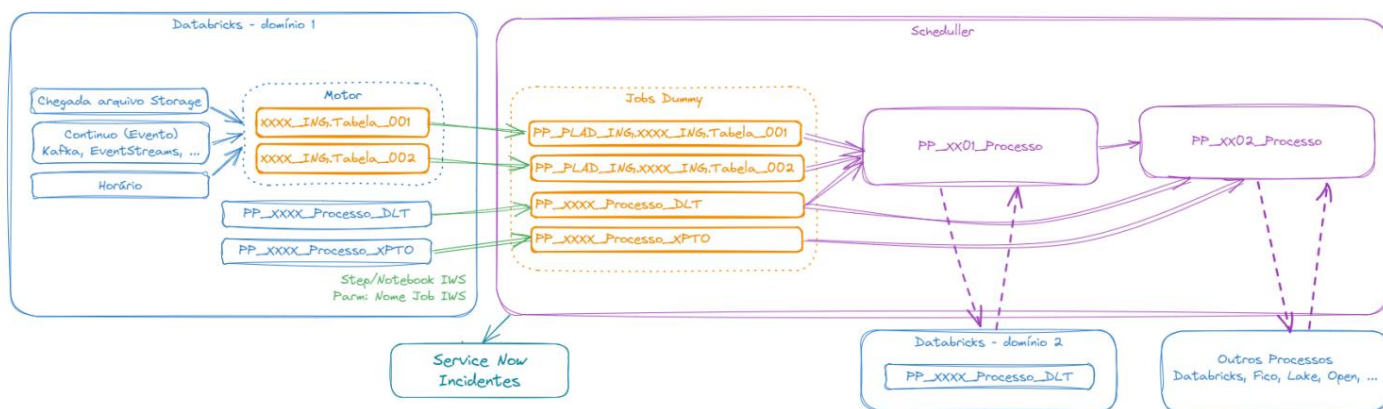


Below is a list of process scenarios involving Databricks that require Support and Monitoring.

DLT (Delta Live Tables) Processes.....	2 Start of process after arrival
of Storage Account & Volume file	3 Process start after loading
Tables.....	4 View Steps,
Logs,	5 Restart of
Jobs.....	6
Model Serving	7
Links.....	8



Databricks + IWS

Considerações:

- Tem que criar um job "dummy" por tabela/processo
- Jobs Databricks executados não devem abrir IN no Service Now
- Não podem ter configuração via Databricks
- Não permite re-run
- Realizar re-run pelo Databricks
- Executor Force OK depois da finalização

Deploy Necessita de 2 esteiras

- 1o Deploy do Workflow no Databricks, obter ID
- 2o Deploy do IWS, com ID Databricks

Todo: Verificar possibilidade de criar Workflow e IWS na mesma Esteira

Databricks + ADF

Problemas

- Tem que estar tudo no mesmo PIPE, não cria dependência entre pipe's distintos
- Não permite dependência entre ADF's diferentes

Databricks (somente)

- Atende processos mais simples, sem integrações entre workspaces e/ou sucessores externos
- Contorno: execução processos externos via API, porém não atende todos os casos

- Não tem como criar dependência job's

Contorno: ter um Job contendo outros Jobs, porém existem limitações

- Não consegue incluir mais de uma dependência (ex: n file arrival's) no mesmo Job

Obs.: Dependência de carga "Via File arrival" observando o storage da tabela pode gerar problemas se a carga não finalizou no tempo esperado.

DLT Processes (Delta Live Tables)

DLT are automatic processes created in Databricks to load tables via streaming, chaining the other processes (pipeline)

Can be •

Continuous or •

Triggered - Storage Gen2 File Reception

- Triggered - Receiving Volumes Files

- Triggered – Load Databricks Tables

Served by IWS?

How?

