



cloud  
lunch & learn

# Marathon



Is there a way that we can build our Azure Synapse Pipelines all with parameters based on MetaData?

Erwin de Kreuk

Lead Data and AI @InSpark

13 May 10:00 - 11:00 (UTC)

13 May 12:00 - 13:00 (UTC+01:00) CET

All sessions available on [www.cloudlunchlearn.com/marathon](http://www.cloudlunchlearn.com/marathon)

Sponsored by



Microsoft





<https://www.linkedin.com/in/erwindekreuk/>



<https://erwindekreuk.com>



@erwindekreuk

# Erwin de Kreuk

---

Lead Data and AI @InSpark

# We Are InSpark

We help organizations  
**accelerating their digital  
transformation with impactful  
Microsoft solutions & expertise**

# Azure Synapse Analytics



Data integration

Data warehousing



Big data analytics

- Pipeline Parameter
- Dataset Parameter
- Notebook Parameter
- Linked Service Parameter
- Dataflow Parameter



# Linked Services Parameters

Connect to different Database  
on same Server

Connect to different Logical  
Servers

Parameters

New | Delete

Name	Type	Default value
Database	String	WideWorldImporters

Advanced

Add dynamic content

@linkedService().Database

Clear contents

Filter...

Learn about linked service parameterization [here](#)

Parameters

Database

# Dataset Parameters

Create 1 dataset for all your  
Linked Services activities

The screenshot shows the 'Parameters' tab of a dataset configuration. At the top, there are three dark blue rounded rectangular boxes labeled 'FileSystem', 'Directory', and 'FileName'. Below them, the dataset details are shown: 'Parquet' file format and 'DS\_ADLS\_RAW' name. The 'Parameters' tab is selected, showing three input fields: 'Linked service \*' (set to 'LS\_ADLS\_DLS2'), 'File path \*' (containing three sub-fields: 'FilePath' (empty), 'Directory' (selected), and 'File' (empty)), and 'Compression type' (set to 'snappy'). Arrows from the 'FileSystem', 'Directory', and 'FileName' labels point to the corresponding 'FilePath', 'Directory', and 'File' fields respectively.

The screenshot shows the 'Parameters' tab of a dataset configuration, specifically the 'Parameters' section. It includes a header row with columns for 'Name', 'Type', and 'Default value'. Below this, there are two rows of parameter definitions:

Name	Type	Default value
FilePath	String	Value
Filename	String	Value

# Dataset Parameters

- Create 1 dataset for all your Linked Services activities

The screenshot shows the 'Parameters' tab for a Parquet dataset named 'DS\_ADLS\_RAW'. The 'File path' field contains the expression 'raw / @dataset().FilePath / @dataset().Filename'. Three arrows point from the labels 'FileSystem', 'Directory', and 'FileName' to the segments of the file path expression.

Parameter	Value
FileSystem	raw
Directory	@dataset().FilePath
FileName	@dataset().Filename

Add dynamic content

@dataset().FilePath

Clear contents

The screenshot shows the 'Parameters' tab of the 'Add dynamic content' dialog. It displays two parameters: 'FilePath' and 'Filename', both defined as 'String' type with a 'Value' placeholder.

Name	Type	Default value
FilePath	String	Value
Filname	String	Value

# Dataflow Parameters

## Pass parameters to Dataflow

The screenshot shows the Azure Data Factory pipeline editor interface. At the top, a data flow diagram is visible, consisting of several stages: a source (Azure Blob Storage), followed by a series of transformation steps including 'DataStoreExcludeColumns', 'IsCurrent', 'RemoveAuditColumnsActive', and 'MarkAsInsert'. Below the diagram, the 'Parameters' tab is selected in the 'Settings' section of the pipeline properties.

**Parameters Tab (Pipeline Settings):**

Name	Type	Default value
TargetFilePath	string	'DeltaLake/' + \$TargetFilePath + '/' + \$TargetFileName + '/Original'
TargetFileName	string	
ProcessType	string	
DataStoreExcludeColumns	string	

**Data Flow Parameters:**

Name	Value	Type	Expression
TargetFilePath	@pipeline().parameters.FilePath	string	
TargetFileName	@pipeline().parameters.FileName	string	
ProcessType	@pipeline().parameters.ProcessType	string	
DataStoreExcludeColum...	@pipeline().parameters.DataStoreExcl...	string	

An arrow points from the 'Value' column of the Data Flow Parameters table to the 'Default value' field in the Parameters tab, indicating the mapping between pipeline parameters and data flow parameters.

