

AZURE DATA FACTORY

Monitorar Pipelines

SSIS IR - Integration Services Integration Runtime

Gatilhos

Redes

Laboratório Prático

TRANSFORMANDO PROFISSIONAIS EM
ESPECIALISTAS EM CLOUD

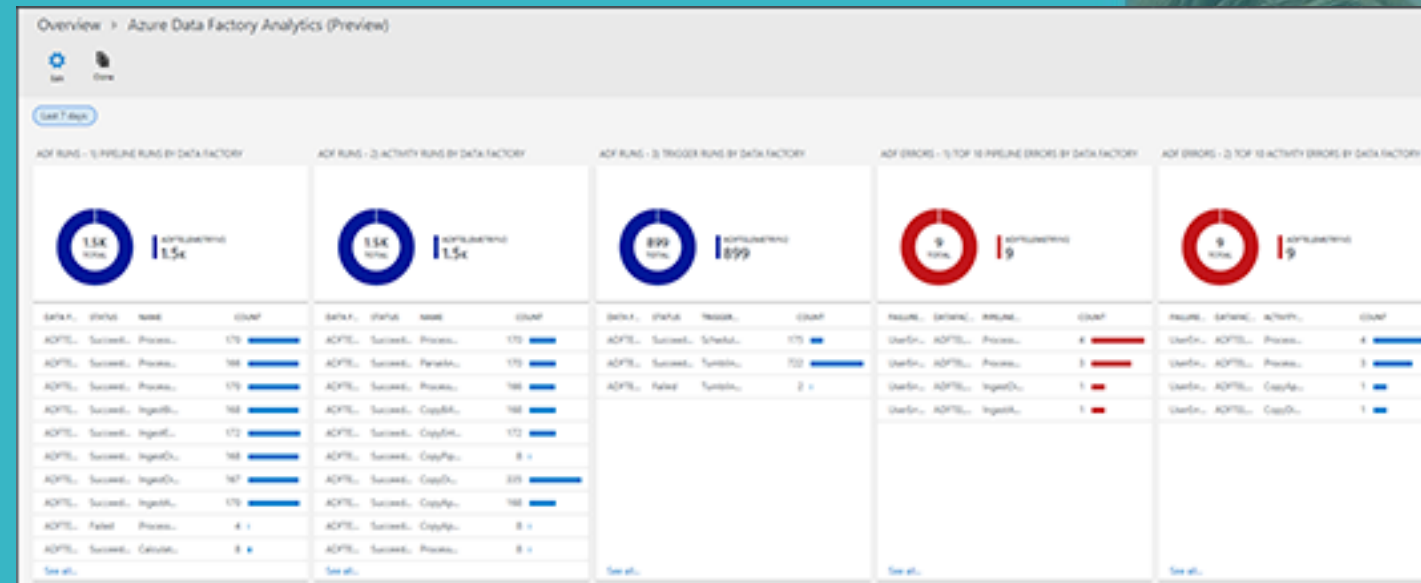


Azure
Academy

www.AzureAcademy.com.br

Data Factory Analytics

Serviço de extensão para análise visual dos logs gerados pelo Data Factory através de workspaces no Azure.



workadf1 | Pastas de trabalho | AzureDataFactoryAnalytics ✕

Workspace do Log Analytics

Pesquisar (Ctrl+/) << Pastas de trabalho Editar

Marcações

Diagnosticar e resolver proble...

Configurações

Bloqueios

Gerenciamento de agentes

Isolamento de Rede

Configurações avançadas

Geral

Resumo de workspace

Exibir Designer

Pastas de trabalho

Pesquisar

ResourceId	↑↓	Count	↑↓	Trend	Status	↑↓	ActivityName
DAHFFHG		1			Queued		LookupNewWaterMarkActivi
DAHFFHG		1			Succeeded		StoredProceduretoWriteWat
DAHFFHG		1			Succeeded		Web1
DAHFFHG		1			Succeeded		LookupOldWaterMarkActivit
DAHFFHG		1			Queued		LookupOldWaterMarkActivit
DAHFFHG		1			Succeeded		LookupNewWaterMarkActivi

Laboratório

Data Factory Analytics

Assuntos:

- Log Analytics
- Workspace
- Logs

Tecnologias:

- Log Analytics
- Azure Monitor
- Data Factory Analytics



Lab – Data Factory Analytics

1. Acesse o **Log Analytics** do Azure e crie um novo workspace.

Painel > Workspaces do Log Analytics >

Criar workspace do Log Analytics

Noções Básicas Tipo de preço Marcas Examinar + Criar

1 Um workspace do Log Analytics é a unidade básica de gerenciamento de logs do Azure Monitor. Há considerações específicas que você deve levar em conta ao criar um workspace do Log Analytics. [Saiba mais](#)

Com os Logs do Azure Monitor, você pode facilmente armazenar, reter e consultar dados coletados dos seus recursos monitorados no Azure e em outros ambientes para obter insights valiosos. Um workspace do Log Analytics é a unidade de armazenamento lógico em que seus dados de log são coletados e armazenados.