Workflows > Delta Live Tables >

Create pipeline

General

* Pipeline name

Product edition

Advanced

[Help me choose](#)

Pipeline mode ⓘ

☒ Triggered ☐ Continuous

[Configure pipeline settings for Delta Live Tables - Azure Databricks | Microsoft Learn](#)

Start of process after arrival of Storage Account & Volume file

Processes started upon arrival of Storage Account v2 files

Processes started upon arrival of Volume files (managed or not)

Served by IWS?
How?

Workflows > Jobs >
New Job 2024-08-19 16:53:21 ☆

RunsTasks

Schedules & Triggers

Trigger Status

Active

Paused

Trigger type

File arrival

This job does not have any failure notifications. Consider adding email or webhook notifications to alert if trigger evaluation fails.

File arrival triggers monitor cloud storage paths of up to 10,000 files for new files. These paths are either Unity Catalog volumes or external locations managed through Unity Catalog.

Storage location

e.g. '/Volumes/mycatalog/myschema/myvolume/path_within_volume/' or 'abfss://filesystem@accountname.dfs.core.windows.net/path/'

Advanced

Test connectionCancelSave

Type*
Source*
Path*
Cluster*
Dependent libraries
Parameters

+ Add

ANDRE DE ANDRADE

Process start after loading Tables

Processes started after loading a table into Databricks

Served by IWS?
How?

The screenshot displays the Databricks workspace interface. A modal dialog titled "Schedules & Triggers" is open, allowing configuration for a job. The dialog includes the following fields and options:

- Trigger Status:** Radio buttons for "Active" (selected) and "Paused".
- Trigger type:** A dropdown menu with options: "Scheduled" (selected), "File arrival", "Table update" (highlighted in yellow), and "Continuous".
- Schedule:** A text input field containing the cron expression "0 0 11 ? * MON-FRI".
- Timezone:** A dropdown menu set to "(UTC-03:00) Brasilia".
- Show cron syntax:** A checked checkbox.
- Buttons:** "Cancel" and "Save" buttons at the bottom right of the dialog.

The background interface shows the job details for "PP_AAES_FinOps_Rateio_Diaria", including its ID (816556308857791), creator (ANDRE DE ANDRADE FERREIRA), and a list of upstream and downstream tables.

View Steps, Logs, ...

View Databricks Workflow Steps

- Metrics • Logs

- Execution times

Served by IWS?
How?

Ex.: Job with 17 tasks (notebooks)

According to HCL Documentation, IWS does not run Notebooks, only Jobs

The screenshot displays the Databricks Jobs interface. The left sidebar contains navigation options: New, Workspace, Recents, Catalog, Workflows, Compute, SQL, SQL Editor, Queries, Dashboards, Genie, Alerts, Query History, SQL Warehouses, Data Engineering, Job Runs (highlighted), Data Ingestion, Delta Live Tables, Machine Learning, Playground, Experiments, Features, Models, and Serving. The main area shows a workflow graph for the job 'PP_AAES_FinOps_Rateio_Diaria run'. The graph consists of a sequence of tasks, with the first task being a cluster creation. The right sidebar provides 'Job run details' for the specific run, including the Job ID, Job run ID, Launch method (By scheduler), Start and End times, Duration (28m 34s), Queue duration, Status (Succeeded), and Lineage (28 upstream tables, 21 downstream tables). Below this, the 'Compute' section shows the job cluster configuration: Driver: Standard_D4s_v5, Workers: Standard_D4s_v5 - 1 worker - from pool - DBR: 14.3 LTS (includes Apache Spark 3.5.0, Scala 2.12). Buttons for 'View details', 'Spark UI', 'Logs', and 'Metrics' are available.

