Pós-graduação em Data Science & Business Analytics Formato Blended 2ª Edição - 2022

Executive Education

Data Warehousing

Artur Vieira

Index

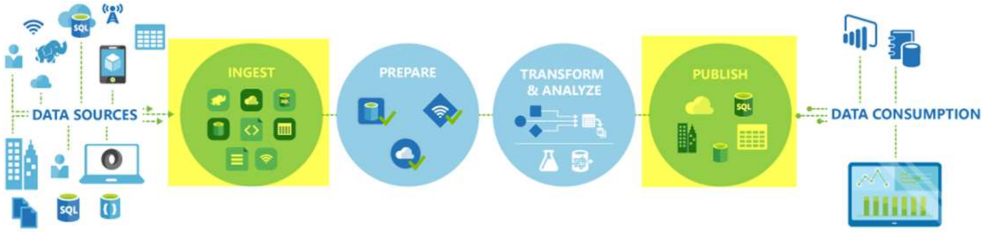
Índice

- Azure Data Factory Concepts
- Azure Data Factory Practical Experience
  - Azure Portal Resources – DW Resource Group
  - Navigate to Azure Data Factory
  - ADF Copy Activity – Configure Dataset and Linked Service
  - ADF Copy Activity – Configure Pipeline
  - ADF Copy Activity – Configure Source
  - ADF Copy Activity – Configure Sink
  - Execute the Azure Data Factory pipeline

Azure Data Factory Concepts

Data Factory Copy Activity

In Azure Data Factory, you can use the Copy activity to copy data among data stores located on-premises and in the cloud. After you copy the data, you can use other activities to further transform and analyze it. You can also use the Copy activity to publish transformation and analysis results for business intelligence (BI) and application consumption.



The Copy activity is executed on an integration runtime. You can use different types of integration runtimes for different data copy scenarios:

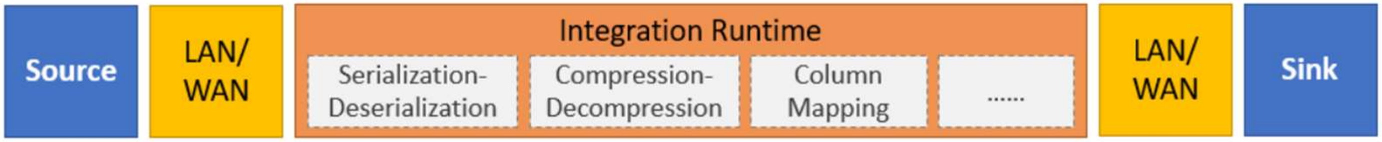
- When you're copying data between two data stores that are publicly accessible through the internet from any IP, you can use the **Azure integration runtime** for the copy activity. This integration runtime is secure, reliable, scalable, and globally available.
- When you're copying data to and from data stores that are located on-premises or in a network with access control (for example, an Azure virtual network), you need to set up a **self-hosted integration runtime**.

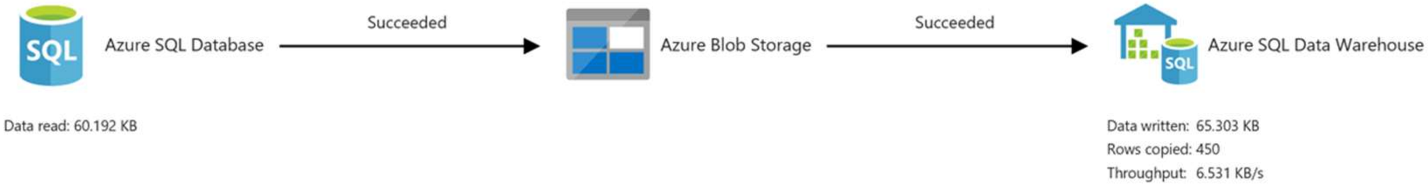
Azure Data Factory Concepts

Data Factory Copy Activity

To copy data from a source to a sink, the service that runs the Copy activity performs these steps:

- **Reads data from a source data store.**
- **Performs serialization/deserialization, compression/decompression, column mapping, and so on**. It performs these operations based on the configuration of the input dataset, output dataset, and Copy activity.
- **Writes data to the sink/destination data store.**





Data read: 60.192 KB

Data written: 65.303 KB
Rows copied: 450
Throughput: 6.531 KB/s

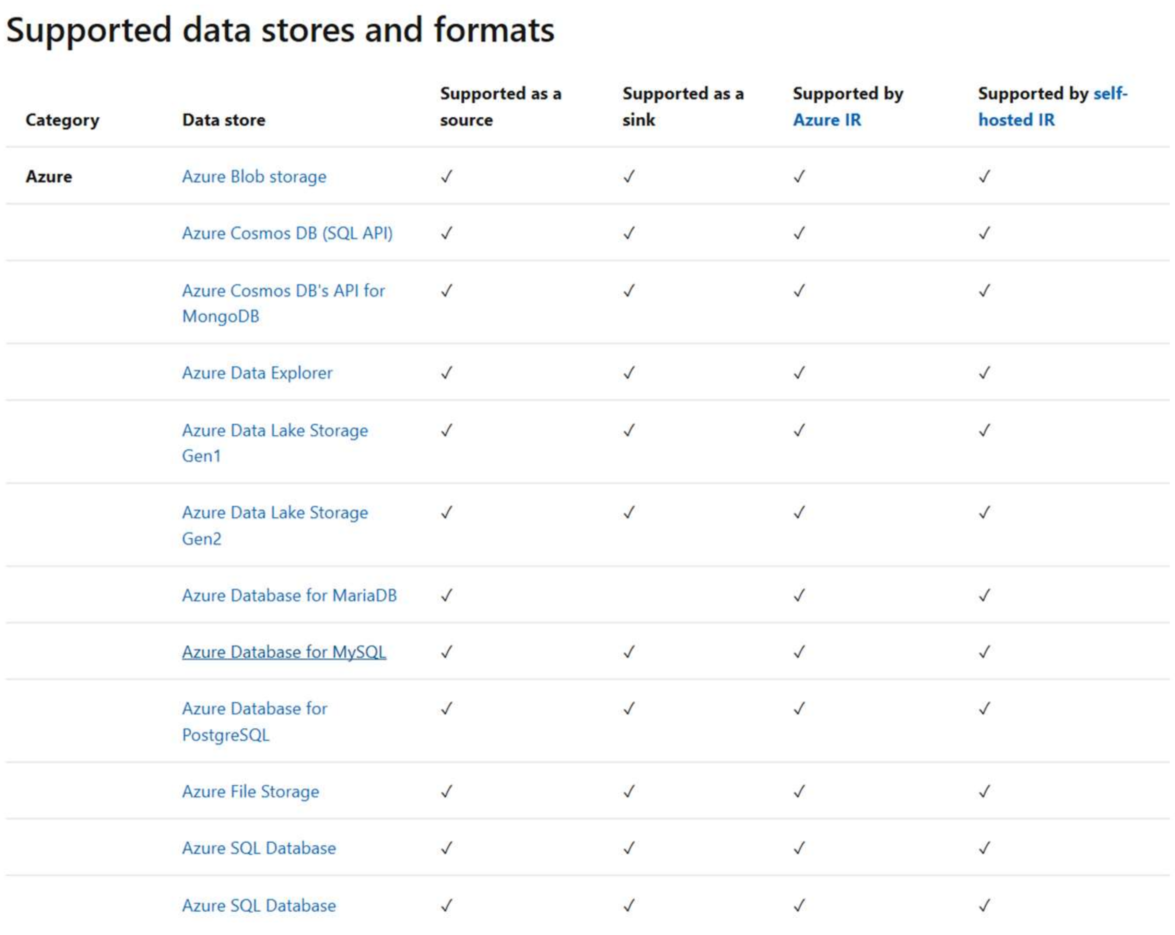
Azure Data Factory Concepts

Copy Activity - Supported file formats

You can use the Copy activity to copy files as is between two file-based data stores. In this case, the data is copied efficiently without any serialization or deserialization.

Azure Data Factory support the following file formats. Refer to each article on format-based settings.

