

AZURE DATA FACTORY

Tratamento e Transformação de dados

Laboratório Prático

TRANSFORMANDO PROFISSIONAIS EM
ESPECIALISTAS EM CLOUD

A smiling man with a beard, wearing a white shirt, is sitting at a desk. He is holding a white mug and looking towards the left. A laptop is open on the desk in front of him. The background is a blurred office environment.

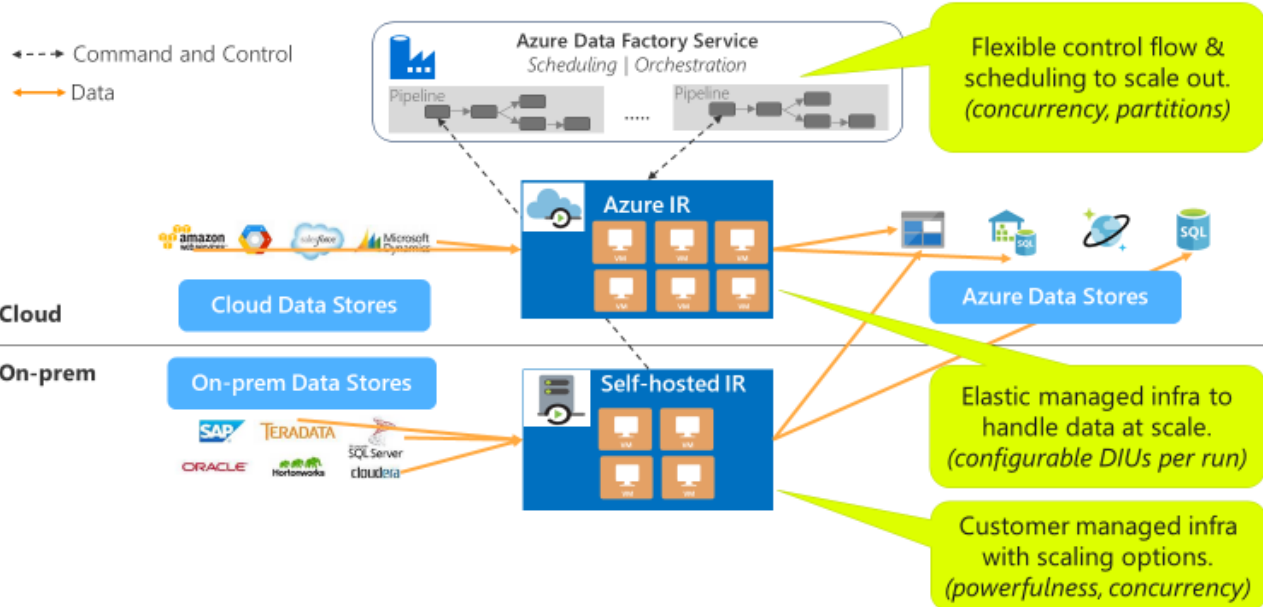
Azure
Academy

www.AzureAcademy.com.br

Conceitos

Performance e Processamento

Understand How ADF Copy Scales



Integration Runtime:

DIUs – Quantidade de computação escalada para execuções. Pode ser do Azure ou Self-hosted.

Consulte <https://docs.microsoft.com/pt-br/azure/data-factory/copy-activity-performance>

Preços: <https://azure.microsoft.com/pt-br/pricing/details/data-factory/data-pipeline/>

Conceitos

Performance e Processamento

Integration Runtime:
Pode ser gerenciado na guia MANAGE.

Integration runtime setup

Network environment:

Choose the network environment of the data source / destination or external compute to which the integration runtime will connect to for data flows, data movement or dispatch activities:



Azure

Use this for running data flows, data movement, external and pipeline activities in a fully managed, serverless compute in Azure.



Self-Hosted

Use this for running activities in an on-premise / private network

[View more](#) ▾

External Resources:

You can use an existing self-hosted integration runtime that exists in another resource. This way you can reuse your existing infrastructure where self-hosted integration runtime is setup.



Linked Self-Hosted

[Learn more](#)

Integration runtime setup

The Data Factory manages the integration runtime in Azure to connect to required data source/destination or external compute in public network. The compute resource is elastic allocated based on performance requirement of activities.

Name *

integrationRuntime1

Description

Enter description here...

Type

Azure

Virtual network configuration (Preview)

☒ Disable ☐ Enable

Region *

Auto Resolve ▾

▲ Data flow run time

Compute type *

General purpose ▾

Core count *

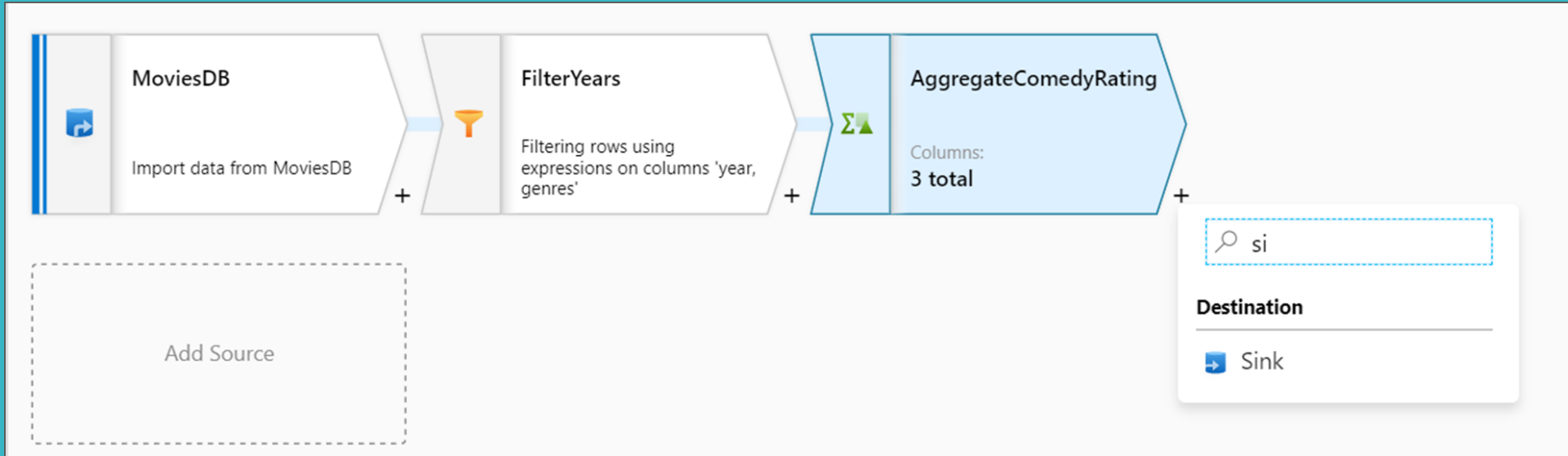
4 (+ 4 Driver cores) ▾

Time to live

30 minutes ▾

Conceitos

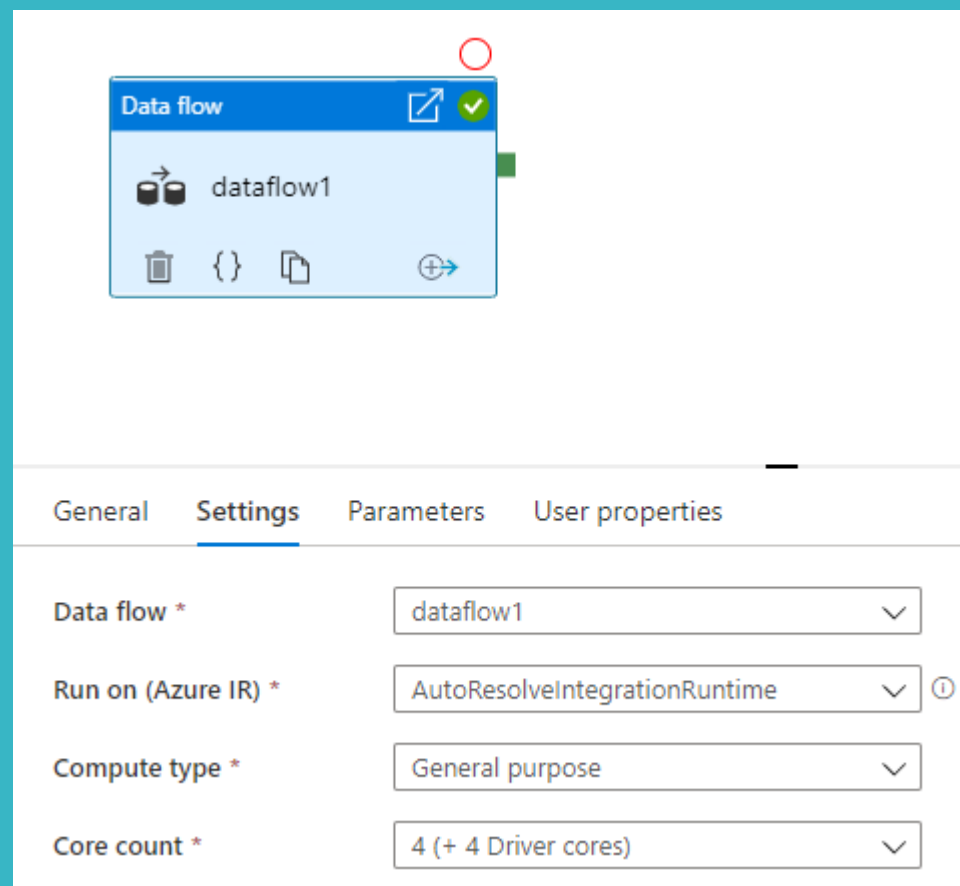
Transformação e limpeza de dados com DataFlows



Permite utilizar connections de diversas arquiteturas simultaneamente controlando visualmente o processo de transformação de dados.

