Pós-graduação em Data Science & Business Analytics

Executive Education

Data Warehousing

Artur Vieira

Executive Education

Índice

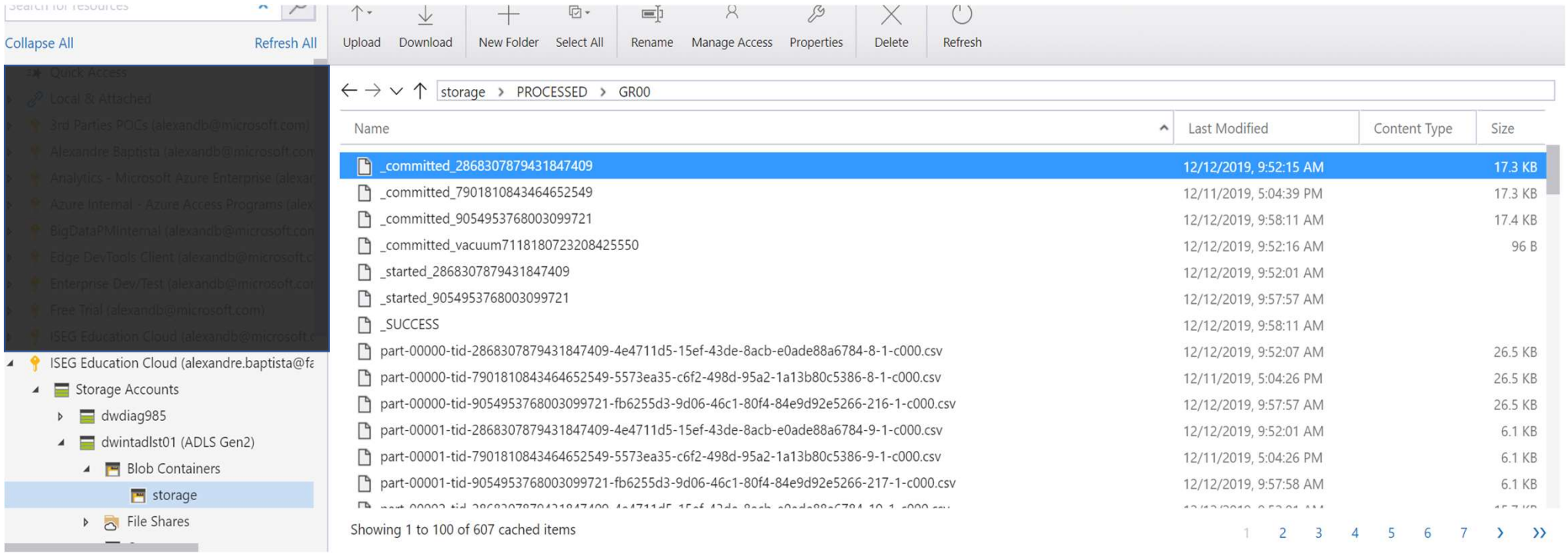
- Azure Data Factory Concepts
- Azure Data Factory Practical Experience
  - ADLS to DW process
- Final Project Q&A

