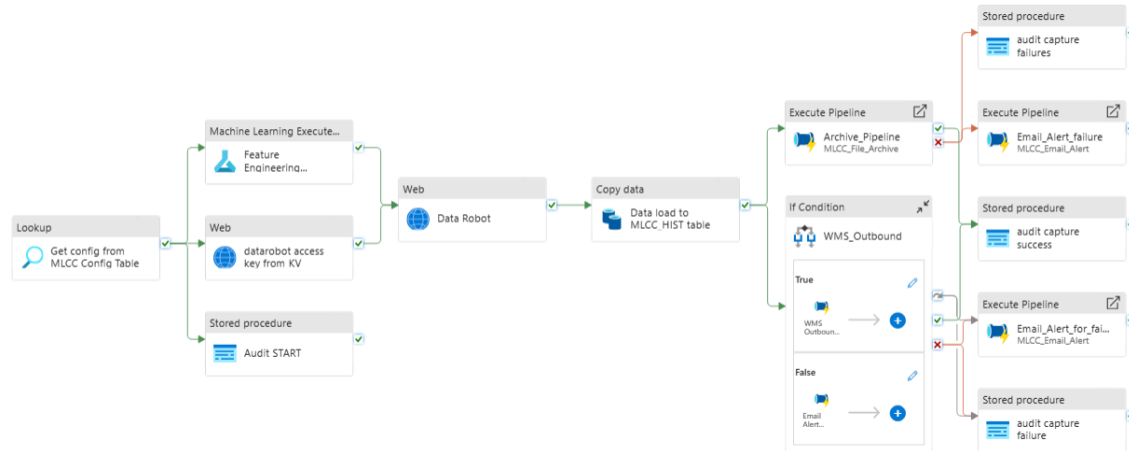


## --MLCC ADF pipeline



As a first activity in this pipeline, it checks if a row of data with the given process identifier exists on the ml\_config table. Then feature engineering notebook(ML studio pipeline) is triggered and alongside Datarobot key is accessed from key vault. When both these activities completed datarobot job definition is triggered for prediction using web activity in ADF. Also, the audit activity creates a record of the pipeline start time, process identifier etc. Once the datarobot prediction is complete the copy activity loads data to mlcc hist table.

### Activity 1 - Get config from MLCC Config Table

General	Settings	User properties
<b>Source dataset *</b> <span>MLCC_Synapse_Tables</span> <span>Open</span> <span>New</span> <span>Preview data</span> <span>Learn more</span>		
<b>First row only</b> <input checked="" type="checkbox"/>		
<b>Use query</b> <input type="radio"/> Table <input checked="" type="radio"/> Query <input type="radio"/> Stored procedure		
<b>Query</b> <span>select * from dbo.ml_integration_con...</span>		
<b>Query timeout (minutes)</b> <span>120</span>		
<b>Isolation level</b> <span>Select...</span>		
<b>Partition option</b> <input checked="" type="radio"/> None <input type="radio"/> Physical partitions of table <input type="radio"/> Dynamic range		
<span>Please preview data to validate the partition settings.</span>		

For this we need to create a datastore. Go to pencil icon > Factory resources > datasets > plus icon for new dataset and choose azure synapse.

Select the datastore in the settings tab of the activity.



Then add query to check if the process identifier is there in the ml config table

```
select * from dbo.ml_integration_config_table where  
process_identifer='@{pipeline().parameters.process_identifier}'
```

## Activity 2 - Feature Engineering Notebook Execution

On the on-success button of activity 1 add this. In the general set retry and retry interval. In the **Azure Machine Learning linked service** we need to add a linked service Azure ML.

### Edit linked service

 Azure Machine Learning [Learn more](#) 

Name \*

AzureMLService

Description

Connect via integration runtime \* 

 AutoResolveIntegrationRuntime 

Authentication method

System-assigned managed identity 

Azure Machine Learning workspace selection method 

☐ From Azure subscription ☒ Enter manually

Subscription ID \*

81a662f6-14b7-4f82-90b4-ae2608e244c4

Resource group name \*

DSCUSNORAMDataWarehouseMLdev

Azure Machine Learning workspace name \*

DSCNORAMMLCycleCountDev

Managed identity name: **dscusnoramdwmladfddev**

Managed identity object ID: **537ca37e-855c-405f-b8ef-56d9d9d222ce**

Grant Data Factory service managed identity access to your Azure Machine Learning.