**Expression Elements:** All, Functions, **Parameters**, Cached lookup

**Expression values:** Filter by keyword, Create new, abc TargetFilePath, abc TargetFileName

# Notebook Parameters

Pass Parameters from  
Synapse to Databricks

```
Cmd 5
1 # Creating widgets for leveraging parameters, and printing the parameters
2
3 dbutils.widgets.text("datalake_connection", "", "")
4 datalake_connection = dbutils.widgets.get("datalake_connection")
5 #if not datalake_connection or not datalake_connection.isstring()
6 print("Data Lakedata Connection:\n" + datalake_connection)
7
8 dbutils.widgets.text("source_file", "", "")
9 source_file = dbutils.widgets.get("source_file")
10 print("\nSource File Path:\n" + source_file)
11
12 dbutils.widgets.text("IsIncremental", "", "")
13 IsIncremental = dbutils.widgets.get("IsIncremental")
14 print("\nIs Incremental loading =\n" + IsIncremental)
15
16 dbutils.widgets.text("target_file", "", "")
17 target_file = dbutils.widgets.get("target_file")
18 print("\nTarget File:\n" + target_file)
19
20 dbutils.widgets.text("target_file_path", "", "")
```

<input type="checkbox"/> datalake_connection	@activity('Get EnvironmentSettings').o...
<input type="checkbox"/> target_file	@{pipeline().parameters.TargetFileNa...}
<input type="checkbox"/> target_PKColumns	@{pipeline().parameters.PKColumns}
<input type="checkbox"/> source_file	@{pipeline().parameters.FilePath}/@{pi...
<input type="checkbox"/> target_file_path	@pipeline().parameters.TargetFilePath
<input type="checkbox"/> HashExcludeColumns	@pipeline().parameters.HashExcludeC...
<input type="checkbox"/>	