- Avro format
- Binary format
- Delimited text format
- JSON format
- ORC format
- Parquet format



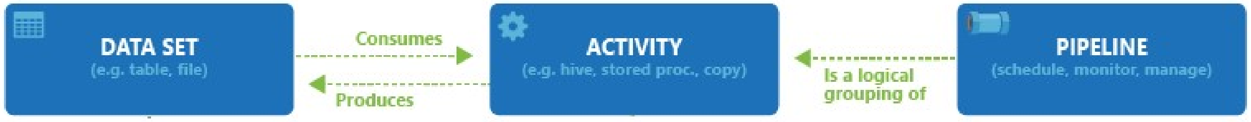
Supported data stores and formats

Category	Data store	Supported as a source	Supported as a sink	Supported by Azure IR	Supported by self-hosted IR
Azure	Azure Blob storage	✓	✓	✓	✓
	Azure Cosmos DB (SQL API)	✓	✓	✓	✓
	Azure Cosmos DB's API for MongoDB	✓	✓	✓	✓
	Azure Data Explorer	✓	✓	✓	✓
	Azure Data Lake Storage Gen1	✓	✓	✓	✓
	Azure Data Lake Storage Gen2	✓	✓	✓	✓
	Azure Database for MariaDB	✓		✓	✓
	Azure Database for MySQL	✓	✓	✓	✓
	Azure Database for PostgreSQL	✓	✓	✓	✓
	Azure File Storage	✓	✓	✓	✓
	Azure SQL Database	✓	✓	✓	✓
	Azure SQL Database	✓	✓	✓	✓

Azure Data Factory Concepts

Data Factory Pipeline, Activities Datasets and Linked Services

- A data factory can have one or more pipelines.
- **A pipeline is a logical grouping of activities that together perform a task**. For example, a pipeline could contain a set of activities that ingest and clean log data, and then kick off a mapping data flow to analyze the log data. The pipeline allows you to manage the activities as a set instead of each one individually. You deploy and schedule the pipeline instead of the activities independently.
- **The activities in a pipeline define actions to perform on your data**. For example, you may use a copy activity to copy data from an on-premises SQL Server to an Azure Blob Storage. Then, use a data flow activity or a Databricks Notebook activity to process and transform data from the blob storage to an Azure Synapse Analytics pool on top of which business intelligence reporting solutions are built.



Azure Data Factory Concepts

Data Factory Pipeline, Activities Datasets and Linked Services

- **Data Factory has three groupings of activities: data movement activities, data transformation activities, and control activities**. An activity can take zero or more input datasets and produce one or more output datasets.
  - **Data movement activities:** Copy Activity in Data Factory copies data from a source data store to a sink data store. Data Factory supports the data stores listed in https://docs.microsoft.com/en-us/azure/data-factory/concepts-pipelines-activities. Data from any source can be written to any sink.
  - **Data transformation activities:** Azure Data Factory supports a great set of transformation activities that can be added to pipelines either individually or chained with another activity. (ex. Azure Function, Spark , Databricks Notebook, etc)
  - **Control flow activities:** Azure Data Factory supports a great set of control flow activities that can be added to pipelines (ex. For Each, Get Metadata, etc)

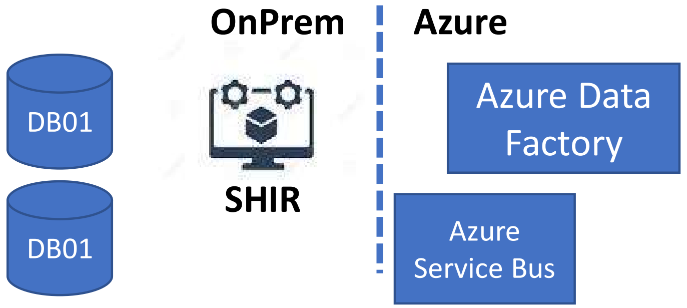
- An **input dataset** represents the input for an activity in the pipeline and an **output dataset** represents the output for the activity. **Datasets identify data within different data stores, such as tables, files, folders, and documents.** After you create a dataset, you can use it with activities in a pipeline.

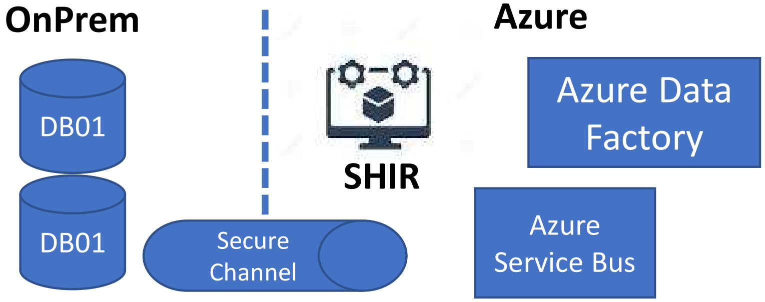
Azure Data Factory Concepts

Data Factory Self-Hosted Integration Runtime

- The Integration Runtime is a **customer managed data integration infrastructure used by Azure Data Factory to provide data integration capabilities across different network environments**. It was formerly called as Data Management Gateway.

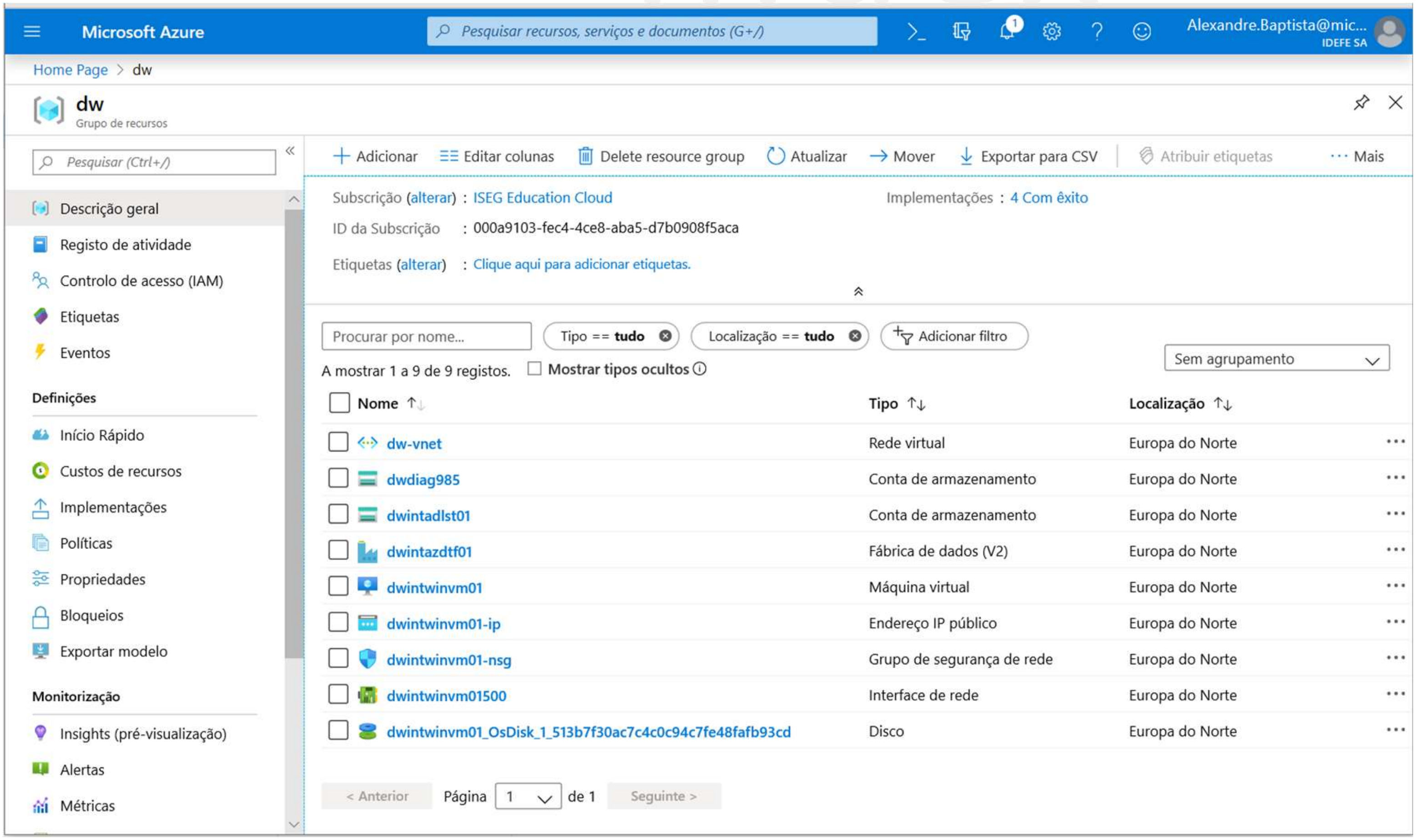
- The integration runtime is **capable of moving data in and out of data stores within private network**, as well as dispatching activities against compute service within private network. **You can install a self-hosted integration runtime on an on-premises machine or a virtual machine inside a private network**. This was formerly called the Data Management Gateway (DMG) and is fully backward compatible. Note: An Integration Runtime instance can be registered with only one of the versions of Azure Data Factory (version 1 -GA or version 2 -GA).





Executive Education

Azure Portal Resources – DW Resource Group


Microsoft Azure
Pesquisar recursos, serviços e documentos (G+/)
Alexandre.Baptista@mic...
IDEFE SA

Home Page > dw

dw
Grupo de recursos

- Descrição geral
- Registo de atividade
- Controlo de acesso (IAM)
- Etiquetas
- Eventos

Definições

- Início Rápido
- Custos de recursos
- Implementações
- Políticas
- Propriedades
- Bloqueios
- Exportar modelo

Monitorização

- Insights (pré-visualização)
- Alertas
- Métricas

Subscrição (alterar) : ISEG Education Cloud

ID da Subscrição : 000a9103-fec4-4ce8-aba5-d7b0908f5aca

Etiquetas (alterar) : Clique aqui para adicionar etiquetas.