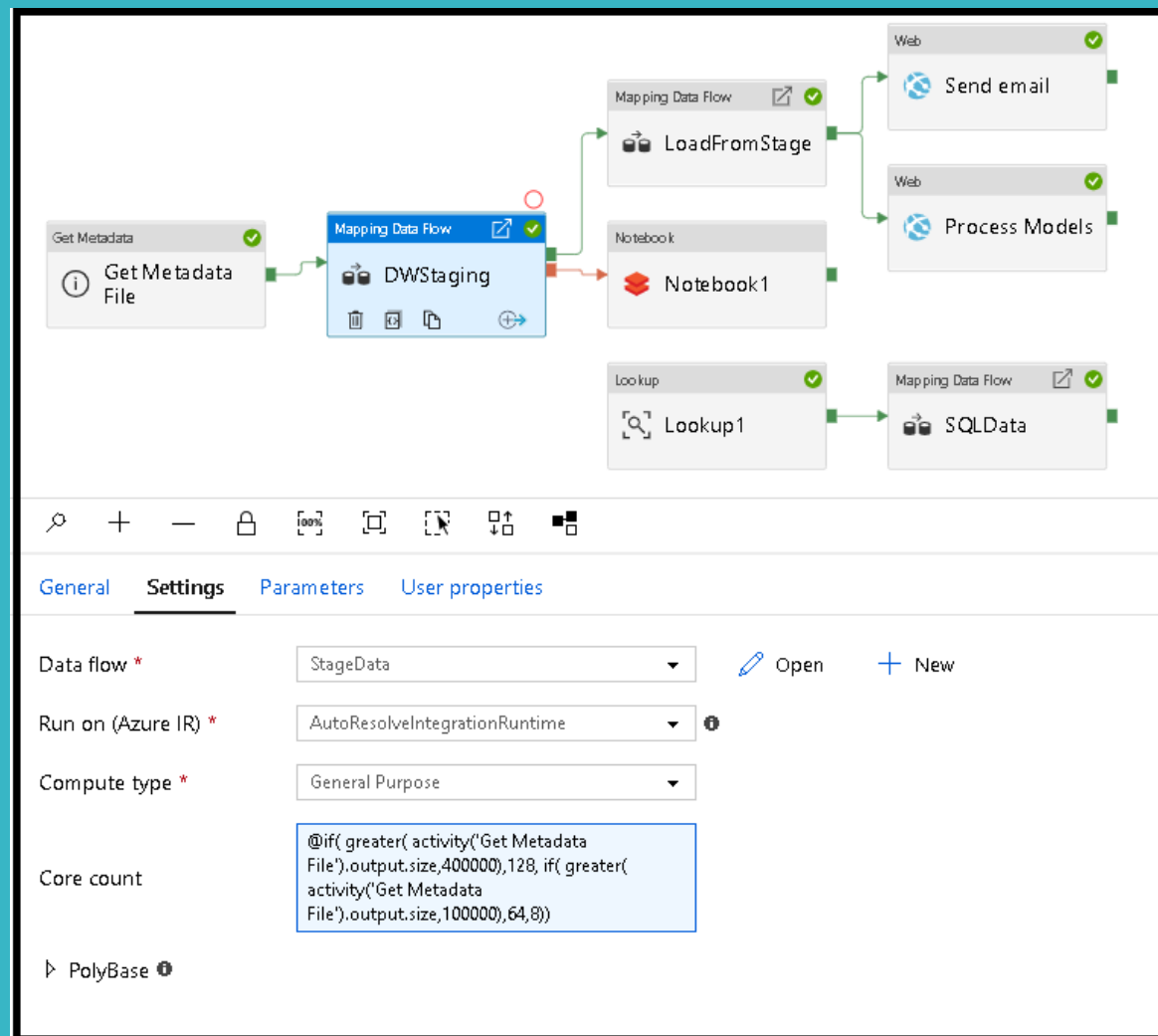
Conceitos Performance e Processamento

Personalize a computação na atividade Data Flow:



The screenshot shows the configuration panel for a Data Flow activity named 'dataflow1'. The 'Settings' tab is selected, showing the following configuration:

- Data flow ***: dataflow1
- Run on (Azure IR) ***: AutoResolveIntegrationRuntime
- Compute type ***: General purpose
- Core count ***: 4 (+ 4 Driver cores)



The screenshot shows a complex data flow diagram in the Azure Data Factory interface. The diagram includes activities such as 'Get Metadata File', 'Mapping Data Flow', 'LoadFromStage', 'Notebook1', 'Web', 'Send email', 'Process Models', 'Lookup1', and 'SQLData'. The 'Settings' tab is selected, showing the following configuration:







- Data flow ***: StageData
- Run on (Azure IR) ***: AutoResolveIntegrationRuntime
- Compute type ***: General Purpose
- Core count**: `@if(greater(activity("Get Metadata File").output.size,400000),128, if(greater(activity("Get Metadata File").output.size,100000),64,8))`

Conceitos

Performance e Processamento

Computação com parametrização:

General **Settings** Parameters User properties

Data flow *	<div>ADLSGen2DF</div>	 Open	 New
Run on (Azure IR) *	<div>AutoResolveIntegrationRuntime</div>		
Compute type	<div>@activity('LookupComputeTypes').output.firstRow.computeType</div>		
Core count	<div>@int(activity('LookupComputeTypes').output.firstRow.coreCount)</div>		
<div>▶ PolyBase </div>			



Laboratório

DataFlows

Assuntos:

- DataFlows
- Funções
- Agregações
- Filtros
- Expressões

Tecnologias:

- Storage Account - Data Lake Storage



Lab – DataFlow

1. Crie uma nova storage account do tipo V2 e crie um container de Blob chamado **arquivos**.
2. Faça o upload para o container, do arquivo **moviesDB.csv** disponível no Portal do Aluno.
3. Abra o Data Factory e crie um novo Pipeline.
4. Verifique se já existe uma connection para a storage account. Caso não, crie uma nova.
5. Crie um Dataflow. Configure o Source conforme as imagens a seguir e clique na opção para criar um novo dataset.

✓ Validate

MoviesDb
Columns:
0 total

Add Source

Source settings Source options Projection Optimize Inspect Data description

Output stream name * MoviesDb [Learn more](#)

Source type * Dataset

Dataset * Select... [+ New](#)

New dataset

Select a data store

Search

All Azure Database File Generic protocol NoSQL Services and apps

Azure Blob Storage Azure Data Lake Storage Gen1 Azure Data Lake Storage Gen2

Azure SQL Data Warehouse Azure SQL Database Amazon Marketplace Web Service

Select format

Choose the format type of your data

Parquet DelimitedText Json

Avro ORC Binary

Set properties

Name
DelimitedText1

Linked service *
AzureDataLakeStorage1

File path
arquivos / Directory / moviesDb.csv

First row as header ☐

Import schema
☒ From connection/store ☐ From sample file ☐ None



VÍDEO

Lab – DataFlow

1. Acione o **Data Preview** para conferir a conexão.

Source settingsSource optionsProjectionOptimizeInspectData preview

Number of rows INSERT 100UPDATE 0DELETE 0UPSERT 0

RefreshTypecastModifyMap driftedStatisticsRemove

↕	movie	abc	title	abc	genres	abc	year	abc	Rating	abc
+	108583		Fawlty Towers (1975		Comedy		-1980		1	
+	32898		Trip to the Moon, A (Voyage dan...		Action Adventure Fantasy Sci-Fi		1902		7	
+	7065		Birth of a Nation, The		Drama War		1915		6	
+	7243		Intolerance: Love's Struggle Thr...		Drama		1915		4	
+	62383		20,000 Leagues Under the Sea		Action Adventure Sci-Fi		1915		9	
+	8511		Immigrant, The		Comedy		1917		4	

2. Caso não tenha importado o título das colunas, edite o dataset e importe o schema.
3. Insira uma atividade 'Filter' e configure o filtro com a expressão: **toInteger(year) >= 1910 && toInteger(year) <= 2000**

MoviesDb

Import data from

Filter1

Columns:

Filter settings Optimize Inspect **Data preview** ●

Output stream name *

Filter1

Incoming stream *

MoviesDb ▾

Filter on *

toInteger(year) >= 1910 && toInteger(year) <= 2000 ✓

Para descobrir quais filmes são Comedies, você pode usar a `rlike()` função para localizar o padrão 'comédia' nos gêneros de coluna. Union a expressão `RLIKE` com a comparação de anos para obter:

toInteger(year) >= 1910 && toInteger(year) <= 2000 && rlike(genres, 'Comedy')

Lab – DataFlow

1. Adicione uma atividade do tipo 'agregação'. Configure conforme as imagens:

Aggregate settings | Optimize | Inspect | Data preview ●

Output stream name * [Learn more](#)

Incoming stream *

☒ Group by ☐ Aggregates

Columns	Name as
<input type="text" value="abc"/> <input type="text" value="year"/>	<input type="text" value="year"/>

Aggregate settings | Optimize | Inspect | Data preview ●

Output stream name * [Learn more](#)

Incoming stream *

☐ Group by ☒ Aggregates

Grouped by: year

[+ Add](#) [Clone](#) [Delete](#) [Open expression builder](#)

Column	Expression
<input type="text" value="AverageComedyRating"/>	<input type="text" value="avg(toInteger(Rating))"/> 1.2

`avg(toInteger(Rating))`

Mais sobre agregação: <https://docs.microsoft.com/pt-br/azure/data-factory/data-flow-aggregate>

Lab – DataFlow

1. Adicione uma atividade do tipo 'Sink'. Configure um novo dataset:

The screenshot shows a DataFlow pipeline with three activities: 'MoviesDb' (Import data from DelimitedText1), 'Filter1' (Filtering rows using expressions on columns 'year'), and 'Aggregate1' (Aggregating data by 'year' producing columns 'AverageComedyRating'). The 'sink1' activity is highlighted, showing its configuration:

- Output stream name ***: sink1
- Incoming stream ***: Aggregate1
- Sink type ***: Dataset

The 'New dataset' dialog shows a search bar and tabs for 'All', 'Azure', 'Database', 'File', 'Generic protocol', 'NoSQL', and 'Services and apps'. Under the 'Azure' tab, 'Azure Data Lake Storage Gen2' is selected.