Executive Education

Tópico

Azure Data Factory Concepts

Data Flow Generated Data for Customers

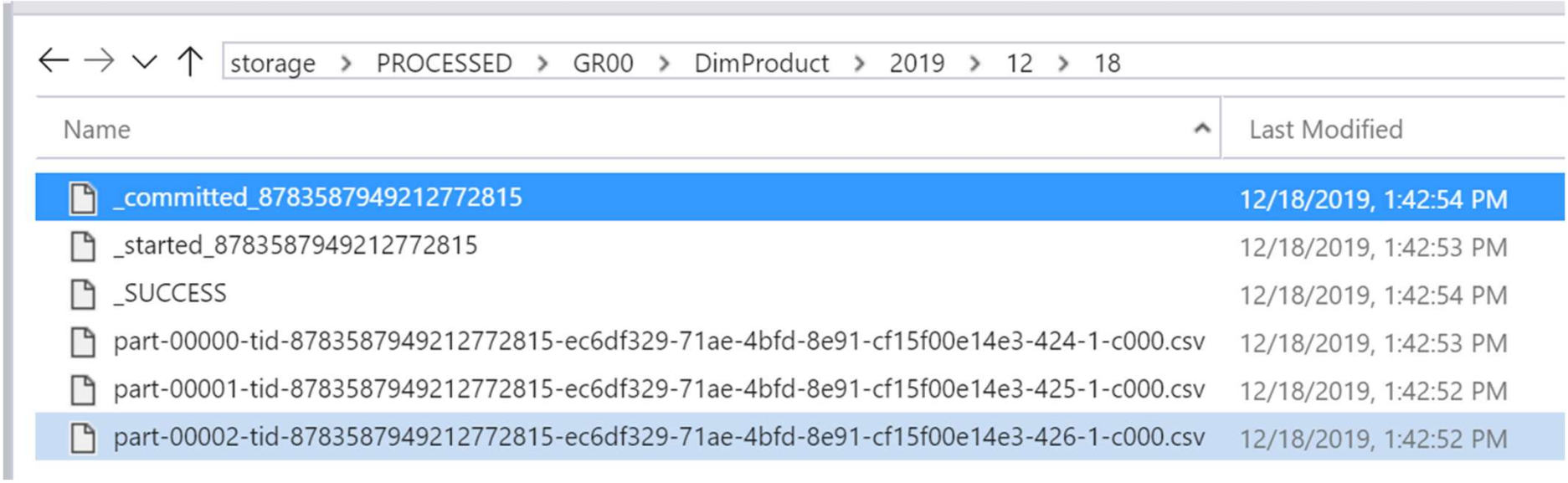


Executive Education

Tópico

Azure Data Factory Concepts

Partition by date exemple



Executive Education

Tópico

Azure Data Factory Concepts

Overall ETL Process

STORAGE

OPERATIONAL DB

AZURE DATA LAKE STORAGE

RAW | PROCESSED

Data Warehouse DB

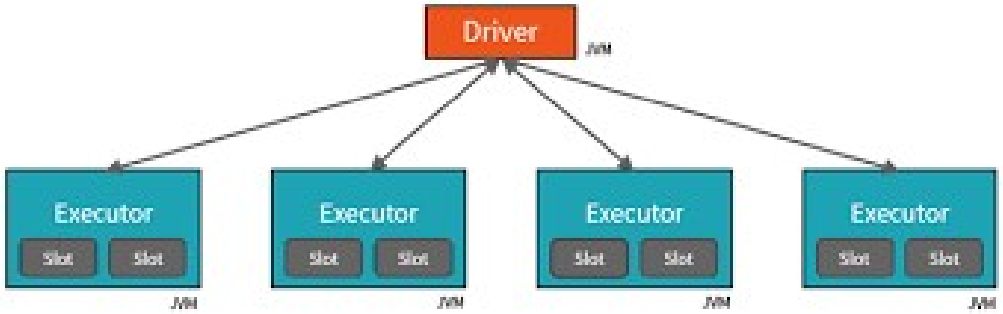
PROCESS

DATA FACTORY

1. Copy Activity INGEST
FROM -> TO

2. DATA FLOW TRANSFORM
SOURCE -> DERIVED COLUMN -> SELECT -> SORT -> SINK

3. Copy Activity EGRESS
FROM -> TO



Executive Education

Tópico

Azure Data Factory Concepts

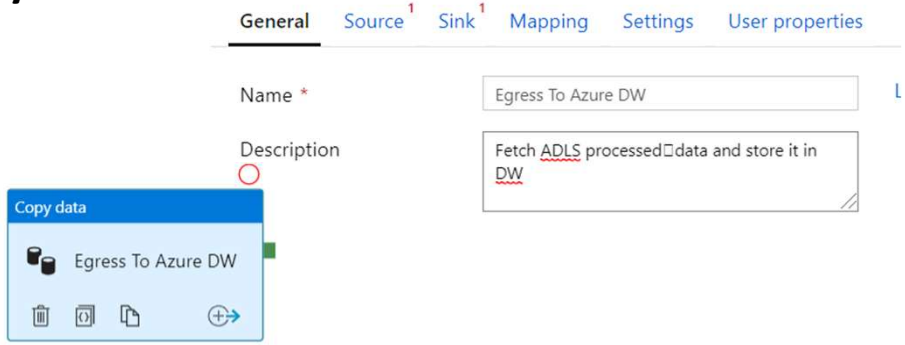
ConfigParameters

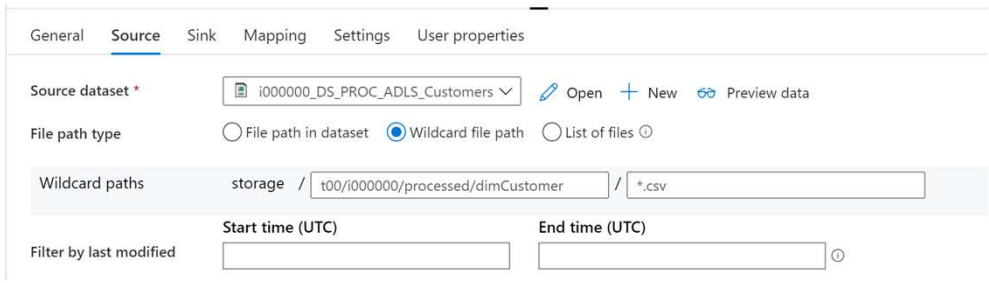
	SourceTable	DestinationTable	active	lastExecution	lastExecutionDW
1	DimCurrency	DimCurrency	1	NULL	NULL
2	DimGeography	DimGeography	1	NULL	NULL
3	DimDate	DimDate	1	NULL	NULL
4	DimProduct	DimProduct	1	NULL	NULL
5	DimProductCategory	DimProductCategory	1	NULL	NULL
6	DimProductSubCate…	DimProductSubCate…	1	NULL	NULL
7	FactInternetSales	FactInternetSales	1	NULL	NULL

ADF Copy Activity – Configure Source for DW

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

1. In the Data Factory pane pick the **…** option right to the **Pipelines** section
2. Select **New pipeline** option
3. Rename the new Data flow **IXXXXXX - Egress To Azure DW**
4. Add **Fetch ADLS processed data and store it in DW** to the description
5. Drag a **Copy Data Activity** available in the Move & Transform section
6. Set the Name for the Copy activity to **Egress Customers To Azure DW**
7. Set the Description for the Copy activity to **Fetch ADLS processed Customers data and store it in DW**
8. Navigate to the **Source TAB**
9. Select the **iXXXXXX_DS_PROC_ADLS_Customers** for the Dataset na
10. Make sure the **Recursively** option is selected
11. Set the value **i000000/processed/dimCustomer** **and** ***.csv** in the Wildcard file name and **"wildcard file path"** option