[Learn more](#) 

In the settings tab we use - `@activity('Get config from MLCC Config Table').output.firstRow.ML_pipeline_id` - to get the `mlcc_id` as input for this activity. To get the experiment name we use the query - `@pipeline().parameters.process_identifier`. in the below section Machine Learning pipeline parameters, we have added parameters as shown below.

Experiment name

`@pipeline().parameters.process_iden...`

Machine Learning pipeline parameters

Import parameters



+ New |  Delete

<input type="checkbox"/>	Name	Value
<input type="checkbox"/>	output_path	<code>@activity('Get config from MLCC Con...</code>
<input type="checkbox"/>	output_file	<code>@activity('Get config from MLCC Con...</code>
<input type="checkbox"/>	container_name	<code>@activity('Get config from MLCC Con...</code>
<input type="checkbox"/>	scoring_py_script	<code>@activity('Get config from MLCC Con...</code>
<input type="checkbox"/>	utils_file	<code>@activity('Get config from MLCC Con...</code>
<input type="checkbox"/>	config_file	<code>@activity('Get config from MLCC Con...</code>

All the details like container name, scripts for scoring config and utils are given in the `ml_config` table.

### Activity 3 – Parallel to activity 2 – Datarobot web

datarobot access key from KV – This activity is used to get the datarobot secret from key vault. We are using GET method. The authentication method used is managed identity.

URL	<div>@concat(pipeline().globalParameters.ke...</div> <div> Information will be sent to the URL specified. Please ensure you trust the URL entered.</div>
Method * ⓘ	<div>GET</div>
Authentication ⓘ	<div>System-assigned managed identity</div>
Resource * ⓘ	<div>https://vault.azure.net</div>
Headers ⓘ	<div>+ New</div>
▽ Advanced	
Integration runtime * ⓘ	<div> AutoResolveIntegrationRuntime</div>
HTTP request timeout ⓘ	<div></div>
Disable async pattern ⓘ	<div><input type="checkbox"/></div>
Disable certificate validation ⓘ	<div><input type="checkbox"/></div>
Datasets ⓘ	<div>+ Add dataset reference</div>
Linked services ⓘ	<div>+ Add linked service reference</div>

```
@concat(pipeline().globalParameters.keyvault_url,activity('Get config from MLCC Config Table').output.firstRow.DR_KV_secret,pipeline().globalParameters.keyvault_api_version)
```

Here the `DR_KV_secret` is the secret name which is 'datarobot-access-key'. Here this link will be used in a Get request to get the secret from the key vault. The key vault url given is <https://dscusnoramamdwm/vaultdev.vault.azure.net/secrets/> which is saved as a global parameter. The API version is ?api-version=7.0. So the final url is <https://dscusnoramamdwm/vaultdev.vault.azure.net/secrets/datarobot-access-key/?api-version=7.0>

## Activity 4 - Audit START – Parallel to Activity 3

General **Settings** User properties

Linked service \* ⓘ dscnoramdwynapsepooldev [Test connection](#) [Edit](#) [+ New](#)

Stored procedure name \* [dbo].[MLCC\_Audit\_Status]  
☒ Enter manually

Stored procedure parameters ⓘ


[← Import](#) [+ New](#) [Delete](#)

<input type="checkbox"/>	Name	Type	Value	
<input type="checkbox"/>	batch_copy_trigger_time		@utcNow()	
<input type="checkbox"/>	batch_id	Int32	1	<input type="checkbox"/> Treat as null
<input type="checkbox"/>	batch_type	String	dailyload	<input type="checkbox"/> Treat as null
<input type="checkbox"/>	error_message	String	Value	<input checked="" type="checkbox"/> Treat as null
<input type="checkbox"/>	lastmodified		@utcNow()	
<input type="checkbox"/>	pipeline_name	String	@pipeline().TriggerName	
<input type="checkbox"/>	pipeline_run_id	String	@pipeline().RunId	
<input type="checkbox"/>	process_identifier	String	@pipeline().parameters.process_iden...	
<input type="checkbox"/>	status	String	STARTED	<input type="checkbox"/> Treat as null

[dbo].[MLCC\_Audit\_Status] – This stored procedure gets executed and it creates an entry in the audit table with the pipeline name run id, run date etc.

## Activity 5 – Web activity - Data Robot – Starts after activity 2 and 3

General **Settings** User properties

URL @activity('Get config from MLCC Con...)  
 Information will be sent to the URL specified. Please ensure you trust the URL entered.

