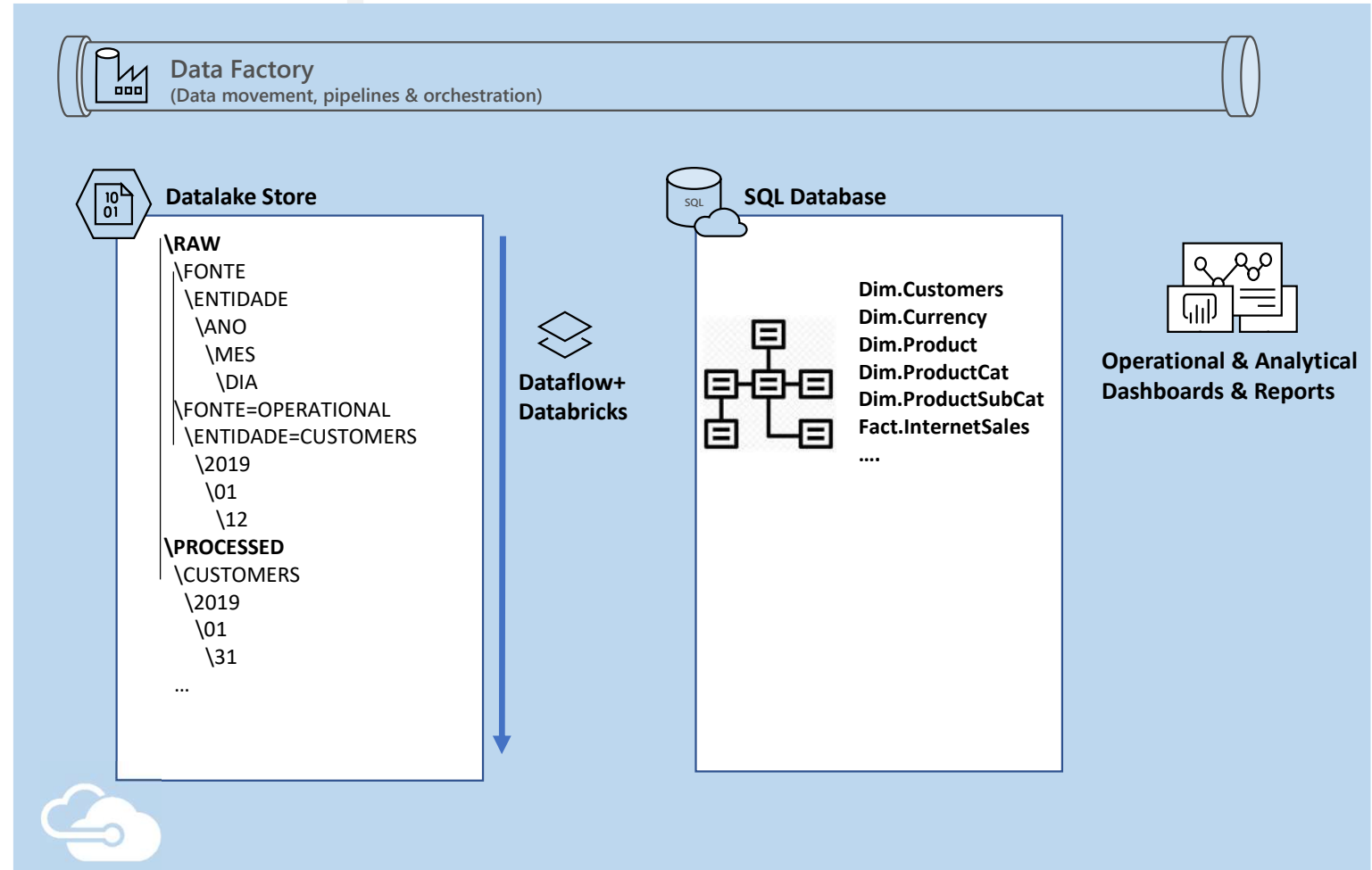
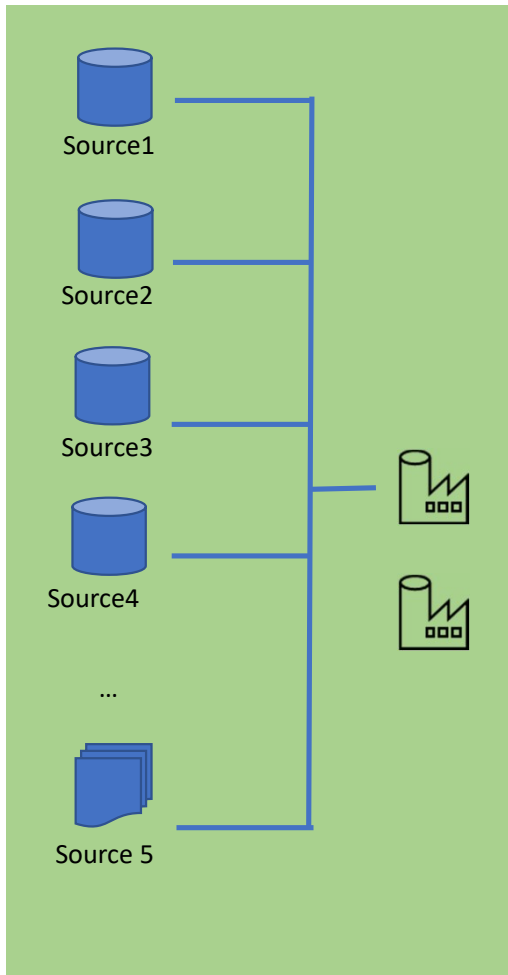


