

AZURE DATA FACTORY

Introdução

Cópia de dados

Conversão de formatos

Laboratório Prático

TRANSFORMANDO PROFISSIONAIS EM
ESPECIALISTAS EM CLOUD

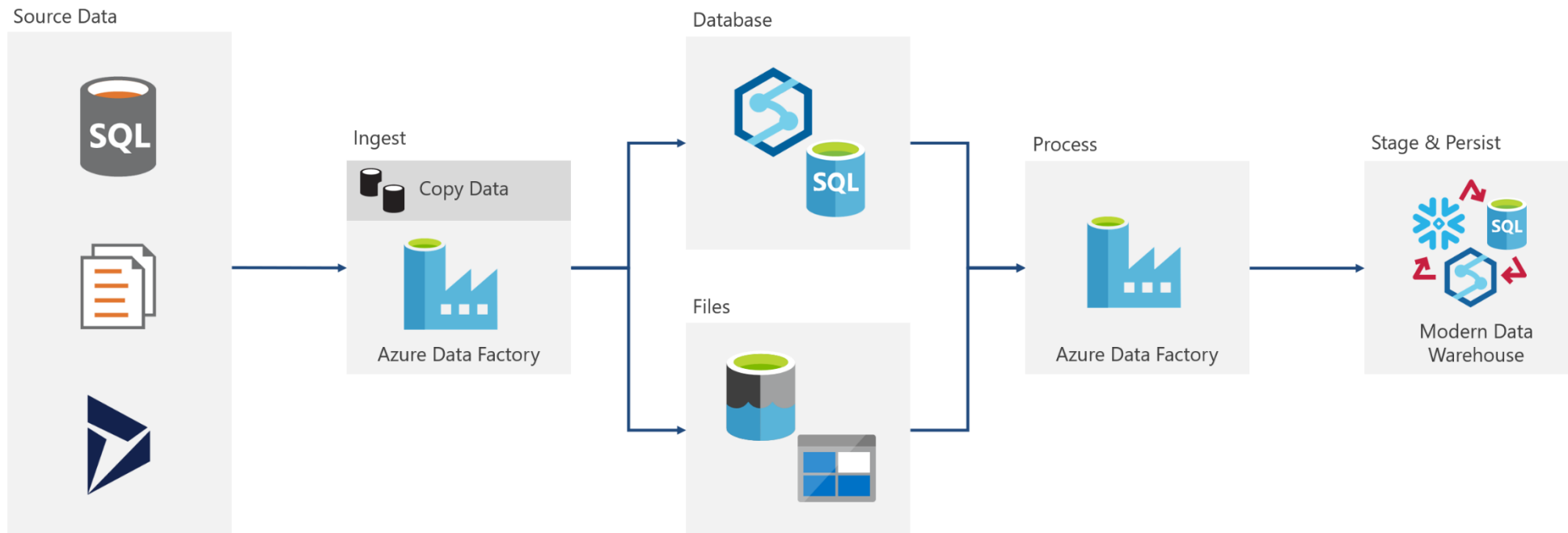
Azure
Academy

www.AzureAcademy.com.br

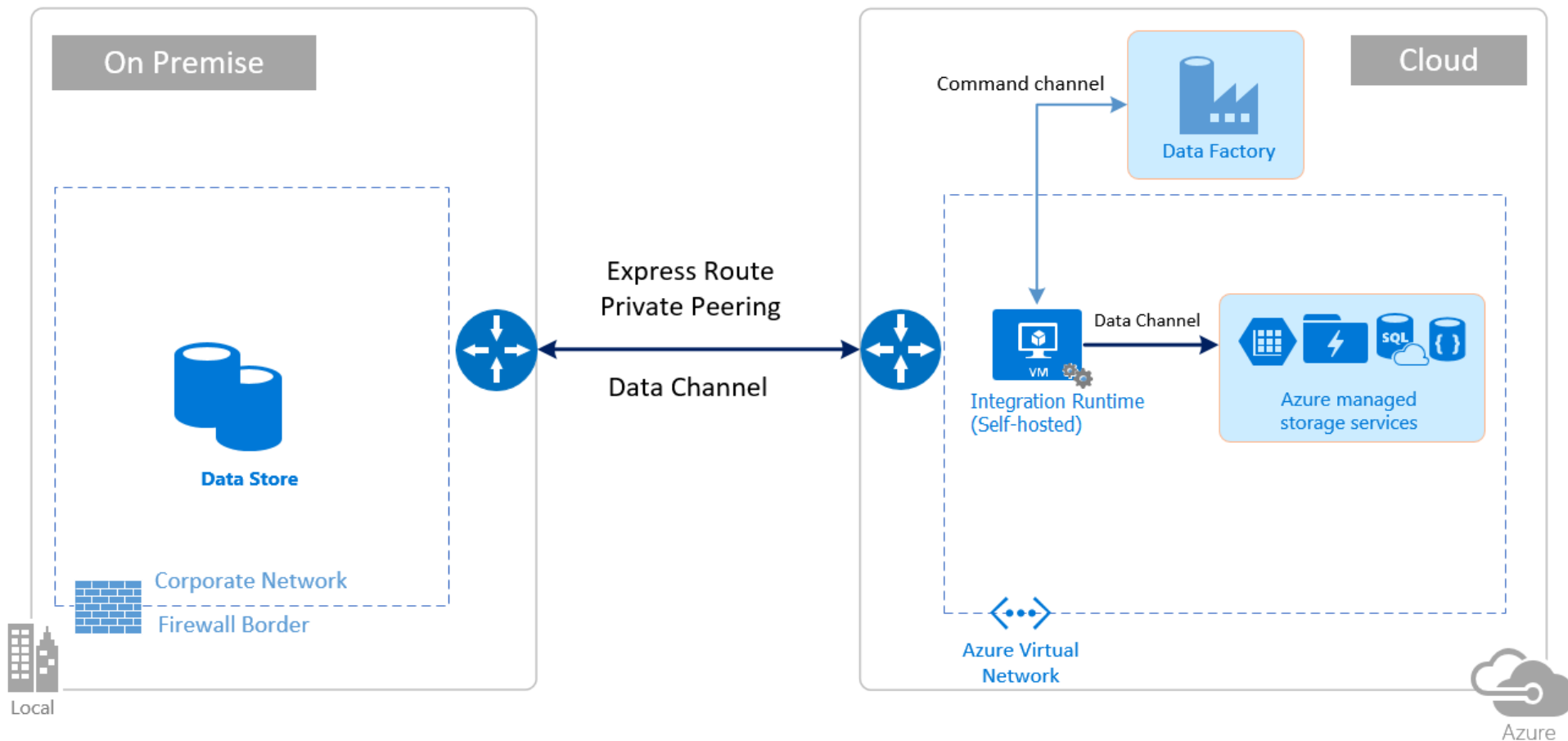


AZURE DATA FACTORY

Orquestrador de dados com capacidade para BIG Data. Ferramenta para engenheiros e cientistas de dados.



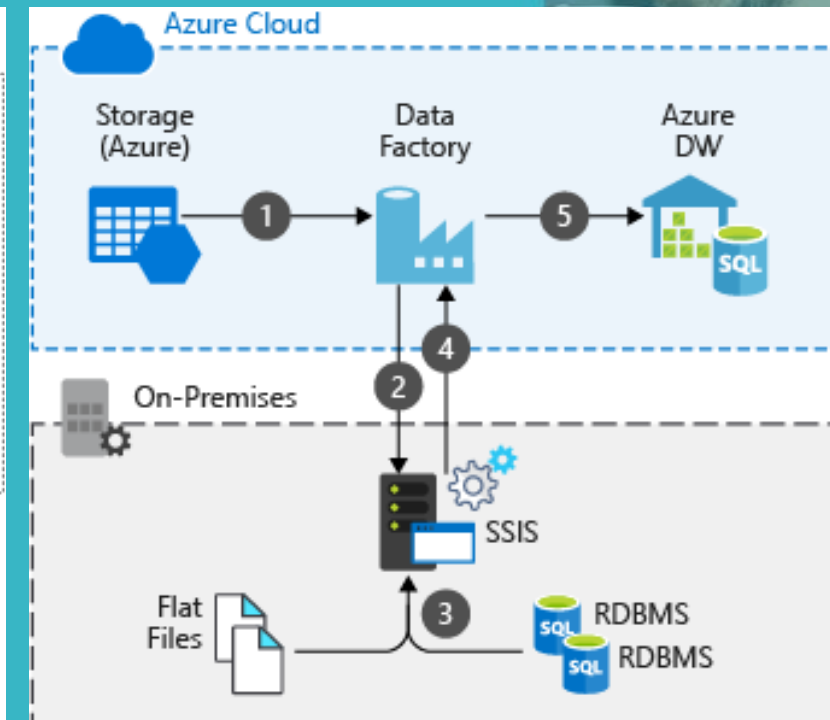
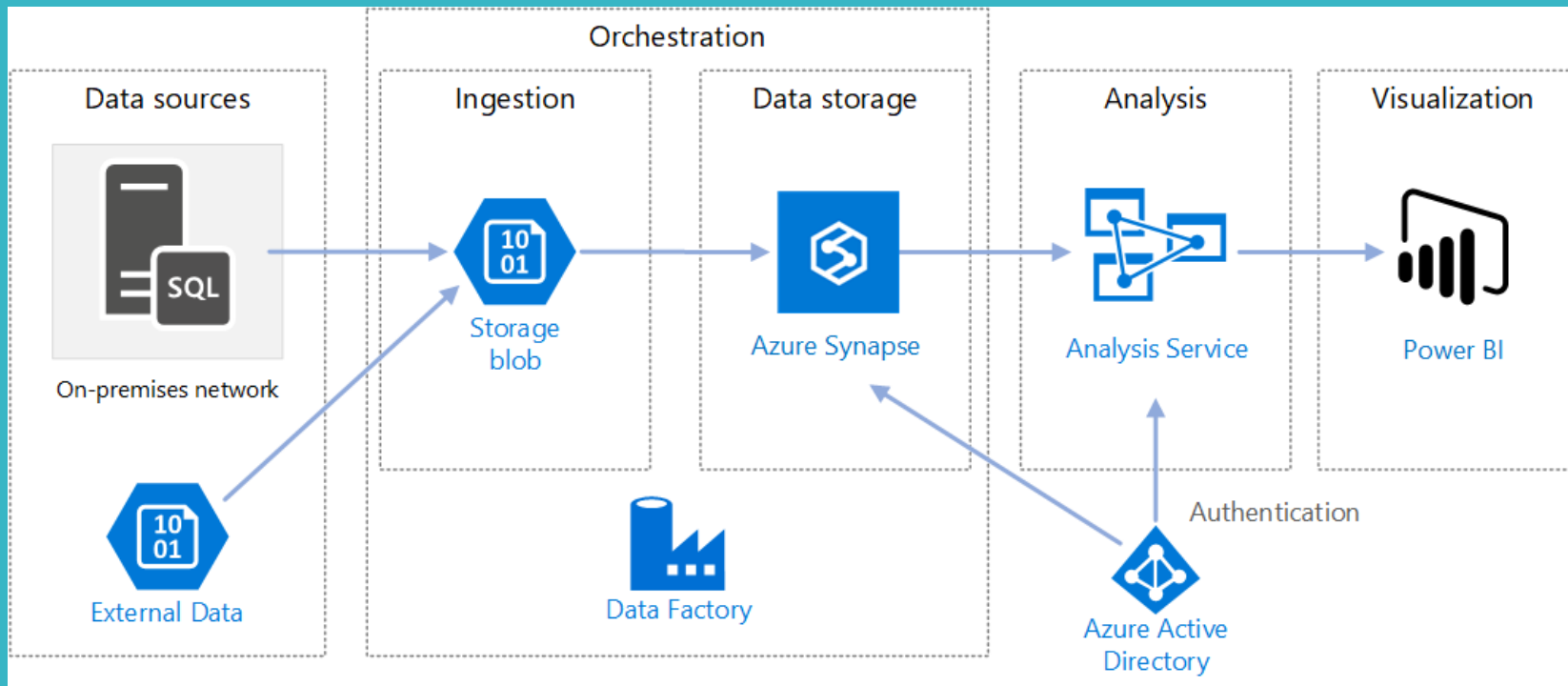
ARQUITETURA HÍBRIDA



[Documentação do Azure Data Factory - Azure Data Factory | Microsoft Docs](#)

AZURE DATA FACTORY

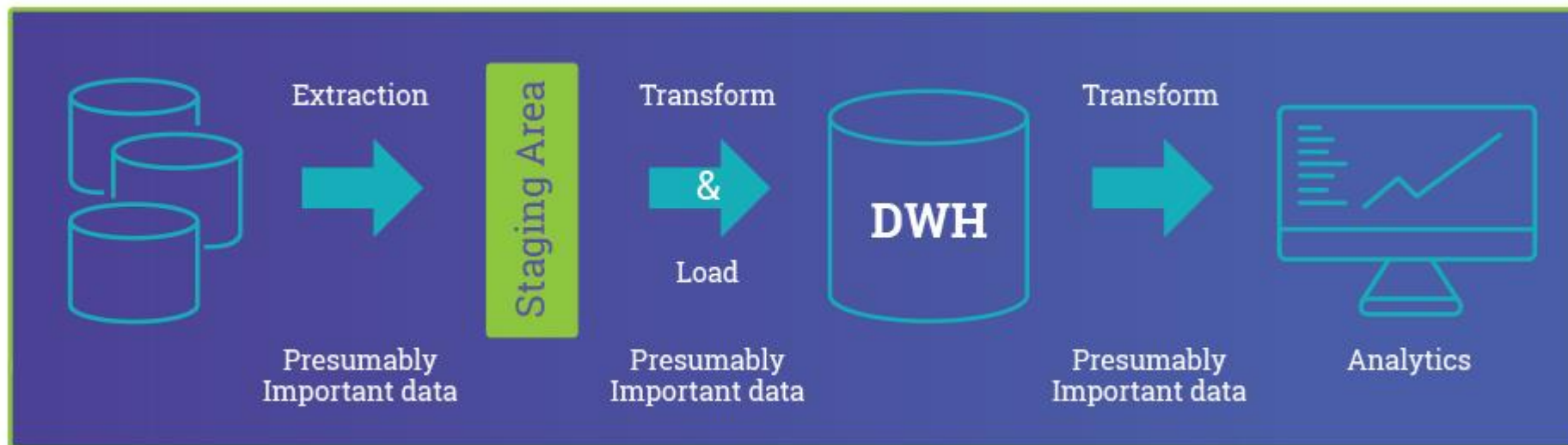
- ETL livre de código em nuvem como serviço.
- Trabalha com Gatilhos (batch, API, etc)
- Suporte multi formatos.