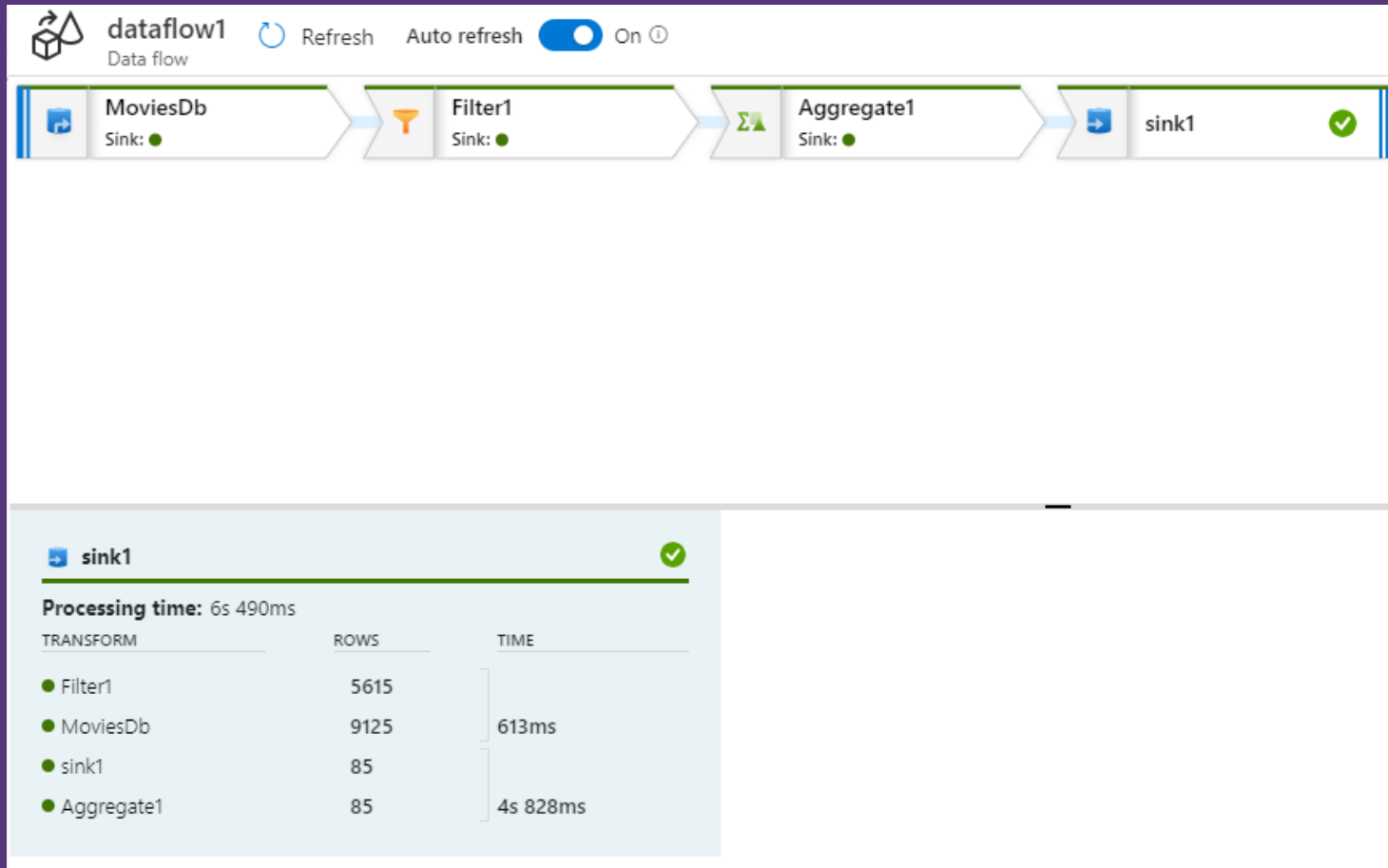
The 'Select format' dialog shows a grid of format options. 'DelimitedText' is selected, indicated by a blue border.

The 'Set properties' dialog for the Sink activity shows the following configuration:

- Name**: DelimitedText2
- Linked service ***: AzureDataLakeStorage1
- File path**: arquivos / Directory / File
- First row as header**: ☒
- Import schema**: ☐ From connection/store ☐ From sample file ☒ None

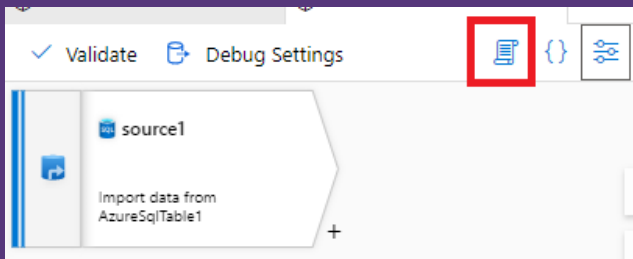
Lab – DataFlow

1. Retorne ao Pipeline e insira a atividade do DataFlow.
2. Teste a execução e confira os resultados.



Lab 2 – DataFlow – Remover duplicidades ou nulos

1. Crie um novo Dataflow e conecte o Source a um dataset do SQL do Azure com uma tabela de exemplo.
2. Em seguida, clique no botão '**scripts**' conforme imagem:



3. Insira as linhas a seguir para procurar e remover duplicidades:
`source1 aggregate(groupBy(mycols = sha2(256,columns())),
 each(match(true()), $$ = first($$)) ~> DistinctRows`
4. Teste os resultados utilizando o Data Preview.
5. O Código a seguir procura e elimina valores nulos:
`source1 split(contains(array(columns()),isNull(#item)),
 disjoint: false) ~> LookForNULLs@(hasNULLs, noNULLs)`

Consulte outros scripts de transformação de dados:

<https://docs.microsoft.com/pt-br/azure/data-factory/data-flow-script#distinct-row-using-all-columns>



Mapping data flows performance tuning

Avalie o resultado das execuções para determinar possíveis melhorias:

Activity runs

Pipeline run ID e1cd063f-b5e3-40a9-869c-39d3770c6f67

All status ▾

Showing 1 - 1 of 1 items

ACTIVITY NAME	ACTIVITY TYPE	RUN START ↑↓	DURATION	STATUS	INTEGRATION RUNTIME	USER PROPERTI...	ERROR
GenericSCDT...	ExecuteDataFlov	7/23/20, 5:06:14 PM	00:07:09	✓ Succeeded	DefaultIntegrationRuntime (West US)		

BestBeersByState Data flow

Refresh Auto refresh ☒ On ⓘ

1. Cluster startup time
Cluster startup time: 4m 12s 157ms Number of transforms: 12 ✕

Reviews Sink: ● ConvertTypes Sink: ● AggregateByBeer Sink: ● FilterInvalidData Sink: ●

Breweries Sink: ●

Beers Sink: ●

4. Sink processing time

2. Source read time

OutputToADLS ✓

Processing time: 1m 35s 864ms

TRANSFORM	ROWS	TIME
Breweries	-	-
ConvertTypes	9m	
Reviews	9m	1m 9s 447ms
Beers	358k	6s 379ms
FilterInvalidData	305k	

ConvertTypes Derive

Total columns	10
New columns	0
Updated columns	6
Dropped columns	0
Drifted columns	0

Stream information

Rows calculated	9,073,128
Total partition	18
Stage time	1m 9s 447ms
Last update (PDT)	7/20/2020, 5:05:58 AM

Partition chart
800000

Mapping data flows performance tuning

Os fluxos de dados **utilizam um otimizador Spark** que reordena e executa sua lógica de negócios em 'estágios' para executar o mais rápido possível.

- Para cada coletor no qual seu fluxo de dados grava, a saída de monitoramento lista a duração de cada estágio de transformação, junto com o tempo que leva para gravar dados no coletor.
- Se o estágio de transformação que leva o maior contém uma fonte, você pode deve otimizar ainda mais seu tempo de leitura.
- **Se uma transformação estiver demorando muito, talvez seja necessário reparticionar ou aumentar o tamanho do tempo de execução do IR.**
- Se o tempo de processamento do coletor for grande, pode ser necessário aumentar seu banco de dados ou verificar se você não está gerando um único arquivo.



Mapping data flows performance tuning

Guia Otimização

Aggregate settings
Optimize
Inspect
Data preview

Partition option *
☐ Use current partitioning
☐ Single partition
☒ Set Partitioning

Partition type *

Round Robin

Hash

Dynamic Range

Fixed Range

Key

Number of partitions *

Configurações para o esquema de particionamento do cluster Spark. Essa guia existe em todas as transformações do fluxo de dados e permite reparticionar os dados após a conclusão da transformação. Ajustar o particionamento fornece controle sobre a distribuição de seus dados entre nós de computação e otimizações de localidade de dados que podem ter efeitos positivos e negativos no desempenho geral do fluxo de dados.



Mapping data flows performance tuning

Por padrão, **Usar particionamento atual** é selecionado, o que instrui o Data Factory a manter o particionamento de saída atual da transformação. Como o reparticionamento de dados leva tempo, este modo é recomendado na maioria dos cenários. Os cenários em que você pode reparticionar dados incluem agregações e junções que distorcem significativamente seus dados ou ao usar o particionamento de origem em um banco de dados SQL.

Round robin

O round robin distribui dados igualmente entre as partições. Use quando você tiver implementado boa estratégia de chaves. Você pode definir o número de partições físicas.

Hash

O ADF produz um hash de colunas para produzir partições uniformes, de forma que as linhas com valores semelhantes caiam na mesma partição.

Range dinâmico

O intervalo dinâmico usa intervalos dinâmicos do Spark com base nas colunas ou expressões que você fornece.

Alcance fixo

Construa uma expressão que forneça um intervalo fixo para valores em suas colunas de dados particionadas. Para evitar distorção da partição, você deve ter um bom conhecimento de seus dados antes de usar esta opção. Os valores inseridos para a expressão são usados como parte de uma função de partição. Você pode definir o número de partições físicas.

Mapping data flows performance tuning

Para Fontes SQL do Azure

O Banco de Dados SQL do Azure tem uma opção de particionamento exclusiva chamada particionamento de 'Origem'. Habilitar o particionamento de origem pode melhorar o tempo de leitura do Banco de Dados, habilitando conexões paralelas no sistema de origem.