# Pipeline Parameters

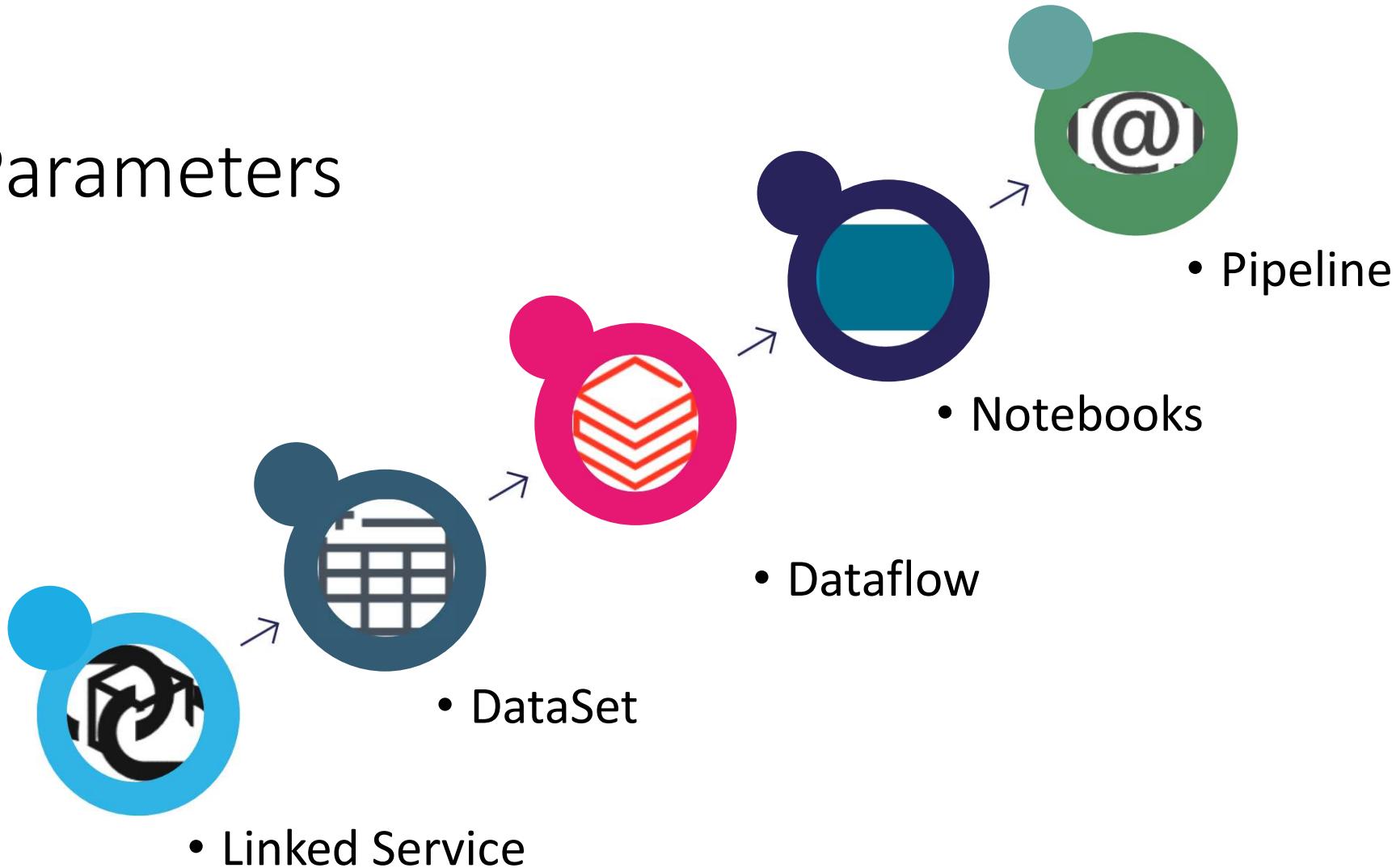
Can be used across all your Pipelines

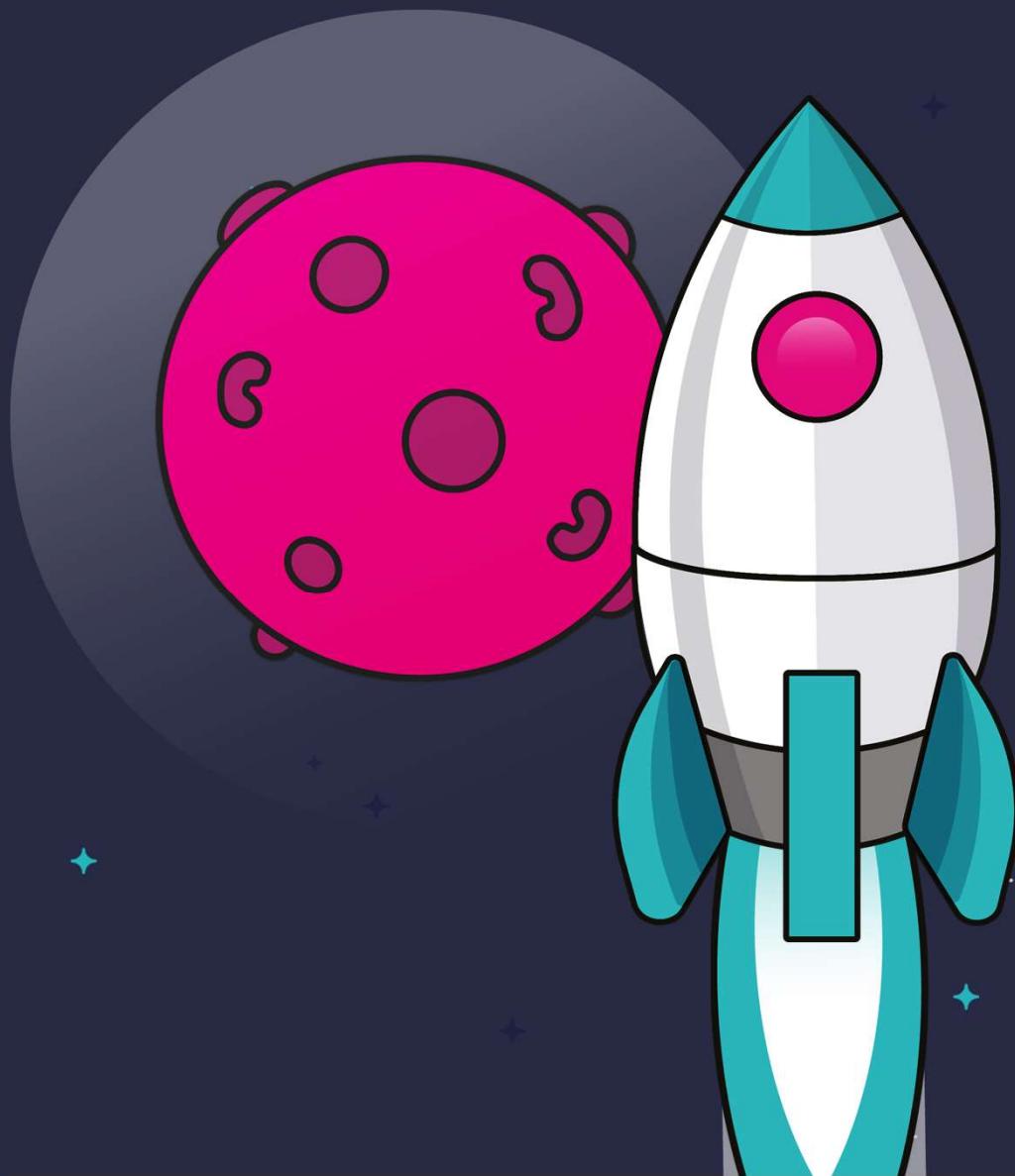
The screenshot shows the 'Parameters' tab in the Azure Data Factory pipeline configuration. A new parameter named 'FilePath' has been created, defined as a String type with a dynamic value of '@pipeline().parameters.FilePath'. The 'Sink' tab is selected, and the 'DS\_ADLS\_RAW' dataset is chosen as the sink dataset. The 'Dataset properties' section shows two parameters: 'FilePath' and 'Filename', both mapped to 'Value'.