Implementações : 4 Com êxito

Procurar por nome... | Tipo == **tudo** | Localização == **tudo** | Adicionar filtro

Sem agrupamento

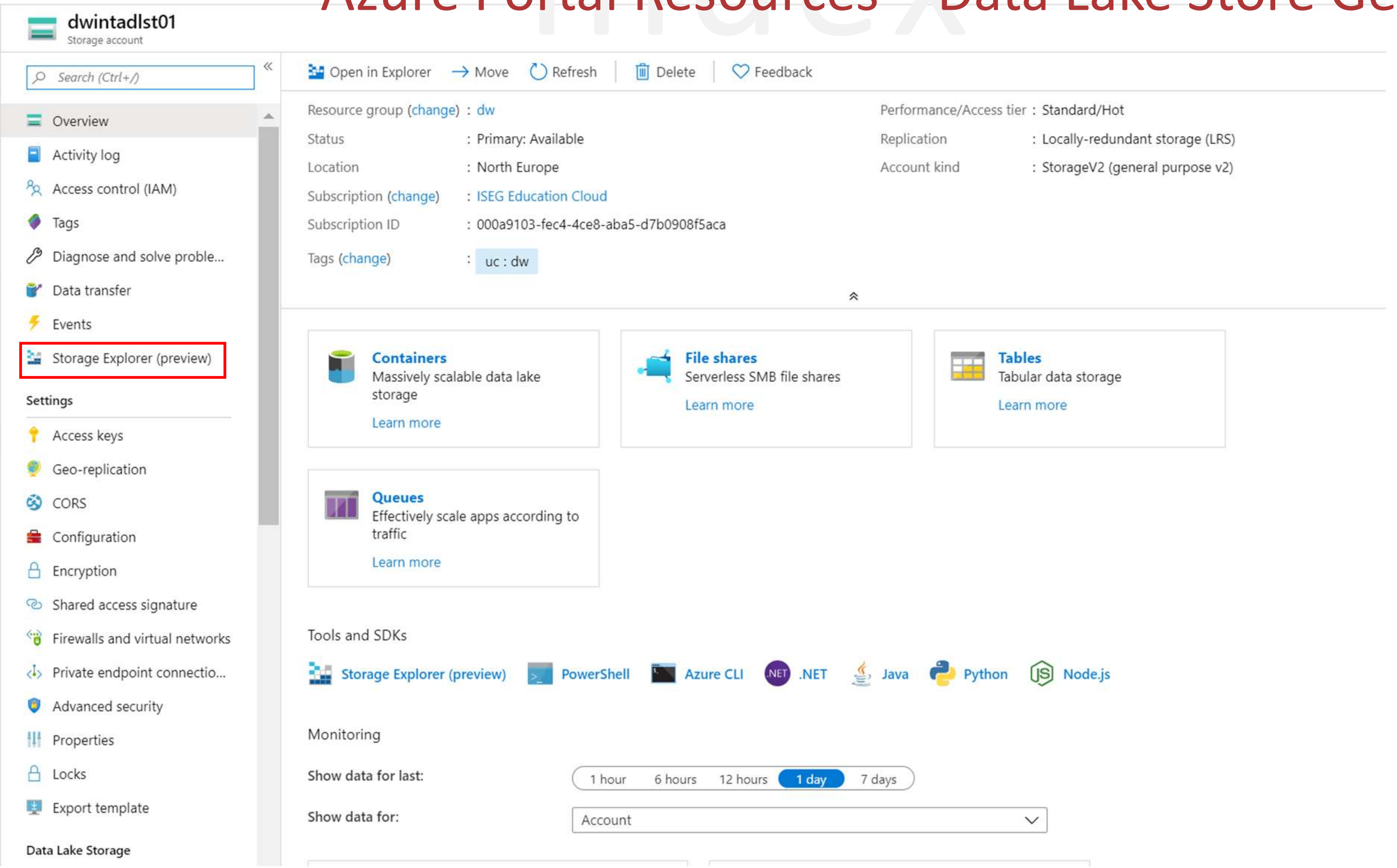
A mostrar 1 a 9 de 9 registos. □ Mostrar tipos ocultos ⓘ

□ Nome ↑↓	Tipo ↑↓	Localização ↑↓	
□ dw-vnet	Rede virtual	Europa do Norte	···
□ dwdiag985	Conta de armazenamento	Europa do Norte	···
□ dwintadlst01	Conta de armazenamento	Europa do Norte	···
□ dwintazdtf01	Fábrica de dados (V2)	Europa do Norte	···
□ dwintwinvm01	Máquina virtual	Europa do Norte	···
□ dwintwinvm01-ip	Endereço IP público	Europa do Norte	···
□ dwintwinvm01-nsg	Grupo de segurança de rede	Europa do Norte	···
□ dwintwinvm01500	Interface de rede	Europa do Norte	···
□ dwintwinvm01_OsDisk_1_513b7f30ac7c4c0c94c7fe48fafb93cd	Disco	Europa do Norte	···

< Anterior   Página 1 de 1   Seguinte >

- Azure DataLake Store gen 2
- Azure Data Factory
- Self Hosted Virtual Machine
  - Disk
  - Network
  - Storage
- SQL Azure Operational Db
- SQL Azure Data Warehouse DB

Azure Portal Resources – Data Lake Store Gen 2



dwintadlst01
Storage account

Search (Ctrl+/)

Open in Explorer · Move · Refresh · Delete · Feedback

- Overview
- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve proble...
- Data transfer
- Events
- Storage Explorer (preview)

Settings

- Access keys
- Geo-replication
- CORS
- Configuration
- Encryption
- Shared access signature
- Firewalls and virtual networks
- Private endpoint connectio...
- Advanced security
- Properties
- Locks
- Export template

Data Lake Storage

Resource group (change)	: dw	Performance/Access tier	: Standard/Hot
Status	: Primary: Available	Replication	: Locally-redundant storage (LRS)
Location	: North Europe	Account kind	: StorageV2 (general purpose v2)
Subscription (change)	: ISEG Education Cloud		
Subscription ID	: 000a9103-fec4-4ce8-aba5-d7b0908f5aca		
Tags (change)	: uc : dw		

Containers
Massively scalable data lake storage
Learn more

File shares
Serverless SMB file shares
Learn more

Tables
Tabular data storage
Learn more

Queues
Effectively scale apps according to traffic
Learn more

Tools and SDKs

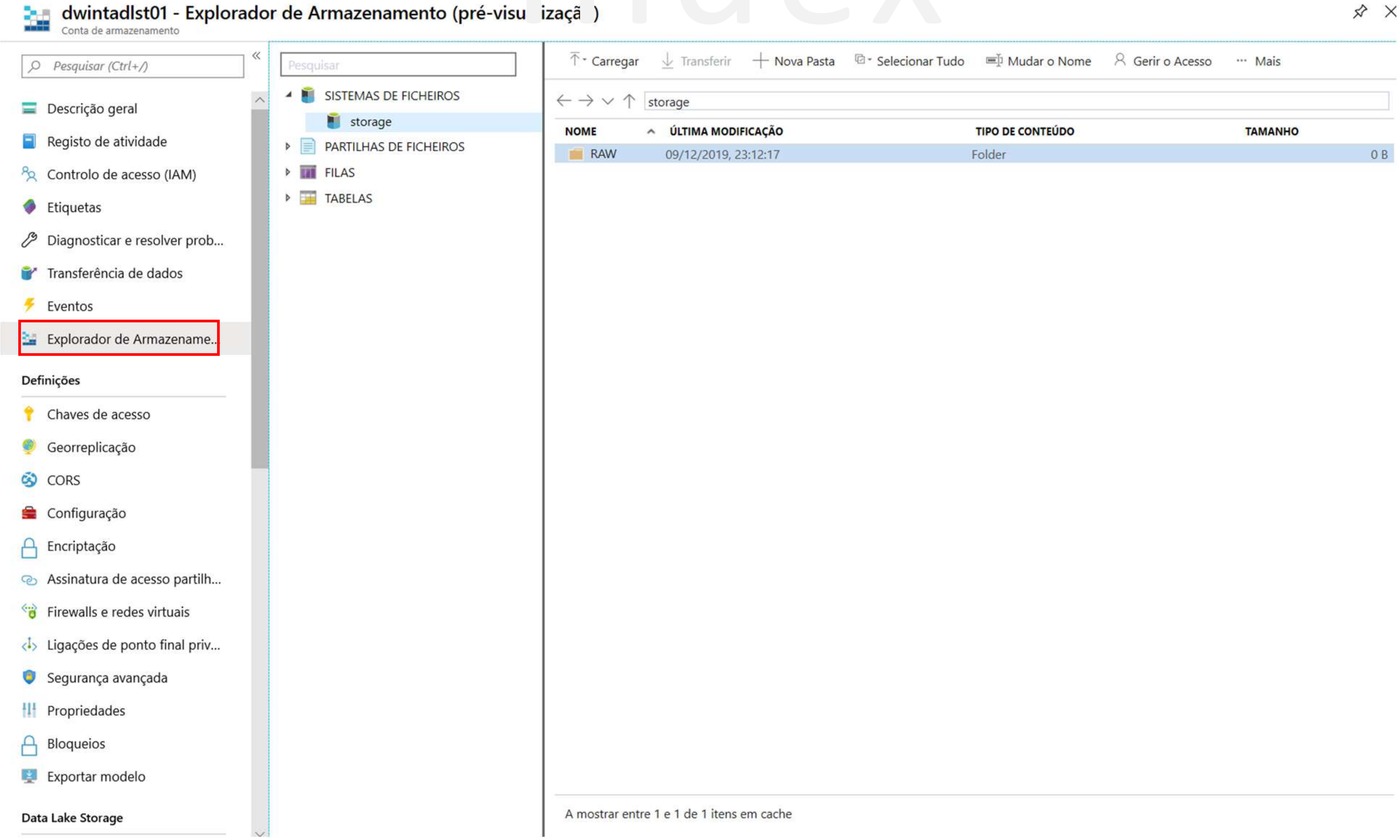
Storage Explorer (preview) · PowerShell · Azure CLI · .NET · Java · Python · Node.js