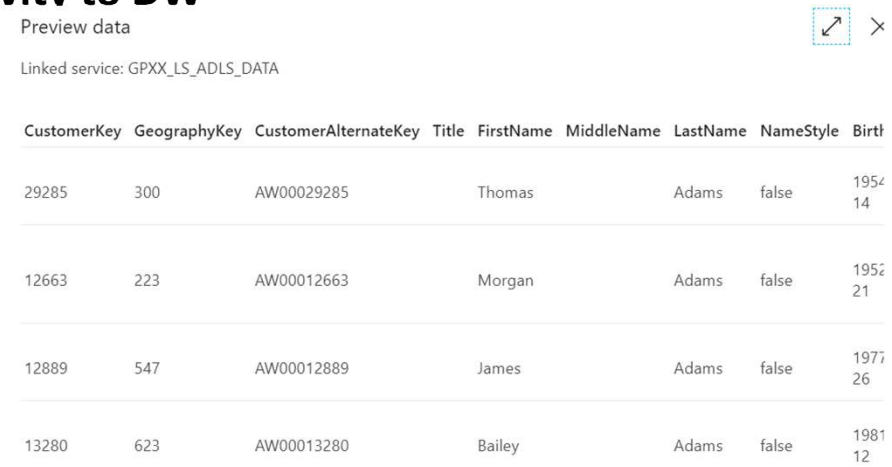


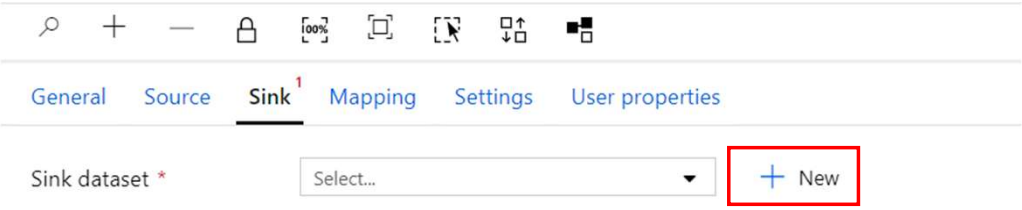


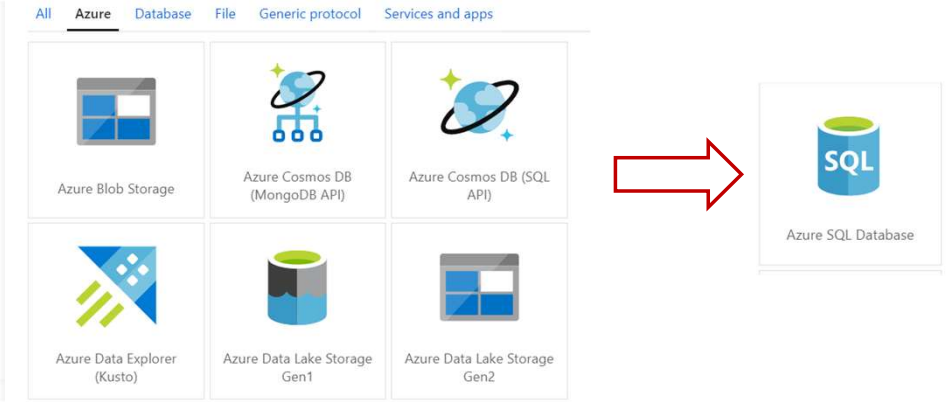
ADF Copy Activity – Configure Source for DW

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

1. Press the **Preview Data** button and ensure Customers Data Is readable and ready to be used
2. Choose the **Sink TAB Window**
3. Press **New** button next to the Sink Dataset Drop Down List
4. In the New Dataset Window select the **Azure Tab**
5. Select the **Azure SQL Database** and Select **Continue**
6. Press the **Continue** button