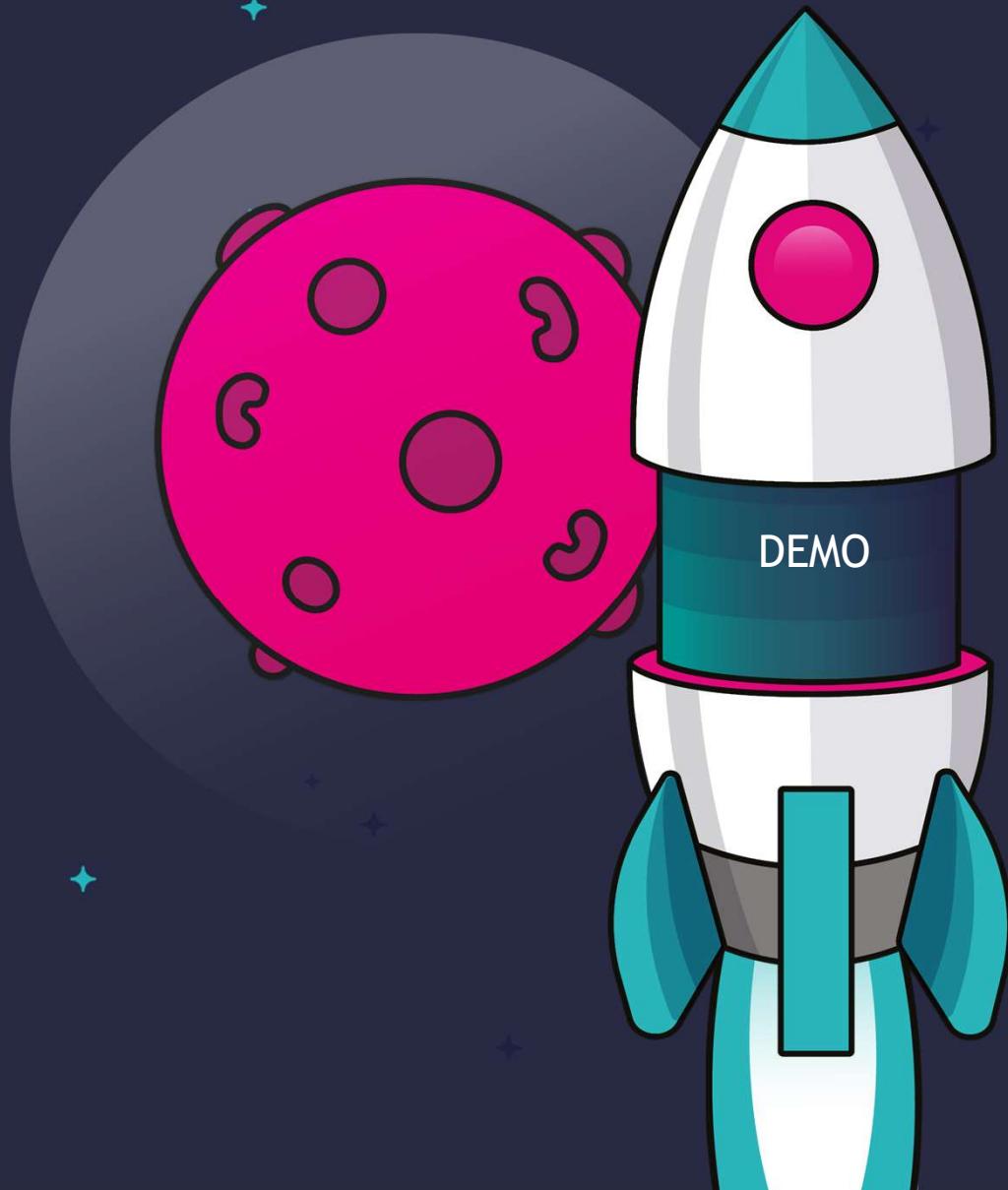
Name	Type	Default value
FilePath	String	@pipeline().parameters.FilePath
FileName	String	