Monitoring

Show data for last: 1 hour · 6 hours · 12 hours · [✓ 1 day] · 7 days

Show data for: Account

Azure Portal Resources – Data Lake Store Gen 2



dwintadlst01 - Explorador de Armazenamento (pré-visu izaçã)
Conta de armazenamento

Pesquisar (Ctrl+/)

- Descrição geral
- Registo de atividade
- Controlo de acesso (IAM)
- Etiquetas
- Diagnosticar e resolver prob...
- Transferência de dados
- Eventos
- Explorador de Armazename...

Definições

- Chaves de acesso
- Georreplicação
- CORS
- Configuração
- Encriptação
- Assinatura de acesso partilh...
- Firewalls e redes virtuais
- Ligações de ponto final priv...
- Segurança avançada
- Propriedades
- Bloqueios
- Exportar modelo

Data Lake Storage

Pesquisar

- SISTEMAS DE FICHEIROS
  - storage
- PARTILHAS DE FICHEIROS
- FILAS
- TABELAS

storage

NOME	ÚLTIMA MODIFICAÇÃO	TIPO DE CONTEÚDO	TAMANHO
RAW	09/12/2019, 23:12:17	Folder	0 B

A mostrar entre 1 e 1 de 1 itens em cache

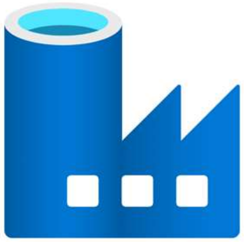
Executive Education

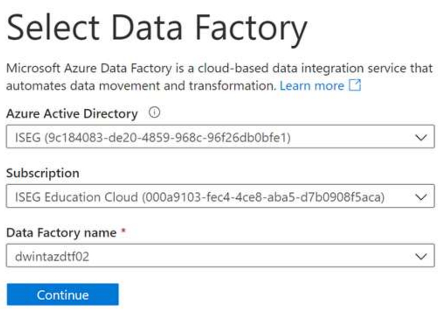
Navigate to the Azure Data Factory

In this task, you will Navigate to the Azure Data Factory.

1. Go to https://adf.azure.com and select your new Azure Data Factory:
2. In Azure Active Directory choose **IDEFE SA**
3. In Subscriptions Choose **--subscription--**
4. In Data Factory Name choose **--adf--**
5. Press the **Continue Button**
6. You will see the let's get started page

Azure Data Factory UI features currently only support the Microsoft Edge e Google Chrome. Make sure you are using one of this browsers





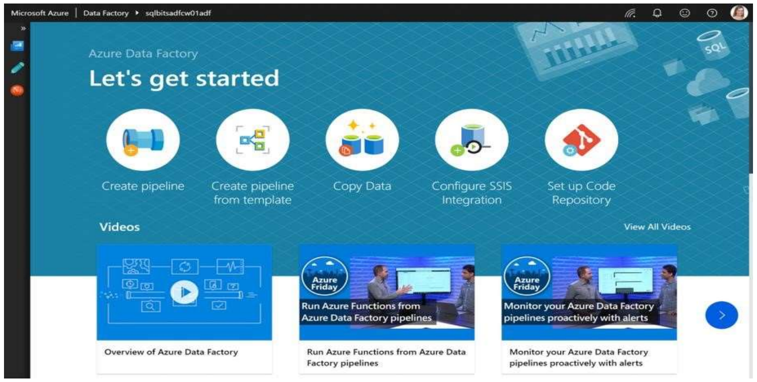
Select Data Factory

Microsoft Azure Data Factory is a cloud-based data integration service that automates data movement and transformation. Learn more

Azure Active Directory: ISEG (9c184083-de20-4859-968c-96f26db0bfe1)

Subscription: ISEG Education Cloud (000a9103-fec4-4ce8-aba5-d7b0908f5aca)

Data Factory name *: dwintazdtf02



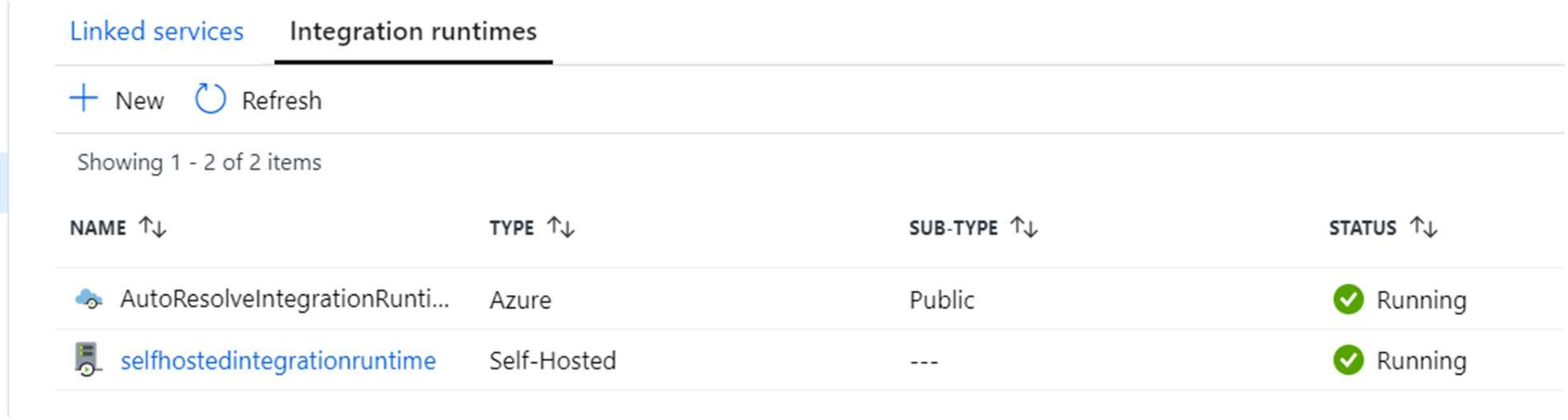
Microsoft Azure | Data Factory ▸ sqlbitsadfcw01adf
Azure Data Factory
Let's get started
Create pipeline
Create pipeline from template
Copy Data
Configure SSIS Integration
Set up Code Repository
Videos
View All Videos
Overview of Azure Data Factory
Run Azure Functions from Azure Data Factory pipelines
Monitor your Azure Data Factory pipelines proactively with alerts

Navigate to the Azure Data Factory

In this task, you will validate that the Self-hosted integration runtime is set and ready to use

1. Click on the **Manage** button in the left panel
2. In the Connections tab, click on **Integration Runtime**,
3. Validate that an Azure IR is set and that the actual status is running

if a Self-Hosted IR is also running:



Showing 1 - 2 of 2 items

NAME ↑↓	TYPE ↑↓	SUB-TYPE ↑↓	STATUS ↑↓
AutoResolveIntegrationRunti...	Azure	Public	✓ Running
selfhostedintegrationruntime	Self-Hosted	---	✓ Running

ADF Copy Activity – Configure Pipeline

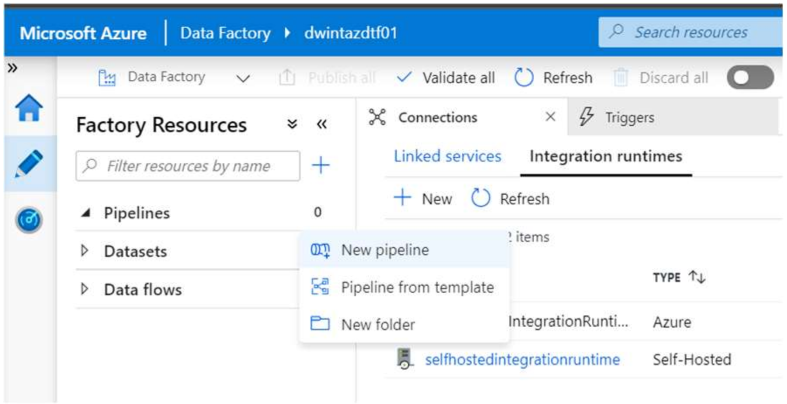
In this task, you will create your first Pipeline