Method \* ⓘ POST

Body @concat(["jobDefinitionId":"","activit...])

Authentication ⓘ None

Headers \* ⓘ [+ New](#) [Delete](#)

<input type="checkbox"/>	Name	Value
<input type="checkbox"/>	Authorization	<span>Bearer @activity('datarobot access k...)</span>

Advanced

Integration runtime \* ⓘ dscusclouddbts [Edit](#)

HTTP request timeout ⓘ

Disable async pattern ⓘ ☐

Disable certificate validation ⓘ ☐

The URL part in this activity gets URL from ml config table using the first activity output using query below

```
@activity('Get config from MLCC Config Table').output.firstRow.datarobot_URL
```

This activity uses a post method to pass the job definition id of the job definition defined in the data robot using the body format of the request given below

```
@concat('{"jobDefinitionId": "', activity('Get config from MLCC Config Table').output.firstRow.jobdefinitionid, '" }')
```

For authentication to DR the below expression is used to get the access key

Here the Bearer is a header

Headers that are sent to the request. For example, to set the language and type on a request: "headers": { "Accept-Language": "en-us", "Content-Type": "application/json" }.

```
Bearer @{activity('datarobot access key from KV').output.value}
```

Also, we are using a specific integration runtime named dscuscloudbtst for this activity to run

## Activity 6 - Copy activity - Data load to MLCC\_HIST table

### Source properties

General **Source** Sink Mapping Settings User properties

Source dataset \* CSVtoMLCCtable [Open](#) [New](#) [Preview data](#) [Learn more](#)

Dataset properties ⓘ

Name	Value	Type
ContainerName	@activity('Get config from MLCC Config Table').output.firstRow.datarobot_URL	string
FilePath	@activity('Get config from MLCC Config Table').output.firstRow.datarobot_URL	string
FileName	@activity('Get config from MLCC Config Table').output.firstRow.datarobot_URL	string

File path type ⓘ

☐ File path in dataset ☐ Prefix ☒ Wildcard file path ☐ List of files ⓘ

Wildcard paths ⓘ

{source container} / @activity('Get config from MLCC Config Table').output.firstRow.datarobot\_URL / @activity('Get config from MLCC Config Table').output.firstRow.jobdefinitionid

Start time (UTC)  End time (UTC)

Filter by last modified ⓘ

Recursively ⓘ ☒

Enable partitions discovery ⓘ ☐

Max concurrent connections ⓘ

Max concurrent connections ⓘ

Skip line count

Additional columns ⓘ + New 🗑 Delete

<input type="checkbox"/> Name	Value
<input type="checkbox"/> Run_ID	<input type="text" value="@pipeline().RunId"/>
<input type="checkbox"/> Predict_Dttm	<input type="text" value="@utcnow()"/>
<input type="checkbox"/> Lst_Upd_Tms_Dttm	<input type="text" value="@utcnow()"/>

Container name value from ML config table

```
@activity('Get config from MLCC Config Table').output.firstRow.Container_Name
```

File path value from ML config table

```
@activity('Get config from MLCC Config Table').output.firstRow.Container_Name
```

File name value from ML config table

```
@activity('Get config from MLCC Config Table').output.firstRow.datarobot_outputFilename
```

## Sink properties

On the sink side we are providing the dataset(manually created) details and the copy method wen need to use. Here we are using bulk insert for copying large amount of data

General Source **Sink** Mapping Settings User properties

Sink dataset \*  [Open](#) [+ New](#) [Learn more](#)

Copy method ☐ Copy command ⓘ ☐ PolyBase ⓘ ☒ Bulk insert ☐ Upsert

Bulk insert table lock ⓘ ☐ Yes ☒ No

Table option ☒ Use existing ☐ Auto create table ⓘ

Pre-copy script ⓘ

Write batch timeout ⓘ

Write batch size ⓘ

Max concurrent connections ⓘ

Disable performance metrics analytics ⓘ ☐

Mapping

Mapping gets created automatically usually and we can verify the data type if needed

GeneralSourceSinkMappingSettingsUser properties

⚠ One of the column names contains special characters like '\n', '\r', etc., which might cause issues if it doesn't match your real data. Please click 'View source code' to review the column: 'Run\_ID'.