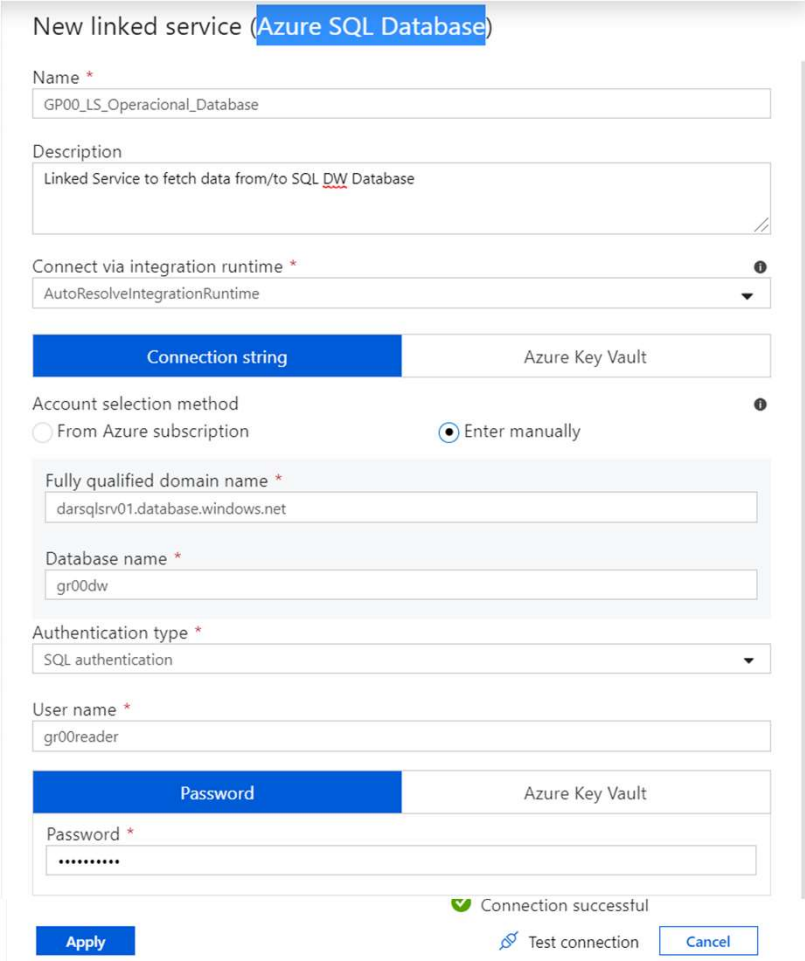


Executive Education

ADF Copy Activity – Configure Sink

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

1. In the set properties window set the name to **IXXXXXX_DS_PROC_ADW_Customers**
2. In the Linked Service Drop Down List select **New** option
3. In the New linked Service (Azure SQL Database) set Name value to **ixxxxxx_LS_DW_DATA**
4. Set Description value to **Linked Service to manage data inside Azure SQL DW Database**
5. In the Connect Via Integration Runtime select **AutoResolveIntegrationRuntime**
6. In Account Selection Method select **Enter manually**
7. In the Fully Qualified Domain Name insert --sqlserver dw--
8. Set the Database Name field value to **--db dw--** and Authentication Type to **SQL Authentication**
9. Set the username to **--user db dw--** and the Password to --psw db dw--
10. Press **Test Connection**
11. Press **Create** button