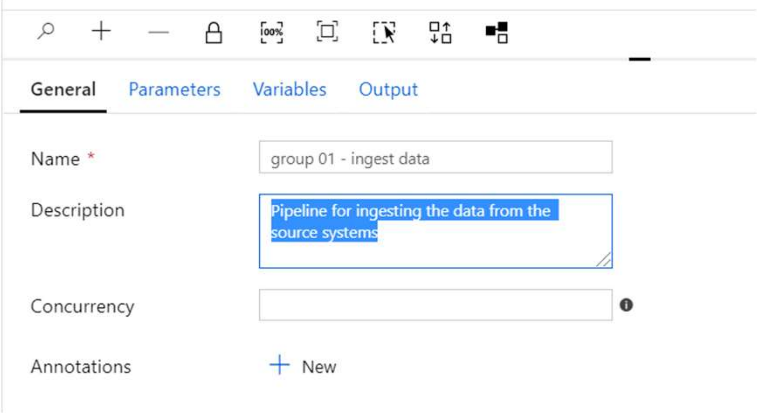
1. On the left pane click on the three dots **...** Right to the pipeline section and select **New Pipeline - if necessary create your personal working folder to work with /ixxxxxx**

2. On the General TAB set the pipeline name property to **iXXXXXX - ingest data**

(names on images may be different)

3. In the Description field insert the value **Pipeline for ingesting the data from the source systems**





Name * group 01 - ingest data

Description Pipeline for ingesting the data from the source systems

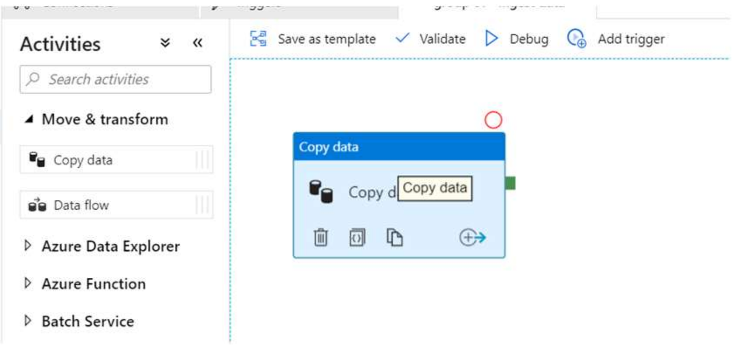
Concurrency

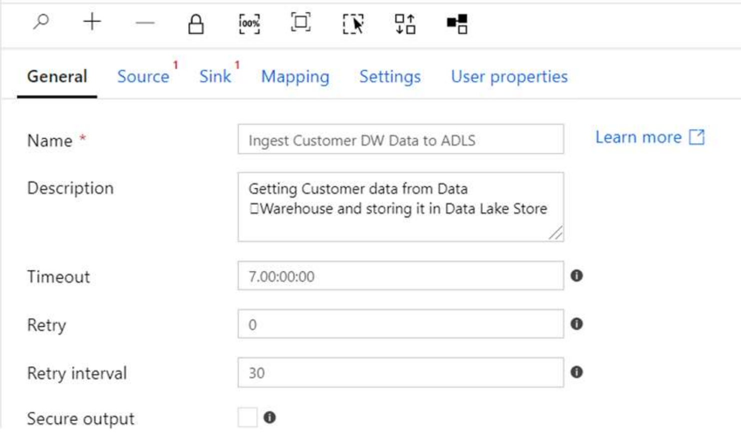
ADF Copy Activity – Configure Pipeline

In this task, you will add activities to the Pipeline created before

1. In the activities list open the **Move & Transform Section**
2. Drag the **Copy Data Activity** into the **Main Pane on the Right**
3. **Select the Copy data Activity** on the main pane
4. Make sure the **General Tab Window** is selected
5. On the **Name Property** set the value **Ingest Customer DW Data to ADLS**
6. On the **Description Property** set the value **Getting Customer data from Data Warehouse and storing it in Data Lake Store**

The Copy activities allow





Name * Ingest Customer DW Data to ADLS

Description Getting Customer data from Data
Warehouse and storing it in Data Lake Store

Timeout 7.00:00:00

Retry 0

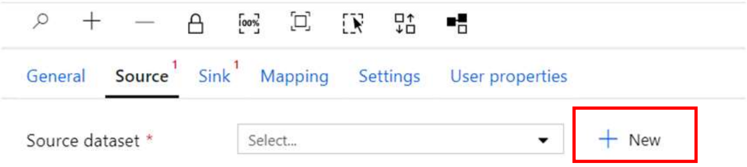
Retry interval 30

Secure output □

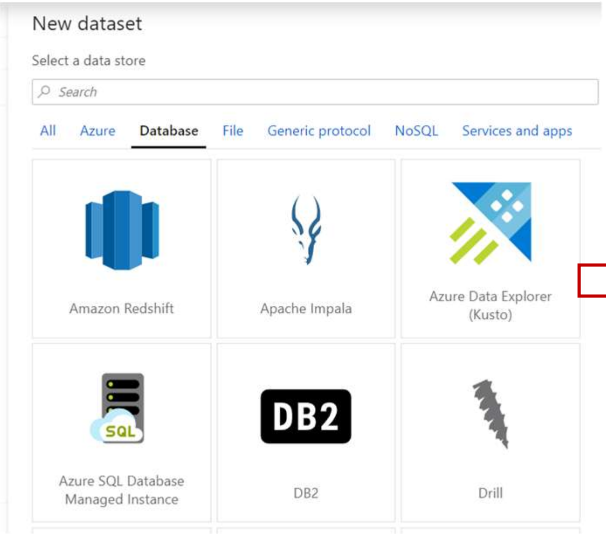
ADF Copy Activity – Configure Source

In this task, you will set the Source settings for Copy Activity

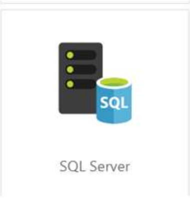
1. Choose the **Source TAB Window**
2. Press **New** button next to the Source Dataset Drop Down List
3. In the New Dataset Window select the **Database Tab**
4. Select the **SQL Server** Dataset and Select **Continue**
5. In the Set Properties Window insert **ixxxx_DS_Operational_SQLDB_Customers**
6. In the Linked Service Drop Down List select **New** option

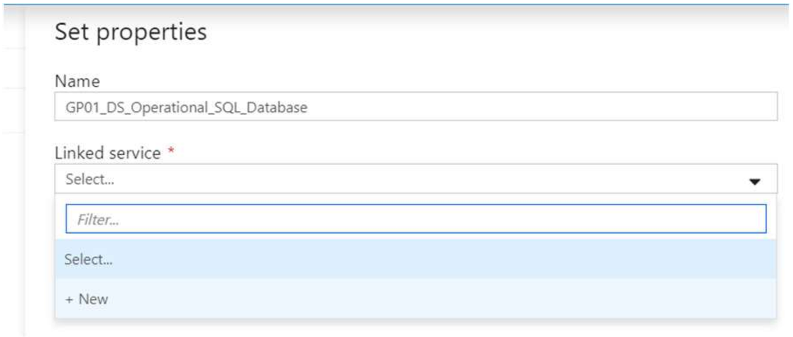


Source dataset * Select...


New dataset

Select a data store




Set properties

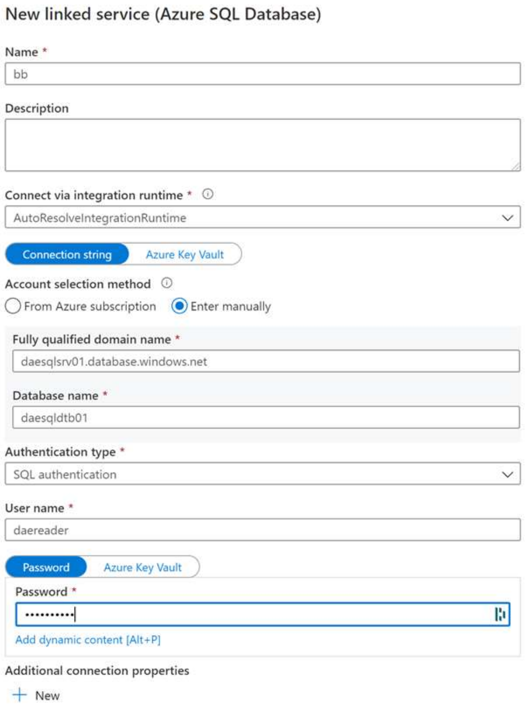
Name: GP01_DS_Operational_SQL_Database

Linked service * Select...