# Pós-graduação em Data Science & Business Analytics

# Index

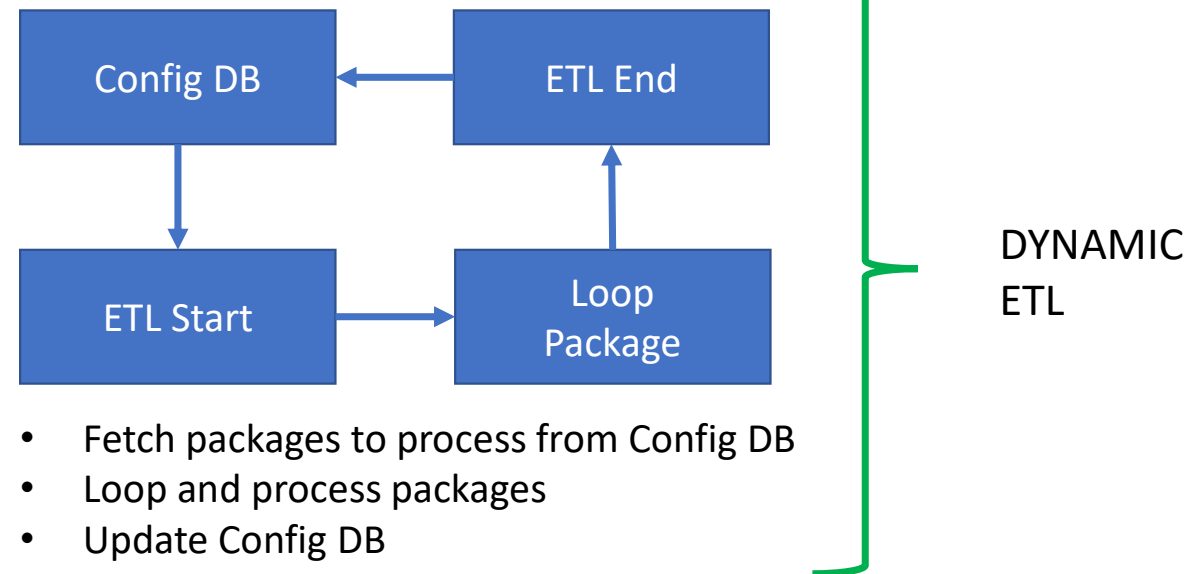
- Azure Data Factory Concepts
- Azure Data Factory Practical Experience
  - Configuration Tables
  - Azure Data Lake Storage Structure
  - Lookup Activity to Ingest Config table
  - Foreach Loop Activity to ingest multiple tables
  - Recursive Copy Activity
  - Update the Config table with last ingestion date

# Esboço da Arquitetura Alto-Nível



## ADF Concepts – ETL Execution Context

- Execution configuration Tables
  - ID Source
  - SourceType
  - SourceTable
  - Status
  - LastExecutionDate
  - Priority



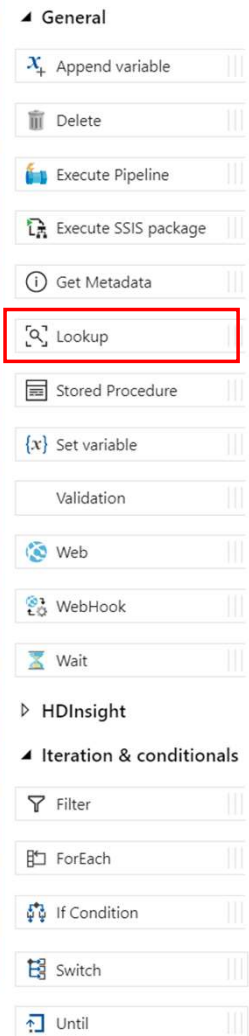
# ADF Concepts – Lookup Activity

## Lookup Activity

Lookup activity can **retrieve a dataset from any of the Azure Data Factory-supported data sources**. Use it in the following scenario:

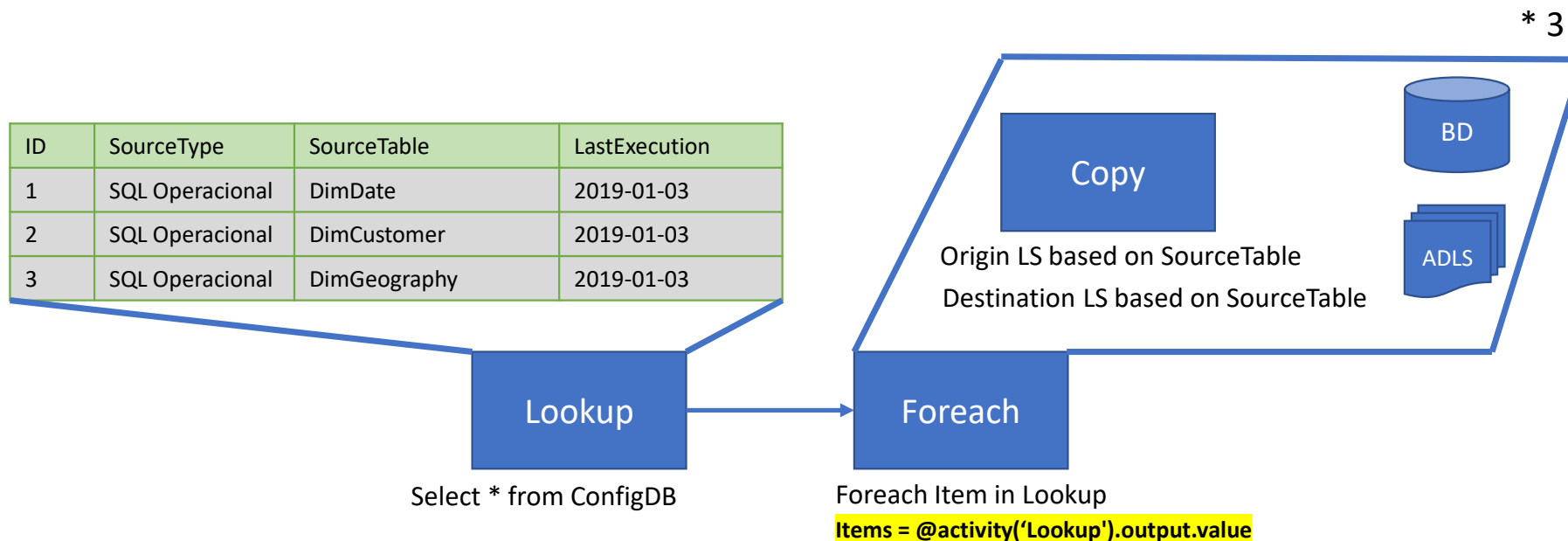
- **Dynamically determine which objects to operate on in a subsequent activity, instead of hard coding the object name.** Some object examples are files and tables.

Lookup activity reads and returns the content of a configuration file or table. It also returns the result of executing a query or stored procedure. The output from Lookup activity can be used in a subsequent copy or transformation activity if it's a singleton value. The output can be used in a ForEach activity if it's an array of attributes.



## ForEach Activity

The ForEach Activity defines a repeating control flow in your pipeline. This activity is used to iterate over a collection and executes specified activities in a loop. The loop implementation of this activity is similar to ForEach looping structure in programming languages.



### General

- Append variable
- Delete
- Execute Pipeline
- Execute SSIS package
- Get Metadata
- Lookup
- Stored Procedure
- Set variable
- Validation
- Web
- WebHook
- Wait
- HDInsight
- Iteration & conditionals
  - Filter
  - ForEach
  - If Condition
  - Switch
  - Until

# ADF Concepts – Store Procedure Activity

## Store Procedure Activity

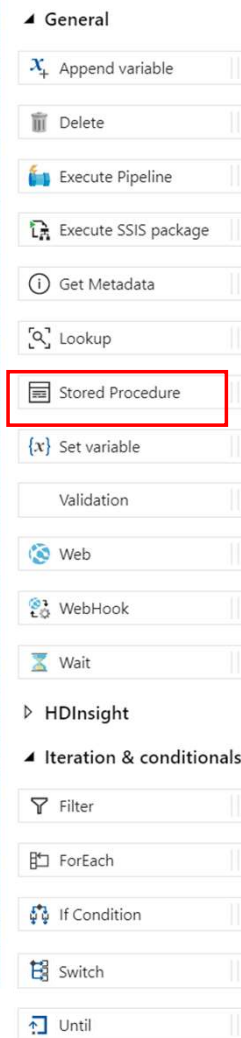
You can use the Stored Procedure Activity to invoke a stored procedure in one of the following data stores in your enterprise or on an Azure virtual machine (VM):

- Azure SQL Database | Azure SQL Data Warehouse | SQL Server Database.