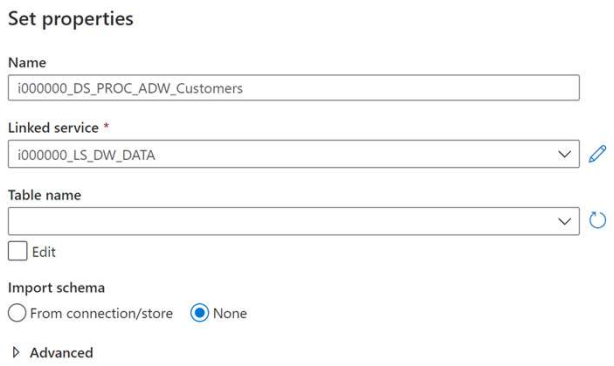
Executive Education

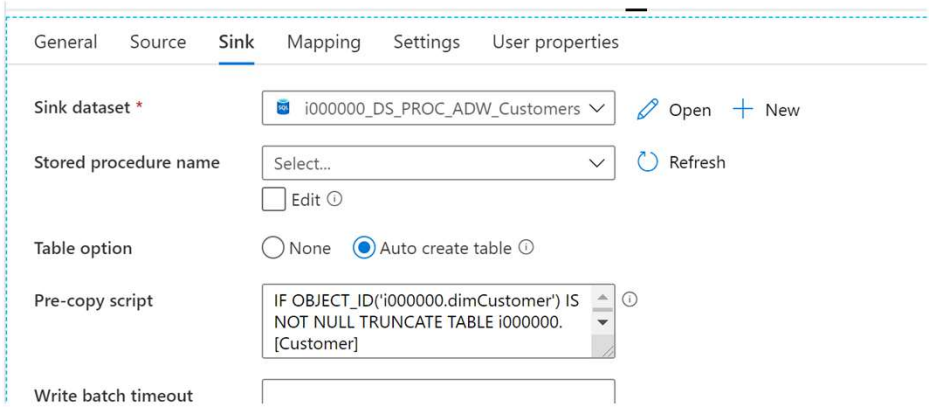
ADF Copy Activity – Configure Sink

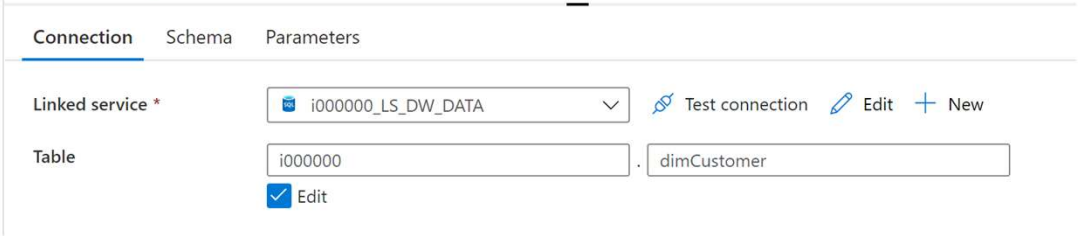
In this task, you will set the Sink settings for Copy Activity

1. Press **OK living the Table Name empty**
2. In the Sink Pane in the Store Procedure name field select **None**
3. In the Table Options choose **Auto create table**
4. Open Sink dataset and check edit chechbox and in the table textbox enter iXXXXXX and dimCustomer
5. Get back to Sink Tab
6. In the Pre-Copy Script write **IF OBJECT_ID('iXXXXXX.dimCustomer') IS NOT NULL TRUNCATE TABLE iXXXXXX.dimCustomer**

(don't forget to change XXXXXX by real number)





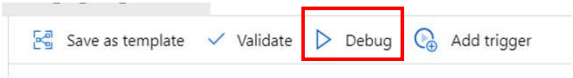


Execute the Azure Data Factory pipeline

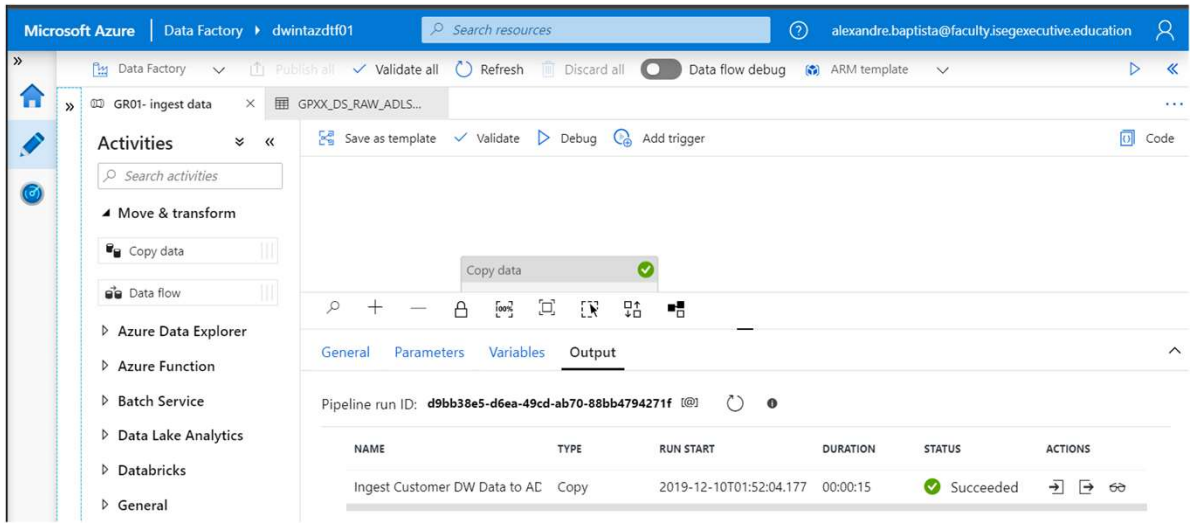
In this task, you will run the pipeline and