ADF Copy Activity – Configure Source

In this task, you will set the Source settings for Copy Activity

1. In the New linked Service (SQL Server Database) set Name value to **ixxxxxx_LS_Operacional_Database**
2. Set Description value to **Linked Service to fetch data from SQL Operational Database**
3. In the Connect Via Integration Runtime select **AutoResolveIntegrationRuntimee**
4. In servername field type **--sqlserver operacional--**
5. In Database field name type **--db operacional--**
6. Make sure the authentication is set to **SQL Authentication**
7. In the field username type **--user db operacional--**
8. Set **--psw db operacional--** As the Password field value
9. Click on the **Create** button



New linked service (Azure SQL Database)

Name *

bb

Description

Connect via integration runtime *

AutoResolveIntegrationRuntime

Account selection method

○ From Azure subscription ● Enter manually

Fully qualified domain name *

daesqlsrv01.database.windows.net

Database name *

daesqldtb01

Authentication type *

SQL authentication

User name *

daereader

Password *

••••••••••

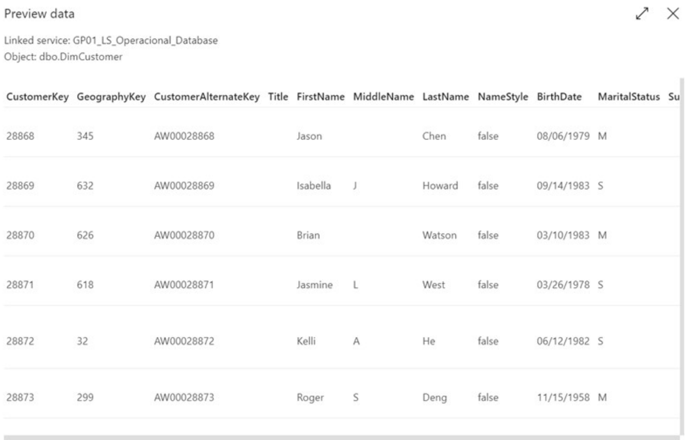
Additional connection properties

Executive Education

ADF Copy Activity – Configure Source

In this task, you will set the Source settings for Copy Activity

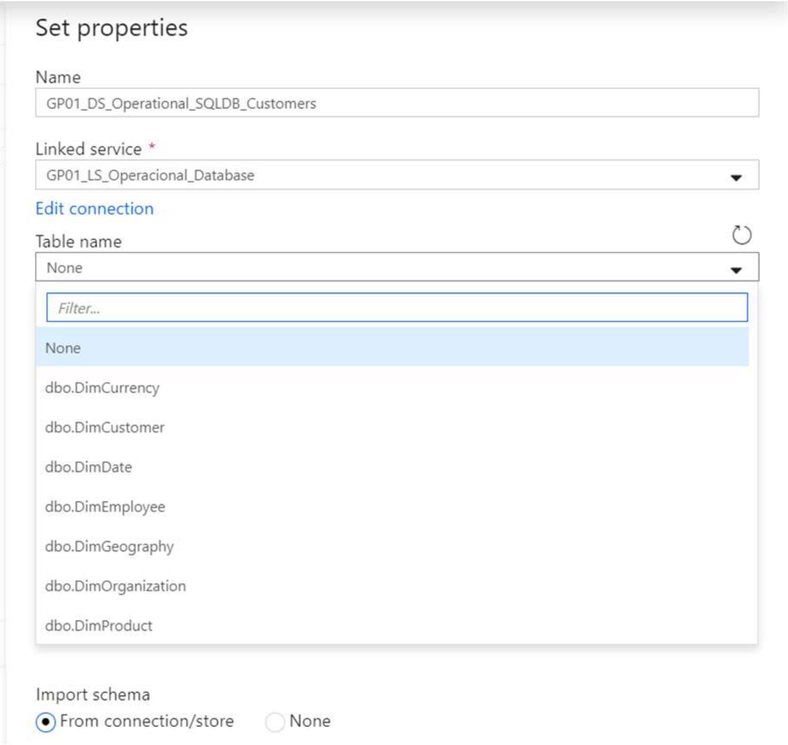
1. In the Table name select **dbo.DimCustomer**
2. Make sure the option **From Connection/store** is selected in Import Schema field
3. Press the **OK** button
4. In the Source Tab press the **Preview Data** button and ensure data is displayed
5. Close the **preview data window**



Preview data

Linked service: GP01_LS_Operacional_Database
Object: dbo.DimCustomer

CustomerKey	GeographyKey	CustomerAlternateKey	Title	FirstName	MiddleName	LastName	NameStyle	BirthDate	MaritalStatus	Su
28868	345	AW00028868		Jason		Chen	false	08/06/1979	M	
28869	632	AW00028869		Isabella	J	Howard	false	09/14/1983	S	
28870	626	AW00028870		Brian		Watson	false	03/10/1983	M	
28871	618	AW00028871		Jasmine	L	West	false	03/26/1978	S	
28872	32	AW00028872		Kelli	A	He	false	06/12/1982	S	
28873	299	AW00028873		Roger	S	Deng	false	11/15/1958	M	



Set properties

Name: GP01_DS_Operational_SQLDB_Customers

Linked service *: GP01_LS_Operacional_Database

Edit connection

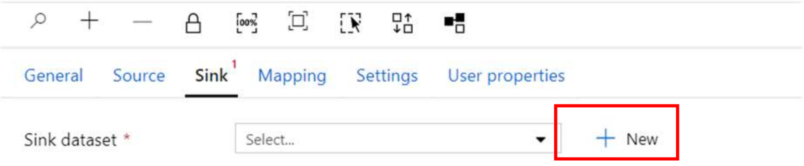
Table name: None

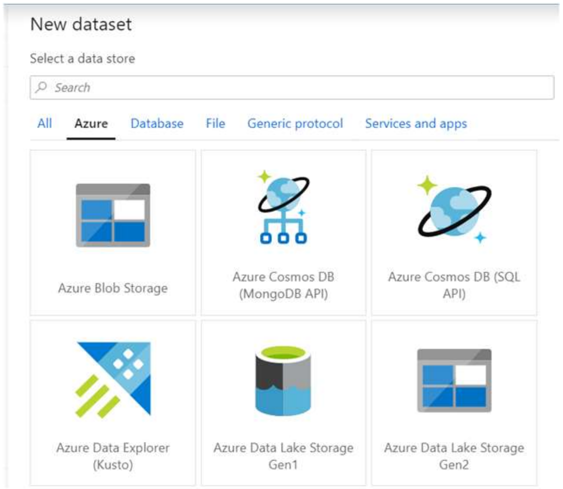
Import schema: ● From connection/store ○ None

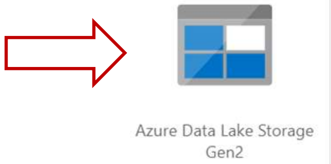
ADF Copy Activity – Configure Sink

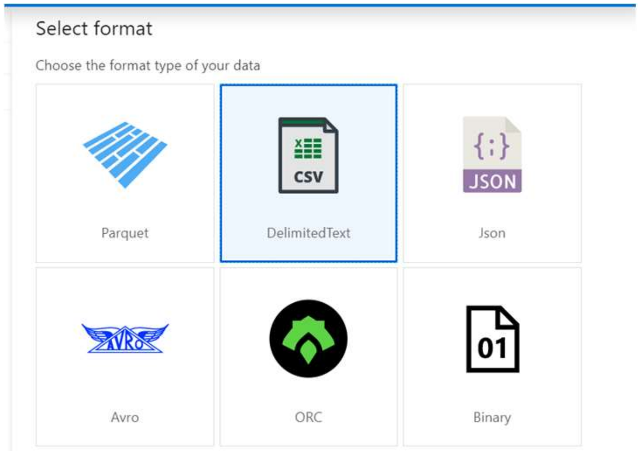
In this task, you will set the Sink settings for Copy Activity

1. Choose the **Sink TAB Window**
2. Press **New** button next to the Sink Dataset Drop Down List
3. In the New Dataset Window select the **Azure Tab**
4. Select the **Azure Data Lake Storage Gen 2** and Select **Continue**
5. In the select format Window select **delimited**
6. Press the **Continue** button




New dataset
Select a data store
Search
All
Azure
Database
File
Generic protocol
Services and apps
Azure Blob Storage
Azure Cosmos DB (MongoDB API)
Azure Cosmos DB (SQL API)
Azure Data Explorer (Kusto)
Azure Data Lake Storage Gen1
Azure Data Lake Storage Gen2