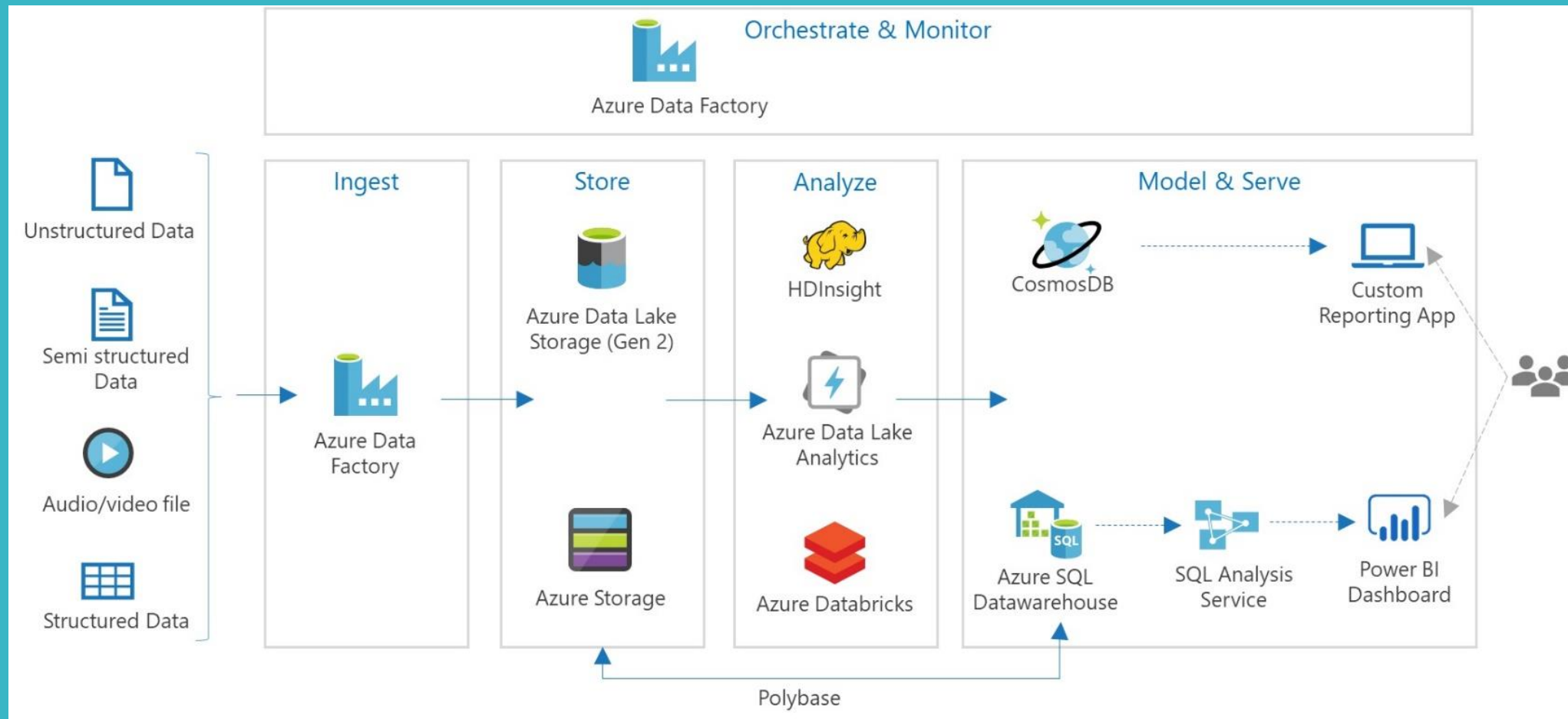
ETL

ATUA NAS CAMADAS BIG DATA PARA CENTRALIZAR E NORMALIZAR OS DADOS.

ETL



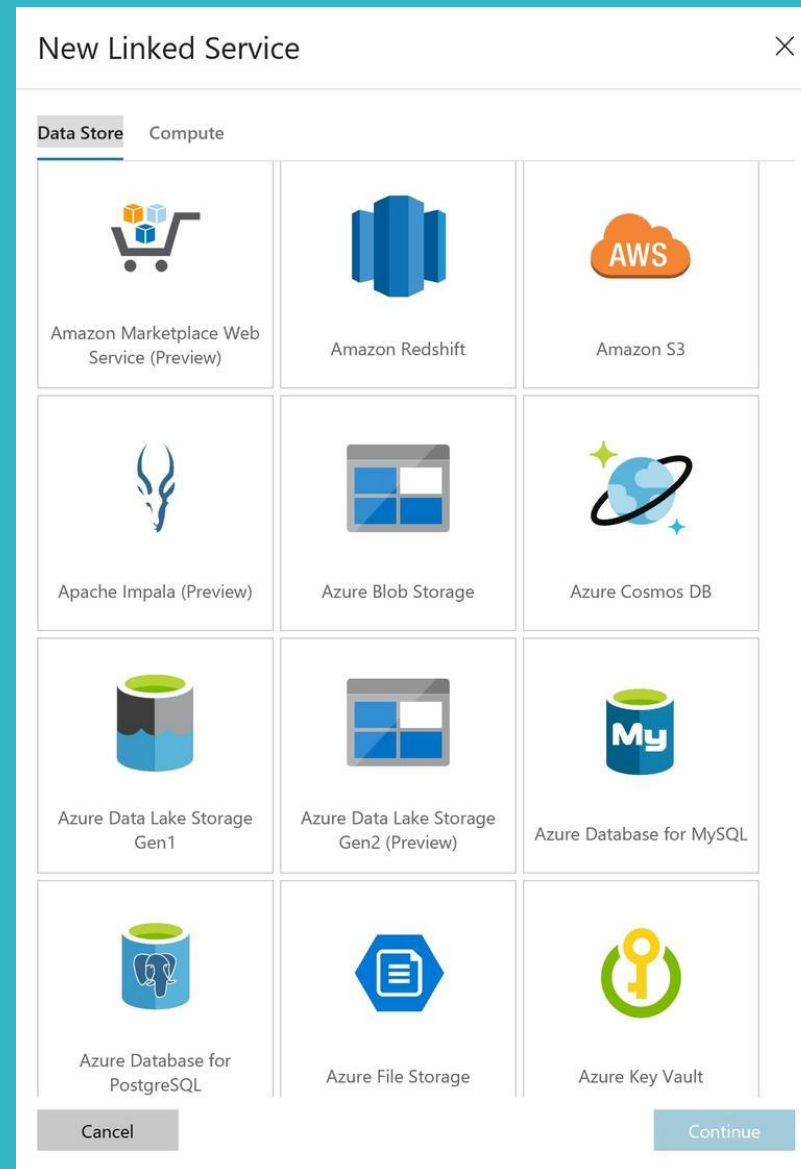
COMBINAÇÃO DE USO



Conectores

- Chamadas para Procedures;
- Chamadas para Runbooks;
- Chamadas com triggers externas;
- Integração de Azure Functions no pipeline;
- Integration Runtime para on-premises;
- Conversão de formatos.

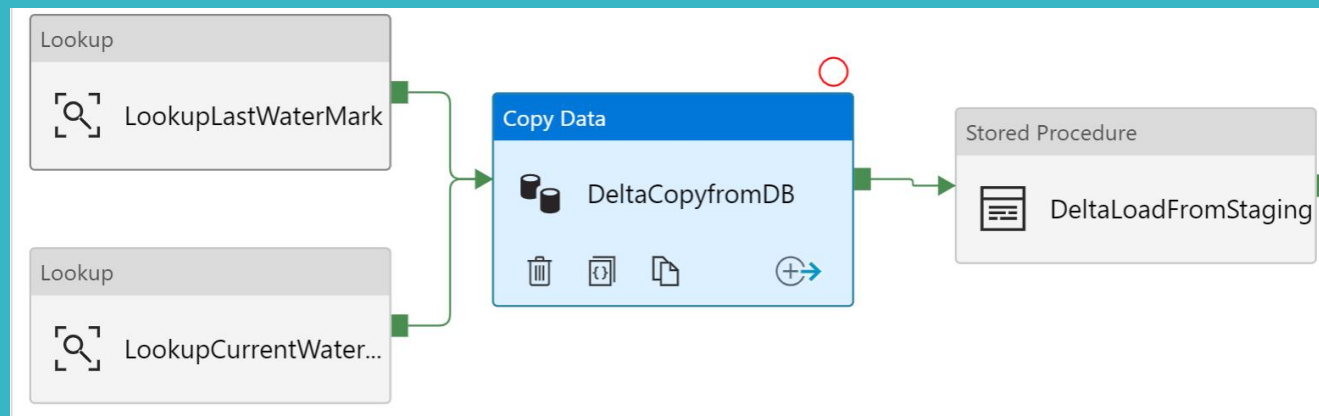
[Visão geral do conector - Azure Data Factory & Azure Synapse | Microsoft Docs](#)



Conceitos

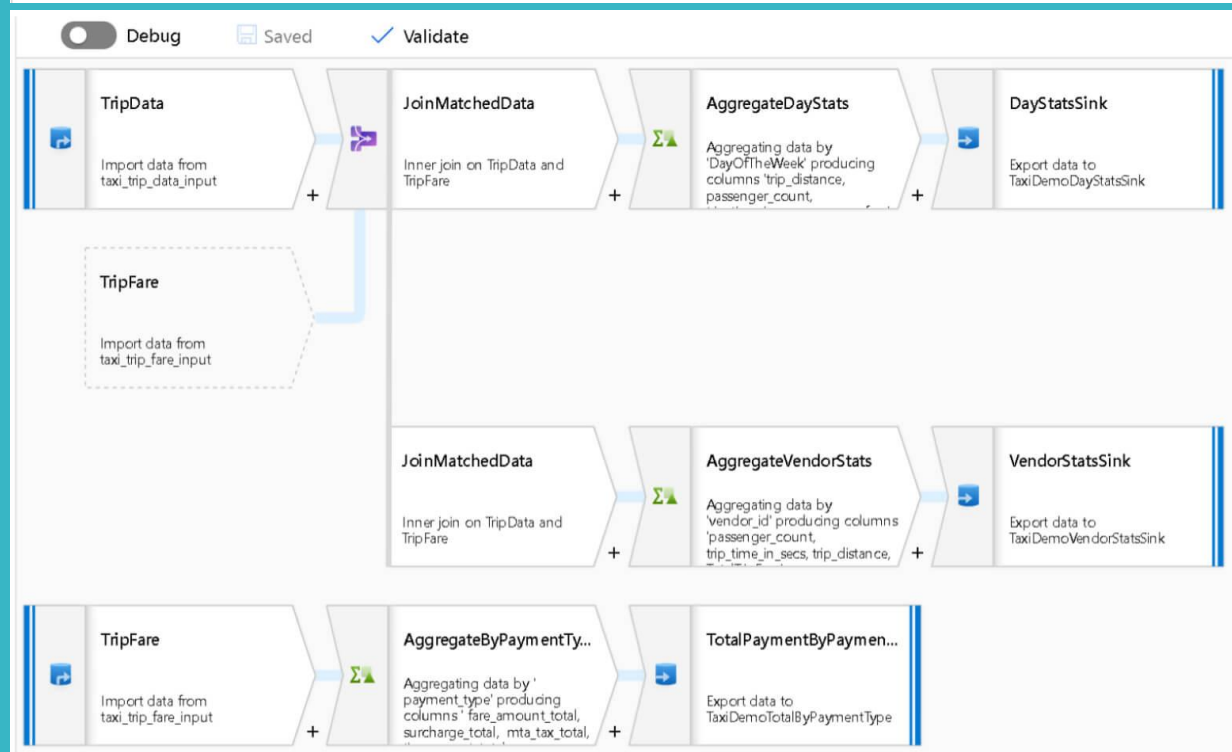
Pipelines

Esteiras com ações agrupadas de cópia, processamento, gravação e outros procedimentos.



Dataflows

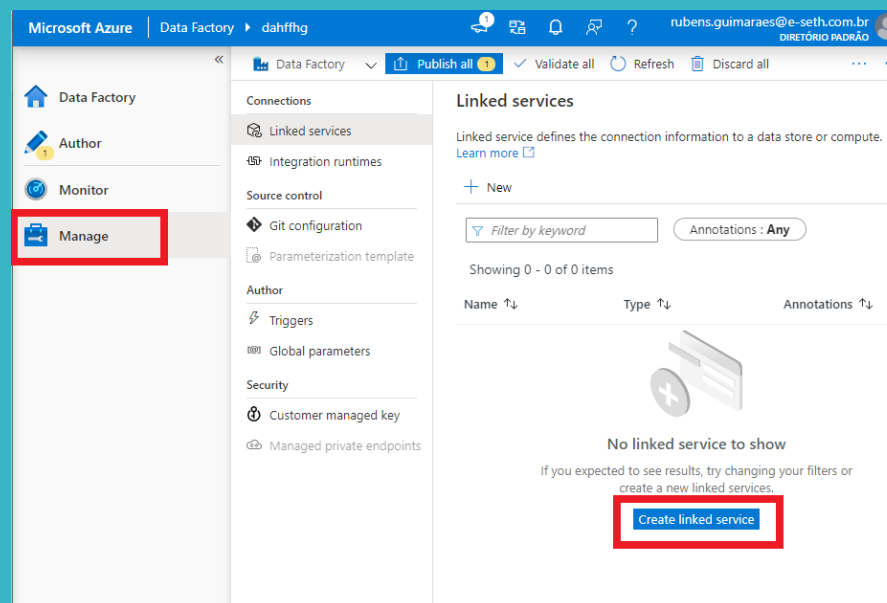
Lógica de transformação de dados gráficos sem escrever códigos.



Conceitos

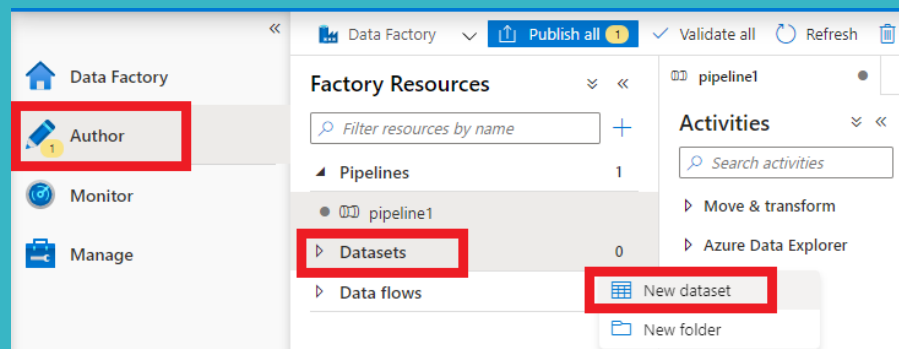
Linked services (Conexões):

Estabelece a **conexão com os drivers** das fontes/destinos de dados externos tais como Discos, SQL Server, Cosmos DB, etc. Estas conexões serão consumidas na esteira do Pipeline.