Detalhes do projeto

Selecione a assinatura para gerenciar os custos e os recursos implantados. Use grupos de recursos como pastas para organizar e gerenciar todos os seus recursos.

Assinatura * ⓘ

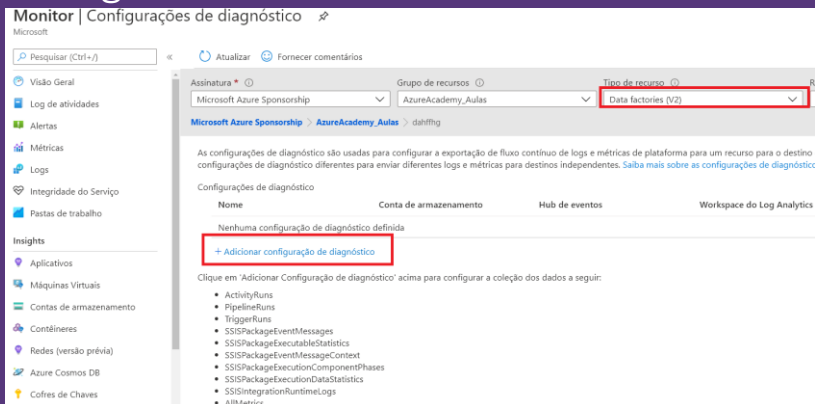
Grupo de recursos * ⓘ
[Criar novo](#)

Detalhes da instância

Nome * ⓘ

Região * ⓘ

2. Acesse o **Azure Monitor**, Configurações de diagnóstico. Localize o Data Factory e ative as configurações conforme a imagem:



Painel > Monitor >

Configuração de diagnóstico

Salvar Descartar Excluir Fornecer comentários

Uma configuração de diagnóstico específica uma lista de categorias de logs e/ou métricas de plataforma que você deseja coletar de um recurso e um ou mais destinos para os quais transmitir. Os encargos de uso normal para o destino serão cobrados. [Saiba mais sobre os diferentes conteúdos e categorias de log desses logs](#)

Nome da configuração de diagnóstico *

Detalhes da categoria

log
<input checked="" type="checkbox"/> ActivityRuns
<input checked="" type="checkbox"/> PipelineRuns
<input checked="" type="checkbox"/> TriggerRuns
<input checked="" type="checkbox"/> SSISPackageEventMessages
<input checked="" type="checkbox"/> SSISPackageExecutableStatistics
<input checked="" type="checkbox"/> SSISPackageEventMessageContext
<input checked="" type="checkbox"/> SSISPackageExecutionComponentPhases
<input checked="" type="checkbox"/> SSISPackageExecutionDataStatistics
<input checked="" type="checkbox"/> SSISIntegrationRuntimeLogs

Detalhes do destino

☒ Enviar para o Log Analytics

Assinatura

Workspace do Log Analytics

Tabela de destino ⓘ [Específico do recurso](#)