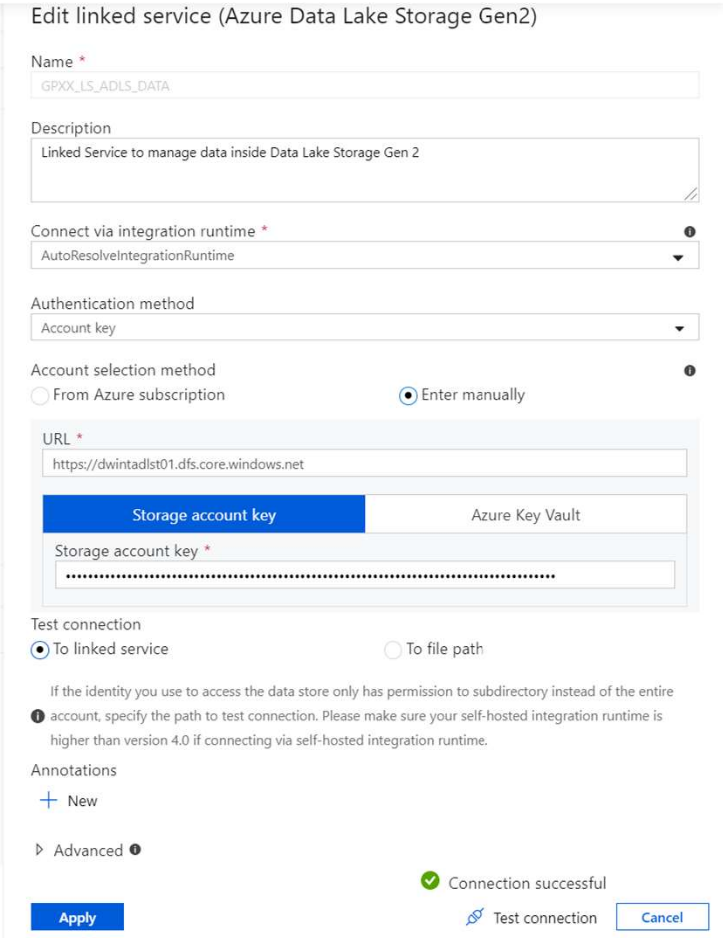




ADF Copy Activity – Configure Sink

In this task, you will set the Sinc settings for Copy Activity

1. In the set properties window set the name to **ixxxxxx_DS_RAW_ADLS_Customers**
2. In the Linked Service Drop Down List select **New** option
3. In the New linked Service (Azure Data Lake Storage Gen2) set Name value to **ixxxxxx_LS_ADLS_DATA**
4. Set Description value to **Linked Service to manage data inside Data Lake Storage Gen 2**
5. In the Connect Via Integration Runtime select **AutoResolveIntegrationRuntime**
6. In Authentication Method select **Account Key**
7. In Account Selection Method select **Enter manually**
8. In the URL set the value to **--adls--**
9. In the Storage Account Key insert **--adls key--**
10. Ensure Test Connection is set to **To Linked Service**
11. Press **Test Connection**
12. Press **Create** button



Edit linked service (Azure Data Lake Storage Gen2)

Name *
GPXX_LS_ADLS_DATA

Description
Linked Service to manage data inside Data Lake Storage Gen 2

Connect via integration runtime *
AutoResolveIntegrationRuntime

Authentication method
Account key

Account selection method
○ From Azure subscription ● Enter manually

URL *
https://dwintadlst01.dfs.core.windows.net

Storage account key *
••

Test connection
● To linked service ○ To file path

If the identity you use to access the data store only has permission to subdirectory instead of the entire account, specify the path to test connection. Please make sure your self-hosted integration runtime is higher than version 4.0 if connecting via self-hosted integration runtime.

Annotations

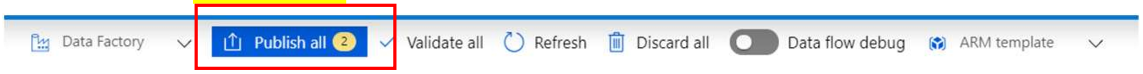
✓ Connection successful

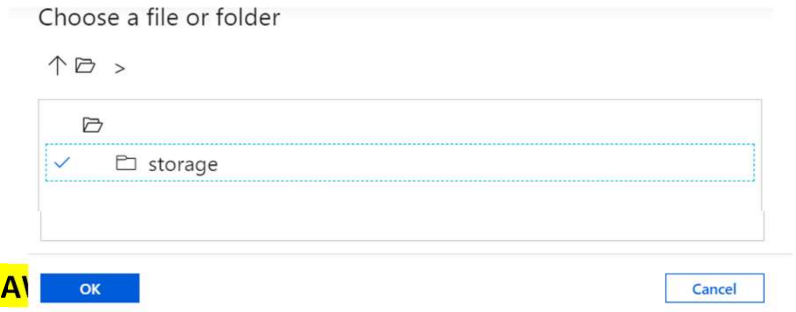
ADF Copy Activity – Configure Sink

In this task, you will set the Sink settings for Copy Activity

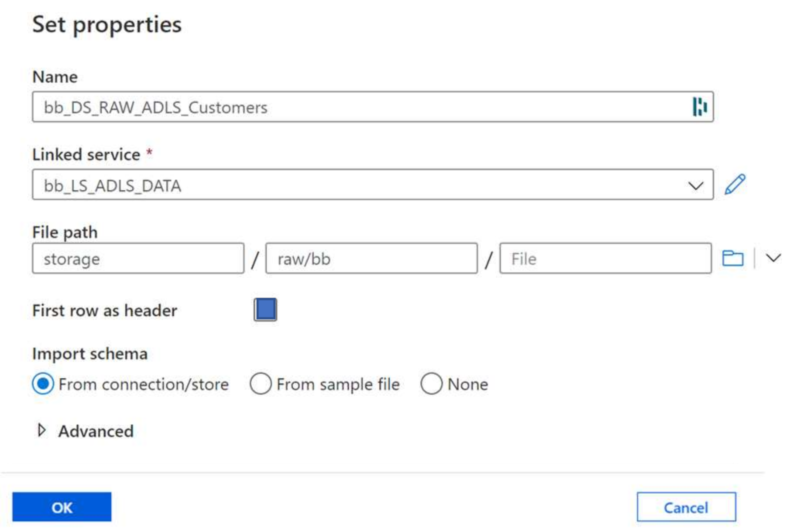
1. In the set properties press the **Browse** button
2. Select the **shared-storage-dw** Path - in the image (raw/bb) –

   path must be pre-created
3. Press the **OK** button
4. In the Directory field make sure it has the **shared-storage-dw/iXXXXXX/RA**
5. Let the **File field** empty
6. Select the **First row as header** option
7. In the Import Schema select **From Connection/store**
8. Press **OK**
9. Inside the Sink TAB choose **Open**
10. In the Connection TAB set the File Field to **dimCustomer.csv**
11. Press **Publish All** button




Choose a file or folder

✓ storage


Set properties

Name: bb_DS_RAW_ADLS_Customers

Linked service *: bb_LS_ADLS_DATA

File path: storage / raw/bb / File

First row as header ☑

Import schema: ● From connection/store ○ From sample file ○ None

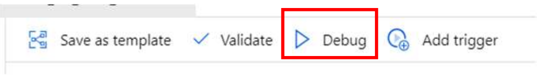
Executive Education

Execute the Azure Data Factory pipeline

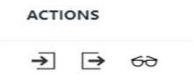
In this task, you will run the pipeline and

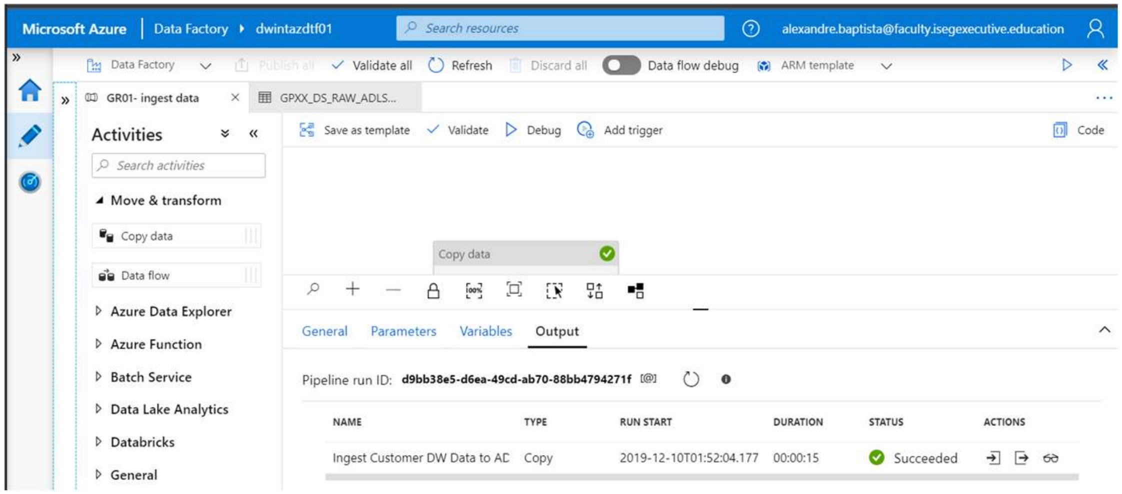
Validate the outcome

1. Open the Pipeline from the left panel
2. Press the **Debug** button and run the package



3. Make sure the **pipeline runs successfully**
4. Open the Actions buttons and explore its content





Pipeline run ID: **d9bb38e5-d6ea-49cd-ab70-88bb4794271f**

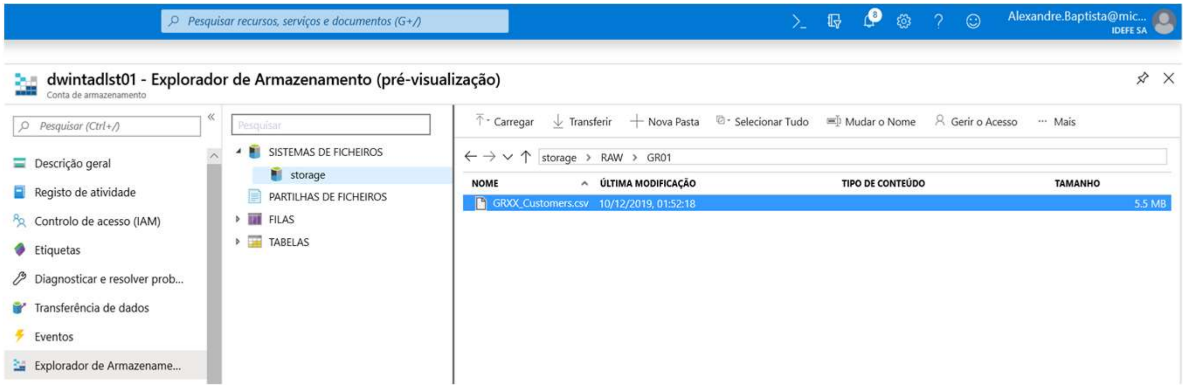
NAME	TYPE	RUN START	DURATION	STATUS	ACTIONS
Ingest Customer DW Data to AD	Copy	2019-12-10T01:52:04.177	00:00:15	✓ Succeeded	