Job ID	Job run ID
81655630857791	155895163121822

Launched	Started	Ended	Duration	Queue duration
By scheduler	08/19/2024, 11:00:00 AM	08/19/2024, 11:28:33 AM	28m 34s	-

Status: Succeeded

Lineage: 28 upstream tables, 21 downstream tables

Compute

- Job cluster

Driver: Standard_D4s_v5 - Workers: Standard_D4s_v5 - 1 worker - from pool - DBR: 14.3 LTS (includes Apache Spark 3.5.0, Scala 2.12)

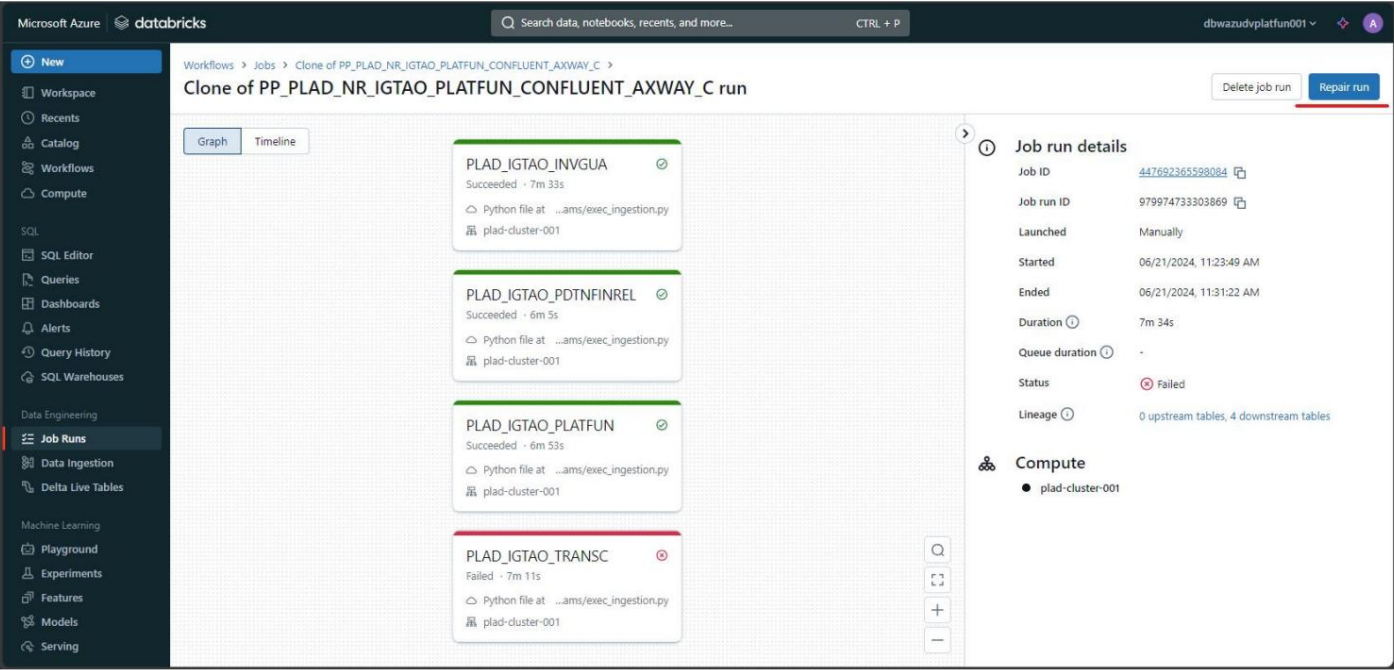
View details | Spark UI | Logs | Metrics

Job Restart

Restart processes from the Step with error/problem, without processing steps without error

Served by IWS?
How?