When copying data into Azure SQL Database or SQL Server, **you can configure the `SqlSink` in copy activity to invoke a stored procedure by using the `sqlWriterStoredProcedureName` property.**

When copying data from Azure SQL Database or SQL Server or Azure SQL Data Warehouse, **you can configure `SqlSource` in copy activity to invoke a stored procedure to read data from the source database by using the `sqlReaderStoredProcedureName` property.**

ID	SourceType	SourceTable	LastExecution
1	SQL Operacional	DimDate	2019-01-03
2	SQL Operacional	DimCustomer	2019-01-03
3	SQL Operacional	DimGeography	2019-01-03



# ADF Concepts – Systems + Context Variables

## Pipeline Scope

**@pipeline().DataFactory** => Name of the data factory the pipeline run is running within

**@pipeline().Pipeline** => Name of the pipeline

**@pipeline().RunId** => ID of the specific pipeline run

**@pipeline().TriggerType** => Type of the trigger that invoked the pipeline (Manual, Scheduler)

**@pipeline().TriggerId** => ID of the trigger that invokes the pipeline

**@pipeline().TriggerName** => Name of the trigger that invokes the pipeline

**@pipeline().TriggerTime** => Time when the trigger that invoked the pipeline.

## Schedule Trigger scope

**@trigger().scheduledTime** => Time when the trigger was scheduled to invoke the pipeline run. **@trigger().startTime** => Time when the trigger actually fired to invoke the pipeline run.

## Tumbling Window Trigger scope

**@trigger().outputs.windowStartTime** => Start of the window when the trigger was scheduled to invoke the pipeline run.

**@trigger().outputs.windowEndTime** => End of the window when the trigger was scheduled to invoke the pipeline run.

## Item scope

**@item().SourceTable** => Field SourceTable from the Item (ex. Foreach Item)

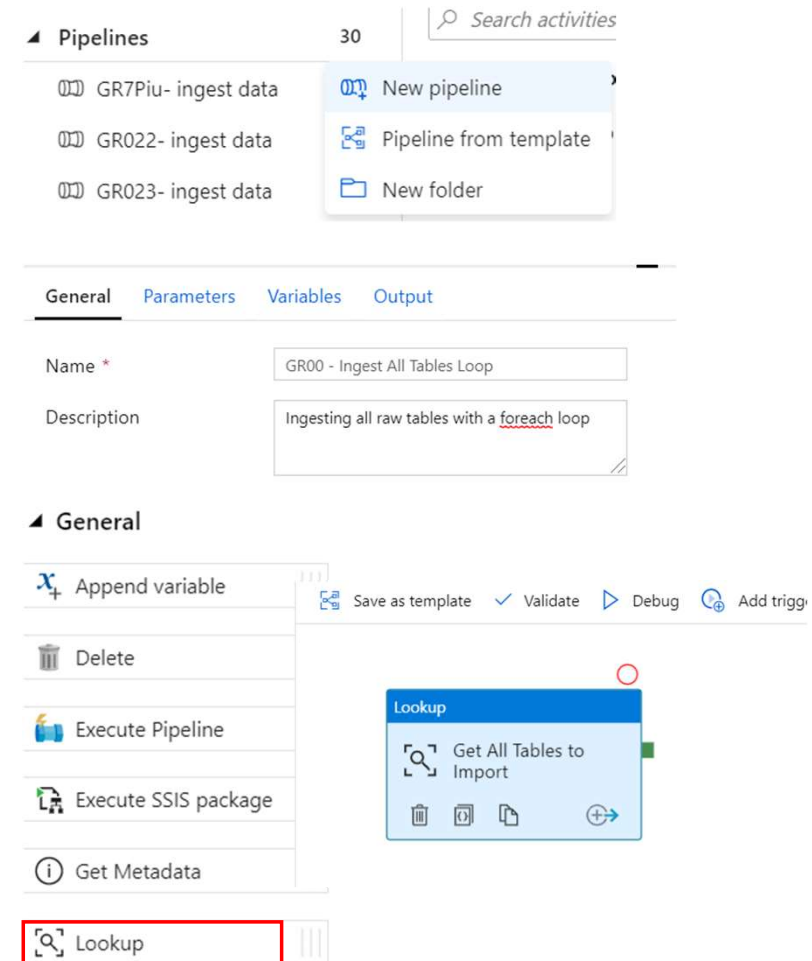


# ADF Data Factory – Lookup Activity

## In this task, you will create a Pipeline and add a Lookup Activity

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 **GRXXXXX - Ingest All Tables Loop**
4. Add **Ingesting all raw tables with a foreach loop** to the description
5. In the Activities pane open **General**
6. Drag the **Lookup Activity** into the main pane window
7. Set the name of the activity to **Get All Tables to Import**
8. Set the description field of the activity to **Fetch the list of tables to import**
9. Select the Settings Tab
10. Press the **+ New** icon next to the Source dataset to add a new dataset

Source dataset \*  [+ New](#)