☐ Arquivar em uma conta de armazenamento

☐ Transmitir para um hub de eventos

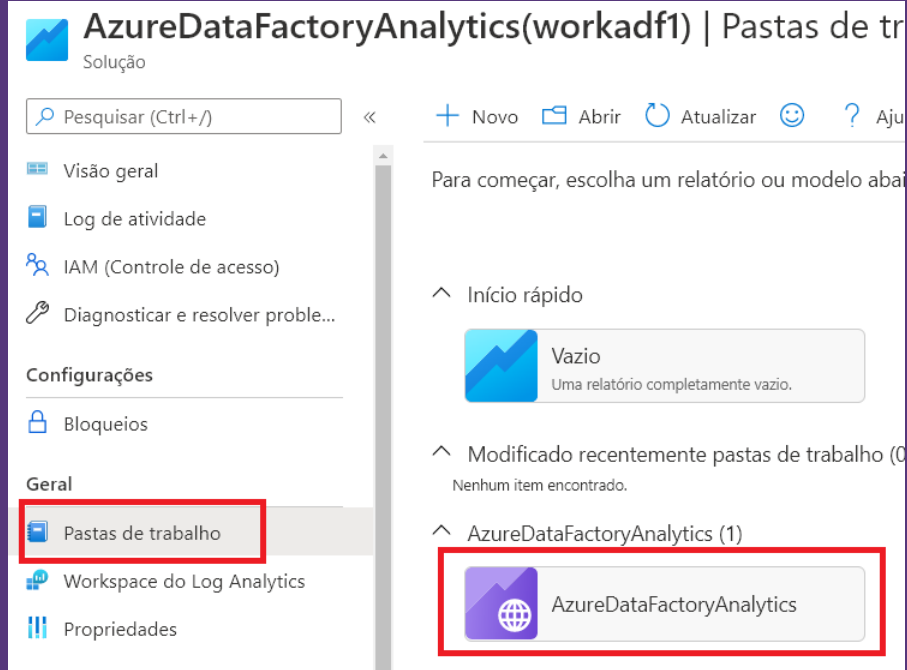
Azure
Academy



VÍDEO

Lab – Data Factory Analytics

1. Acesse o marketplace do Azure e instale o serviço **Data Factory Analytics** apontando para o mesmo workspace de coleta do ADF.
2. Após a instalação, acesse a solução e acompanhe os gráficos de monitoramento.



3. Para melhor acompanhamento, configure alertas de Logs do Data Factory:
<https://docs.microsoft.com/pt-br/azure/data-factory/monitor-using-azure-monitor>

Lab extra – consumir logs do Application Insights no Data Factory:

<https://www.ben-morris.com/using-azure-data-factory-with-the-application-insights-rest-api/>



Gatilhos de Pipelines

Os pipelines do Data Factory são disparados através de gatilhos.

Manual / sob demanda

API Rest: através de post. Exemplo:

<https://management.azure.com/subscriptions/mySubId/resourceGroups/myResourceGroup/providers/Microsoft.DataFactory/factories/myDataFactory/pipelines/copyPipeline/createRun?api-version=2017-03-01-preview>

PowerShell:

```
Invoke-AzDataFactoryV2Pipeline -DataFactory $df -PipelineName "Adfv2QuickStartPipeline" -ParameterFile .\PipelineParameters.json
```

SDK .NET

Agenda: horários pré-definidos ou recorrência.

Gatilho de eventos: executa pipelines em resposta a um evento, tal como a chegada ou a exclusão de um arquivo no Armazenamento de Blobs do Azure.



Lab – Gatilho por eventos

1. Acesse um pipeline do Data Factory e adicione uma nova trigger.
2. Configure o gatilho de evento, baseado no monitoramento da criação de arquivos em containers de Blob. Veja exemplo:

New trigger

Description

Type *

☐ Schedule ☐ Tumbling window ☒ Event

Account selection method *

☒ From Azure subscription ☐ Enter manually

Azure subscription

Select all

Storage account name *

teste1231234

Container name *

arquivos

Blob path begins with

Blob path ends with

Event *

☒ Blob created ☐ Blob deleted

Ignore empty blobs *



SSIS Integration Runtime

Permite executar pacotes SSIS.

Execução de Jobs.

Gerenciamento através do SQL Data Tools, SSMS e outros.

Criar DR de bancos locais e sincronizar dados.

Siga os passos para criar as conexões:

<https://docs.microsoft.com/pt-br/azure/data-factory/tutorial-deploy-ssis-packages-azure>

Utilizar SSIS com Data Factory:

<https://docs.microsoft.com/pt-br/azure/data-factory/how-to-invoke-ssis-package-ssdt?view=sql-server-ver15>

Saiba mais sobre arquitetura SSIS:

<https://docs.microsoft.com/pt-br/sql/integration-services/packages/deploy-integration-services-ssis-projects-and-packages?view=sql-server-ver15>

NAME	TYPE	SUB-TYPE	STATUS	REGION
AutoResolveIntegrationRuntime	Azure	Public	Running	Auto Resolve
myAzureSSISIntegrationRuntime	Azure-SSIS	---	Stopped	East US



Integração do Office 365

Permite a criação de Datasets para extração de dados do Microsoft 365.

Office 365
Office365Table1

General **Connection** Parameters

Linked service * LS_Office365 Test connection Edit + New

Table None Refresh

Filter...