← Import schemas

🔗 Preview source

+ New mapping

🧼 Clear

↺ Reset

🗑 Delete

<input type="checkbox"/>	Source	Type		Destination	Type		
⋮ <input type="checkbox"/>	site_sk	abc String	→	Site_SK	123 int	+	🗑
<input type="checkbox"/>	prod_sk	abc String	→	Prod_SK	123 int	+	🗑
<input type="checkbox"/>	loc_sk	abc String	→	Loc_SK	123 int	+	🗑
<input type="checkbox"/>	MLCC_ID	abc String	→	MLCC_ID	abc varchar	+	🗑
<input type="checkbox"/>	Dummy_Inv_Trxn_Cnt	abc String	→	Inv_Trxn_Cnt	123 int	+	🗑
<input type="checkbox"/>	Dummy_cnt_dt	abc String	→	Cnt_Dt	📅 date	+	🗑
<input type="checkbox"/>	Dummy_Cnt_Hist_Id	abc String	→	Cnt_Hist_ID	abc varchar	+	🗑
<input type="checkbox"/>	loc_nme	abc String	→	Loc_Nme	abc varchar	+	🗑
<input type="checkbox"/>	ar_cd	abc String	→	Ar_Cd	abc varchar	+	🗑
<input type="checkbox"/>	wrk_zn_cd	abc String	→	Wrk_Zn_Cd	abc varchar	+	🗑
<input type="checkbox"/>		abc String	→		abc varchar	+	🗑

Settings



General
Source
Sink
Mapping
**Settings**
User properties

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

Maximum data integration unit ⓘ

Auto
Use custom value

Degree of copy parallelism ⓘ

Auto

Fault tolerance ⓘ

Enable logging ⓘ

☐

Enable staging ⓘ

☐

## Activity 7 – Execute Pipeline - Archive\_Pipeline

General
**Settings**
User properties

Invoked pipeline \*

MLCC\_File\_Archive
Open
+ New

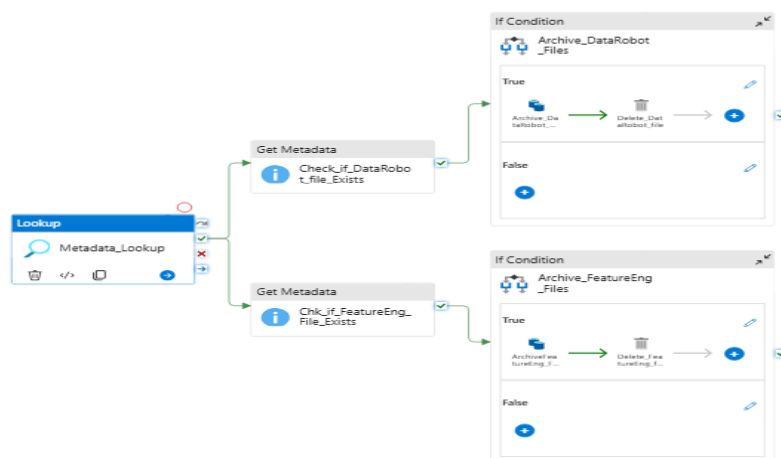
Wait on completion

☒

Parameters

Name	Type	Value	Default value
process_identifier	string	@pipeline().parameters.process_iden...	

Here a pipeline MLCC\_File\_Archive is invoked, and process identifier is the same parameter that we pass while debugging pipeline. The process identifier is needed as a parameter to run this pipeline also.



Here in the first lookup activity we are getting values of the output file name file paths etc by running a query as shown below in the dataset MLCC\_Synapse\_tables

```
@concat('select ArchiveFilePath, Container_name,datarobot_outputFilename,
datarobot_outputFilePath,FeatureEngOutputFilePath, FeatureEngOutputFileName,
WMSOutputFilePath,WMSOutputFileName from dbo.ml_integration_config_table where
process_identfier='',pipeline().parameters.process_identfier, ''')
```

## Lookup activity

General **Settings** User properties

---

Source dataset \* MLCC\_Synapse\_Tables Open New Preview data Learn more

First row only ☒

Use query ☐ Table ☒ Query ☐ Stored procedure

Query @concat('select ArchiveFilePath, Con...

Query timeout (minutes) ⓘ 120

Isolation level ⓘ Select...

Partition option ⓘ ☒ None ☐ Physical partitions of table ⓘ ☐ Dynamic range ⓘ

ⓘ Please preview data to validate the partition settings.