Source Settings Source Options Projection **Optimize** Inspect Data Preview

Partition option *
☐ Use current partitioning ☐ Single partition ☒ Set Partitioning

Partition type *

Round Robin

Hash

Dynamic Range

Fixed Range

Key

Source

Number of partitions *

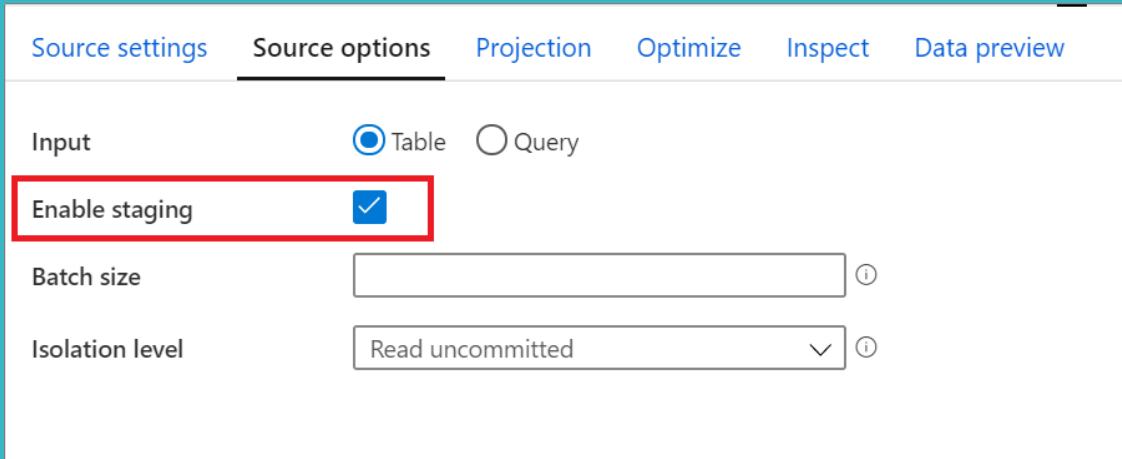
Partition read via *
☒ Column ☐ Query condition ⓘ

Partition column *
 ⓘ

Mapping data flows performance tuning

Para Fontes Synapse Analytics – Pool DW

Ao usar o Azure Synapse Analytics, existe uma configuração chamada **Habilitar preparação** nas opções de origem. Isso permite que o ADF leia o Synapse usando o Staging, o que melhora muito o desempenho de leitura. Habilitar o teste requer que você especifique um local de teste do Azure Blob Storage ou do Azure Data Lake Storage gen2 nas configurações de atividade de fluxo de dados.



Source settings **Source options** Projection Optimize Inspect Data preview

Input ☒ Table ☐ Query

Enable staging ☒

Batch size ⓘ

Isolation level ⓘ

Mapping data flows performance tuning

Otimizar operações de Sink – Gravação em SQL

Desativar índices antes de carregar dados em um banco de dados SQL pode melhorar muito o desempenho de gravação na tabela. Execute o comando abaixo antes de gravar em seu coletor de SQL.

Sink
Settings
Mapping
Optimize
Inspect
Data preview

Update method
☒ Allow insert Add dynamic content [Alt+P]
☐ Allow delete
☐ Allow upsert
☐ Allow update

Table action
☐ None
☒ Recreate table
☐ Truncate table

Batch size
 ⓘ

Pre SQL scripts

ALTER INDEX ALL ON dbo. [Table Name]
DISABLE

Post SQL scripts

ALTER INDEX ALL ON dbo. [Table Name]
REBUILD

DATA FACTORY + SERVERLESS MICROSERVICES

Serverless Components in Azure



Functions

Serverless Compute



Logic Apps

Serverless Workflows



Event Grid

Serverless Events

Oportunidade para integrar Data Factory com Serverless aumentando a capacidade de limpeza e processamento de dados.



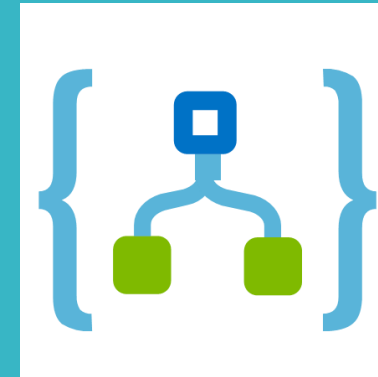
MICROSERVICES NO AZURE

Logic Apps:

Simplificam a criação de **fluxos de trabalho em nuvem**. Permitem modelagem e automação com fácil escala. Possuem gatilhos para disparo por agenda, APIs e outros procedimentos.

Podem orquestrar diferentes Azure Functions em um processo, especialmente quando o processo requer interação com um sistema externo ou uma API.

[Visão geral dos Aplicativos Lógicos do Azure - Azure Logic Apps | Microsoft Docs](#)



Azure Functions:

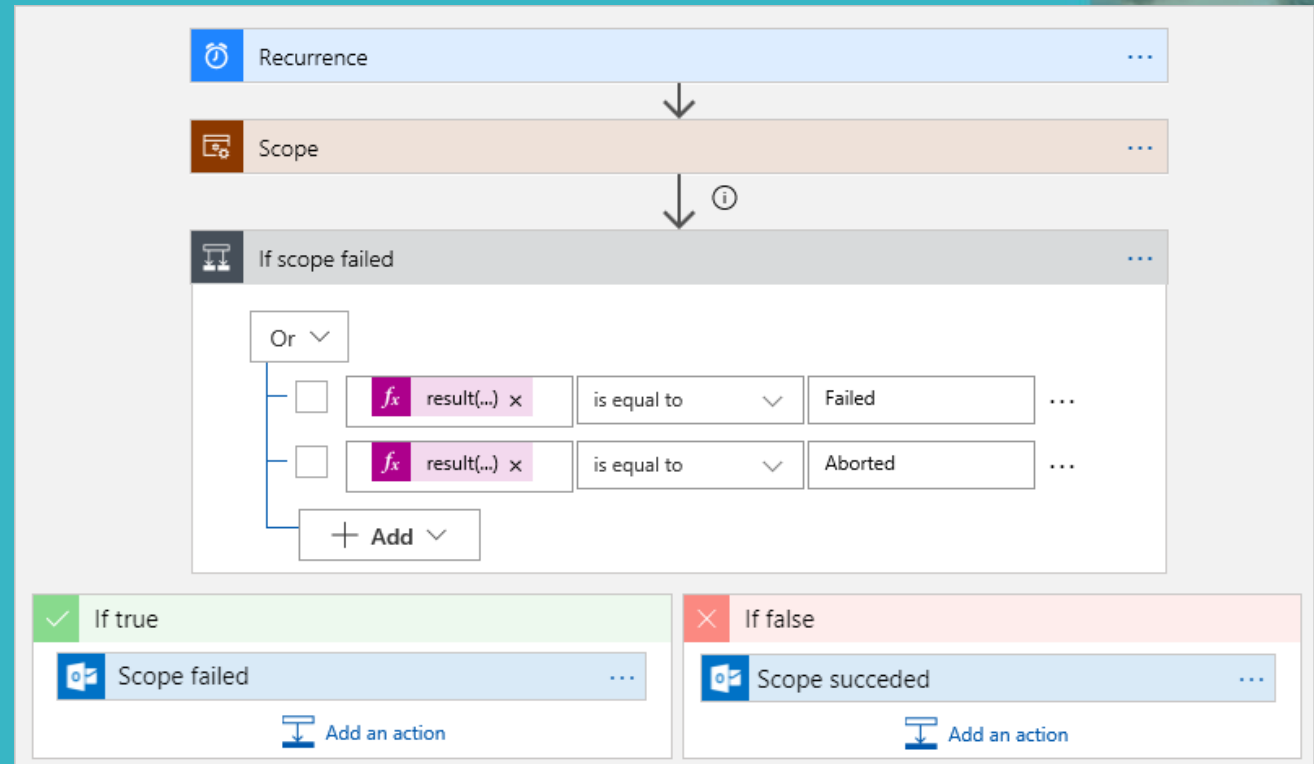
Serviço para **executar partes de código ou "funções" em nuvem**. Você pode escrever apenas o código necessário para o problema atual, sem se preocupar com um aplicativo inteiro ou a infraestrutura necessária. Suporta C#, F#, Node.js, Python ou PHP. Você paga apenas pelo tempo em que seu código é executado.



AZURE LOGIC APPS

Plataforma para criar e executar fluxos de trabalho automatizados que integram **aplicativos, dados, serviços e sistemas**.

- Agende e envie notificações por email usando o Office 365 quando ocorrer um evento específico, por exemplo, um novo arquivo for carregado.
- Encaminhe e processe pedidos de clientes entre sistemas locais e serviços de nuvem.
- Mova arquivos carregados de um servidor SFTP ou FTP para o Armazenamento do Azure.
- Monitore tweets, analise o sentimento e crie alertas ou tarefas para itens que exigem revisão.



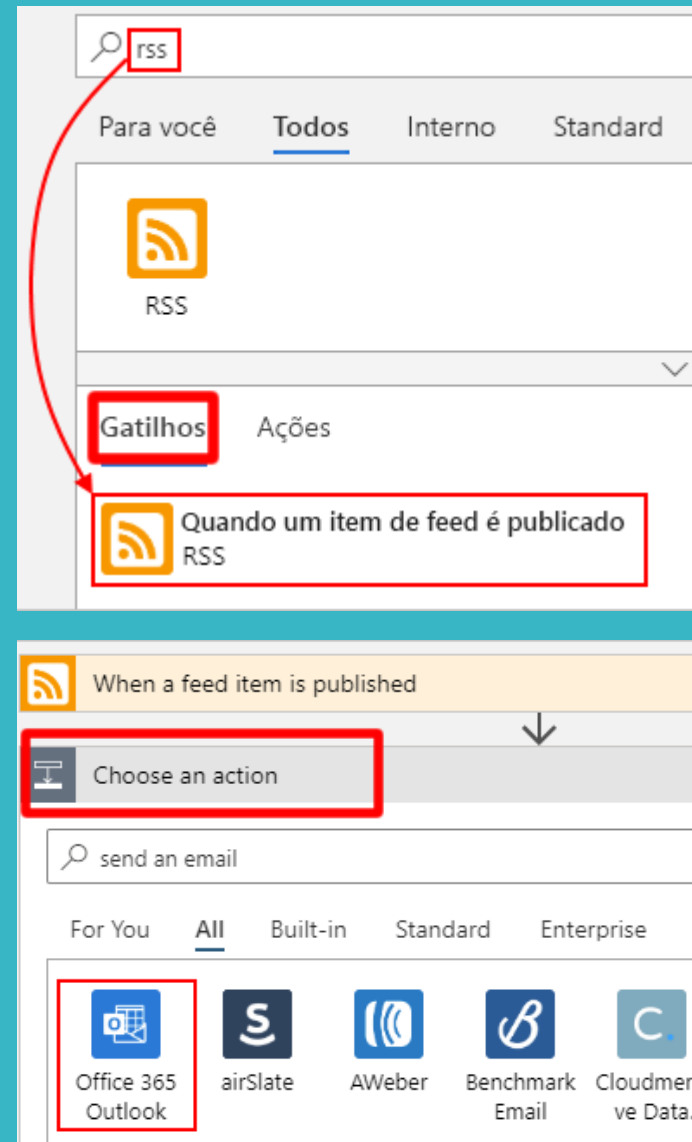
Logic Apps - Conceitos

Gatilhos:

Configuração inicial do Logic Apps. Determine um 'gatilho' para executar o Flow de ações. O gatilho pode receber parâmetros para se comportar como um request de uma API.