- BasicDataSet_v0.MailboxSettings_v0
- BasicDataSet_v0.Manager_v0
- BasicDataSet_v0.Message_v0**
- BasicDataSet_v0.Message_v1
- BasicDataSet_v0.SentItem_v0
- BasicDataSet_v0.SentItem_v1
- BasicDataSet_v0.User_v0
- BasicDataSet_v0.User_v1

Lab – Office 365

1. No Azure Data Factory, crie um novo Pipeline e insira uma atividade 'Copy'.
2. Siga os passos descritos no link a seguir para criar a conexão e o dataset do Office 365 para criar uma atividade de extração de dados de emails apontando para uma storage account:

<https://docs.microsoft.com/pt-br/azure/data-factory/load-office-365-data>

Save as template Validate Debug Add trigger

Copy data

CopyFromOffice365To Blob

General Source Sink¹ Mapping Settings User properties

Source dataset * Office365Table1 Edit + New

User scope Select groups from the office 365 ten...

+ Add user groups

GROUP NAME	GROUP ID	REMOVE
test3099	00168f80-df7d-4aa1-88d6-39bf9042cedc	X

Date filter

Column name * Start time (UTC) * End time (UTC) *

ReceivedDateTime 09/30/2019 00:00:00

Output columns * Import schema + New column Clear Delete Total number of columns: 26 Filter columns

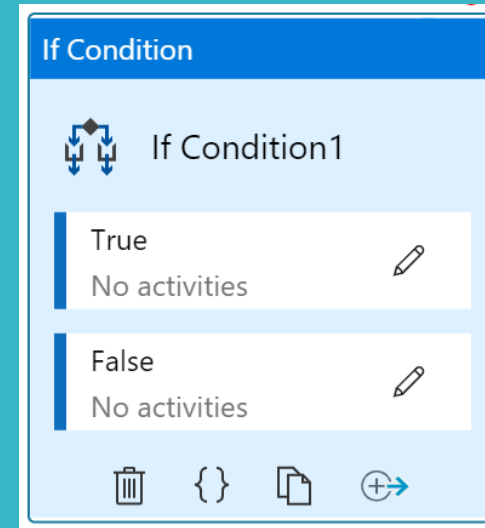
COLUMN NAME	TYPE	DESCRIPTION
ReceivedDateTim	DateTime	The date and time the message was received.
SentDateTime	DateTime	The date and time the message was sent.
HasAttachments	Boolean	Indicates whether the message has attachments.
InternetMessageId	String	The internet message id
Subject	String	The subject of the message.

<https://docs.microsoft.com/pt-br/onedrive/find-your-office-365-tenant-id>

Conceitos

Atividade: If

- Condicional para executar sequencia de atividades.
- Semelhante a estrutura do ForEach.



Lab sugerido:

Verifica alterações nos logs de tabelas do SQL Server do Azure - Instancia gerenciada e dispara triggers aninhadas para atividades de extração de dados somente das alterações.

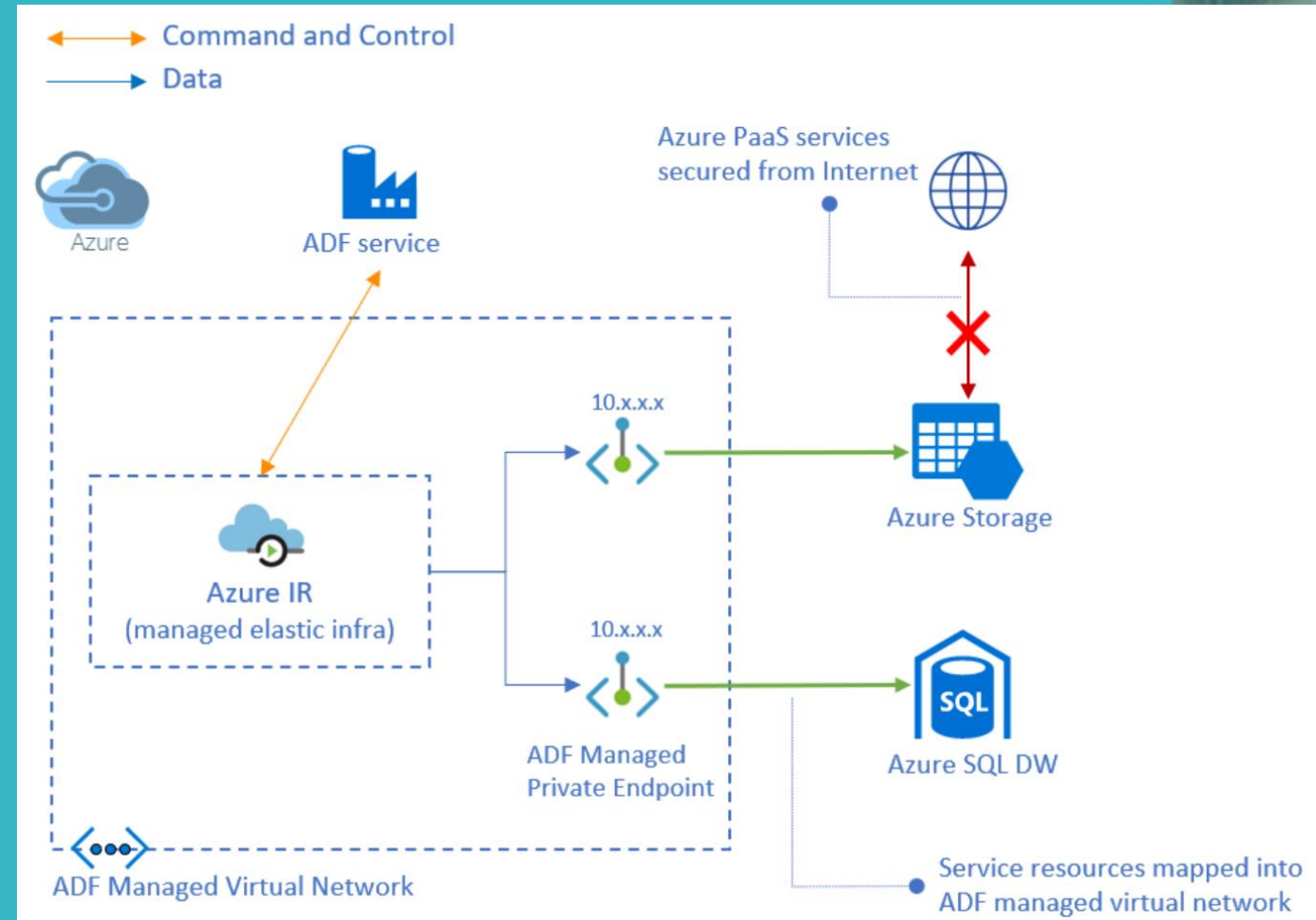
<https://docs.microsoft.com/pt-br/azure/data-factory/tutorial-incremental-copy-change-data-capture-feature-portal>