## MLCC\_Synapse\_tables



Azure Synapse Analytics  
MLCC\_Synapse\_Tables

Connection **Schema** Parameters

---

Linked service \* dscnoramdwsynapsepooldev Test connection Edit New Learn more

Table dbo.ml\_integration\_config\_table Refresh Preview data

☐ Enter manually

In the next activity we are checking if the Feature Engineering output file(features.csv) and Data robot output file is present in the specified blob storage which we connect using a linked service. Here the linked service is dscnoramdwmIstoragedev which is defined inside the Dataset part of the Get Metadata activity settings

## Get Metadata activity

General Settings User properties

Name \*  [Learn more](#)

Description

Activity state <sup>①</sup> ☒ Activated ☐ Deactivated

Timeout <sup>①</sup>

Retry <sup>①</sup>

Retry interval (sec) <sup>①</sup>

Secure output <sup>①</sup> ☐

Secure input <sup>①</sup> ☐

General Settings User properties

Dataset \*  [Open](#) [+ New](#) [Learn more](#)

Dataset properties <sup>①</sup>

Name	Value
file_name	<input type="text" value="@activity('Metadata_Lookup').output..."/>
file_path	<input type="text" value="@activity('Metadata_Lookup').output..."/>
container_name	<input type="text" value="@activity('Metadata_Lookup').output..."/>

Field list \* [+ New](#) [Delete](#)

☐ Argument

☐

Filter by last modified <sup>①</sup>  Start time (UTC)  End time (UTC)


Skip line count


file name - `@activity('Metadata_Lookup').output.firstRow.datarobot_outputFilename`

file path - `@activity('Metadata_Lookup').output.firstRow.datarobot_outputFilePath`

Container\_name - `@activity('Metadata_Lookup').output.firstRow.Container_name`

The csv\_source is defined as

 Saved {}

 DelimitedText  
csv\_source

ConnectionSchemaParameters

Linked service \*

File path \*

Compression type

Column delimiter ⓘ

Row delimiter ⓘ

Encoding ⓘ

Quote character ⓘ

Escape character ⓘ

First row as header ⓘ

Null value ⓘ

dscnoramdwmlstoragedev

@(dataset().container\_name) / @(dataset().file\_path) / @(dataset().file\_name)

No compression

Comma (,)

Line feed (\n)

Default(UTF-8)

Double quote (")

Double quote (")

☒

Test connection

Edit

New

Learn more

Browse

Here the file name is taken like this from the first activity of MLCC\_Archive\_pipeline


```
@activity('Metadata_Lookup').output.firstRow.datarobot_outputFilename
```

Similarly, container\_name and file\_path. All these values vary based on the process\_identifier used in the metadata lookup activity

The next activity that happens if the files exist is a copy data activity and a delete activity which archives the files created to an archive folder named MLCCArchiveFolder within the blob storage







Blob containers > mlcc > MLCCArchiveFolder

Authentication method: Access key ([Switch to Microsoft Entra user account](#))

 Search blobs by prefix (case-sensitive)

Only show active blobs

Showing all 34 items

<input type="checkbox"/>	Name	Last modified	Access tier	Blob type
<input type="checkbox"/>	 <a href="#">[..]</a>			
<input type="checkbox"/>	 <a href="#">CampbellFo...</a>	2/4/2025, 10:16:08 AM	Hot (Inferred)	Block blob
<input type="checkbox"/>	 <a href="#">CampbellFo...</a>	2/4/2025, 10:58:22 AM	Hot (Inferred)	Block blob
<input type="checkbox"/>	 <a href="#">CampbellFo...</a>	2/4/2025, 7:42:24 PM	Hot (Inferred)	Block blob
<input type="checkbox"/>	 <a href="#">CampbellFo...</a>	2/5/2025, 10:27:24 AM	Hot (Inferred)	Block blob
<input type="checkbox"/>	 <a href="#">CampbellFo...</a>	2/5/2025, 10:53:49 AM	Hot (Inferred)	Block blob





Here the if condition works on the basis of true or false value returned for the below Expression

```
@activity('Check_if_DataRobot_file_Exists').output.exists
```

General Activities (2) User properties

Expression ⓘ

@activity('Check\_if\_DataRobot\_file\_E...

Case	Activity	
True	 Archive_DataRo...  Delete_DataRob...	
	2 Activities	
False	No activities	

Once the copy data(archive) activity is done the delete activity is performed.