NAME	VALUE	TYPE
FilePath	Value	string
Filename	Value	string

# Parameters





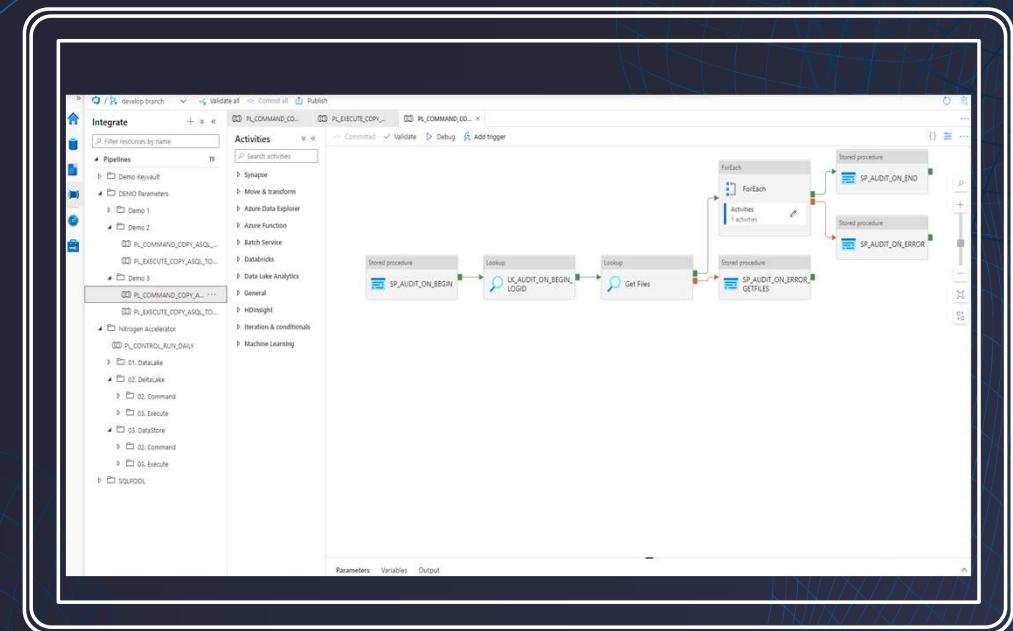


Innovate to accelerate

# DEMO

# Metadata

Load your pipelines  
dynamically



Marathon

# *Can we get answers on the following questions?*



Can we build Synapse Pipelines dynamically?



Can we extract data from my sources based on MetaData?



Can we log the execution of the Pipelines?



Can we load the active(current) or historical records to a DataStore?