Datasets (Conjunto de dados) - exemplos:

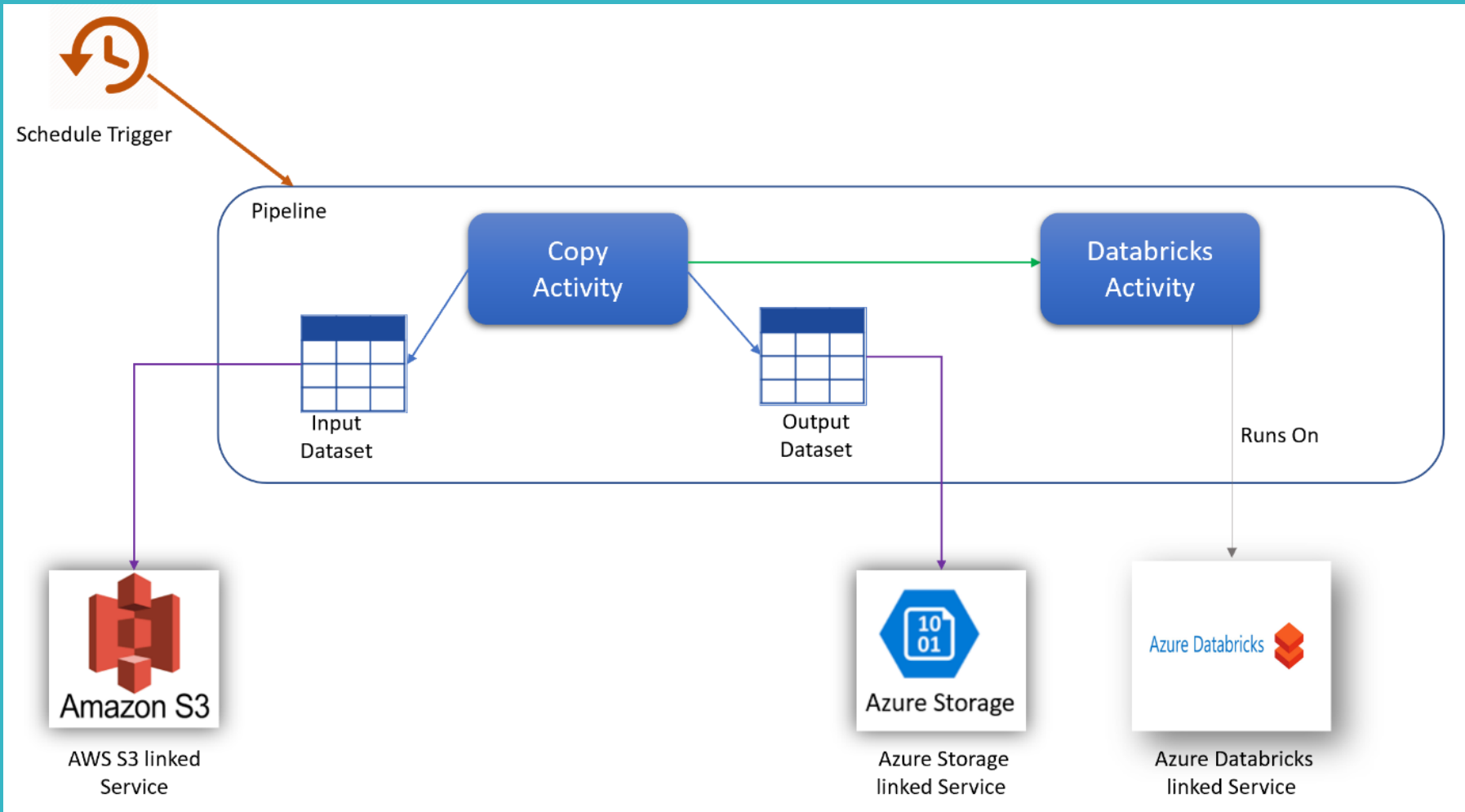
Select de registros da tabela da Connection – SQL.
Selecionar um container dentro da storage account (Connection).



O Dataset depende da Connection. Enquanto a connection se conecta ao driver, o dataset navega nos registros/discos selecionando ou setando os dados como fonte ou destino.

Datasets

Datasets de entrada e saída de dados (Fonte e Destino) para serem consumidos no Pipeline.



Laboratório

Extração e Conversão de Formatos

Assuntos:

- Linked services (Connections)
- Datasets
- Mapping

Tecnologias:

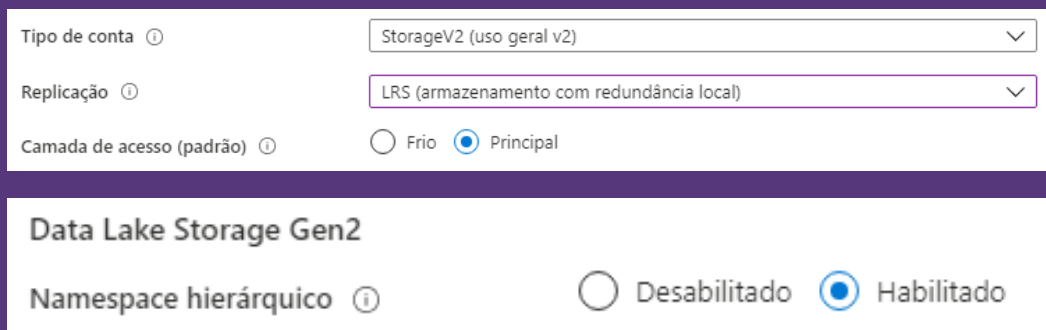
- SQL do Azure
- Storage Account - Data Lake Storage



Lab – Extração e Conversão de Formatos

No Azure:

1. Instale um banco de dados **SQL do Azure** selecionando uma instância de **amostra**.
2. Acesse o **Firewall do SQL** instalado e libere a opção para permitir conexões de serviços do Azure.
3. Instale um **Storage Account** de acordo com as configurações:
 - Storage V2
 - Replicação: LRS
 - Guia Avançado: Data Lake Storage Gen2 – **Habilitado**



The screenshot shows the configuration options for a new Azure Storage Account. It includes three dropdown menus: 'Tipo de conta' set to 'StorageV2 (uso geral v2)', 'Replicação' set to 'LRS (armazenamento com redundância local)', and 'Camada de acesso (padrão)' with radio buttons for 'Frio' and 'Principal' (selected). Below these is a section for 'Data Lake Storage Gen2' with a radio button for 'Namespace hierárquico' set to 'Habilitado'.

Tipo de conta ⓘ	StorageV2 (uso geral v2) ▼
Replicação ⓘ	LRS (armazenamento com redundância local) ▼
Camada de acesso (padrão) ⓘ	<input type="radio"/> Frio <input checked="" type="radio"/> Principal
Data Lake Storage Gen2	
Namespace hierárquico ⓘ	<input type="radio"/> Desabilitado <input checked="" type="radio"/> Habilitado

4. Após a instalação, acesse a Storage Account e crie um novo **Container** (Disco de blobs).
5. Instale o **Data Factory** e abra o serviço (Author & Monitor).



VÍDEO

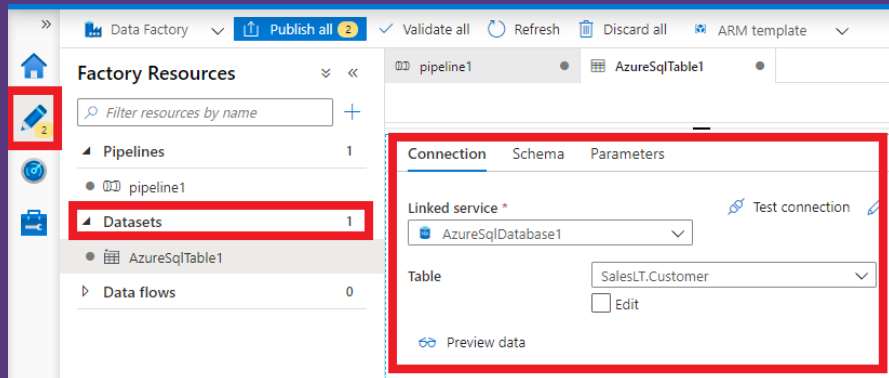
Lab – Extração e Conversão de Formatos

Conexões no Data Factory:

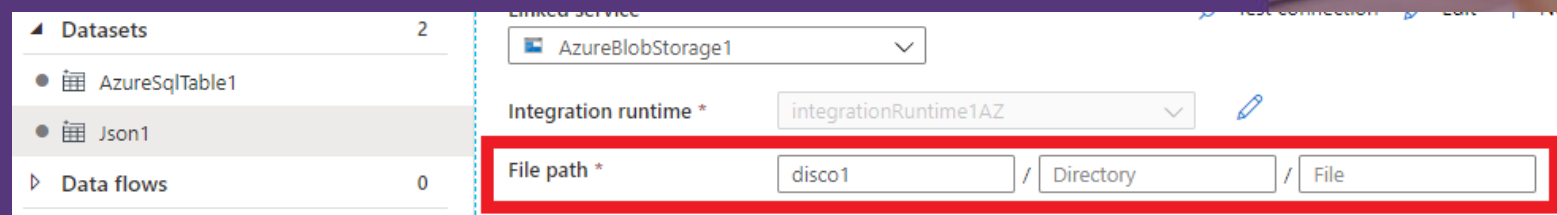
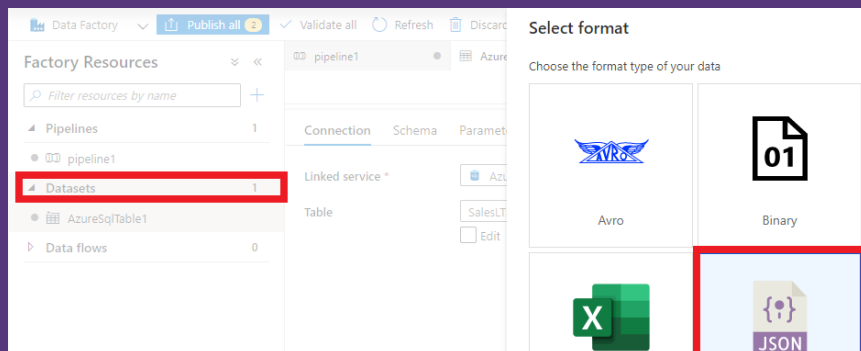
1. Acesse a guia Manage, Linked Services.
2. Configure uma nova Connection para o SQL do Azure.
3. Crie outra Connection, desta vez para a Storage Account (Azure Blob Storage).

Datasets:

1. Acesse a guia Author, Datasets.
2. Configure um Dataset para uma tabela do SQL do Azure.



3. Crie outro dataset apontando para Azure Blob Storage, formato JSON.



Lab – Extração e Conversão de Formatos

Agora que as conexões e datasets estão prontos, vamos criar o pipeline de importação de dados.

1. Crie um novo pipeline.
2. Selecione o objeto **Move & Transform > Copy Data**.
3. Configure o **Source** com o Dataset do SQL do Azure.
4. Aproveite para personalizar a query do select.
5. Configure o **Sink** apontando para o dataset da Storage Account.
6. Ative o modo **Debug** para acompanhar o processamento.
7. Execute o Pipeline e verifique se o Json foi gerado na Storage Account.

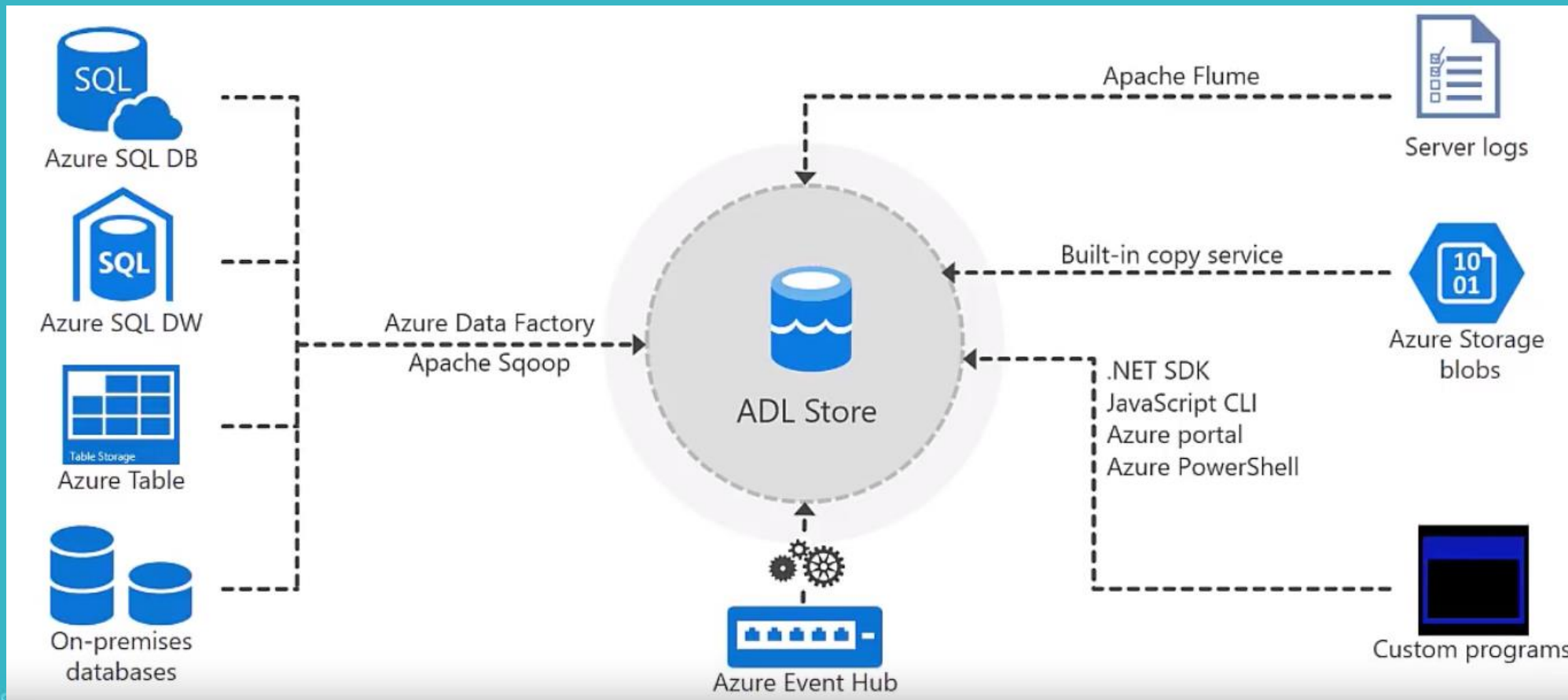
Consulte o link a seguir para compreender como criar cópias em formato Parquet:

<https://docs.microsoft.com/pt-br/azure/data-factory/format-parquet>



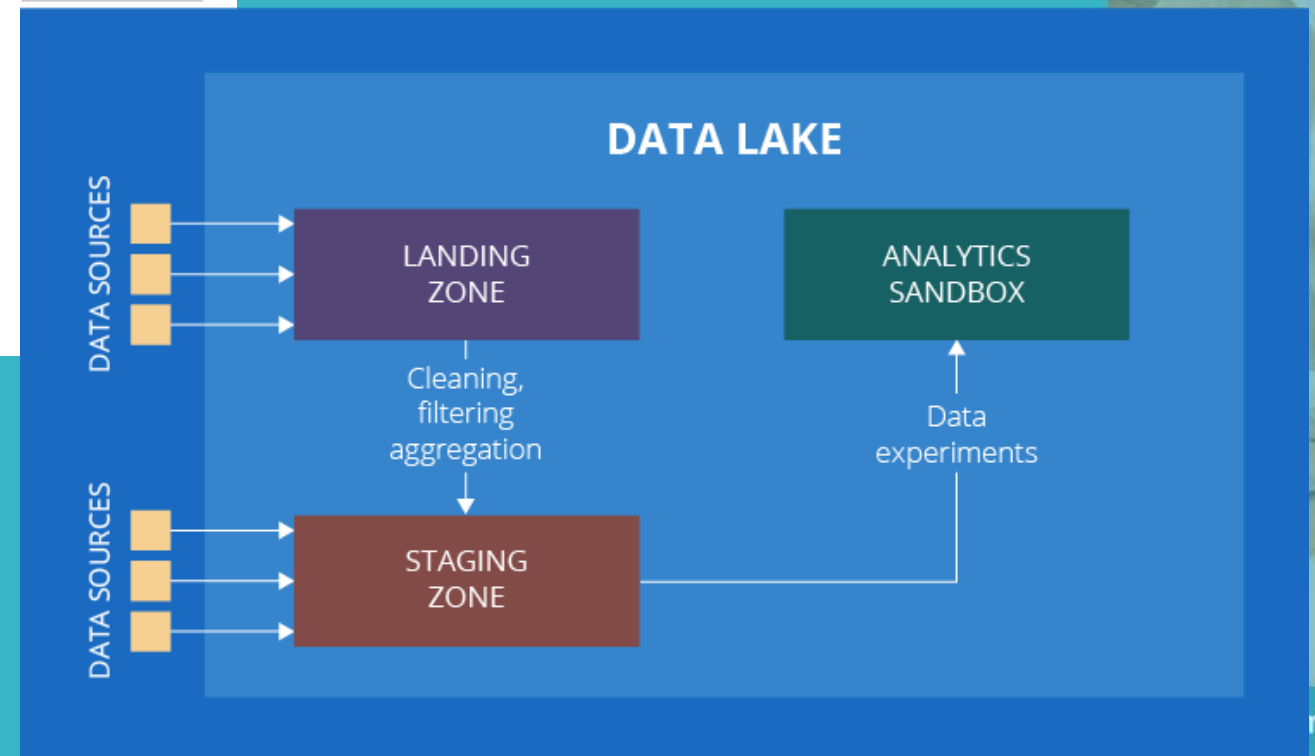
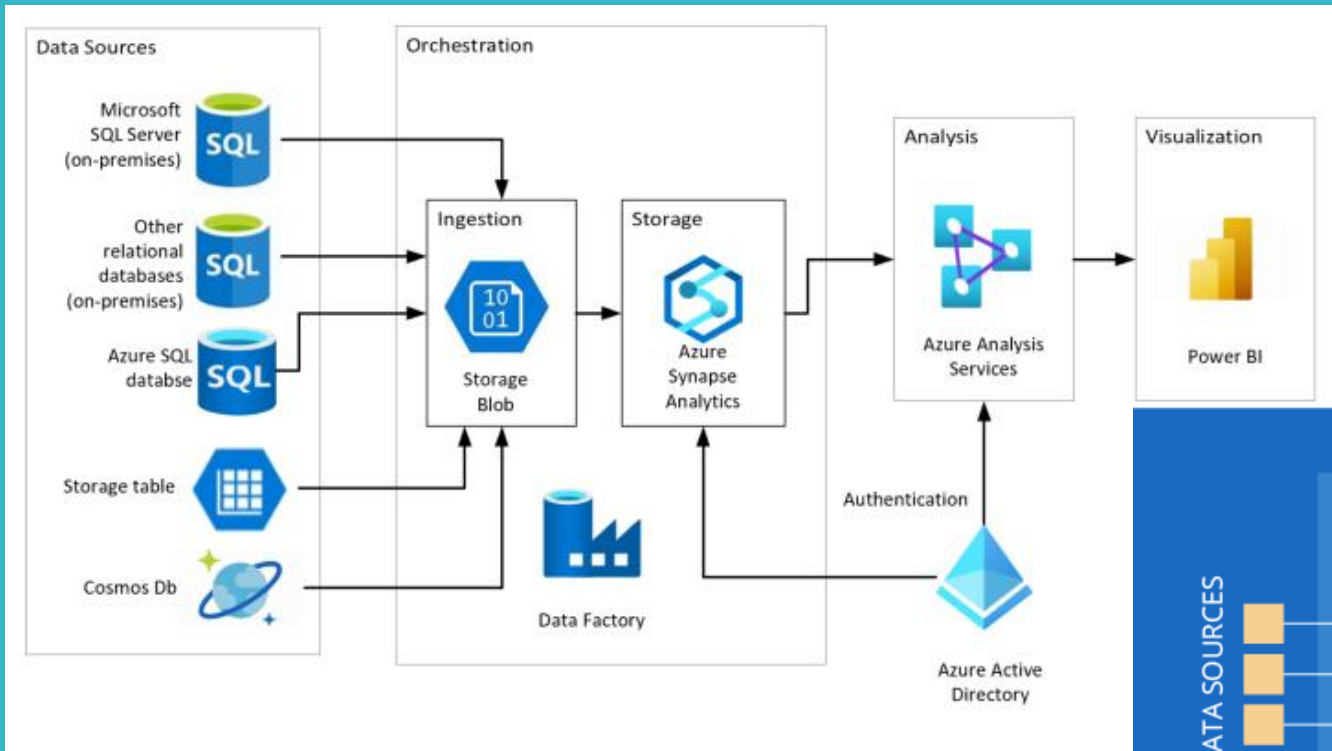
CONTA DE ARMAZENAMENTO - DATA LAKE NO AZURE

- Utilizado com storage para qualquer formato de arquivos.
- Permite armazenar arquivos de petabytes de tamanho e trilhões de objetos.
- Implantação rápida com pagamento por armazenamento.



STAGING

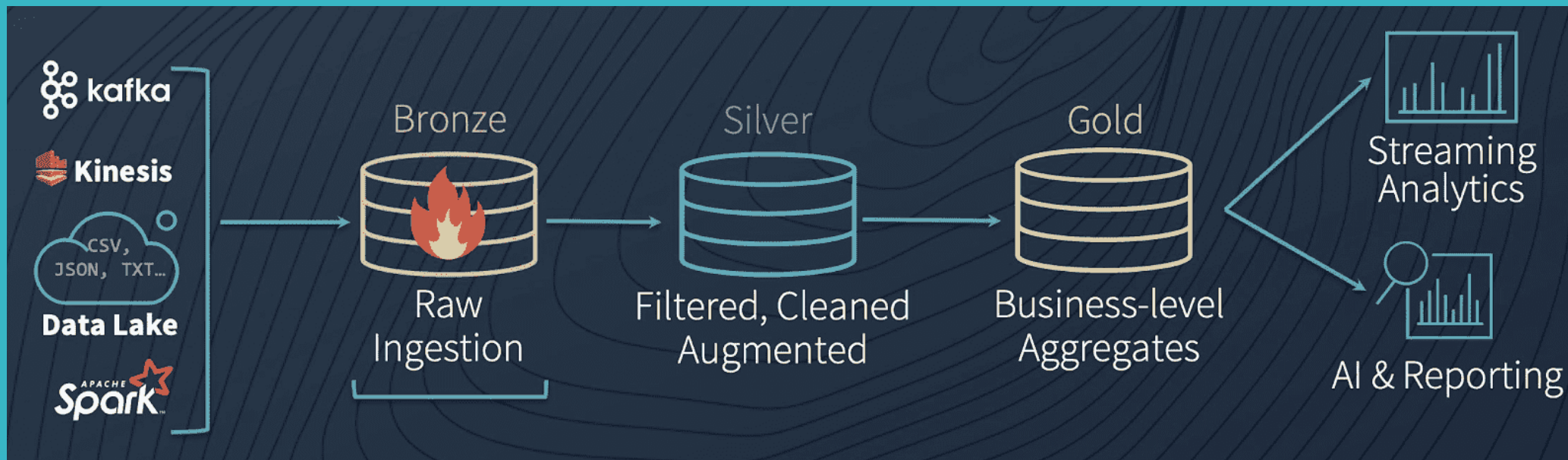
CENTRALIZA OS DADOS **ANTES E APÓS** O PROCESSAMENTO.



ORQUESTRAÇÃO E CONVERSÃO DE FORMATOS

A orquestração de dados possui objetivos como:

- CENTRALIZAR DADOS DE DIVERSAS FONTES EM UM DATA LAKE.
- CONVERSÃO DE DIVERSOS FORMATOS EM UM ÚNICO.
- PREPARO DE PRÉ PROCESSAMENTO.
- MIGRAÇÕES.
- DESCARTE DE DADOS NÃO ÚTEIS E TRATAMENTO.



FORMATOS

CSV é mais rápido para escrever, o JSON o mais fácil de ser entendido por um ser humano e o Parquet o mais rápido de ler.

CSV é o padrão de fato para muitos dados; fácil de compreender para usuários e computadores e tornou-se mais acessível através do Excel. Muitos sistemas compreendem CSV.

JSON é o padrão para comunicação na web. APIs e sites estão se comunicando constantemente usando JSON por causa de suas propriedades de usabilidade, como esquemas personalizáveis.

Parquet é otimizado para o paradigma Write Once Read Many (WORM). É lento para escrever, mas incrivelmente rápido para ler, especialmente quando você está acessando apenas um subconjunto das colunas totais. Para casos de uso que requerem operação em linhas inteiras de dados, um formato como CSV, JSON ou mesmo AVRO deve ser usado.

Manipular dados Parquet:

[Parquet Files - Spark 3.1.2 Documentation \(apache.org\)](#)

Manipular dados Json:

[JSON Files - Spark 3.1.2 Documentation \(apache.org\)](#)

Spark Format Showdown		File Format		
		CSV	JSON	Parquet
A t t r i b u t e	Columnar	No	No	Yes
	Compressable	Yes	Yes	Yes
	Splittable	Yes*	Yes**	Yes
	Human Readable	Yes	Yes	No
	Nestable	No	Yes	Yes
	Complex Data Structures	No	Yes	Yes
	Default Schema: Named columns	Manual	Automatic (full read)	Automatic (instant)
	Default Schema: Data Types	Manual (full read)	Automatic (full read)	Automatic (instant)



Conceitos

Atividade - Lookup:

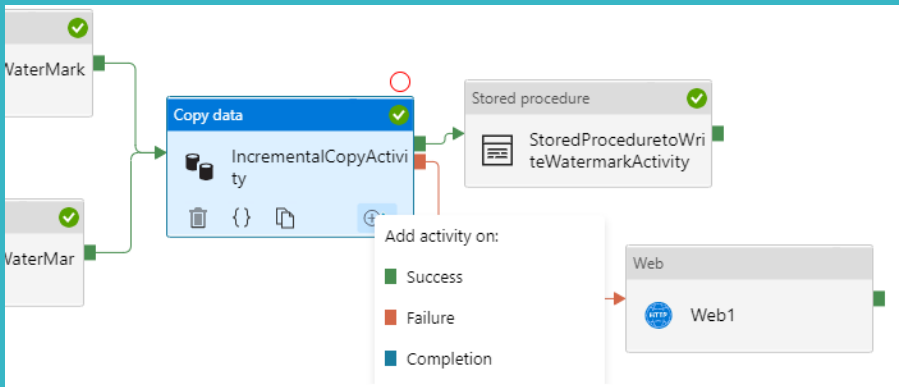
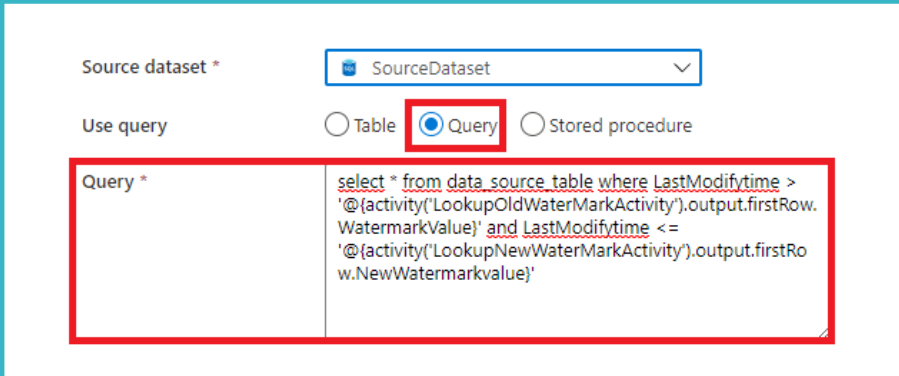
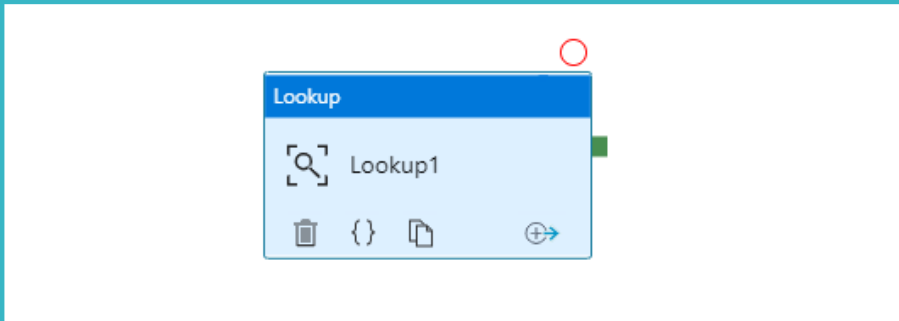
A atividade Lookup lê e retorna o conteúdo de um arquivo de configuração ou tabela. Ela também retorna o resultado da execução de uma consulta ou procedimento armazenado.