General Source Sink Mapping Settings User properties

Name \*  [Learn more](#) 

Description

Activity state ⓘ ☒ Activated ☐ Deactivated

Timeout ⓘ

Retry ⓘ

Retry interval (sec) ⓘ

Secure output ⓘ ☐

Secure input ⓘ ☐

Source dataset \* csv\_source [Open](#) [+ New](#) [Preview data](#) [Learn more](#)

## Dataset properties

Name	Value	Type
file_name	@activity('Metadata_Lookup').output...	string
file_path	@activity('Metadata_Lookup').output...	string
container_name	@activity('Metadata_Lookup').output...	string

File path type ☒ File path in dataset ☐ Prefix ☐ Wildcard file path ☐ List of filesStart time (UTC)  End time (UTC) 

Filter by last modified

Recursively ☒Enable partitions discovery ☐Max concurrent connections Skip line count Additional columns [+ New](#)Sink dataset \* MLCCArchivePath [Open](#) [+ New](#) [Learn more](#)

## Dataset properties

Name	Value	Type
ContainerName	@activity('Metadata_Lookup').output...	string
ArchivePath	@activity('Metadata_Lookup').output...	string
ArchiveFileName	@concat(activity('Metadata_Lookup')...	string

Copy behavior Max concurrent connections Block size (MB) Metadata [+ New](#)Quote all text ☒File extension Max rows per file 


## Mapping is blank

**i** You will be charged # of used DIUs \* copy duration \* \$0.25/DIU-hour. Local currency and separate discounting may apply per subscriptionMaximum data integration unit  ☐ Use custom valueDegree of copy parallelism Data consistency verification ☐Fault tolerance Enable logging ☐Enable staging ☐

The MLCCArchivePath is a datastore as shown below

MLCC\_ADF\_Pipeline MLCC\_File\_Archive csv\_source MLCCArchivePath

Saved

 DelimitedText  
MLCCArchivePath

Connection Schema Parameters

Linked service \*  
File path \*  
Compression type  
Column delimiter  
Row delimiter  
Encoding  
Quote character  
Escape character  
First row as header  
Null value

dscnoramdwmlstoragedev

@dataset().ContainerName / @dataset().ArchivePath / @dataset().ArchiveFileName

No compression

Comma (,)

Default (\r\n, or \n)

Default(UTF-8)

No quote character

Backslash (\)

☒

Test connection Edit + New Learn more

Browse Preview data Detect format

Now the delete activity

General Source Logging settings User properties

Name \*

Delete\_DataRobot\_file

Learn more

Description

Activity state ⓘ

☒ Activated ☐ Deactivated

Timeout ⓘ

0.12:00:00

Retry ⓘ

1

Retry interval (sec) ⓘ

30

Secure output ⓘ

☐

Secure input ⓘ

☐



General **Source** Logging settings User properties

Dataset ⊕  [Open](#) [+ New](#) [Learn more](#)

Dataset properties ⊕

Name	Value	Type
Container_Name	@activity("Metadata_Lookup").output...	string
File_Path	@activity("Metadata_Lookup").output...	string
File_Name	@activity("Metadata_Lookup").output...	string

File path type ☒ File path in dataset ☐ Wildcard file path ☐ Prefix ☐ List of files