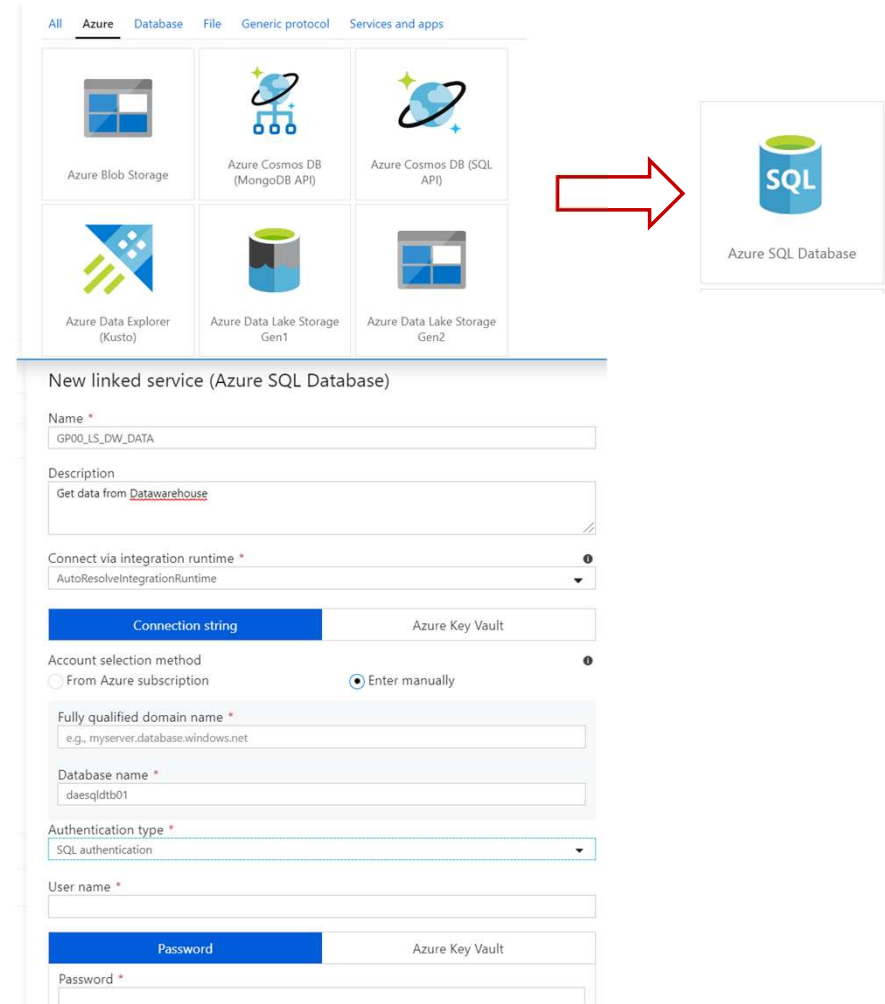


The screenshot shows the Azure Data Factory interface. At the top, the 'Pipelines' section is expanded, showing a list of pipelines: 'GR7Piu- ingest data', 'GR022- ingest data', and 'GR023- ingest data'. A context menu is open next to the 'Pipelines' section, showing options: 'New pipeline', 'Pipeline from template', and 'New folder'. Below this, the 'General' tab is selected for the pipeline configuration. The 'Name' field is set to 'GR00 - Ingest All Tables Loop' and the 'Description' field is set to 'Ingesting all raw tables with a foreach loop'. In the 'Activities' pane on the right, the 'General' tab is selected, showing a list of activities: 'Append variable', 'Delete', 'Execute Pipeline', 'Execute SSIS package', 'Get Metadata', and 'Lookup'. The 'Lookup' activity is highlighted with a red box. A preview of the 'Lookup' activity is shown in the bottom right corner, displaying the name 'Get All Tables to Import' and a description 'Fetch the list of tables to import'.

# ADF Data Factory – Lookup Activity

**In this task, you will configure the Lookup Activity**

1. In the new Dataset window choose **Azure**
2. Chose **Azure SQL Database**
3. Press the **Continue** button
4. In the Name field insert **ixxxxxx\_00\_DS\_DW\_SQLDB\_Config**
5. In the Linked Service select **New**
6. Set the name field to **ixxxxxx\_LS\_DW\_EXECDATA**
7. Set the Description field to **Get data from Datawarehouse**
8. Make sure the **Connect via integration runtime** is set to **AutoResolveIntegrationRuntime**
9. In Account Selection Method select **Enter manually**
10. In the Fully Qualified Domain Name insert **--sqlserver dw--**
11. Set the Database name to **--db dw--**



The screenshot displays the 'New linked service (Azure SQL Database)' configuration window in Azure Data Factory. The window is divided into two main sections: a top section with a grid of service icons and a bottom section with configuration fields. The top section includes icons for Azure Blob Storage, Azure Cosmos DB (MongoDB API), Azure Cosmos DB (SQL API), Azure Data Explorer (Kusto), Azure Data Lake Storage Gen1, and Azure Data Lake Storage Gen2. A red arrow points from the 'Azure SQL Database' icon in the top right to the configuration window. The configuration fields include: Name (GP00\_LS\_DW\_DATA), Description (Get data from Datawarehouse), Connect via integration runtime (AutoResolveIntegrationRuntime), Account selection method (Enter manually), Fully qualified domain name (e.g., myserver.database.windows.net), Database name (daesqldb01), Authentication type (SQL authentication), User name, and Password.

# ADF Data Factory – Lookup Activity

## In this task, you will configure the Lookup Activity

1. In the Authentication Type select **SQL Authentication**
2. Set the username field to **-- user db dw--**
3. Set the password field to **--psw db dw--**
4. Test the connection to ensure all settings are OK
5. Press the **Create** button
6. Select **iXXXXXX.ConfigParameters** in the Table Name field
7. Set the import schema to **From Connection/store**
8. Press the **OK** button
9. In the Settings pane press the **Preview data** Button to ensure configurations are ok
10. Make sure the **First row only** option **is not selected**