Ações:

Sequência de atividades. Pode chamar Funcions, gravar rotinas em bancos de dados, gerar arquivos e mais.



Laboratório

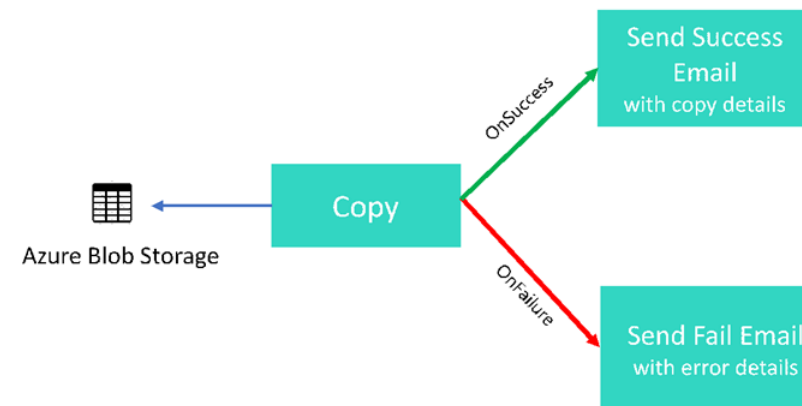
Ramificar atividades Serverless

Assuntos:

- Logs
- E-mails
- Posts no Teams

Tecnologias:

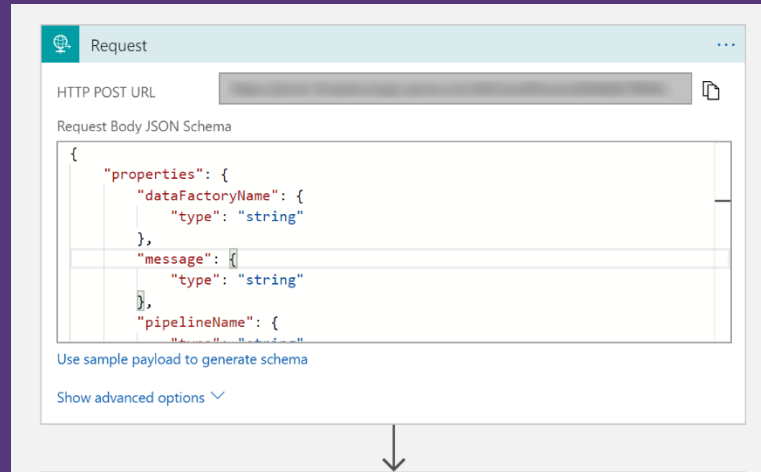
- Logic Apps
- Office 365



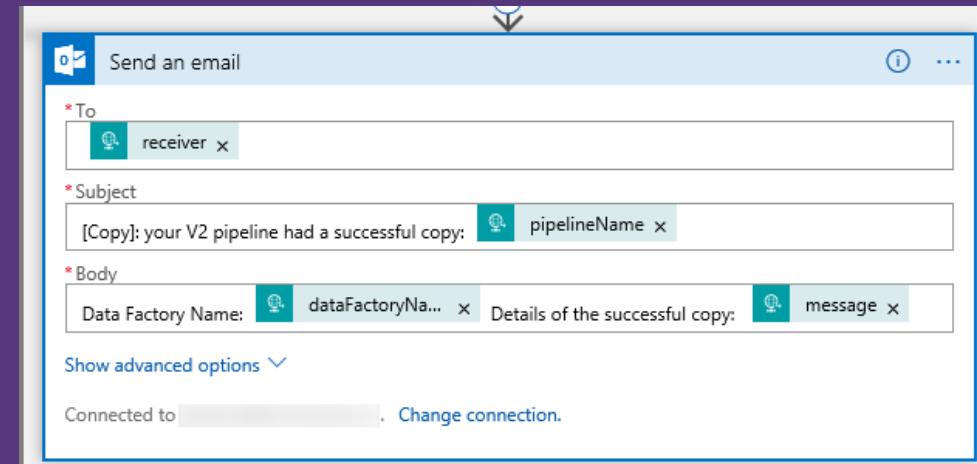
Lab – Logic Apps

1. No Azure, instale o serviço Logic App.
2. Edite o projeto e insira o gatilho: **When an HTTP request is received**
3. Preencha o body do gatilho com os parâmetros que serão enviados dinamicamente pelo Data Factory:

```
{
  "properties": {
    "dataFactoryName": {
      "type": "string"
    },
    "message": {
      "type": "string"
    },
    "pipelineName": {
      "type": "string"
    },
    "receiver": {
      "type": "string"
    }
  },
  "type": "object"
}
```

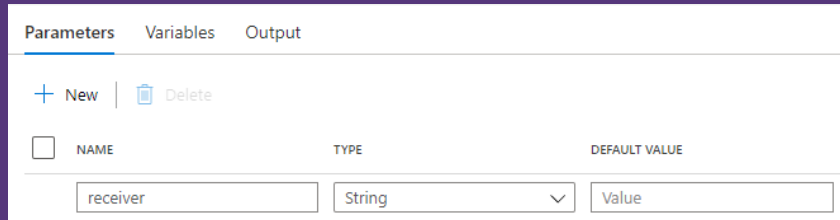


4. Insira uma ação: **Office 365 Outlook – Send an email** e configure conforme a imagem. Salve o projeto e anote o https do gatilho.
5. Crie outro logic app com as mesmas configurações, porém ajustando o assunto/body da mensagem para gerar um email de falha.



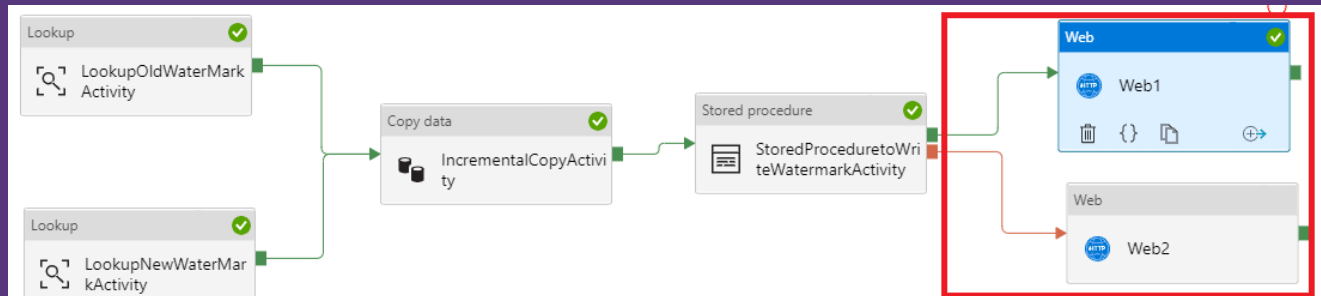
Lab – Logic Apps

1. Retorne ao Data Factory e edite um pipeline existente.
2. Na janela de propriedades do pipeline, alterne para a guia Parâmetros e use o botão Novo para adicionar o parâmetro a seguir do tipo String: **receiver**.

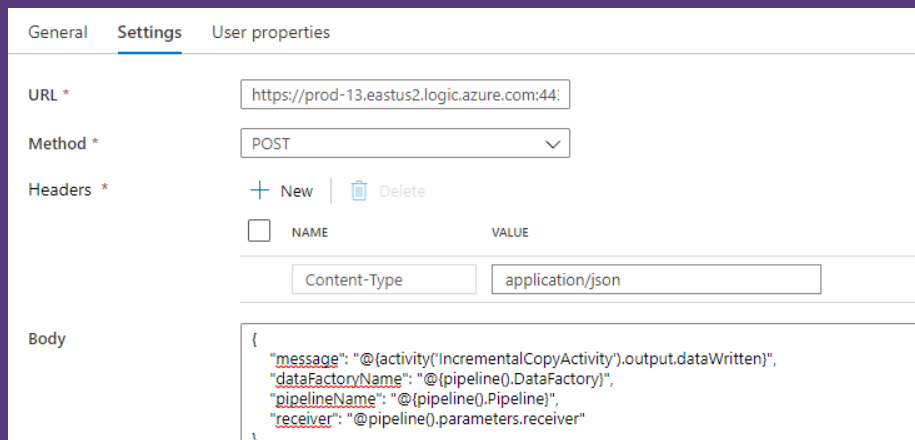


NAME	TYPE	DEFAULT VALUE
receiver	String	Value

3. Insira 02 atividades **WEB**. Uma para o sucesso da execução e outra para a falha.



4. Configure as atividades conforme o exemplo apontando a URL para o logic app.



NAME	VALUE
Content-Type	application/json

Body

```
{  
  "message": "@(activity('IncrementalCopyActivity').output.dataWritten)",  
  "dataFactoryName": "@(pipeline().DataFactory)",  
  "pipelineName": "@(pipeline().Pipeline)",  
  "receiver": "@pipeline().parameters.receiver"  
}
```

```
{  
  "message": "@(activity('Copy1').output.dataWritten)",  
  "dataFactoryName": "@(pipeline().DataFactory)",  
  "pipelineName": "@(pipeline().Pipeline)",  
  "receiver": "@pipeline().parameters.receiver"  
}
```

Lab – Logic Apps

1. Teste a execução e confira os resultados.

Acesse o artigo a seguir e verifique configurações adicionais, além da possibilidade de criar nomes de arquivos genéricos no Sink: <https://docs.microsoft.com/pt-br/azure/data-factory/tutorial-control-flow-portal>

```
@CONCAT(pipeline().RunId, '.txt')
```