Querys personalizadas:

Após configurar o dataset, é possível editar a query de consulta, update ou inserção fornecendo detalhes ou enviando parâmetros do pipeline.

Flows de Atividades no Pipeline:

Conecte o próximo passo para guiar a execução de acordo com o status da atividade.





Laboratório

Migração incremental com Delta

Assuntos:

- Lookup.
- Querys personalizadas.
- Flow de atividades.
- Parâmetros.

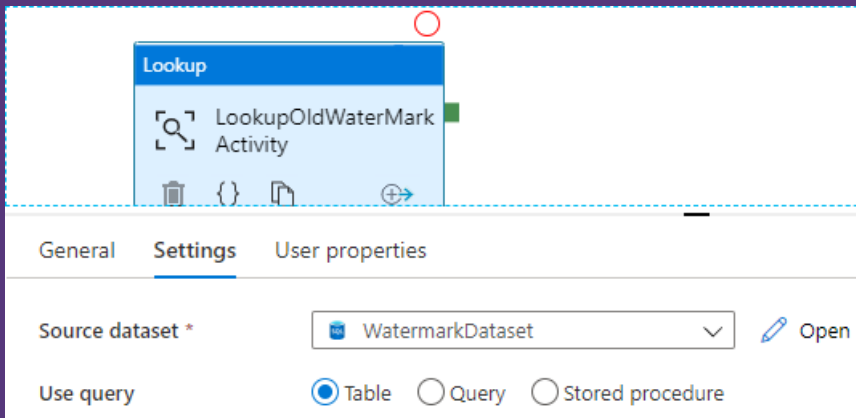
Tecnologias:

- SQL do Azure
- Storage Account - Data Lake Storage

Lab – Migração incremental com Delta

No Azure:

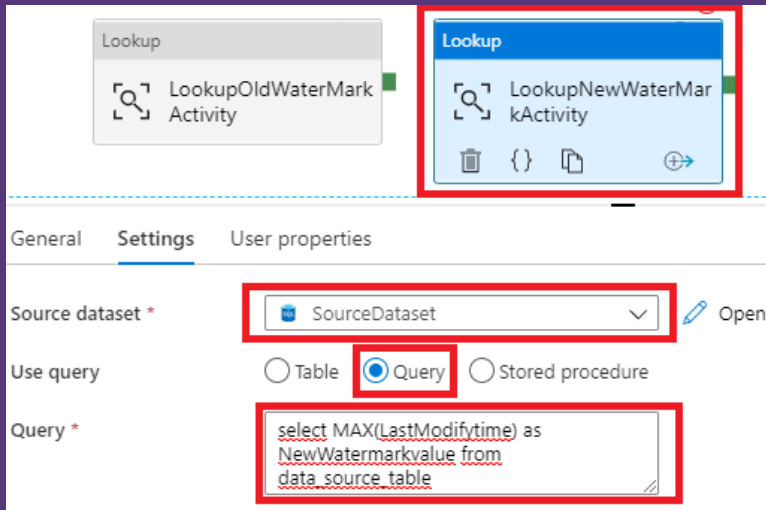
1. Acesse o **SQL do Azure** criado no lab anterior e autorize o seu IP cliente no **Firewall do SQL**.
2. Na sequencia, acesse o **editor de consultas**. Execute os comandos do arquivo 'lab2_SQL.txt' disponível no Portal do Aluno.
3. Abra o Data Factory e crie um novo Pipeline.
4. Insira a atividade **Lookup**. Configure o nome para **LookupOldWaterMarkActivity**.
5. Acesse a guia Settings e crie um novo **dataset** com o nome **WatermarkDataset** apontando para a tabela **watermarktable** do SQL.



VÍDEO

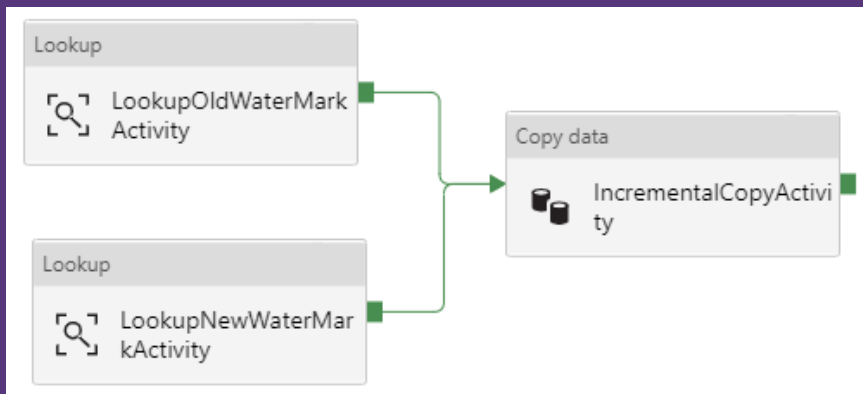
Lab – Migração incremental com Delta

1. Insira outra atividade **Lookup** no Pipeline e configure o nome para **LookupNewWaterMarkActivity**.
2. Configure o **dataset** para a tabela **data_source_table** e personalize a query conforme imagem a seguir:



```
select  
MAX(LastModifytime) as  
NewWatermarkvalue from  
data_source_table
```

3. Insira uma atividade de **Copy**. Configure o nome para **IncrementalCopyActivity** e ligue o fluxo conforme a imagem:



VÍDEO

Lab – Migração incremental com Delta

1. Na atividade Copy, selecione a guia Source e configure conforme a imagem:

The screenshot shows the configuration of a Copy Activity in Azure Data Factory. The 'Source' tab is selected. The 'Source dataset' is set to 'SourceDataset'. The 'Use query' radio button is selected, and the 'Query' text area contains the SQL query for incremental copy. The 'Copy data' activity is highlighted with a red box in the top diagram.

```
select * from data_source_table where
LastModifytime >
'@{activity('LookupOldWaterMarkActivi
ty').output.firstRow.WatermarkValue}'
and LastModifytime <=
'@{activity('LookupNewWaterMarkActivi
ty').output.firstRow.NewWatermarkvalu
e}'
```

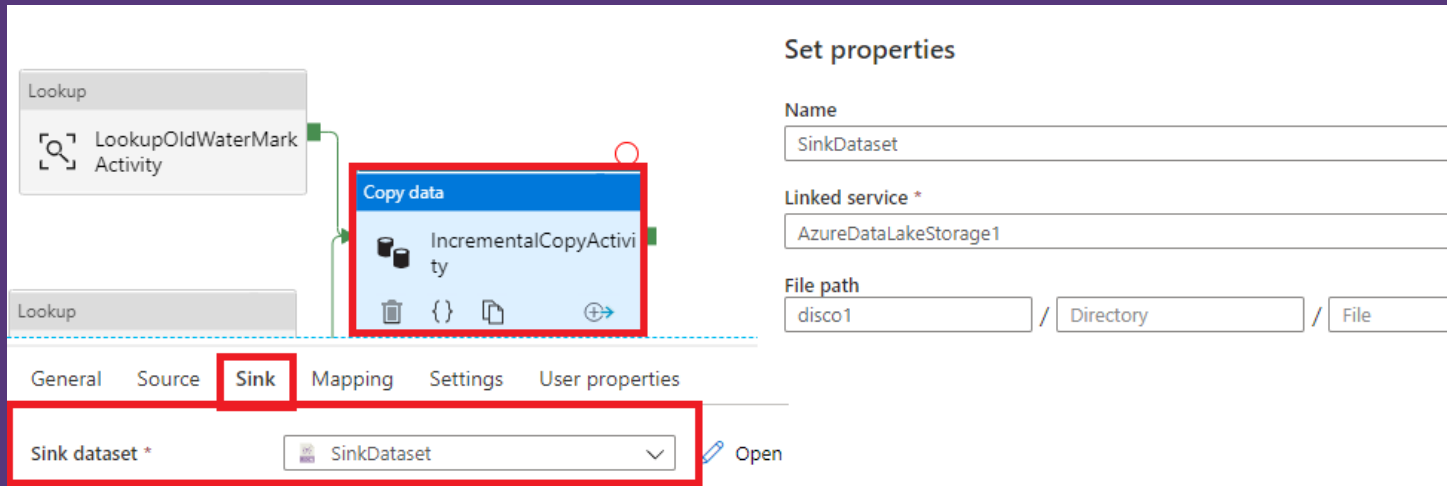
```
select * from data_source_table where
LastModifytime >
'@{activity('LookupOldWaterMarkActivi
ty').output.firstRow.WatermarkValue}'
and LastModifytime <=
'@{activity('LookupNewWaterMarkActivi
ty').output.firstRow.NewWatermarkvalu
e}'
```



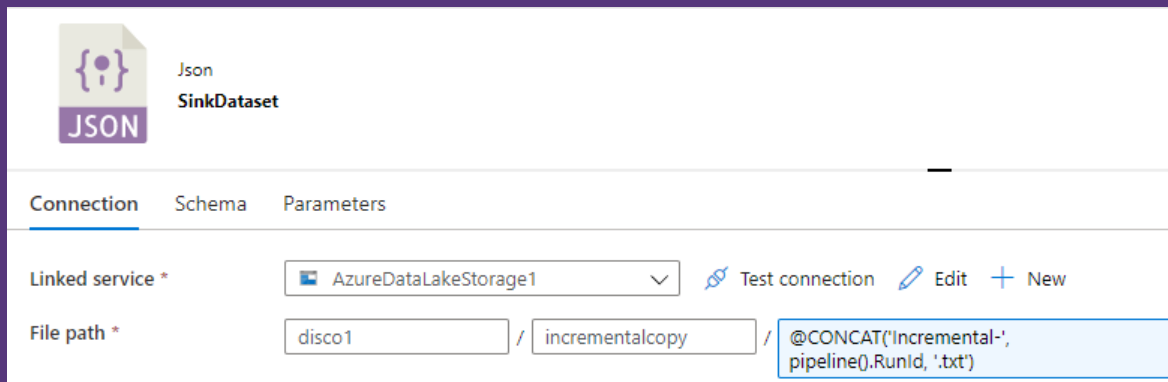
VÍDEO

Lab – Migração incremental com Delta

1. Na atividade Copy, selecione a guia **Sink** e crie um novo dataset apontando para um container de blob da storage account criada no lab anterior.



2. Após Concluir a criação do Dataset, selecione a opção Open e personalize o dataset conforme imagem a seguir. Este procedimento criará arquivos com os dados da extração a cada execução do pipeline.



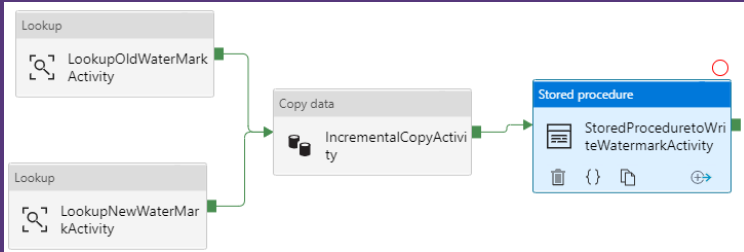
```
@CONCAT('Incremental-',  
pipeline().RunId, '.txt')
```



VÍDEO

Lab – Migração incremental com Delta

1. De volta ao pipeline, insira uma nova atividade de **Stored procedure** com o nome **StoredProceduretoWriteWatermarkActivity** e configure o fluxo:



2. Configure a guia Settings conforme a imagem a seguir. Note a utilização dos parâmetros que serão herdados das atividades do lookup:

General Settings User properties

Linked service * AzureSqlDatabase1

Stored procedure name * [dbo].[usp_write_watermark]

Stored procedure parameters

Import

NAME	TYPE	VALUE
LastModifiedtime	DateTime	@{activity("LookupNewWaterMarkActivit...") Add dynamic content [Alt+P]
TableName	String	@{activity("LookupOldWaterMarkActivity...") Add dynamic content [Alt+P]

```
@{activity('LookupNewWaterMarkActivity')  
.output.firstRow.NewWatermarkvalue}
```

```
@{activity('LookupOldWaterMarkActivity')  
.output.firstRow.TableName}
```



VÍDEO

Lab – Migração incremental com Delta

1. Ative o modo Debug do Pipeline.
2. Clique em Validate para verificar se todas as configurações estão consistentes.
3. Execute o Pipeline e verifique os arquivos criados na Storage Account.
4. Aproveite para verificar se a tabela do SQL Server foi atualizada.

The screenshot displays an Azure Data Factory pipeline with four activities: two 'Lookup' activities, one 'Copy data' activity, and one 'Stored procedure' activity. The 'Lookup' activities are 'LookupOldWaterMarkActivity' and 'LookupNewWaterMarkActivity'. The 'Copy data' activity is 'IncrementalCopyActivity'. The 'Stored procedure' activity is 'StoredProceduretoWriteWatermarkActivity'. The pipeline is shown in the 'Output' tab, displaying the results of a pipeline run with ID 8117c0e9-2767-47f0-847a-5c3acfbe89be. The output table lists the activities and their status, all of which are 'Succeeded'.

Name	Type	Run start	Duration	Status
StoredProceduretoWriteWatermarkActivity	SqlServerStored	2020-10-07T13:43:21.16447	00:00:12	✓ Succeeded
IncrementalCopyActivity	Copy	2020-10-07T13:43:14.75912	00:00:05	✓ Succeeded
LookupNewWaterMarkActivity	Lookup	2020-10-07T13:43:12.18935	00:00:02	✓ Succeeded
LookupOldWaterMarkActivity	Lookup	2020-10-07T13:43:12.17377	00:00:02	✓ Succeeded



Conceitos

Metadata:

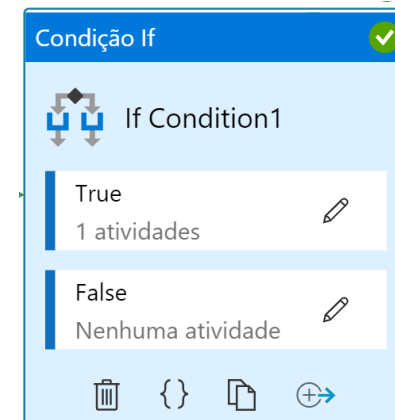
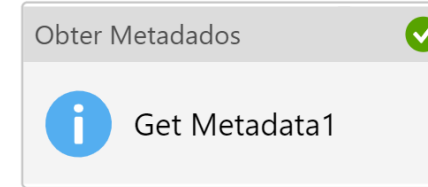
Permite validar a estrutura de fontes e destinos de dados antes ou depois de movimentações. Utilize a propriedade **ESTRUTURA** para retornar uma lista de nomes de coluna e tipos de coluna.