Set properties

Name  
GP00\_DS\_DW\_SQLDB\_Config

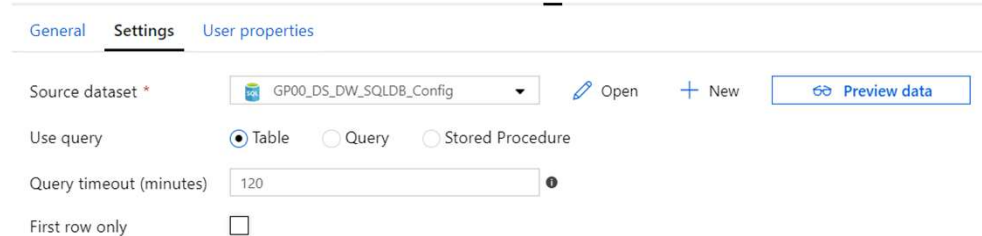
Linked service \*  
GP00\_LS\_DW\_DATA

Edit connection

Table name  
dbo.ConfigParameters

Edit

Import schema  
☒ From connection/store ☐ None



General Settings User properties

Source dataset \*  
GP00\_DS\_DW\_SQLDB\_Config

Open + New Preview data

Use query  
☒ Table ☐ Query ☐ Stored Procedure

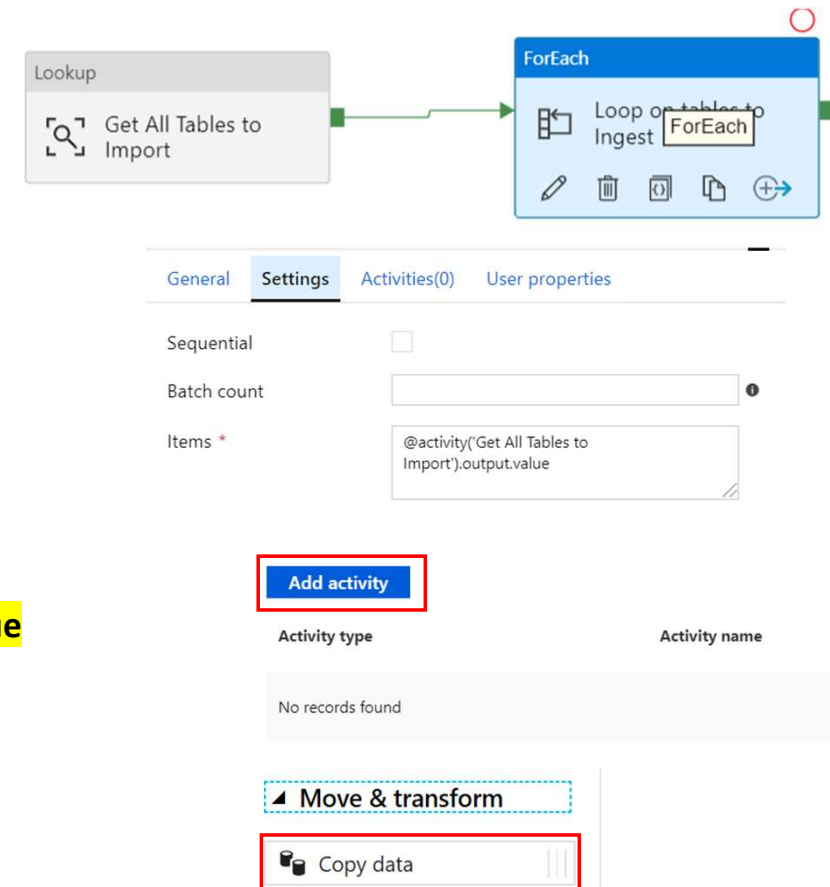
Query timeout (minutes)  
120

First row only  
☐

# ADF Data Factory – Foreach Loop Activity

In this task, you will add the Foreach activity to the pipeline

1. In the Activities pane open **Iteration & Conditionals**
2. Drag the **Foreach Activity** into the main pane window
3. Connect the **Foreach Activity** to the **Lookup activity** created before
4. Make sure the newly added activity is selected
5. Set the Name field to **Loop on tables to Ingest**
6. Set the Description field to **In each loop a new table is ingested**
7. Change to the **settings pane**
8. Select the **Sequential** option available in this pane
9. In the Items property add **@activity('Get All Tables to Import').output.value**
10. Change to the **Activities pane**
11. Press the **Add activity** button
12. In the pane Drag and drop the **Copy Data** activity available inside **Move & Transform** section



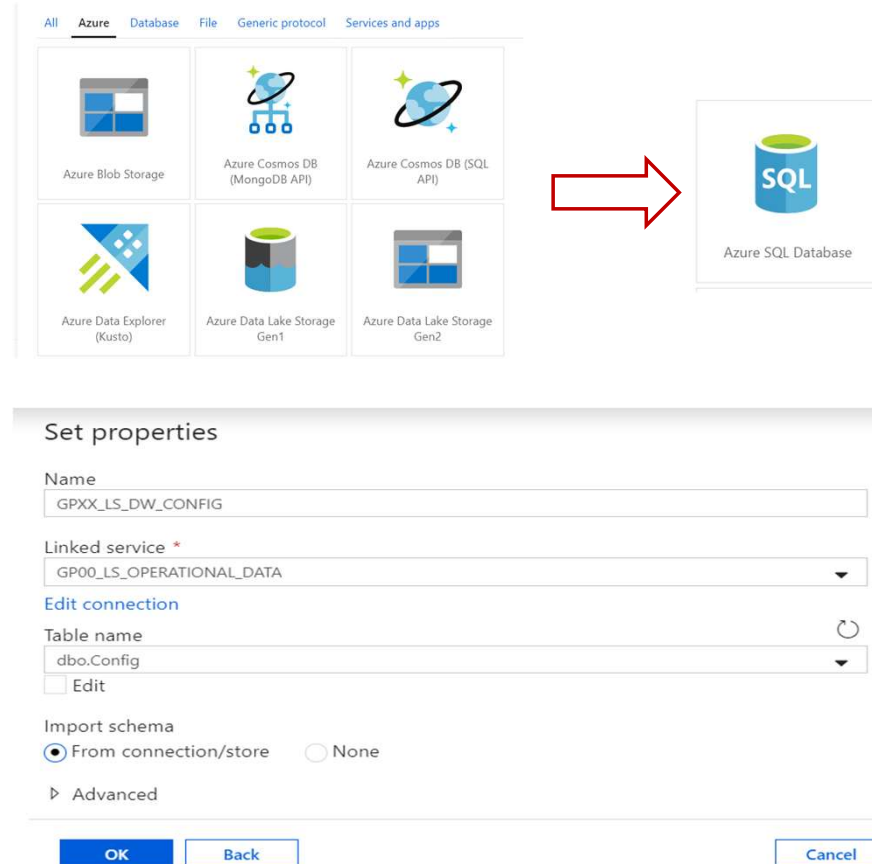
The screenshot shows the Azure Data Factory pipeline editor. A 'Lookup' activity named 'Get All Tables to Import' is connected to a 'Foreach' activity named 'Loop on tables to Ingest'. The 'Foreach' activity is selected, and its 'Settings' tab is active. The 'Sequential' option is selected, and the 'Items' property is set to '@activity('Get All Tables to Import').output.value'. Below the settings pane, the 'Add activity' button is highlighted with a red box. In the 'Activities' pane, the 'Move & transform' section is expanded, and the 'Copy data' activity is highlighted with a red box.

# Tópico

## ADF Data Factory – Foreach Loop Activity

In this task, you will add the ForEach activity to the pipeline

1. Set the Name field to **Copy table data using foreach**
2. Set the Description field to **Copying all data from operational to raw**
3. Change to the **Source** TAB
4. Press the **+ New** next to the Source Dataset
5. In the new Dataset window select **Azure** TAB
6. Select **Azure SQL Database** and select **Continue**
7. In the Set properties window set **iXXXXXX\_DS\_OPERATIONAL\_FROMCONFIG**
8. In the Linked Service select **New**
9. Set the name field to **iXXXXXX\_LS\_OPERATIONAL\_DATA**
10. Set the Description field to **Get data from Operational DB**
11. Make sure the **Connect via integration runtime** is set to
12. **AutoResolveIntegrationRuntime**
13. In Account Selection Method select **Enter manually**
14. In the Fully Qualified Domain Name insert **--sqlserver operacional--**
15. Set the Database name to **--db operacional--**