Lab extra – Integração com Azure Functions:

<https://github.com/Azure/Azure-DataFactory/tree/master/SamplesV2/UntarAzureFilesWithAzureFunction>

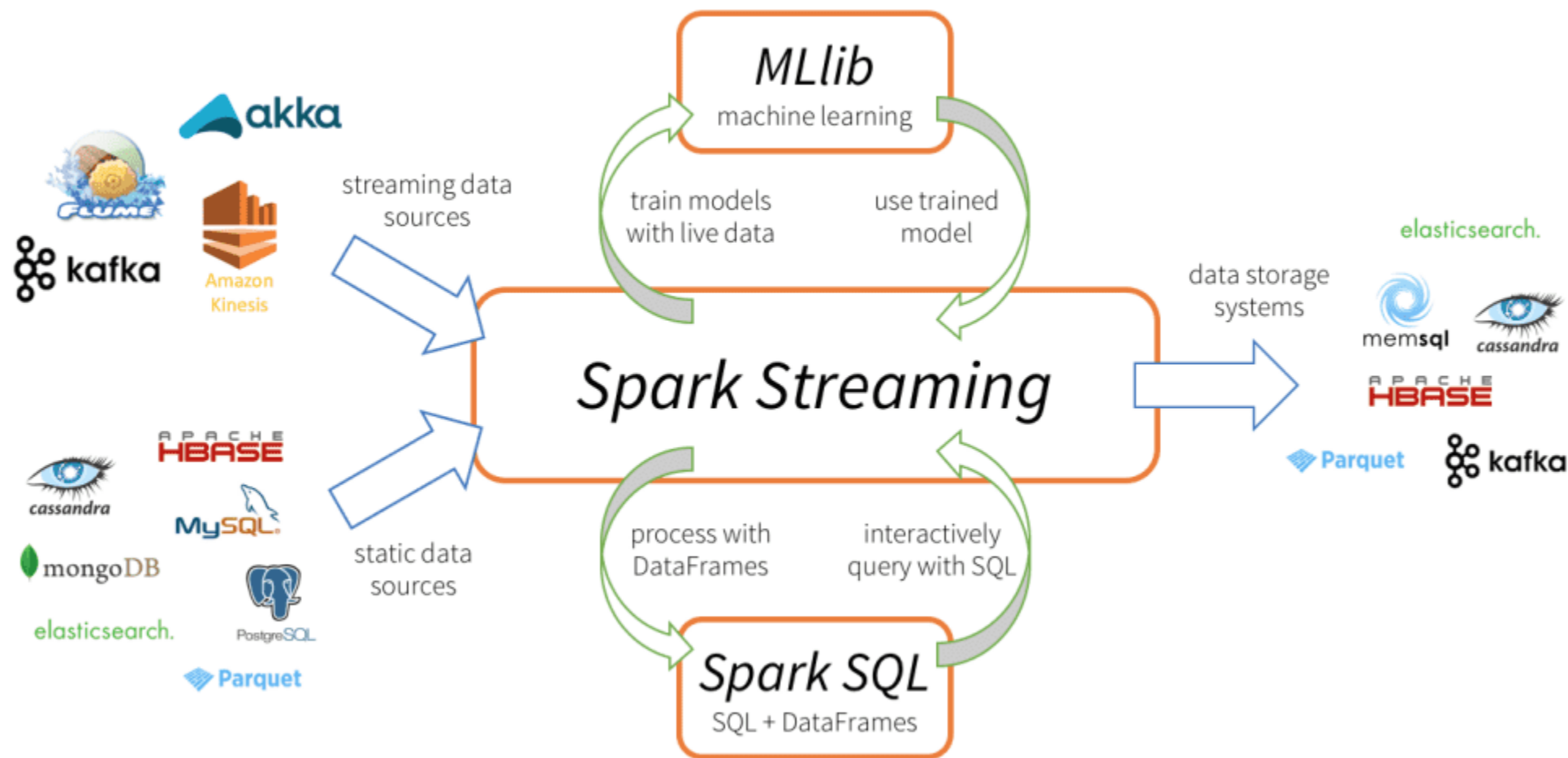


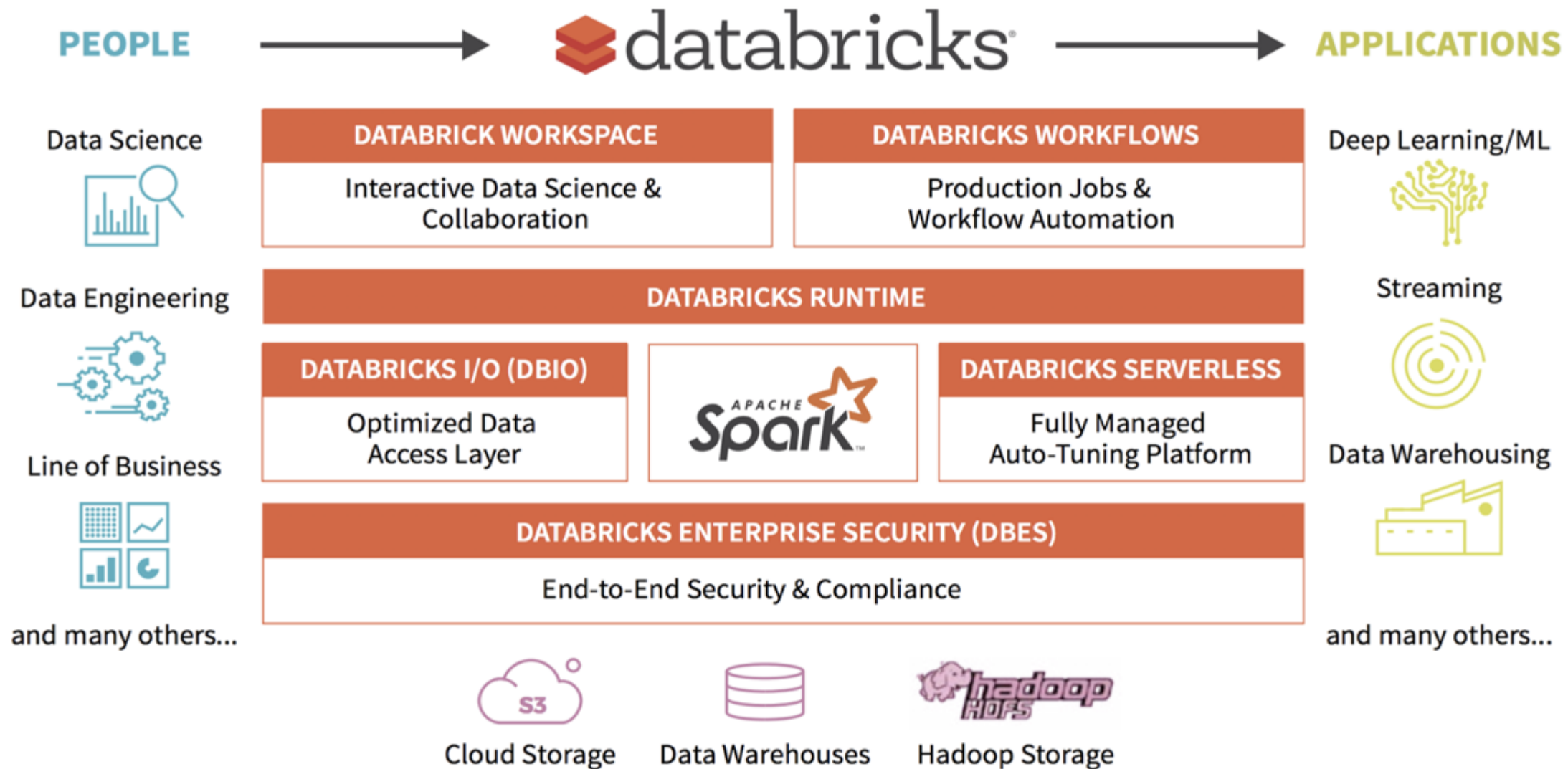
TRANSFORMAR DADOS COM PYTHON



Open Source Ecosystem

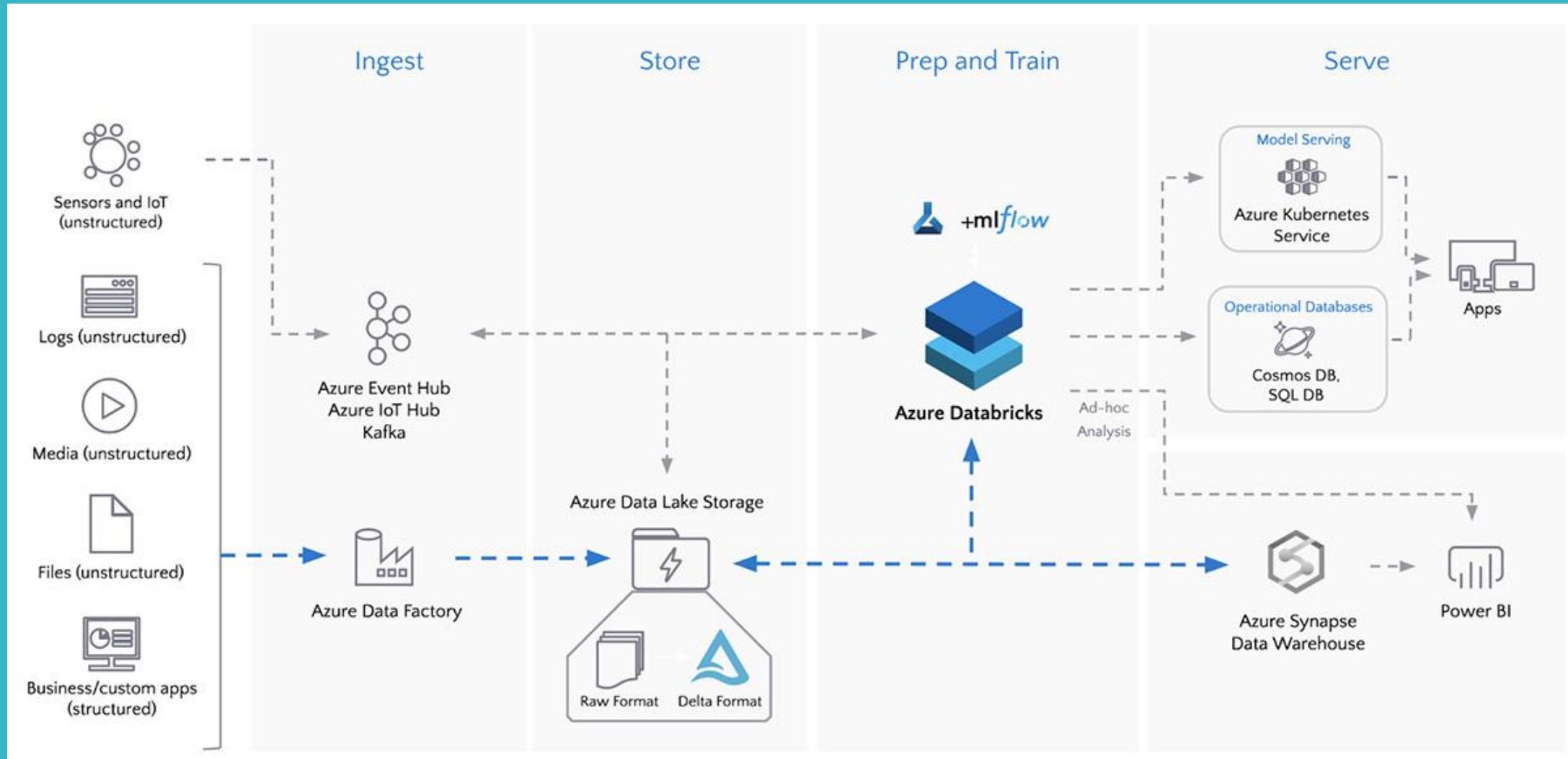






Databricks + Data Factory

- Funciona com Clusters: Interativos integrados à notebooks ou Jobs.