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



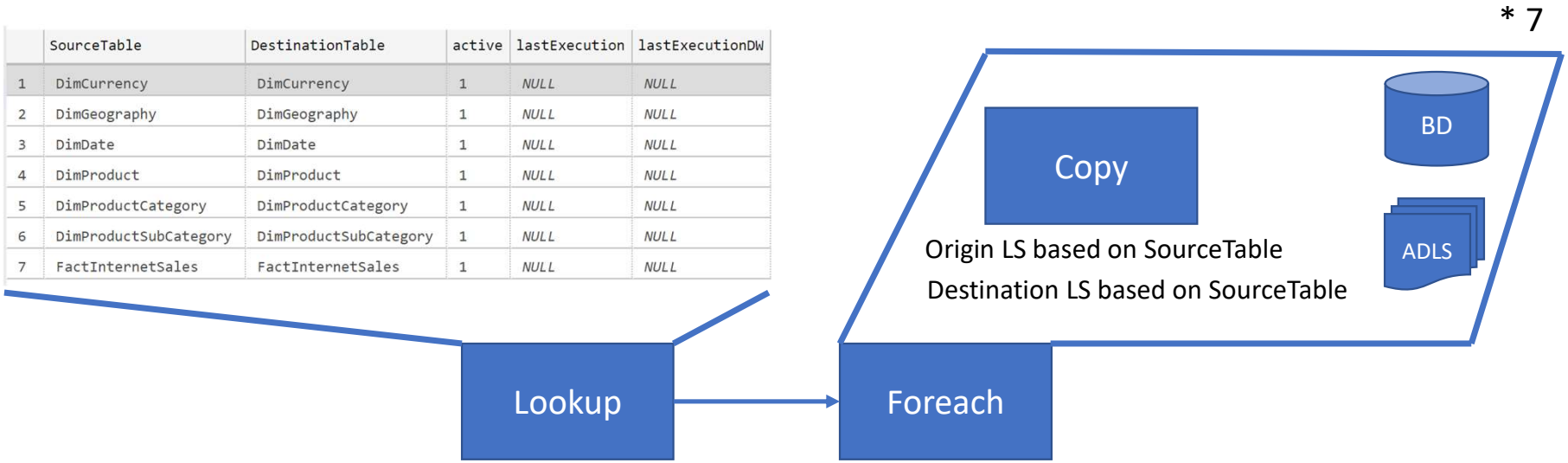
If the pipeline execution fails, use the error action button to understand and correct the problem that occurred. If no error occurred use ghe pipeline details button mentioned above to understand pipeline's statistics (how long did it took, number of lines that were processed, etc.

Automating the Egress to DW Pipeline

In this task, you will egress all other tables to DW using yesterdays learnings

- Lookup activity to fetch data from table **ConfigParameters**
- Foreach activity to loop all tables and fetch information from ADLS and Store in DW

	SourceTable	DestinationTable	active	lastExecution	lastExecutionDW
1	DimCurrency	DimCurrency	1	NULL	NULL
2	DimGeography	DimGeography	1	NULL	NULL
3	DimDate	DimDate	1	NULL	NULL
4	DimProduct	DimProduct	1	NULL	NULL
5	DimProductCategory	DimProductCategory	1	NULL	NULL
6	DimProductSubCategory	DimProductSubCategory	1	NULL	NULL
7	FactInternetSales	FactInternetSales	1	NULL	NULL