Set properties

Name  
GPXX\_LS\_DW\_CONFIG

Linked service \*  
GP00\_LS\_OPERATIONAL\_DATA

Edit connection

Table name  
dbo.Config

Edit

Import schema  
☒ From connection/store
 ☐ None

Advanced

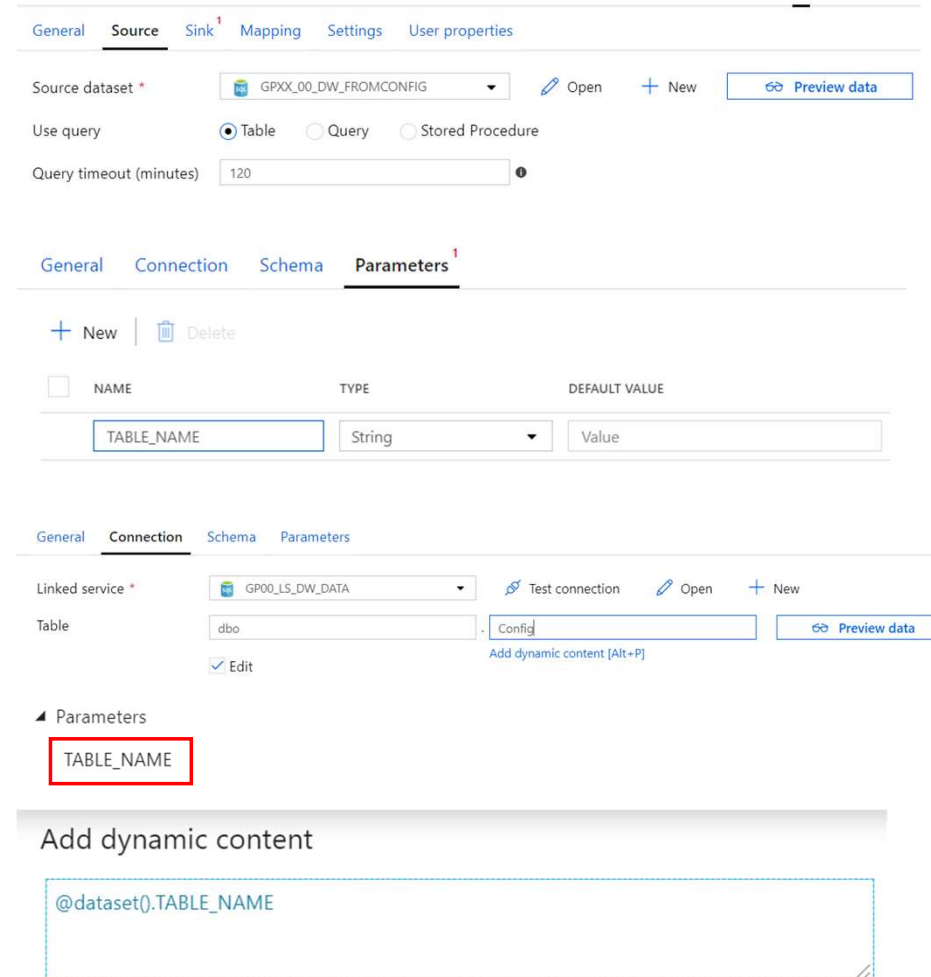
OK Back Cancel

# Tópico

## ADF Data Factory – Foreach Loop Activity

In this task, you will configure the ForEach activity

1. Insert **--user db operacional--** as the username
2. Set password as **--psw db operacional--**
3. Ensure **Test Connection** return success
4. Press **Create** button
5. Select **dbo.DimDate** for the Table Name
6. Make sure the **From Connection/store** is selected in import schema
7. Press **OK** button
8. In the Source Tab press the **Open** next to the Dataset
9. Inside the Dataset properties select the **Parameters** Tab
10. Click on the **+ New** Button
11. Set the new Parameter name to **TABLE\_NAME**
12. Change to the **Connection's** TAB
13. Click on the **Edit** Checkbox bellow the Table
14. Select the Table name **second text box**



The screenshot shows the ADF Data Factory interface with the following configuration:

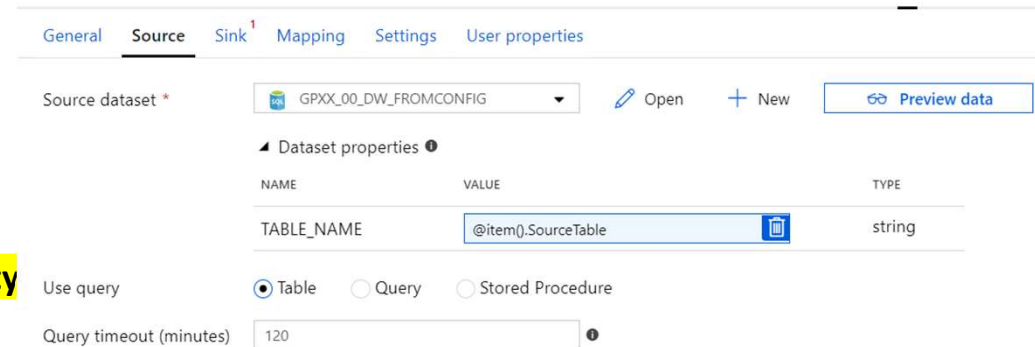
- Source Tab:**
  - Source dataset: GPXX\_00\_DW\_FROMCONFIG
  - Use query: ☒ Table
  - Query timeout (minutes): 120
- Parameters Tab:**
  - Table: TABLE\_NAME
  - Type: String
  - Default Value: Value
- Connection Tab:**
  - Linked service: GP00\_IS\_DW\_DATA
  - Table: dbo
  - Config: ☒ Edit
- Parameters List:**
  - TABLE\_NAME
- Add dynamic content:**
  - @dataset().TABLE\_NAME

# Tópico

## ADF Data Factory – Foreach Loop Activity

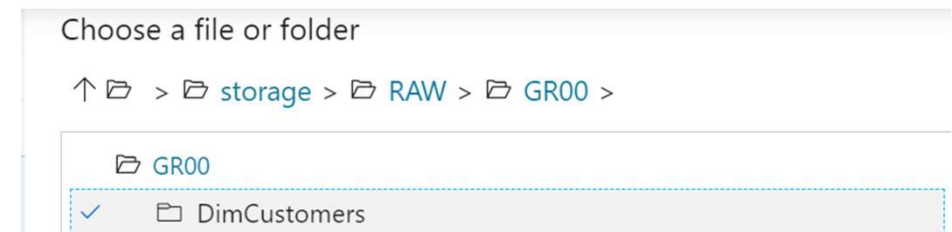
In this task, you will configure the ForEach activity

1. Press the **Add Dynamic Content (Alt + P)** option
2. In the Add Dynamic Content window Click on **TABLE\_NAME** inside the parameters section
3. Press the **Finish** button
4. On the Opened Windows TAB **navigate back to the Copy Activity**
5. In the Source Tab in the Dataset Properties windows
6. Add to the Value Textbox **@item().SourceTable**
7. Change to the **Sink** Tab
8. Press the **+ New** next to the Source Dataset
9. In the new Dataset window select **Azure** TAB
10. Select the **Azure Data Lake Storage Gen 2** and Select **Continue**
11. In the select format Window select **delimited**
12. Press the **Continue** button



The screenshot shows the 'Source' tab of the 'Dataset properties' window in Azure Data Factory. The 'Source dataset' is 'GPXX\_00\_DW\_FROMCONFIG'. The 'Dataset properties' table has columns 'NAME', 'VALUE', and 'TYPE'. The row for 'TABLE\_NAME' has a value of '@item().SourceTable' and a type of 'string'. Below the table, the 'Use query' section has radio buttons for 'Table' (selected), 'Query', and 'Stored Procedure'. The 'Query timeout (minutes)' is set to 120.