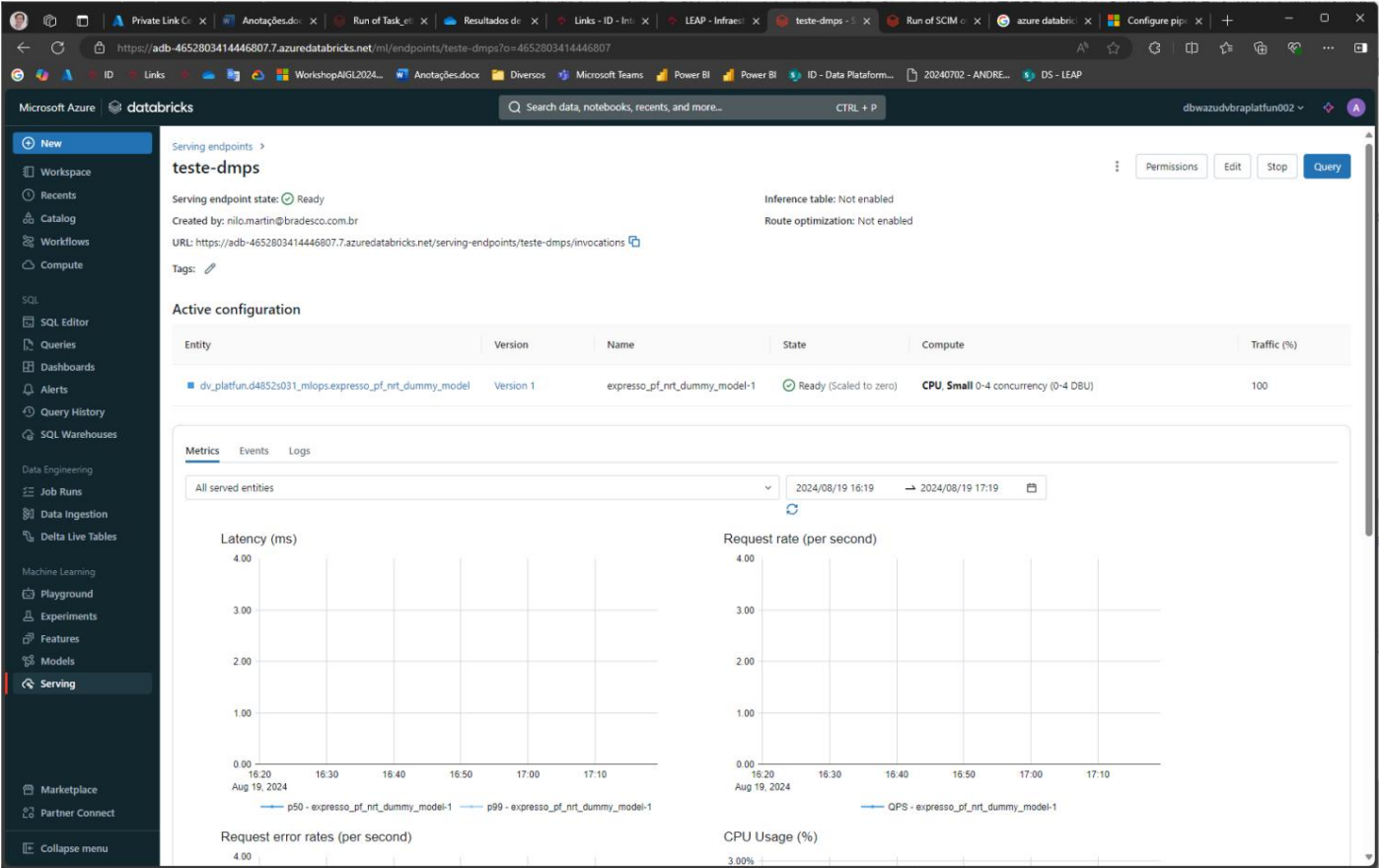
Example, in the case below, if Repair Run is executed, only the process with error will be re-executed.



Model Serving

Monitoring of models served via API endpoint.
If unavailability occurs, the metrics and logs in Databricks must be checked.

Served by IWS?
How?



[Model service with Azure Databricks – Azure Databricks | Microsoft Learn](#)

Links

[Automation Hub!](#)

[HCL \(IWS\) – Azure Databricks integration](#)

[Azure Blob and Data Lake Storage Integration](#)

- Only Upload or Download file from Azure Blob and Data Lake Storage.

[Submitting jobs by using Azure Databricks Plugin with Workload Automation \(hcl-software.com\)](#)