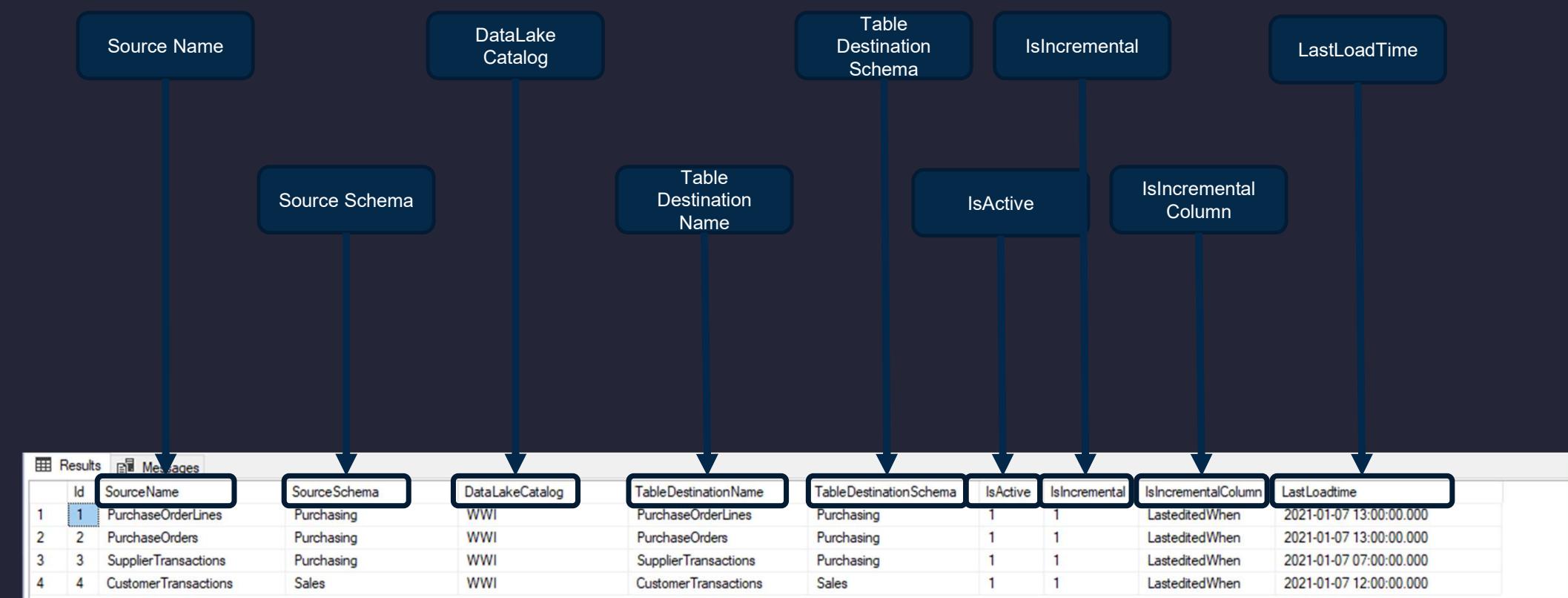
Can we build history from extracted data based on MetaData?



YES WE  
CAN.