Consulte outros tipos de retornos para metadados:

<https://docs.microsoft.com/pt-br/azure/data-factory/control-flow-get-metadata-activity>

Estrutura condicional - IF:

Permite avaliar condições lógicas e resultados de variáveis agrupando atividades para TRUE e FALSE.



Laboratório

Metadados + IF

Assuntos:

- Parâmetros.
- Multi pipelines.
- Schemas

Tecnologias:

- Storage Account
- CSV



Lab: Metadados + IF:

1. Utilize o arquivo [validar_schema.csv](#) disponível no Portal do aluno para este lab.
2. Faça o upload do arquivo para um container em um Data lake do Azure.
3. Já no Data Factory, crie uma nova connection apontando o Data Lake.

Editar serviço vinculado (Azure Data Lake Storage Gen2)

Nome *
AzureDataLakeStorage1

Descrição

Conectar-se via runtime de integração * ⓘ
AutoResolveIntegrationRuntime

Método de autenticação
Chave de conta

Método de seleção de conta ⓘ
☐ Da assinatura do Azure ☒ Inserir manualmente

URL *
https://storageteste123ab.dfs.core.windows.net

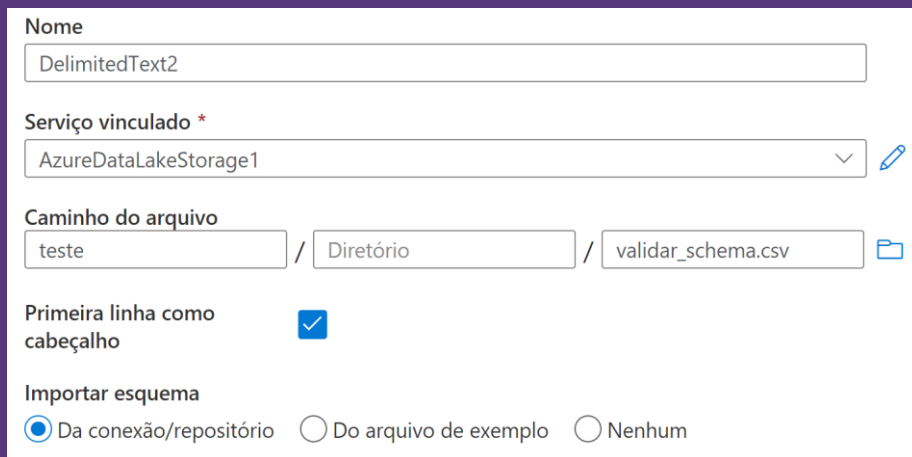
Chave da conta de armazenamento Azure Key Vault

Chave da conta de armazenamento *
.....



Lab: Metadados + IF:

1. Crie um Dataset apontando para o arquivo csv.



Nome
DelimitedText2

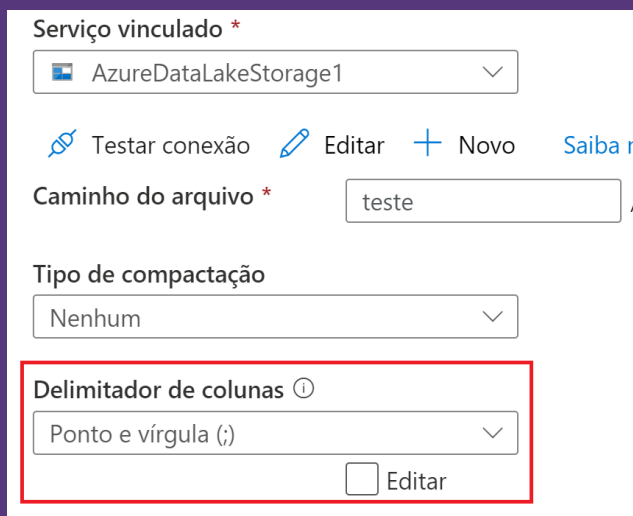
Serviço vinculado *
AzureDataLakeStorage1

Caminho do arquivo
teste / Diretório / validar_schema.csv

Primeira linha como cabeçalho ☒

Importar esquema
☒ Da conexão/repositório ☐ Do arquivo de exemplo ☐ Nenhum

2. Edite o dataset e altere o limitador de colunas para 'Ponto e vírgula':



Serviço vinculado *
AzureDataLakeStorage1

Testar conexão Editar + Novo Saiba mais

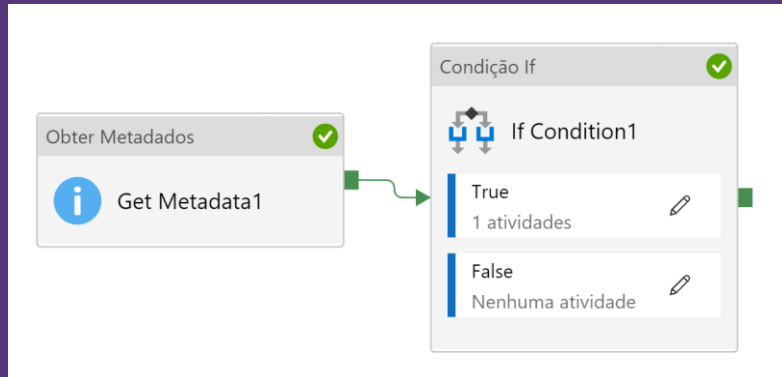
Caminho do arquivo *
teste

Tipo de compactação
Nenhum

Delimitador de colunas ⓘ
Ponto e vírgula (;) Editar

Lab: Metadados + IF:

1. Crie um novo Pipeline com as atividades:

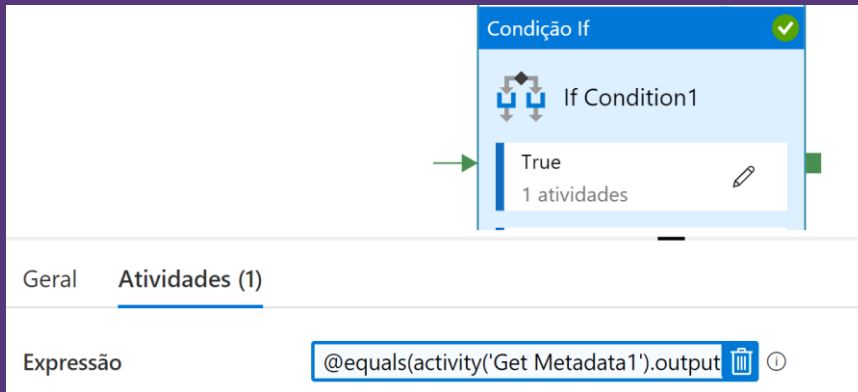


2. Configure a atividade Metadados, conforme a imagem a seguir, selecionando o dataset do arquivo CSV:

The screenshot shows the configuration for the 'Obter Metadados' activity. The 'Conjunto de Dados' (Data Set) is set to 'DelimitedText1'. The 'Filtrar pela hora da última modificação' (Filter by last modification time) is set to 'Hora de início (UTC)' (Start time (UTC)). The 'Ignorar contagem de linhas' (Ignore line count) is set to 'Não' (No). The 'Lista de campos' (List of fields) is set to 'Argumento' (Argument) and 'Estrutura' (Structure).

Lab: Metadados + IF:

1. Configure a expressão a seguir para a atividade Condição IF:

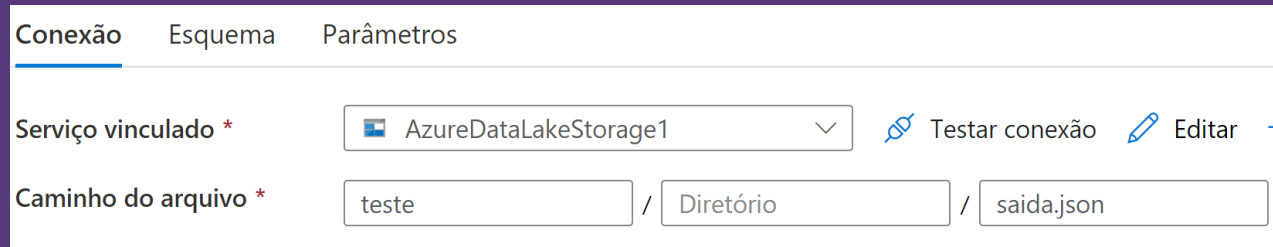


```
@equals(activity('Get Metadata1').output.structure,  
  createArray(  
    json('{"name": "Column1", "type": "String"}'),  
    json('{"name": "Column N", "type": "String"}')  
  )  
)
```



Lab: Metadados + IF:

1. Edite o Pipeline da atividade IF TRUE e configure uma atividade Copy com dataset de origem apontado para o CSV e um novo dataset de destino em formato JSON, conforme exemplo:



Conexão Esquema Parâmetros

Serviço vinculado * AzureDataLakeStorage1 Testar conexão Editar

Caminho do arquivo * teste / Diretório / saída.json

2. Teste a execução do projeto.

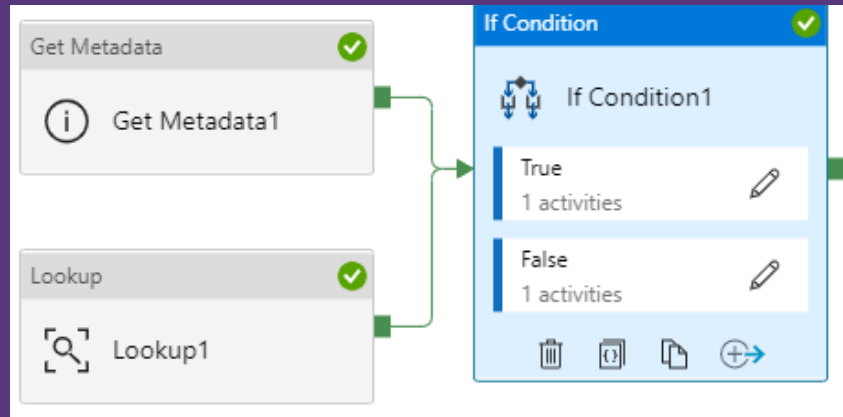


Lab: Metadados + IF:

Opcionalmente, é possível tornar a validação do schema dinâmico, inserindo o formato a ser validado em uma tabela do SQL Server:

ID	object_name	schema_text
1	Test.csv	[{"name":"Column1","type":"String"}, ...

O pipeline, para este cenário seria configurado com as atividades e expressão IF:

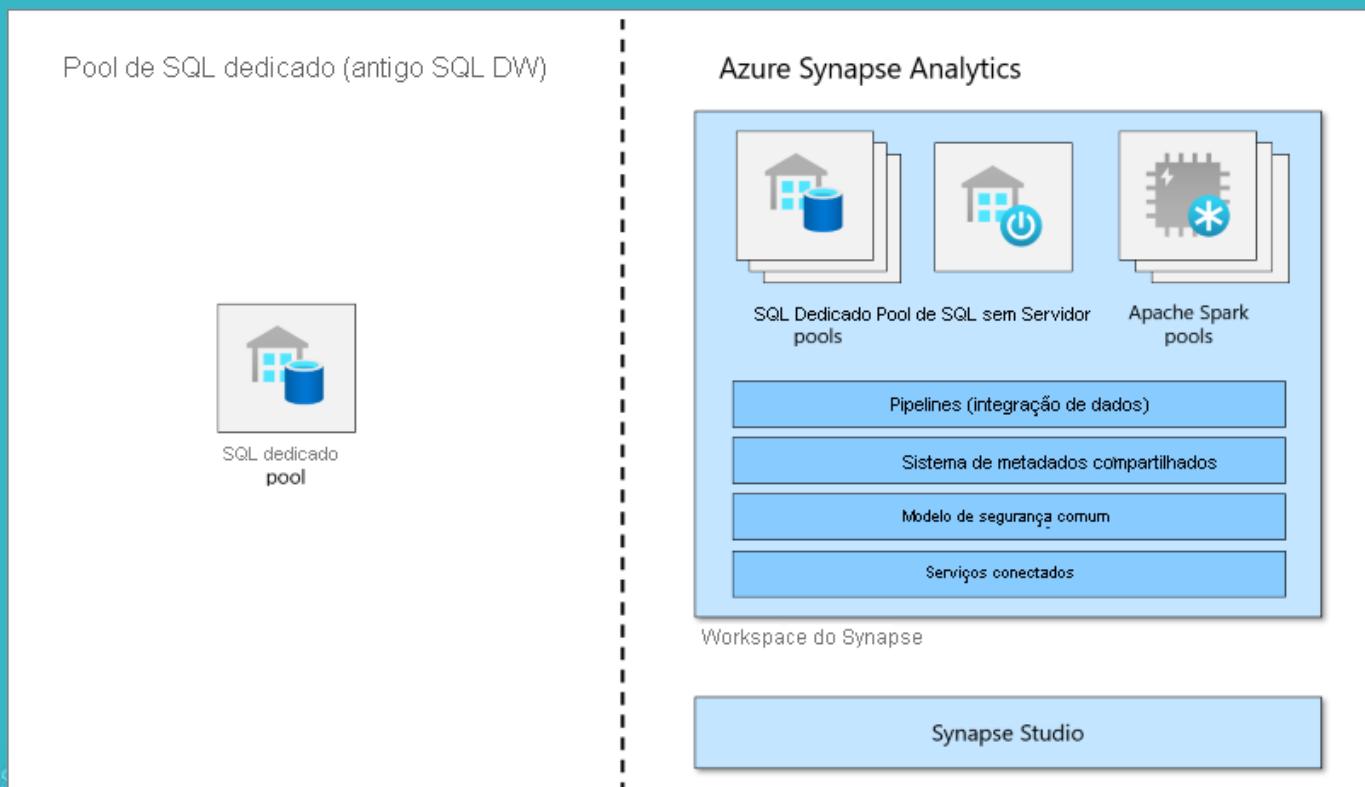


```
@equals(activity('Get Metadata1').output.structure,json(activity('Lookup1').output.firstRow.schema_text))
```

DW - Arquitetura MPP (processamento paralelo maciço)