Select * from ConfigParameters

Foreach Item in Lookup

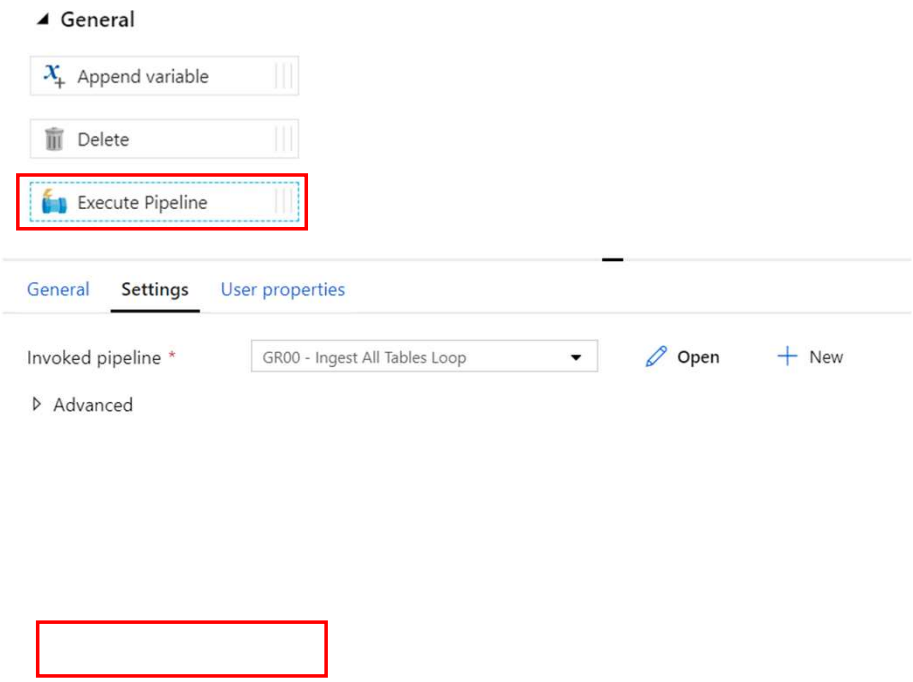
Items = @activity('Lookup').output.value

Executive Education

ADF Data Factory – Main pipeline

In this task, you will create the main pipeline

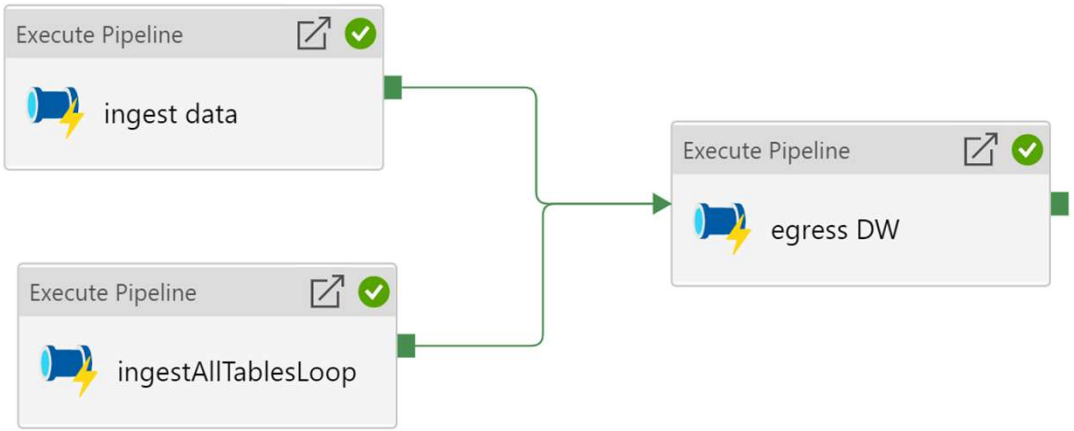
1. Create a new Pipeline using the **new pipeline** option
2. Name the pipeline **iXXXXXX - Main Pipeline**
3. Set the Description value to **Pipeline that will invoke all others**
4. Drag the newly created pipeline to the **correct group folder**
5. From the General Section In the Activities Pane Drag **Execute Pipeline** activity
6. Name the new activity **Ingest data**
7. Set the description field to **Ingest Customers**
8. Navigate to the **Settings TAB**
9. Set the Invoked Pipeline to **iXXXXXX – ingest data**



ADF Data Factory – Main pipeline

In this task, you will create the main pipeline

1. Do the previous step to add execute pipelines for the created pipelines: '… IngestAllTablesLoop and '… egress to Azure DW'
2. **Connect** the new Activity to the previous ones as described



Executive
Education

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

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