NAME	VALUE	TYPE
TABLE_NAME	@item().SourceTable	string



The screenshot shows the 'Choose a file or folder' dialog in the ADF dataset creation wizard. The breadcrumb path is '↑ > storage > RAW > GR00 >'. Under the 'GR00' folder, the 'DimCustomers' folder is selected, indicated by a checkmark and a dashed border.

# ADF Data Factory – Foreach Loop Activity

## In this task, you will configure the ForEach activity

1. In the Set Properties window Name the Dataset  
**iXXXXXX\_DS\_ADLS\_FROMCONFIG**
2. In the Linked Service select **iXXXXXX\_LS\_ADLS\_DATA**
3. Select the **Browse** button
4. In the Choose a File or Folder select the path **Storage > tXX > iXXXXXX > RAW**
5. Press the **OK** button
6. Select the **First Row as Header** Option
7. Ensure the **From Connection/store** option is selected
8. Press the **OK** button
9. In the Sink Tab press the **Open** next to the Sink Dataset
10. Inside the Dataset properties select the **Parameters** Tab
11. Click on the **+ New** Button
12. Set the new Parameter name to **TABLE\_NAME**

Choose a file or folder

↑ > storage > RAW > GR00 > DimCustomers >

DimCustomers

OK Cancel

Set properties

Name  
GP00\_LS\_ADLS\_FROMCONFIG

Linked service \*  
GPXX\_LS\_ADLS\_DATA

Edit connection

File path  
storage / RAW/GR00/DimCustom / File Browse

First row as header ☒

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

Advanced

General Connection Schema Parameters

Linked service \*  
GPXX\_LS\_ADLS\_DATA Test connection Open + New

File path \*  
storage / RAW/GR00 / File Browse Preview data

Add dynamic content [Alt+P]

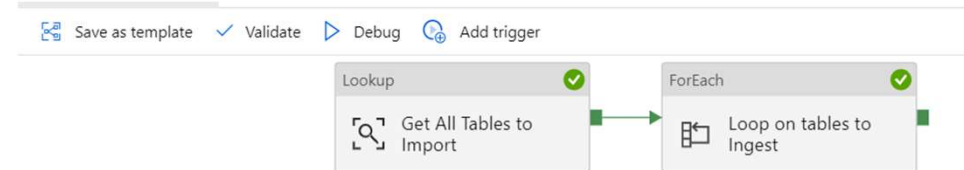
Compression type  
none



# ADF Data Factory – Foreach Loop Activity

In this task, you will configure the ForEach activity

1. Change to the **Connection's** TAB
2. Click on the **file** textbox in the file path
3. Press the **Add Dynamic Content (Alt + P)** option
4. In the Add Dynamic Content window add **@concat(dataset().TABLE\_NAME, '.csv')**  
Press the **Finish** button
5. In the Sink TAB Inside Dataset Properties set value **@item().DestinationTable** for TABLE\_NAME
6. Publish the pipeline using the **publish All** Option
7. Make sure you are at **1000000 - Ingest All Tables Loop** pipeline
8. Press the **Debug** button
9. Make sure the **pipeline runs sucessfull**



Save as template Validate Debug Add trigger

General Parameters Variables **Output**

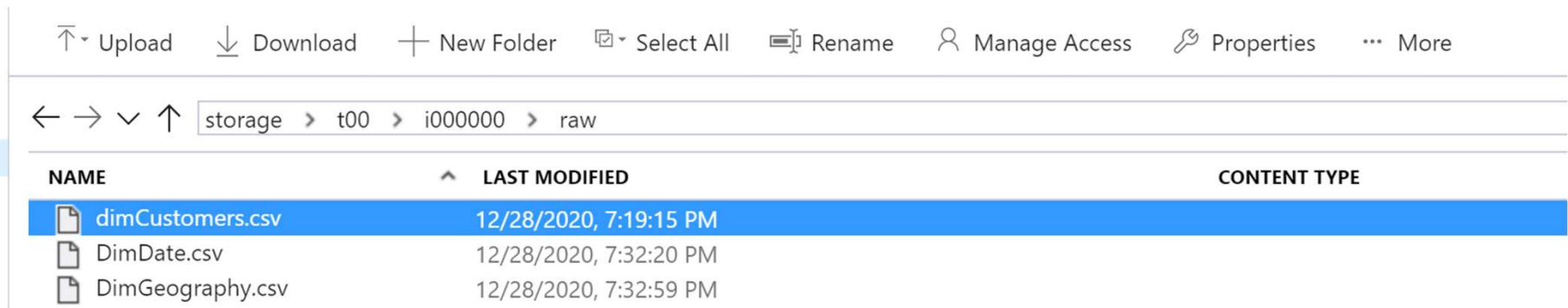
Pipeline run ID: e26fb08e-e1f6-47df-afbc-d360fd65ab2e [@] [Refresh] [Info]




NAME	TYPE	RUN START	DURATION	STATUS	ACTIONS	RUN ID
Copy table data using foreach	Copy	2019-12-16T16:38:00.320	00:00:05	✓ Succeeded	[Refresh] [Copy] [Delete]	942042e
Copy table data using foreach	Copy	2019-12-16T16:38:00.320	00:00:06	✓ Succeeded	[Refresh] [Copy] [Delete]	ed61c17
Copy table data using foreach	Copy	2019-12-16T16:38:00.305	00:00:05	✓ Succeeded	[Refresh] [Copy] [Delete]	6074f37
Loop on tables to Ingest	ForEach	2019-12-16T16:37:59.836	00:00:08	✓ Succeeded	[Refresh] [Copy] [Delete]	6a0abb9
Get All Tables to Import	Lookup	2019-12-16T16:37:46.935	00:00:13	✓ Succeeded	[Refresh] [Copy] [Delete]	69b310d

# ADF Data Factory – Data Lake Store

## In this task, you will validate Data Lake Storage Output

1. Open **Azure Portal**
2. Authenticate using your students account  
[IXXXX@students.isegexecutives.education](mailto:IXXXX@students.isegexecutives.education)
3. Open the Storage **FileSystem**
4. Navigate to the path
5. Check the CSV file is available