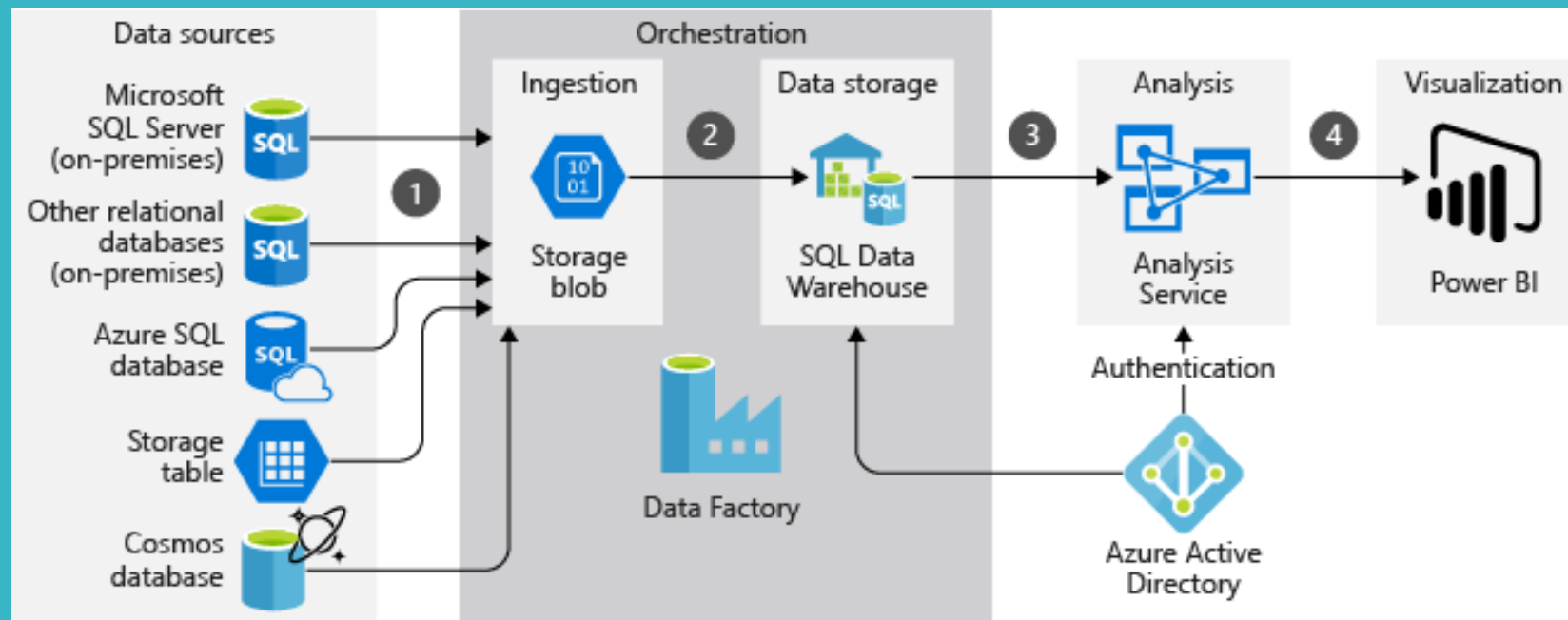
O pool de SQL dedicado (antigo SQL DW) refere-se aos recursos de data warehouse disponíveis no Azure Synapse Analytics.

- Pool do SQL (pagamento por DWU provisionado)
- SQL sob demanda (pagamento por TB processado).
- Pipelines do Synapse: Integração de dados híbridos.



DW - Arquitetura

- ELT (Extrair, Carregar e Transformar).
- Permite distribuir o processamento em vários locais.
- **Trabalha com controle de nós.**
- Datasources externos multiformatos.
- Clusters de alto desempenho.
- Polybase
- DMVs (exibições de gerenciamento dinâmico).



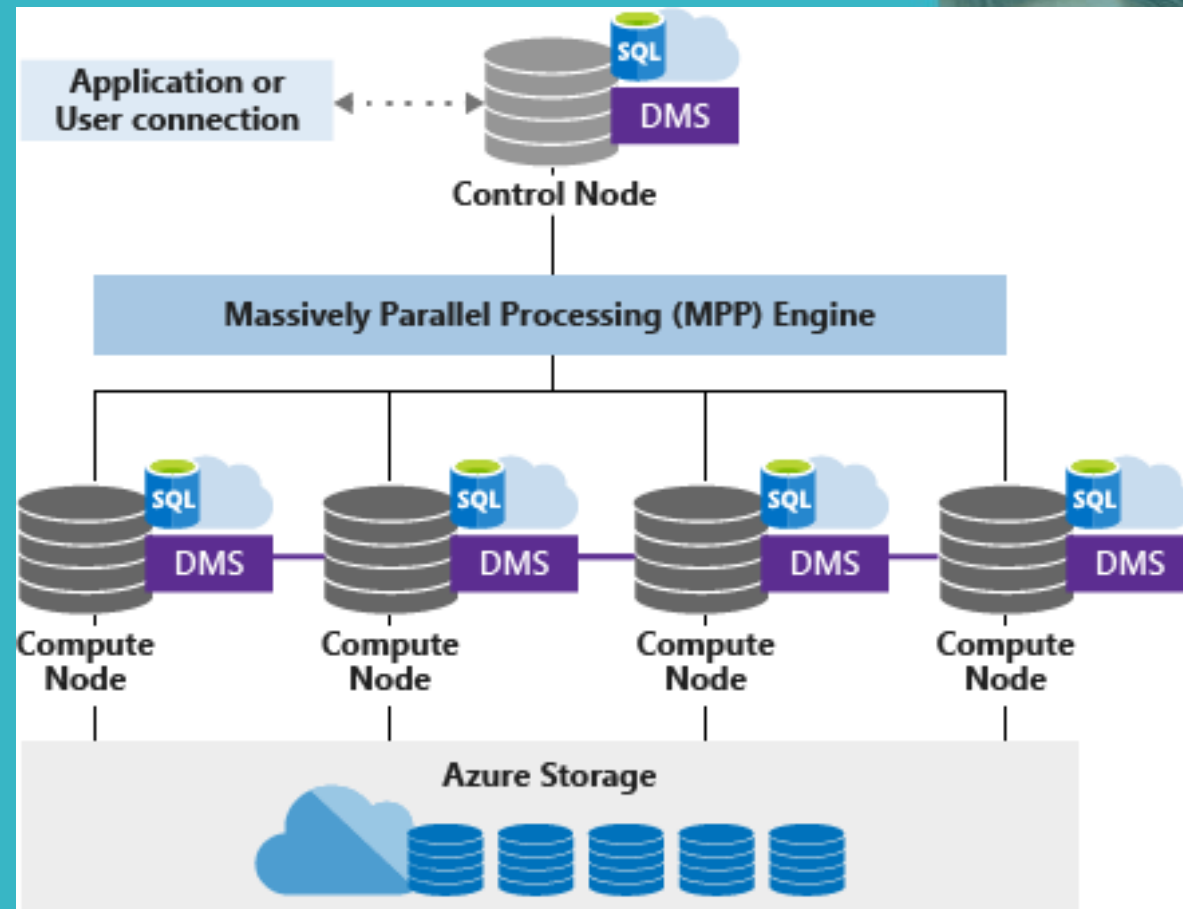
DW - NÓS – Processamento massivo

Nós de controle

Quando você envia uma consulta T-SQL, o nó de Controle a transforma em consultas que são executadas em cada distribuição paralelamente.

Nós de computação

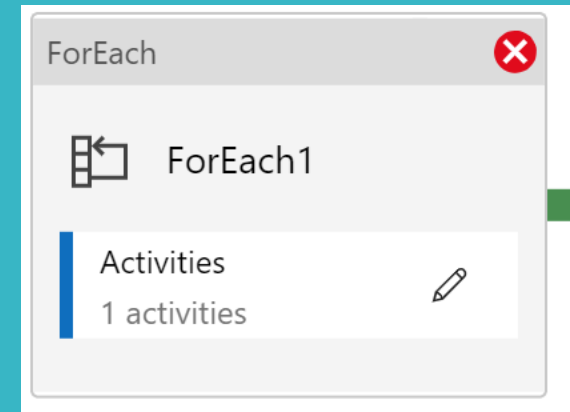
Os nós de computação fornecem capacidade de computação. Distribuições são mapeados para nós de computação para processamento. À medida que você paga por mais recursos de computação, as distribuições são remapeadas para os nós de computação disponíveis.



Conceitos

Atividade ForEach:

A atividade ForEach define um fluxo de controle repetitivo no seu pipeline. Essa atividade é usada para iterar em uma coleção e executa atividades especificadas em um loop. A implementação dessa atividade em loop é semelhante à estrutura em loop Foreach nas linguagens de programação.



Conceitos

Parametrização de Fluxos de Dados:

Permite 'passar' valores estáticos e dinâmicos entre Pipelines e atividades. Controle fluxos de dados, utilize expressões e tome decisões com base em 'parâmetros'.

Consulte:

<https://docs.microsoft.com/pt-br/azure/data-factory/parameters-data-flow>

General	Settings	Parameters	User properties
Data flow parameters ⓘ			
NAME	VALUE	TYPE	EXPRESSION ⓘ
stringParam	@pipeline().Pipeline ⓘ	string	<input type="checkbox"/>
intParam	@pipeline().parameters.pipelineParam ⓘ	integer	<input type="checkbox"/>

DataFlowDemo

Save Save as template Validate Debug

Mapping Data Flow

DataFlowDemo

General Settings Parameters User properties

Parameters ⓘ

VALUE

```
@if(equals(pipeline().parameters.pipelineParam, 'paramValue'), 'trueOutput', 'falseOutput')
```

Clear contents

Filter...

Use [expressions, functions](#) or refer to [system variables](#).

Returns true if two values are equal. For example, if parameter1 is foo, the following returns true:

greater

Returns true if the first argument is greater than the second. Note, values can only be compared if they are of the same data type.

greaterOrEquals

Returns true if the first argument is greater than or equal to the second. Note, values can only be compared if they are of the same data type.

if

Returns a specified value based on if the expression provided results in true or false.

less

Returns true if the first argument is less than the second. Note, values can only be compared if they are of the same data type.

lessOrEquals

Returns true if the first argument is less than or equal to the second. Note, values can only be compared if they are of the same data type.

not

Returns true if the parameter is false. The following returns true: not(contains('200', '200')).

or

Returns true if either of the parameters are true. Both arguments need to be Boolean.

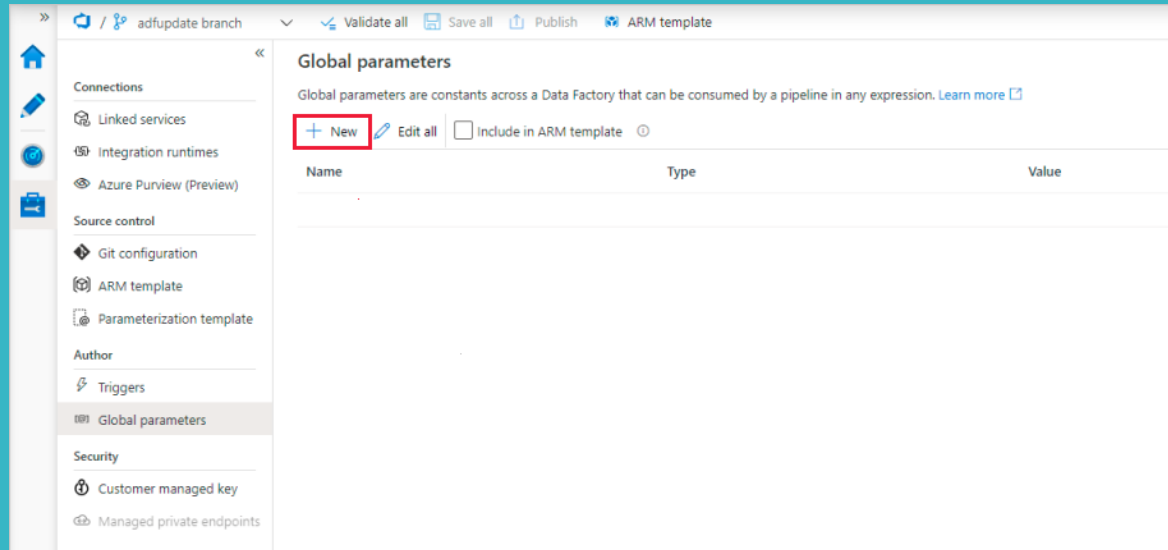
▶ **Math Functions**

▶ **String Functions**

▲ **Parameters**

pipelineParam

Variáveis Globais:



Parâmetros globais: <https://docs.microsoft.com/pt-br/azure/data-factory/author-global-parameters>

Parametrização:

General	Settings	Parameters	User properties
Data flow parameters ⓘ			
NAME	VALUE	TYPE	EXPRESSION ⓘ
stringParam	<input type="text" value="@pipeline().Pipeline"/>	string	<input type="checkbox"/>
intParam	<input type="text" value="@pipeline().parameters.pipelineParam"/>	integer	<input type="checkbox"/>

Parametrização de fluxos: <https://docs.microsoft.com/pt-br/azure/data-factory/parameters-data-flow>

Expressões:

Utilize a linguagem de expressões para construir valores dinâmicos e manipular os dados que transitam entre as atividades do pipeline.