- Apache Spark possui foco em Big Data e tem o objetivo de processar grandes volumes de dados.



Tipos de dados

Estruturado

- Idade
- Faturamento
- Renda
- Núm. De produtos
- Estado

Estatística aplicada

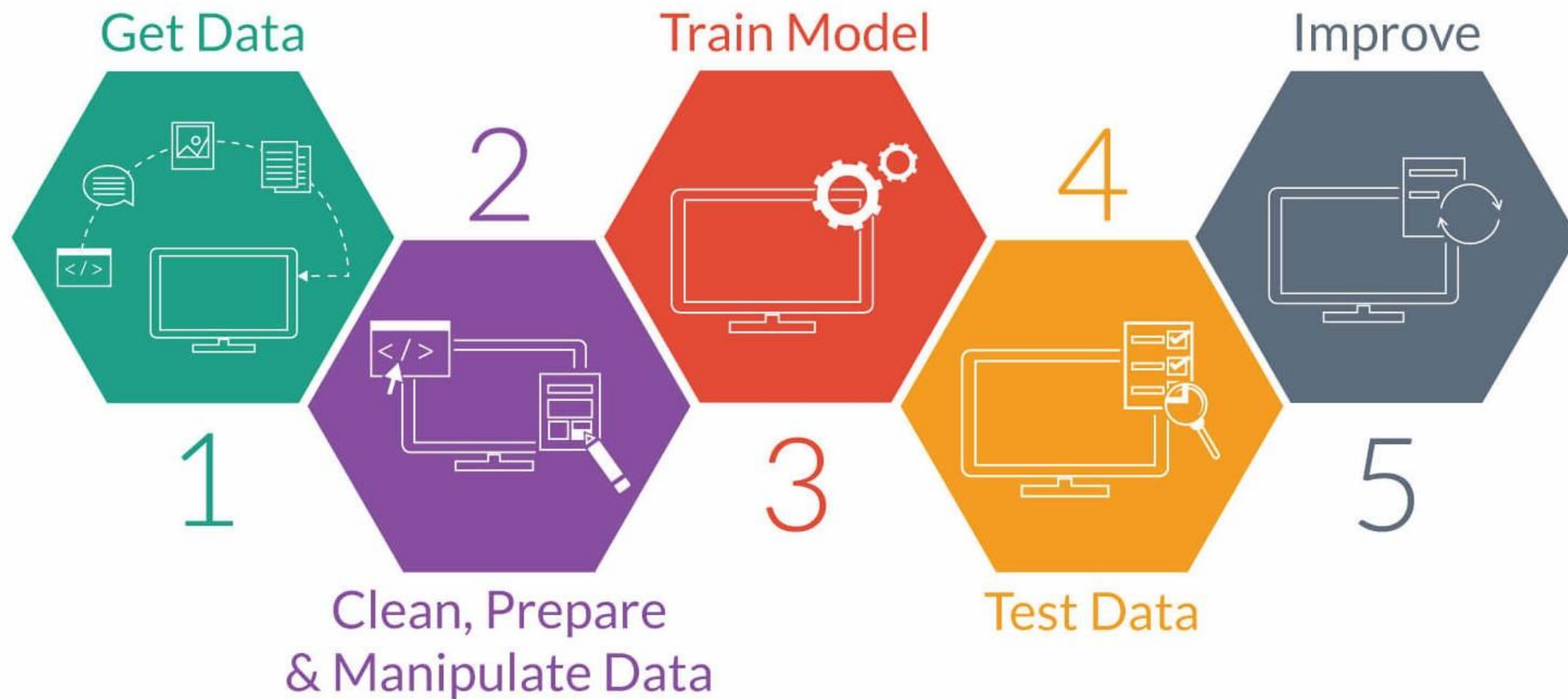
Não Estruturado

- Imagem
- Som
- Texto

Inteligência artificial

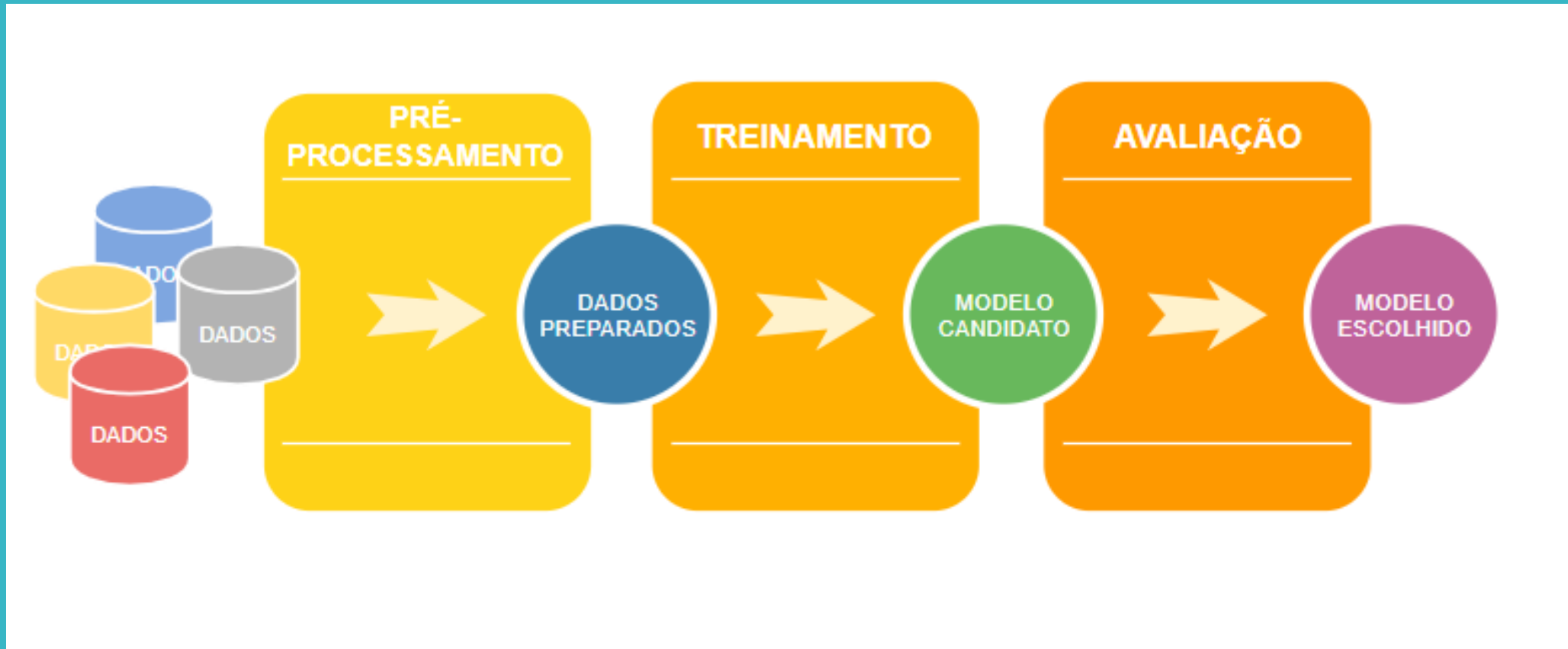


Inteligência Artificial



Arquitetura

Etapas de treinamento ensinam e aprimoram o algoritmo de acordo com os objetivos da detecção e classificação.