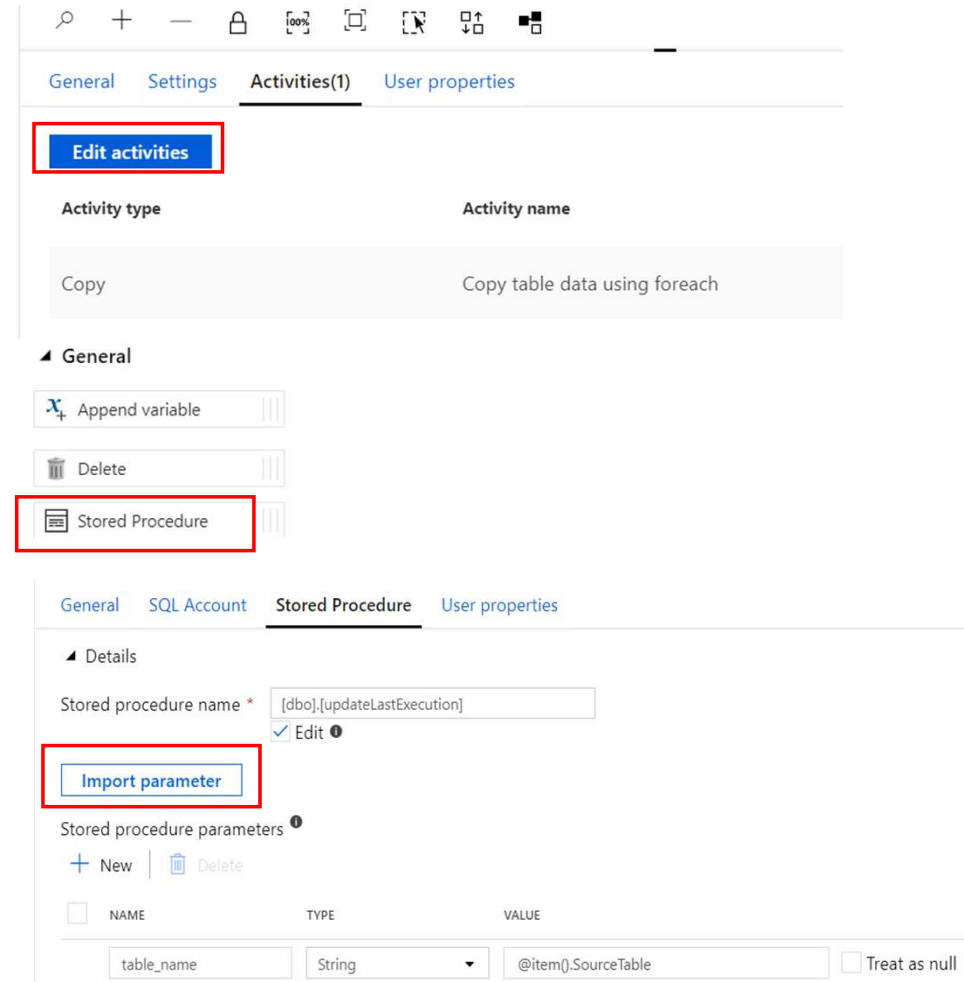


↑ Upload	↓ Download	+ New Folder	☑ Select All	🔄 Rename	👤 Manage Access	🔗 Properties	⋮ More
← → ↕ ↑	storage > t00 > i000000 > raw						
NAME	LAST MODIFIED	CONTENT TYPE					
 dimCustomers.csv	12/28/2020, 7:19:15 PM						
 DimDate.csv	12/28/2020, 7:32:20 PM						
 DimGeography.csv	12/28/2020, 7:32:59 PM						

# ADF Data Factory – Run a Store Procedure

## In this task, you will use a Store Procedure Activity

1. Make sure you are at **iXXXXXX - Ingest All Tables Loop** pipeline
2. Select the Foreach Activity **Loop on tables to ingest**
3. In the Activities TAB select **Edit Activities** button
4. In the Activities Pane inside the General Section choose **Stores Procedure** and drag it to the right of the copy activity
5. Name the Store Procedure **Updating Last Execution Date**
6. Set the Description to **Set success on the last Execution Date**
7. On SQL Account Tab set the Linked Service to **iXXXXXX\_LS\_DW\_EXECDATA**
8. Change to Stored Procedure TAB and set **dbo.UpdateLastExecution**
9. Press the **Import parameter** button
10. In the Store Procedure Parameters on dest add the value **@item().SourceTable**



The screenshot shows the ADF Data Factory interface with the 'Activities(1)' tab selected. The 'Edit activities' button is highlighted with a red box. Below it, the 'Activity type' is set to 'Copy' and the 'Activity name' is 'Copy table data using foreach'. The 'General' section is expanded, showing the 'Stored Procedure' activity type selected with a red box. The 'SQL Account' tab is selected, and the 'Stored Procedure' tab is also selected. The 'Details' section shows the 'Stored procedure name' as '[dbo].[updateLastExecution]' and the 'Edit' button is checked. The 'Import parameter' button is highlighted with a red box. The 'Stored procedure parameters' section shows a table with columns 'NAME', 'TYPE', and 'VALUE'. The first row has 'table\_name' as the name, 'String' as the type, and '@item().SourceTable' as the value. The 'Treat as null' checkbox is unchecked.

NAME	TYPE	VALUE
table_name	String	@item().SourceTable

# ADF Data Factory – Run a Store Procedure

## In this task, you will configure the Store Procedure Activity

1. Publish the pipeline using the **publish All** Option
2. Make sure you are at **iXXXXXX - Ingest All Tables Loop** pipeline
3. Press the **Debug** button
4. Make sure the **pipeline runs successful**
5. **Access to the Config Database** using SQL Server Management Studio
6. Validate that the last execution dates were successfully updates

General Parameters Variables Output					
Pipeline run ID: e26fb08e-e1f6-47df-afbc-d360fd65ab2e [0] [1]					
NAME	TYPE	RUN START	DURATION	STATUS	ACTIONS
Copy table data using foreach	Copy	2019-12-16T16:38:00.320	00:00:05	✓ Succeeded	→ ↗ ⚙
Copy table data using foreach	Copy	2019-12-16T16:38:00.320	00:00:06	✓ Succeeded	→ ↗ ⚙
Copy table data using foreach	Copy	2019-12-16T16:38:00.305	00:00:05	✓ Succeeded	→ ↗ ⚙
Loop on tables to Ingest	ForEach	2019-12-16T16:37:59.836	00:00:08	✓ Succeeded	→
Get All Tables to Import	Lookup	2019-12-16T16:37:46.935	00:00:13	✓ Succeeded	→ ↗

	Configid	SourceType	SourceTable	lastExecution
	1	SQL Operacional	DimDate	2019-01-03 00:0...
	2	SQL Operacional	DimCustomer	2019-01-03 00:0...
▶	3	SQL Operacional	DimGeography	2019-01-03 00:0...
*	NULL	NULL	NULL	NULL





Executive  
Education

[www.isegexecutive.education](http://www.isegexecutive.education)

Rua do Quelhas, 6  
1200-781 Lisboa

(+351) 213 922 891  
[info@executive.education](mailto:info@executive.education)