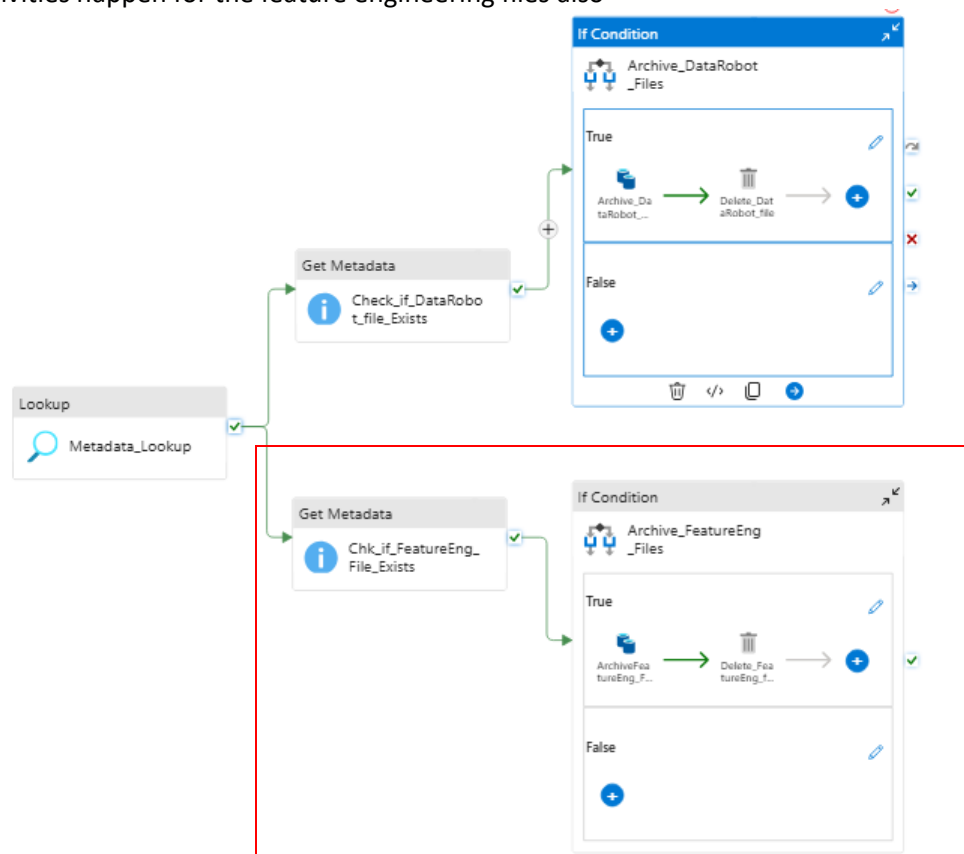
Filter by last modified ⊕ Start time (UTC)  End time (UTC)

Recursively ☒

Max concurrent connections ⊕

Enable logging and User properties are blank

The exact activities happen for the feature engineering files also



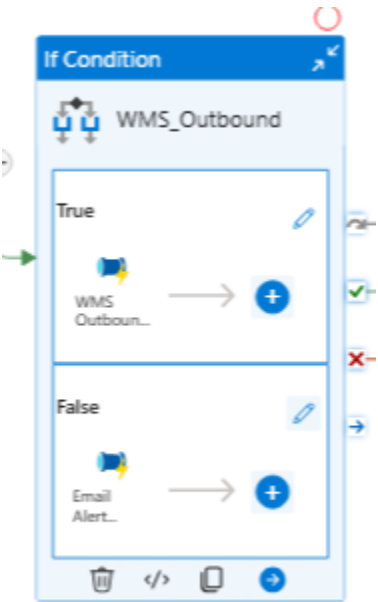
### Activity 8 – WMS\_Outbound - If condition

In this activity if data load has happened to MLCC\_hist\_table the true part which is the WMS outbound pipeline gets executed and if false an email alert pipeline gets executed

Expression used

```
@greaterOrEquals(activity('Data load to MLCC_HIST table').output.rowsCopied,1)
```