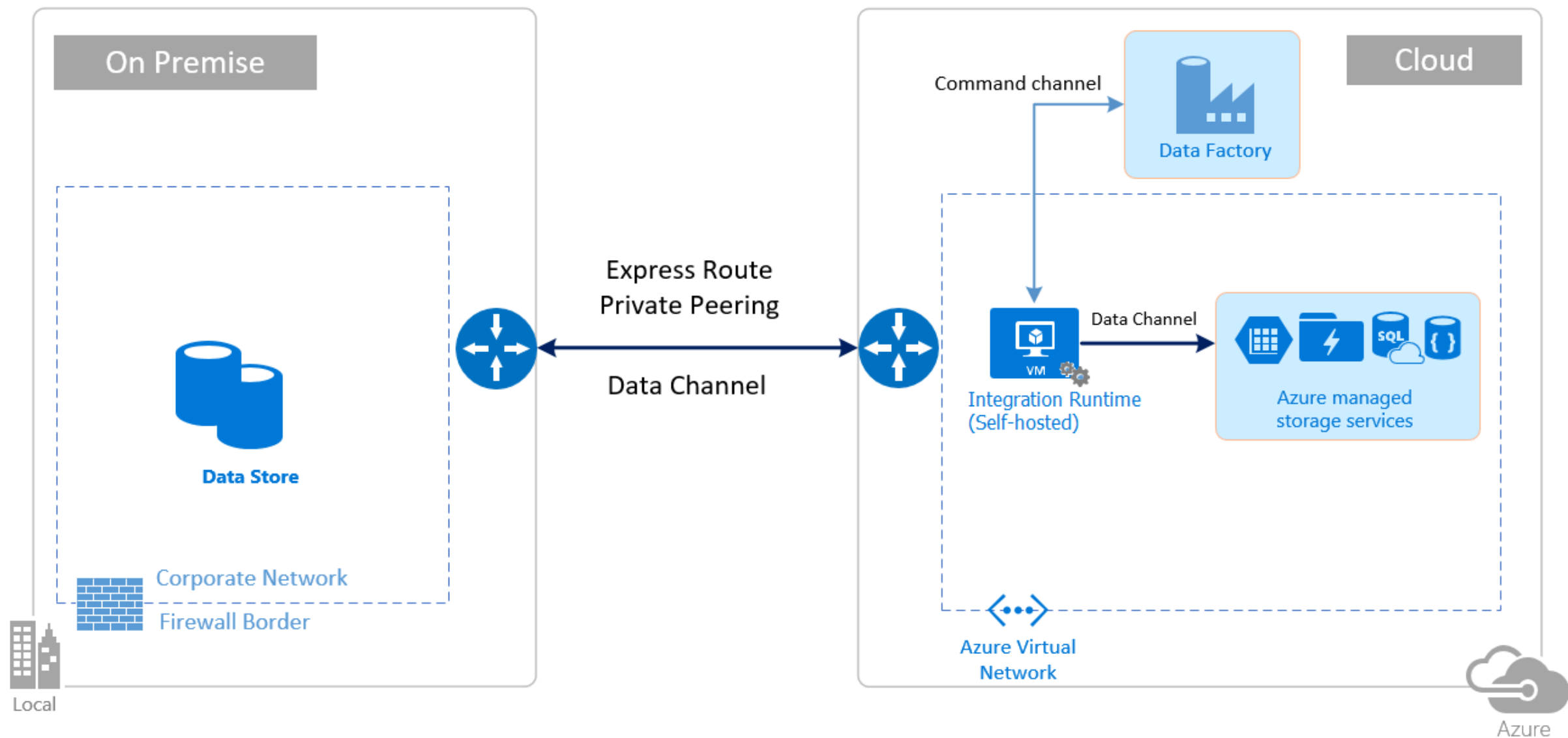


Rede Gerenciada

Quando você cria uma Azure Integration Runtime (IR) no Data Factory rede virtual gerenciada (VNET), o Integration Runtime será provisionado com a rede virtual gerenciada e aproveitará os pontos de extremidade privados para se **conectar com segurança aos armazenamentos de dados** com suporte.

Acesso aos serviços do Azure de forma segura e controlada com autorização de comunicação entre Data Factory e os recursos.





Lab – Rede Gerenciada

1. No Azure Data Factory, acesse a guia Manage e crie um novo Integration Runtime com a configuração de rede gerenciada ativada. Veja:

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 * ⓘ

integrationRuntime1Rede

Description

Enter description here...

Status

✓ Running

Virtual network configuration (Preview) ⓘ

☐ Disable ☒ Enable

Interactive authoring ⓘ

☐ Disable ☒ Enable

☒ Auto termination after 60 minutes of inactivity ⓘ

Region *

East US 2 ▼

▸ Data flow run time

2. Crie um novo Pipeline e insira uma atividade **Copy**.



Lab – Rede Gerenciada

1. Configure o Source com uma nova Connection e Dataset:

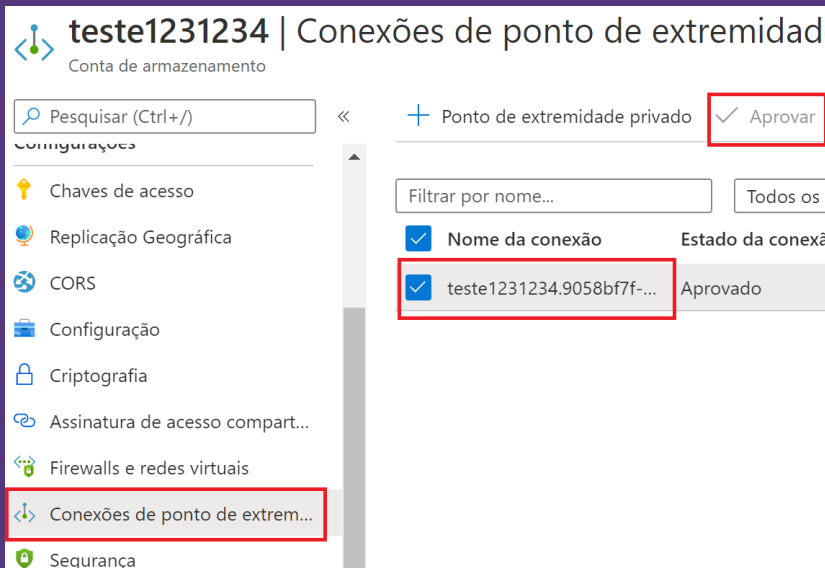
Dataset:

- Tipo: Data Lake Storage Gen 2.
- Formato: JSON.

Connection:

- Selecione o IR que contém a rede gerenciada.
- Selecione a storage account.