Exemplos:

Nome de arquivo de saída:

'Test_' + toString(currentDate()) + '.csv'

Linguagem de expressão: <https://docs.microsoft.com/pt-br/azure/data-factory/control-flow-expression-language-functions>

Variáveis do sistema:

@pipeline().DataFactory Nome dos dados ou do workspace do Synapse em que a execução de pipeline está ocorrendo
 @pipeline().Pipeline Nome do pipeline
 @pipeline().RunId ID da execução do pipeline específica

Variáveis de execução: <https://docs.microsoft.com/pt-br/azure/data-factory/control-flow-system-variables>

Laboratório extra

ForEach com extração e Sink para Synapse Analytics DW

Assuntos:

- Parâmetros.
- Multi pipelines.
- ForEach
- Schemas
- Staging

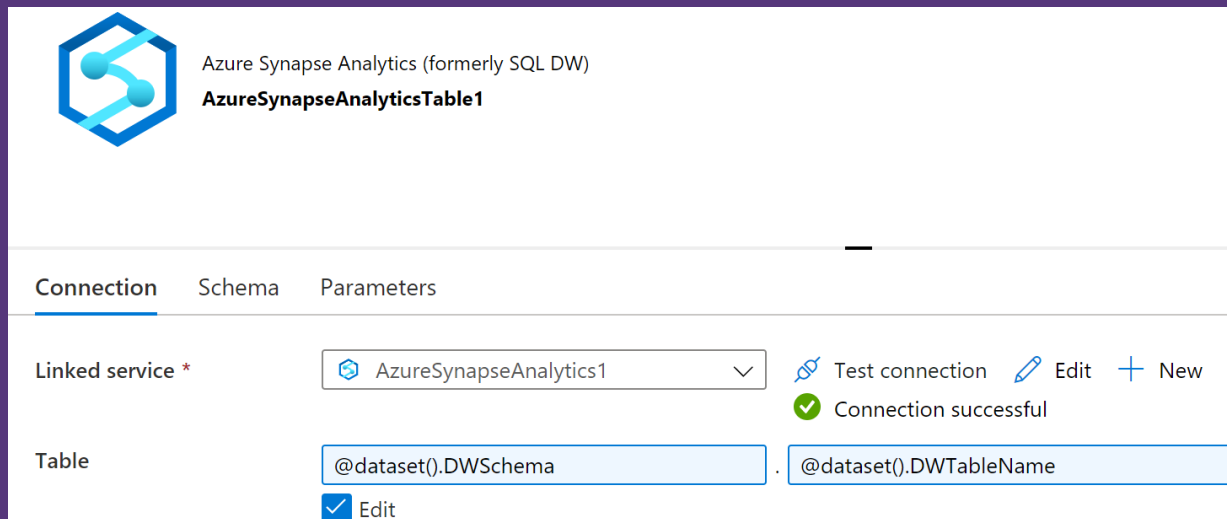
Tecnologias:

- Synapse Analytics
- Storage Account
- SQL do Azure



Lab Desafio: Cópia SQL para DW com parametrização:

1. Instale uma **storage account** com um container.
2. Instale um **SQL do Azure** com banco de dados de amostra.
3. Instale um **SQL Pool DW** no mesmo servidor do SQL.
4. No Data Factory, configure as **03 conexões** para os serviços instalados anteriormente.
5. Configure um dataset para a storage account e outro para o SQL do Azure.
6. Para o dataset do DW, configure a connection e guia parameters conforme as imagens:



Azure Synapse Analytics (formerly SQL DW)

AzureSynapseAnalyticsTable1

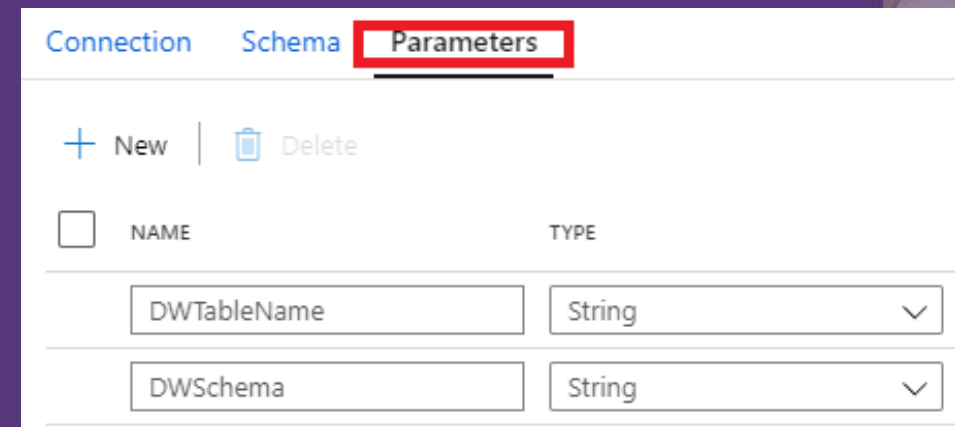
Connection Schema Parameters

Linked service * AzureSynapseAnalytics1 Test connection Edit + New
Connection successful

Table @dataset().DWSchema @dataset().DWTableName
Edit

@dataset().DWSchema

@dataset().DWTableName



Connection Schema **Parameters**

+ New | Delete

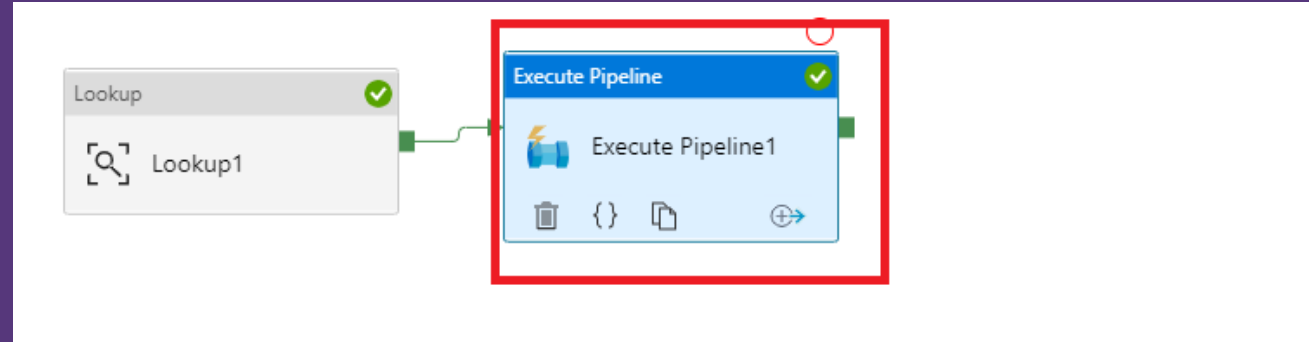
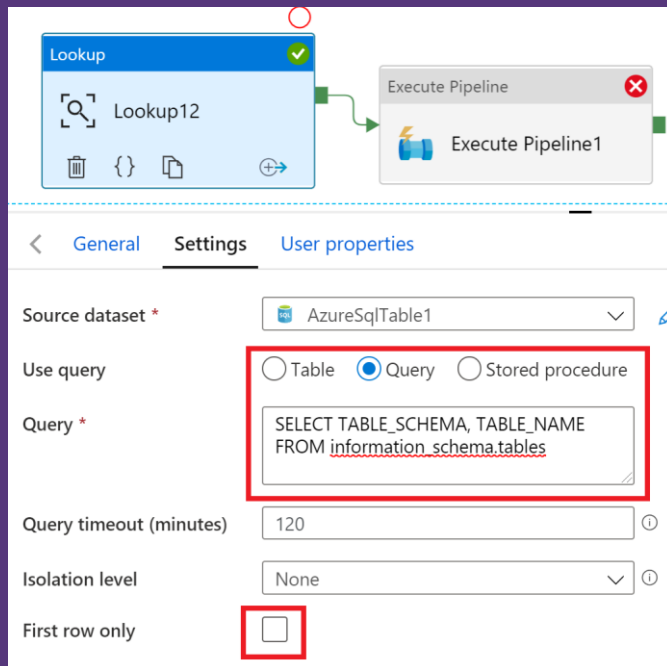
NAME	TYPE
DWTableName	String
DWSchema	String

DWTableName

DWSchema

Lab Desafio: Cópia SQL para DW com parametrização:

1. Crie um novo Pipeline, insira um lookup e um módulo exec pipeline.

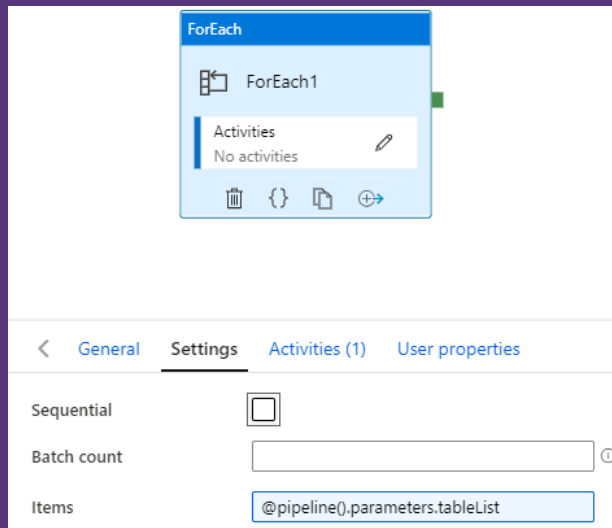


```
SELECT TABLE_SCHEMA,  
TABLE_NAME FROM  
information_schema.TABLES
```

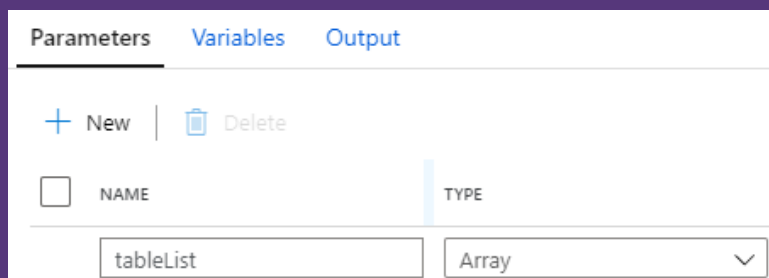

Lab Desafio: Cópia SQL para DW com parametrização:

1. **Crie um segundo pipeline.** Insira uma atividade de ForEach no Pipeline. No campo settings, determine:

`@pipeline().parameters.tableList`

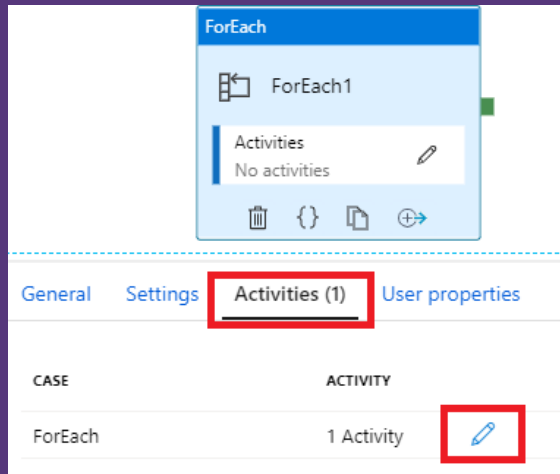


2. Clique em qualquer área do Pipeline e adicione o parâmetro:

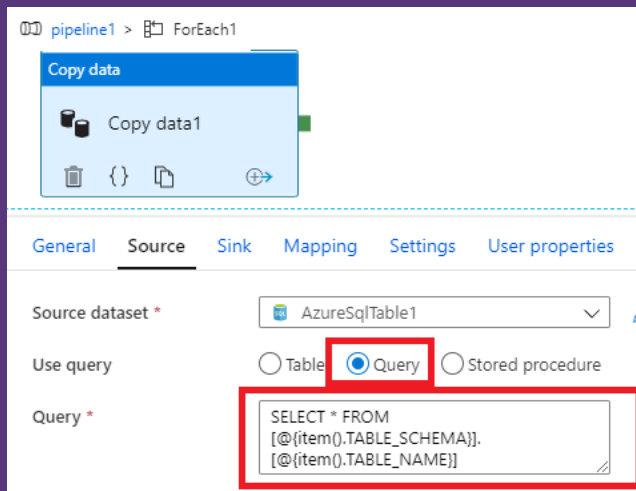


Lab Desafio: Cópia SQL para DW com parametrização:

1. Selecione oForEach, clique na guia de atividades e acesse o modo de edição:



2. Insira uma atividade Copy Data dentro do ForEach. No Source, defina o **SQL Server** com a query personalizada: `SELECT * FROM [{item().TABLE_SCHEMA}].[{item().TABLE_NAME}]`



Lab Desafio: Cópia SQL para DW com parametrização:

1. No Sink, aponte para o DW, definindo o pre copy com o script:

```
TRUNCATE TABLE [{item().TABLE_SCHEMA}].[{item().TABLE_NAME}]
```

General Source Sink Mapping Settings User properties

Sink dataset * Open New

Dataset properties ⓘ

NAME	VALUE
DWTableName	@item().TABLE_NAME
DWSchema	@item().table_schema

Copy method ☒ PolyBase ⓘ ☐ Copy command (Preview) ⓘ ☐ Bulk insert

Allow PolyBase ☒

Reject type Value

Reject value 0

Use type default ☐

Table option ☒ None ☐ Auto create table ⓘ

Pre-copy script ⓘ

```
TRUNCATE TABLE  
[{item().TABLE_SCHEMA}].  
[{item().TABLE_NAME}]
```

General Source Sink Mapping Settings User properties

ⓘ You will be charged # of used DIUs * copy duration * \$0.25/DIU-hour. Local

Data integration unit Auto ⓘ

☐ Edit

Add dynamic content [Alt+P]

Degree of copy parallelism ⓘ

☒ Edit

Fault tolerance ⓘ

Enable staging ☒ ⓘ

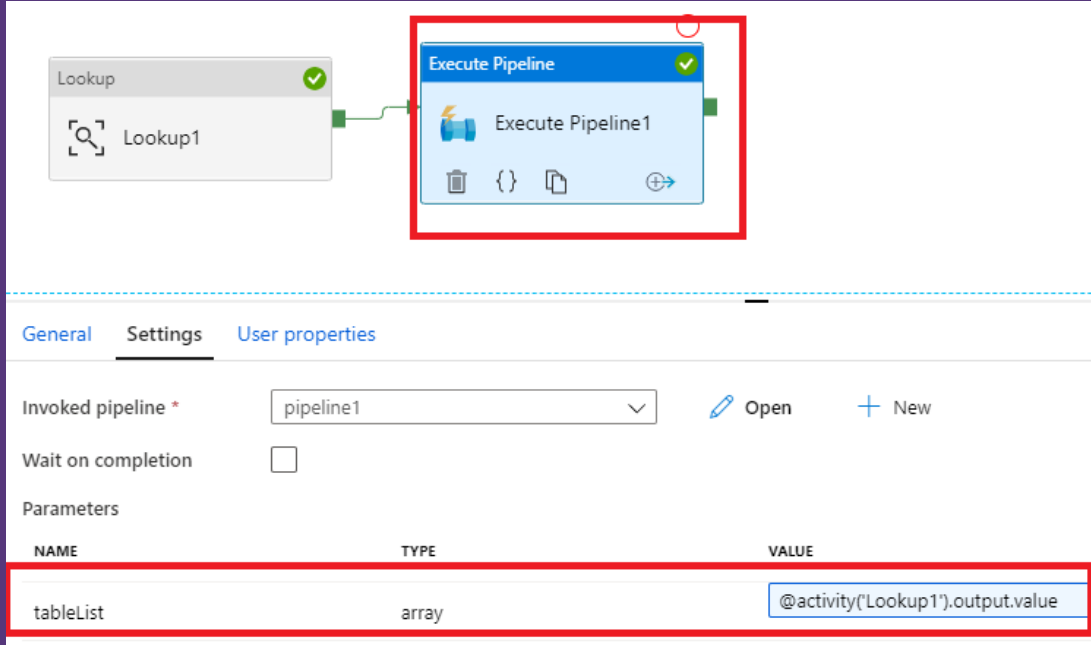
Staging settings

Staging account linked service AzureBlobStorage1 ⓘ

Storage Path vulnerability-assessment ⓘ

Lab Desafio: Cópia SQL para DW com parametrização:

1. Retorne ao Pipeline inicial e configure a chamada para o Pipeline que possui o ForEach:



General Settings User properties

Invoked pipeline * [Open](#) [+ New](#)

Wait on completion ☐

Parameters

NAME	TYPE	VALUE
tableList	array	@activity("Lookup1").output.value

```
@activity('Lookup1').output.value
```

2. Teste a execução do projeto.

Fonte: <https://docs.microsoft.com/pt-br/azure/data-factory/tutorial-bulk-copy-portal>



Azure Academy

Rubens Guimarães

 [/rubensguimaraes](#)



Microsoft®
Most Valuable
Professional

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