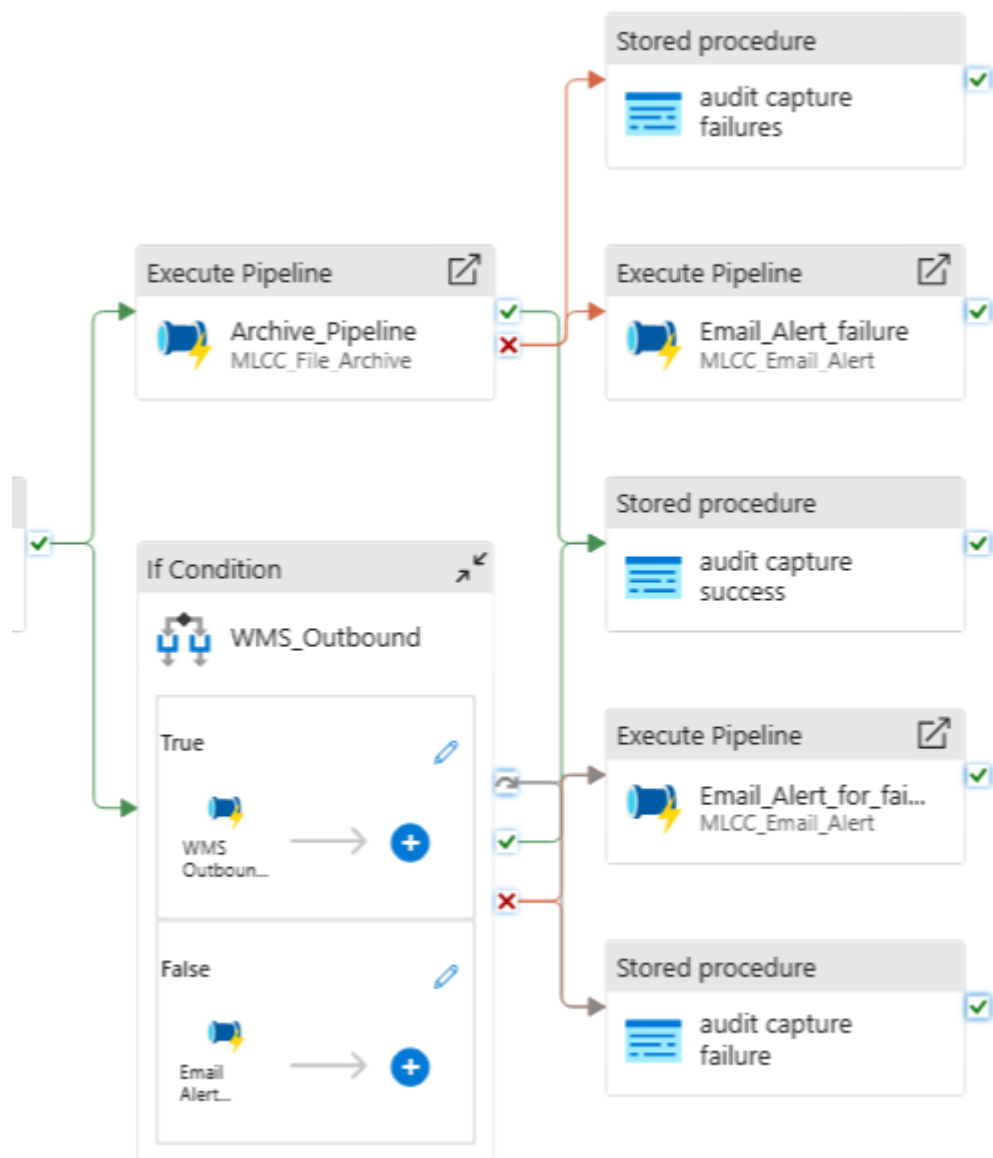
This expression is a condition that checks if the number of rows copied during the execution of an activity called 'Data load to MLCC\_HIST table' is greater than or equal to 1.



General <b>Activities (2)</b> User properties		
Expression ⓘ @greaterOrEquals(activity('Data load...		
Case	Activity	
True	WMS Outbound...	
	1 Activity	
False	Email Alert _no_...	
	1 Activity	

The WMS outbound pipeline loads data to the final oracle DB and if it fails an email alert is sent by triggering email alert pipeline

The last part of the MLCC\_ADF\_pipeline is shown zoomed below



Triggers

MLCC_Trigger_CarharttDC4_SMRU73201_0240	Schedule	⏸ Stopped	1
MLCC_Trigger_CarharttDC4_SMRU73201_0600	Schedule	⏸ Stopped	1
MLCC_Trigger_Clorox_SMRU1846_0545	Schedule	✔ Started	1
MLCC_Trigger_Clorox_SMRU1846_0700	Schedule	⏸ Stopped	1
MLCC_Trigger_CloroxAberdeen_SMRU1042_0545	Schedule	✔ Started	1
MLCC_Trigger_Danone_SMRU0085_0345	Schedule	✔ Started	1
MLCC_Trigger_Duracell_SMRU12951_0545	Schedule	✔ Started	1
MLCC_Trigger_Energizer_Monroe_SMRU80700_0545	Schedule	✔ Started	1
MLCC_Trigger_EnergizerFranklin_SMRU80828_0545	Schedule	✔ Started	1
MLCC_Trigger_EnergizerFranklin_SMRU80828_1745	Schedule	✔ Started	1
MLCC_Trigger_Ferrero_SMRU72129_0545	Schedule	✔ Started	1
MLCC_Trigger_FerreroGoodyear_SMRU80604_0545	Schedule	✔ Started	1
MLCC_Trigger_GeneralMills_Fort_Wayne_SMRU1856_0545	Schedule	✔ Started	1
MLCC_Trigger_GM_Palmyra_SMRU1857_0545	Schedule	✔ Started	1