Para aprovar a conexão, abra outra guia do Navegador, acesse a storage account e Autorize o acesso, conforme imagem:



New linked service (Azure Data Lake Storage Gen2)

Name *

AzureDataLakeStorage3

Description

Connect via integration runtime *

integrationRuntime1Rede (Managed Virtual Network)

✓ Interactive authoring enabled ⓘ

Authentication method

Account key

Account selection method

☒ From Azure subscription ☐ Enter manually

Azure subscription

Select all

Storage account name *

teste1231234

Managed private endpoint

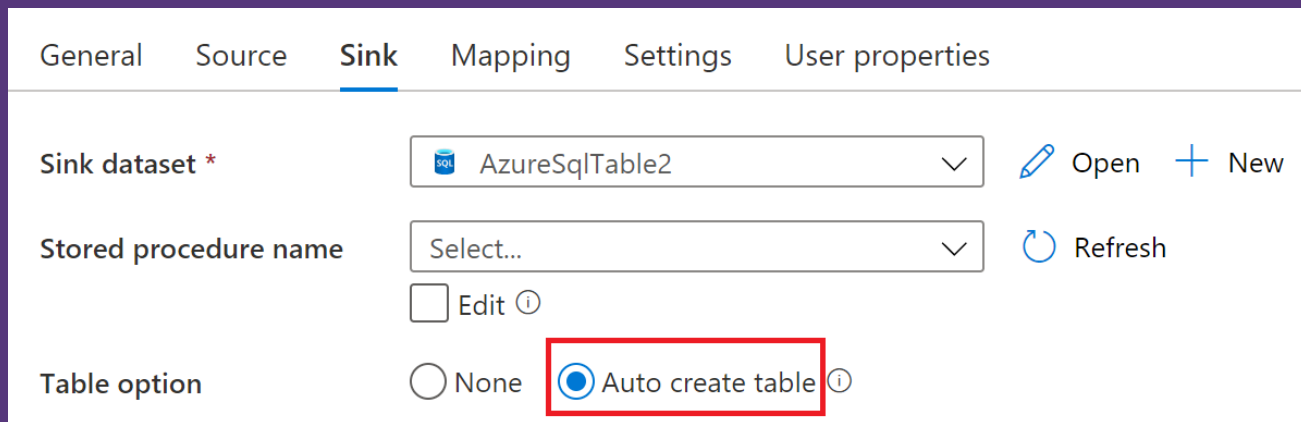
AzureDataLakeStorage1

✓ Approved ↻

Test connection

Lab – Rede Gerenciada

1. Selecione um arquivo JSON no dataset.
2. Para o Sink, repita os procedimentos de configuração, porém selecionando o destino como um SQL do Azure.
3. Acione o Auto create table.
4. Valide e execute o pipeline.



The screenshot shows the 'Sink' configuration tab in Azure Data Factory. The 'Sink dataset' is set to 'AzureSqlTable2'. The 'Stored procedure name' is set to 'Select...'. The 'Table option' is set to 'Auto create table', which is highlighted with a red box. The 'Auto create table' option is selected, and the 'Edit' button is visible next to it.

Descubra mais sobre Pontos de extremidade gerenciados no Azure:

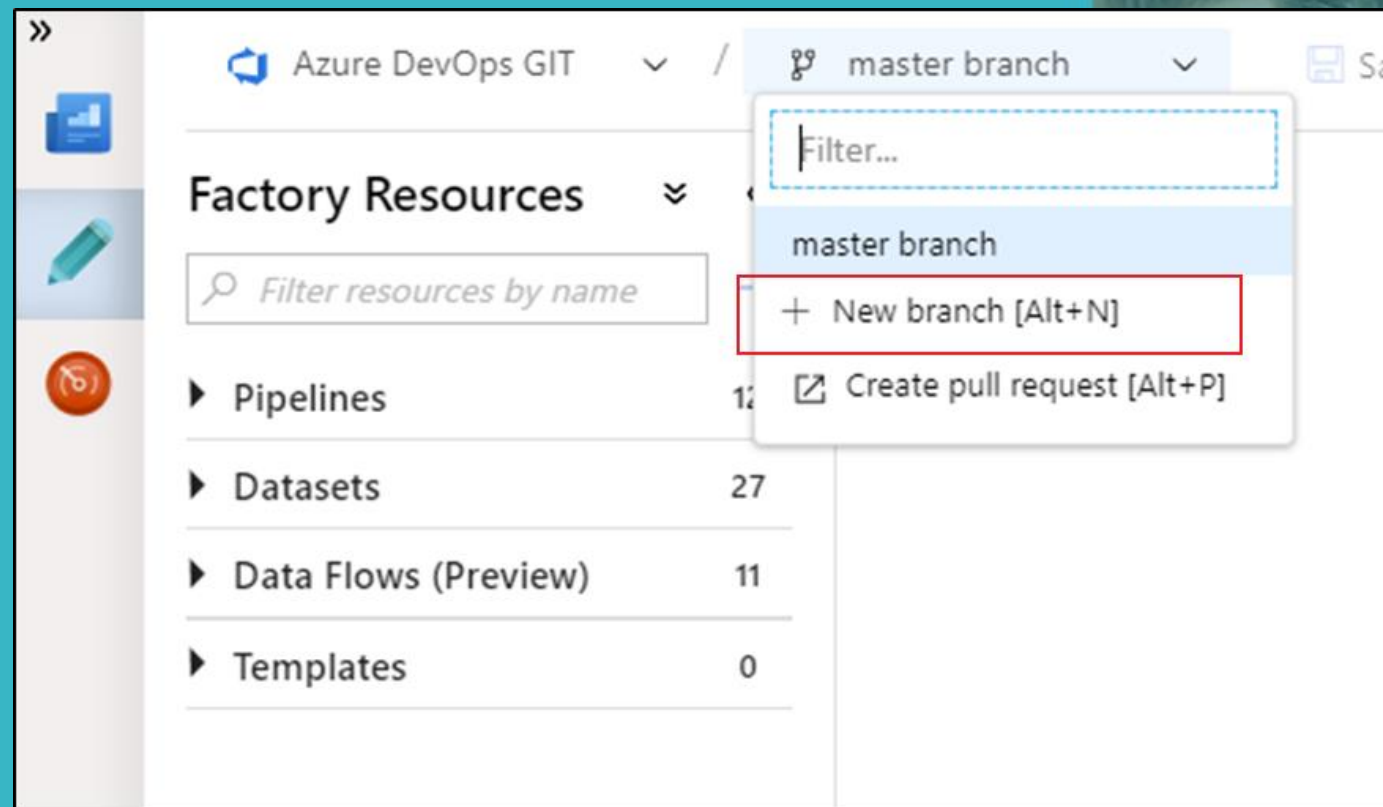
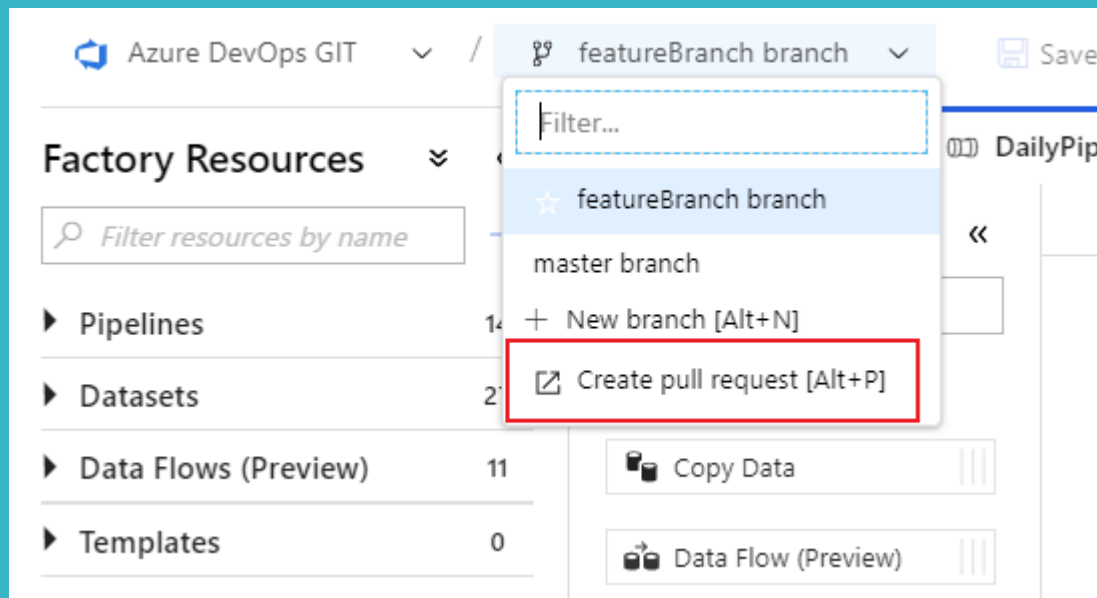
<https://docs.microsoft.com/pt-br/azure/data-factory/managed-virtual-network-private-endpoint>



Versões - GIT

Utilize o Azure Repos ou o GitHub para armazenar e controlar a versão de projetos e esteiras do Data Factory.

<https://docs.microsoft.com/pt-br/azure/data-factory/source-control>



CI/CD completo:

<https://docs.microsoft.com/pt-br/azure/data-factory/continuous-integration-deployment>



Azure Academy

Rubens Guimarães

 [/rubensguimaraes](#)



Microsoft
Regional Director

TRANSFORMANDO PROFISSIONAIS EM
ESPECIALISTAS EM CLOUD

www.AzureAcademy.com.br

PATROCÍNIO E APOIO:



e.Seth Cloud