Laboratório

Processar dados com Databricks notebooks

Assuntos:

- Notebooks
- Machine Learning

Tecnologias:

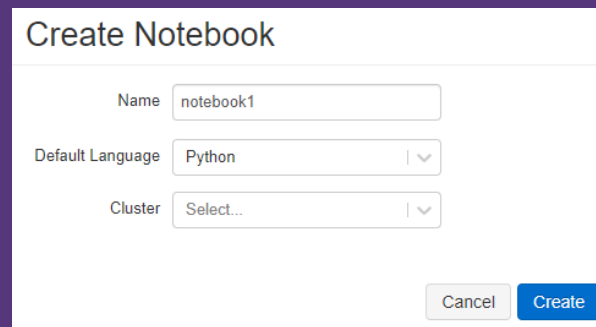
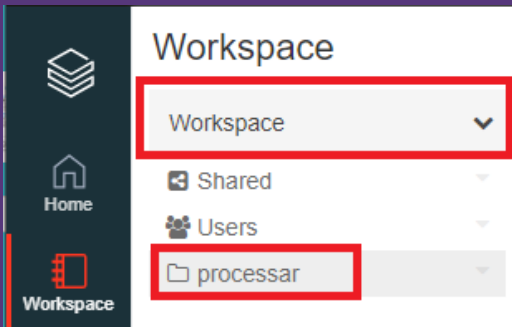
- Azure Databricks



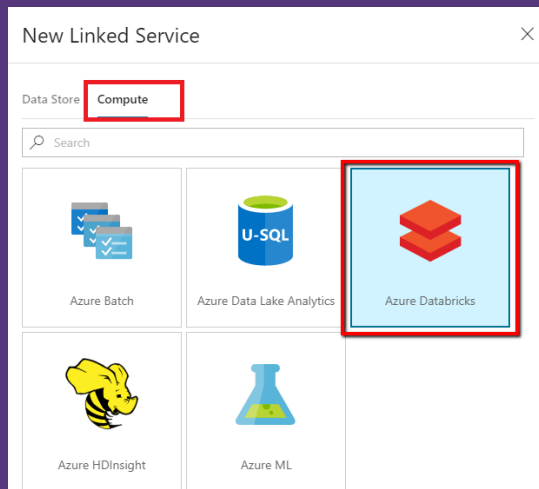
Lab – Databricks Notebooks

No Azure:

1. Instale o serviço **Azure Databricks Premium (14 dias Free)**.
2. Acesse o workspace do Databricks, crie um diretório chamado '**processar**' e crie um novo notebook com a **linguagem SQL**. Utilize o código disponível no arquivo 'notebook1.txt' no Portal do aluno.

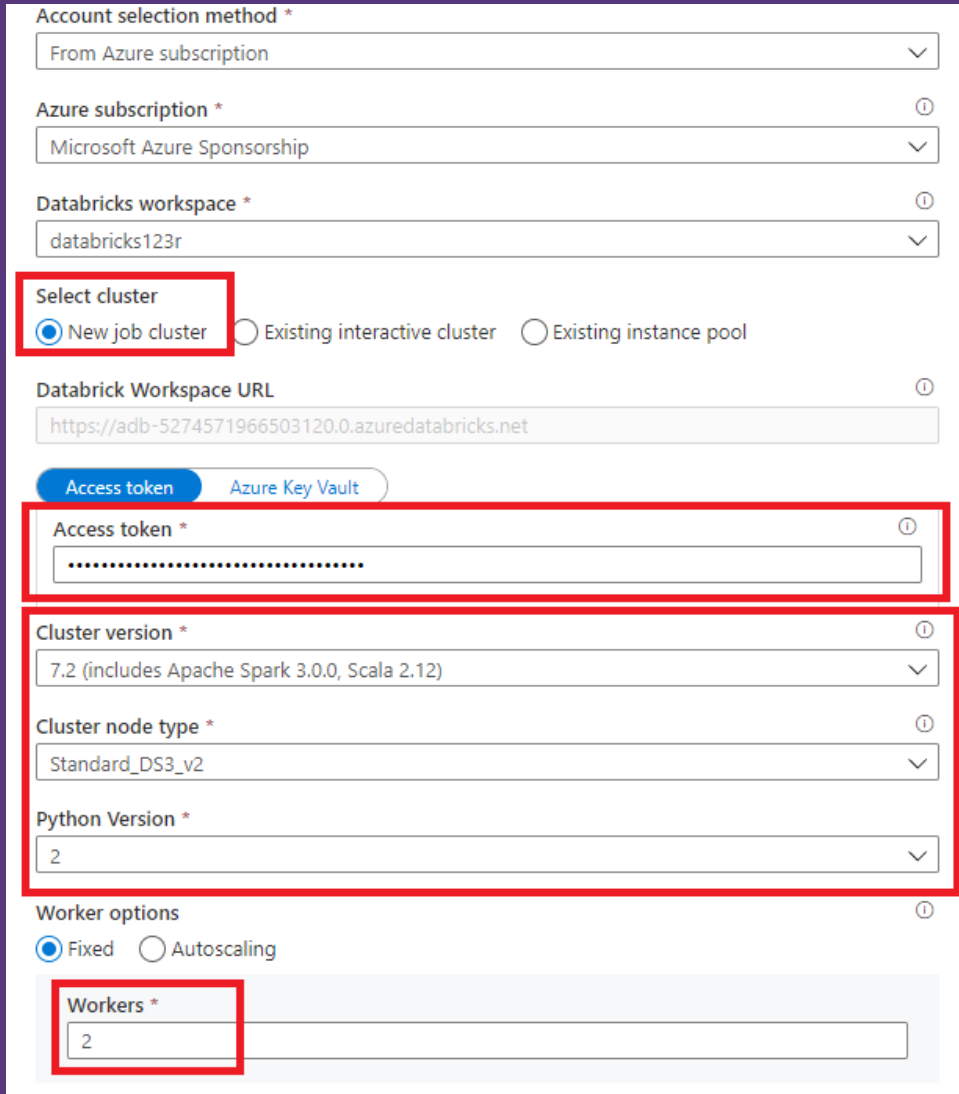


3. Abra um Data Factory e crie um novo pipeline.
4. Crie uma nova **connection** selecionando Databricks na guia de Computação.



Lab – Databricks Notebooks

1. Na janela de conexão, defina os parâmetros conforme a imagem:



Account selection method *

From Azure subscription

Azure subscription *

Microsoft Azure Sponsorship

Databricks workspace *

databricks123r

Select cluster

☒ New job cluster ☐ Existing interactive cluster ☐ Existing instance pool

Databrick Workspace URL

https://adb-5274571966503120.0.azuredatabricks.net

Access token

Access token *

Cluster version *

7.2 (includes Apache Spark 3.0.0, Scala 2.12)

Cluster node type *

Standard_DS3_v2

Python Version *

2

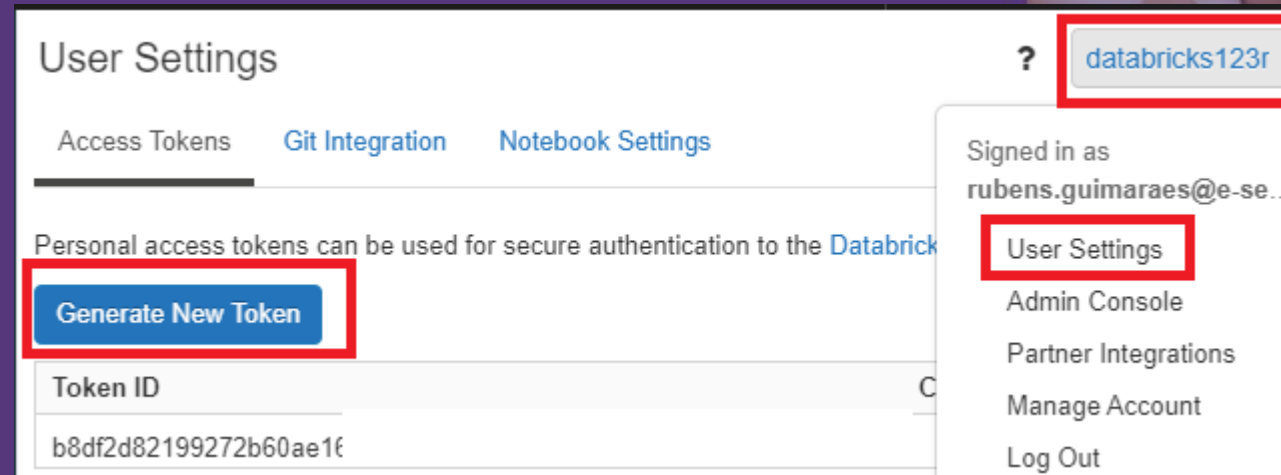
Worker options

☒ Fixed ☐ Autoscaling

Workers *

2

Dica: Para gerar o token, acesse o workspace do Databricks instalado anteriormente, clique no ícone de usuários e gere um token de acesso.



User Settings

Access Tokens Git Integration Notebook Settings

Personal access tokens can be used for secure authentication to the Databricks

Generate New Token

Token ID

b8df2d82199272b60ae1f

Signed in as rubens.guimaraes@e-se...

User Settings

Admin Console

Partner Integrations

Manage Account

Log Out

Lab – Databricks Notebooks

1. Após a criação da connection, retorne ao Pipeline.
2. No pipeline vazio, clique na guia parâmetros, em **Novo** e chame-a de **'name'**.

NAME	TYPE	DEFAULT VALUE
name	String	Value

3. Insira a atividade Databricks, notebooks no pipeline. Selecione a connection criada anteriormente. Na guia Settings, aponte para o notebook criado no início do lab.

General Azure Databricks **Settings** User properties

Notebook path * /processar/notebook1

Base parameters

+ New Delete

NAME	VALUE
input	@pipeline().parameters.name

4. Teste a execução. A caixa de diálogo **Execução de Pipeline** solicita o parâmetro name. Use **/path/filename**. Clique em Concluir.

Labs adicionais

ETL com Databricks:

<https://docs.microsoft.com/pt-br/azure/data-factory/solution-template-databricks-notebook>

Notebook conectado à Storage Account:

<https://docs.microsoft.com/pt-br/azure/azure-databricks/quickstart-create-databricks-workspace-portal>

Dataframes com Python

<https://docs.microsoft.com/pt-br/azure/databricks/spark/latest/dataframes-datasets/introduction-to-dataframes-python>

Machine Learning – Classificação com Databricks:

https://docs.microsoft.com/pt-br/azure/databricks/_static/notebooks/binary-classification.html





Azure Academy

Rubens Guimarães

 [/rubensguimaraes](#)



Microsoft
Regional Director

TRANSFORMANDO PROFISSIONAIS EM
ESPECIALISTAS EM CLOUD

www.AzureAcademy.com.br

PATROCÍNIO E APOIO:



e.Seth Cloud

Azure Academy



@azure-Academy



@azureacademyoficial



@Azure_Academy



@azureacademyBR



AzureAcademy



www.AzureAcademy.com.br

e.SethCloud

Acelere o crescimento da sua empresa

MIGRAÇÕES PARA CLOUD

Metodologia eficiente com foco em segurança,
economia de consumo e agilidade.

www.eSeth.com.br/Cloud



ACESSE NOSSOS SITES:

e.Seth
tecnologia

www.eSeth.com.br

e.SethCloud

www.eSeth.com.br/Cloud

Azure
Academy

www.azureacademy.com.br