If the pipeline execution fails, use the error action button to understand and correct the problem that occurred. If no error occurred use ghe pipeline details button mentioned above to understand pipeline's statistics (how long did it took, number of lines that were processed, etc.

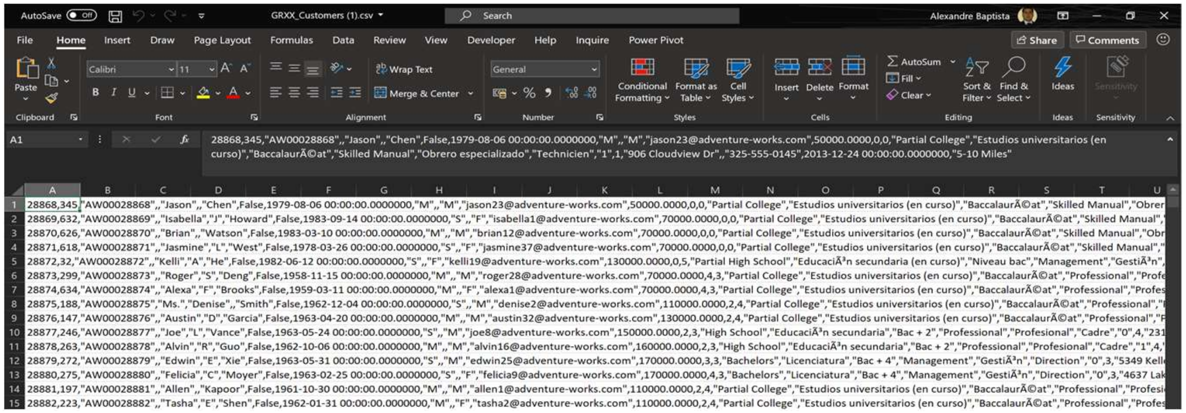
Validate pipeline execution results

In this task, you will access the datalake storage to ensure data was saved

1. In Azure Portal open the **DW Resource Group**
2. Choose the Data Lake Store **dsbafb2sharedstor**
3. On the left Panel select Storage Explorer (preview)
4. Open **FILE SYSTEM**
5. Open **Storage**
6. Open directory
7. Open the **dimCustomers.csv** file, download it and and check its content **using Excel**



NOME	ÚLTIMA MODIFICAÇÃO	TIPO DE CONTEÚDO	TAMANHO
GRXX_Customers.csv	10/12/2019, 01:52:18		5.5 MB



```
28868,345,"AW00028868",,"Jason",,"Chen",False,1979-08-06 00:00:00.0000000,"M",,"M","jason23@adventure-works.com",50000.0000,0,0,"Partial College","Estudios universitarios (en
curso)","BaccalaurÃ©at","Skilled Manual","Obrero especializado","Technicien","1",1,"906 Cloudview Dr",,"325-555-0145",2013-12-24 00:00:00.0000000,"5-10 Miles"
```

```
1  28868,345,"AW00028868",,"Jason",,"Chen",False,1979-08-06 00:00:00.0000000,"M",,"M","jason23@adventure-works.com",50000.0000,0,0,"Partial College","Estudios universitarios (en curso)","BaccalaurÃ©at","Skilled Manual","Obrer
2  28869,632,"AW00028869",,"Isabella","J","Howard",False,1983-09-14 00:00:00.0000000,"S",,"F","isabella1@adventure-works.com",70000.0000,0,0,"Partial College","Estudios universitarios (en curso)","BaccalaurÃ©at","Skilled Manual",
3  28870,626,"AW00028870",,"Brian",,"Watson",False,1983-03-10 00:00:00.0000000,"M",,"M","brian12@adventure-works.com",70000.0000,0,0,"Partial College","Estudios universitarios (en curso)","BaccalaurÃ©at","Skilled Manual","Obr
4  28871,618,"AW00028871",,"Jasmine","L","West",False,1978-03-26 00:00:00.0000000,"S",,"F","jasmine37@adventure-works.com",70000.0000,0,0,"Partial College","Estudios universitarios (en curso)","BaccalaurÃ©at","Skilled Manual",
5  28872,32,"AW00028872",,"Kelli","A","He",False,1982-06-12 00:00:00.0000000,"S",,"F","kelli19@adventure-works.com",130000.0000,0,5,"Partial High School","EducaciÃ³n secundaria (en curso)","Niveau bac","Management","GestiÃ³n",
6  28873,299,"AW00028873",,"Roger","S","Deng",False,1958-11-15 00:00:00.0000000,"M",,"M","roger28@adventure-works.com",70000.0000,4,3,"Partial College","Estudios universitarios (en curso)","BaccalaurÃ©at","Professional","Profe
7  28874,634,"AW00028874",,"Alexa","F","Brooks",False,1959-03-11 00:00:00.0000000,"M",,"F","alexa1@adventure-works.com",70000.0000,4,3,"Partial College","Estudios universitarios (en curso)","BaccalaurÃ©at","Professional","Profes
8  28875,188,"AW00028875","Ms.","Denise",,"Smith",False,1962-12-04 00:00:00.0000000,"S",,"M","denise2@adventure-works.com",110000.0000,2,4,"Partial College","Estudios universitarios (en curso)","BaccalaurÃ©at","Professional",
9  28876,147,"AW00028876",,"Austin","D","Garcia",False,1963-04-20 00:00:00.0000000,"M",,"M","austin32@adventure-works.com",130000.0000,2,4,"Partial College","Estudios universitarios (en curso)","BaccalaurÃ©at","Professional",
10 28877,246,"AW00028877",,"Joe","L","Vance",False,1963-05-24 00:00:00.0000000,"S",,"M","joe8@adventure-works.com",150000.0000,2,3,"High School","EducaciÃ³n secundaria","Bac + 2","Professional","Profesional","Cadre","0",4,"231
11 28878,263,"AW00028878",,"Alvin","R","Guo",False,1962-10-06 00:00:00.0000000,"M",,"M","alvin16@adventure-works.com",160000.0000,2,3,"High School","EducaciÃ³n secundaria","Bac + 2","Professional","Profesional","Cadre","1",4,
12 28879,272,"AW00028879",,"Edwin","E","Xie",False,1963-05-31 00:00:00.0000000,"S",,"M","edwin25@adventure-works.com",170000.0000,3,3,"Bachelors","Licenciatura","Bac + 4","Management","GestiÃ³n","Direction","0",3,"5349 Kell
13 28880,275,"AW00028880",,"Felicia","C","Moyer",False,1963-02-25 00:00:00.0000000,"S",,"F","felicia9@adventure-works.com",170000.0000,4,3,"Bachelors","Licenciatura","Bac + 4","Management","GestiÃ³n","Direction","0",3,"4637 Lak
14 28881,197,"AW00028881",,"Allen",,"Kapoor",False,1961-10-30 00:00:00.0000000,"M",,"M","allen1@adventure-works.com",110000.0000,2,4,"Partial College","Estudios universitarios (en curso)","BaccalaurÃ©at","Professional","Profesi
15 28882,223,"AW00028882",,"Tasha","E","Shen",False,1962-01-31 00:00:00.0000000,"M",,"F","tasha2@adventure-works.com",110000.0000,2,4,"Partial College","Estudios universitarios (en curso)","BaccalaurÃ©at","Professional","Profes
```

Validate pipeline execution results

In this task, you will repeat the bellow steps for all the other Tables

1. Perform the same steps for the following tables:

   - **dbo.DimCurrency**
   - **dbo.DimGeography**
   - **dbo.DimDate**
   - **dbo.DimProduct**
   - **dbo,DimProductCategory**
   - **dbo,DimProductSubCategory**
   - **dbo.FactInternetSales**

2. Validate the sucess of the ingesting these tables

Executive Education

www.isegexecutive.education
Rua do Quelhas, 6
1200-781 Lisboa

(+351) 213 922 891
info@executive.education