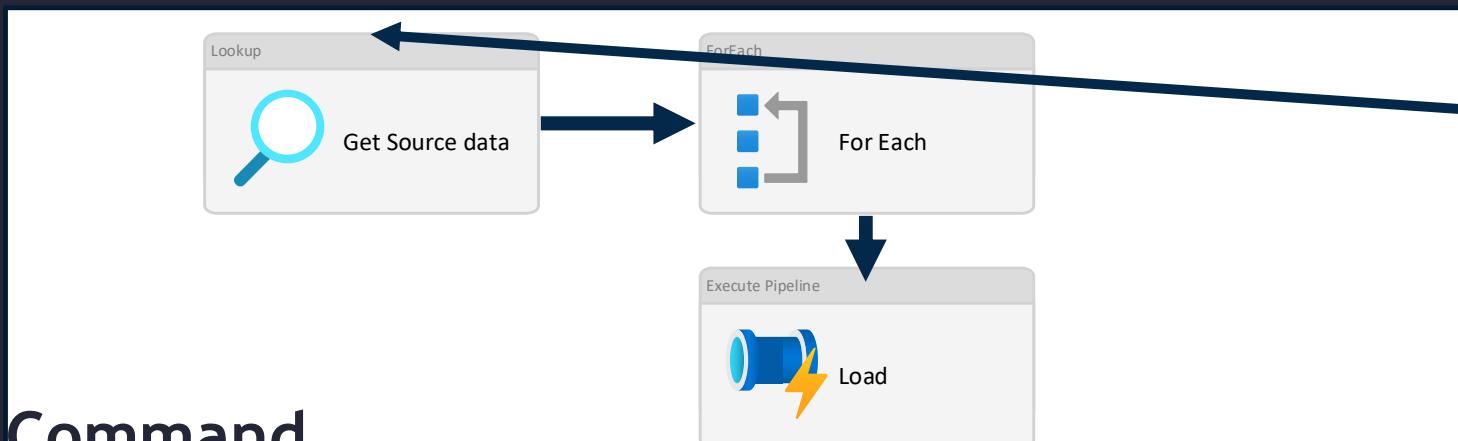
# Metadata

Source Parameter table



# Metadata

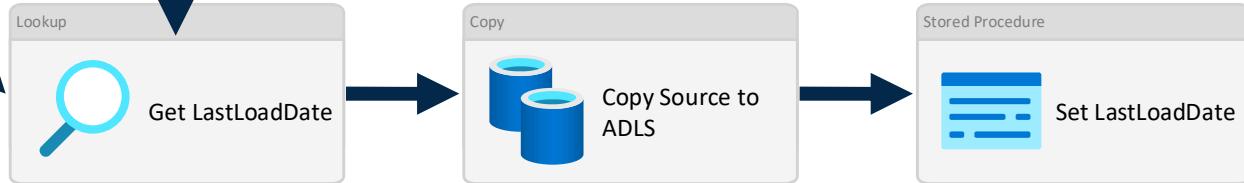
Source Parameter table



```
SELECT [PipelineParameterId],  
       ,[SourceName],  
       ,[SourceSchema],  
       ,[SelectQuery],  
       ,[SelectLastLoadDate],  
       ,[FilePath],  
       ,[FileName],  
       ,[TableDestinationName],  
       ,[ProcessType],  
       ,[IsActive],  
       ,[IsIncremental],  
       ,[IsIncrementalColumn],  
       ,[LastLoadtime]  
  FROM [execution].[Pipeline_DataLake_Files]
```

# Command

```
SELECT case when 1=1 then  
convert(varchar,max>LastEditedWhen),120) else  
convert(varchar,getdate(),120) end as LastLoadDate  
FROM SourceSchema.SourceTable
```



# Execute



Marathon

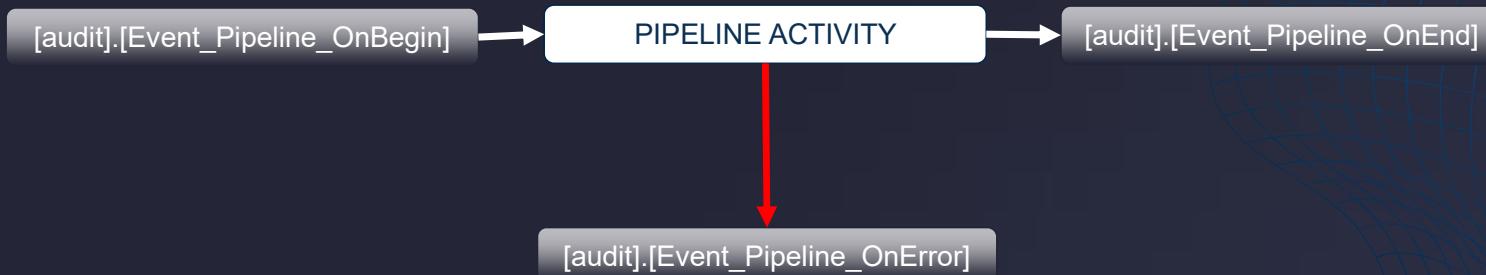
# Logging

Log Start and End Time of records

Log Extracted Records

Log Execution Failure

Create Pipeline\_ExecutionLog table





## Marathon

# Logging

### BEGIN

Insert new Record  
Insert Metadata  
Insert Start time

### END

End Time  
Status(1)  
Row Counts  
Pipeline Details

### ERROR

End Time  
Status(2)  
Failure Message

Log Start and End Time of records

Log Extracted Records

Log Execution Failure

Create Pipeline\_ExecutionLog table

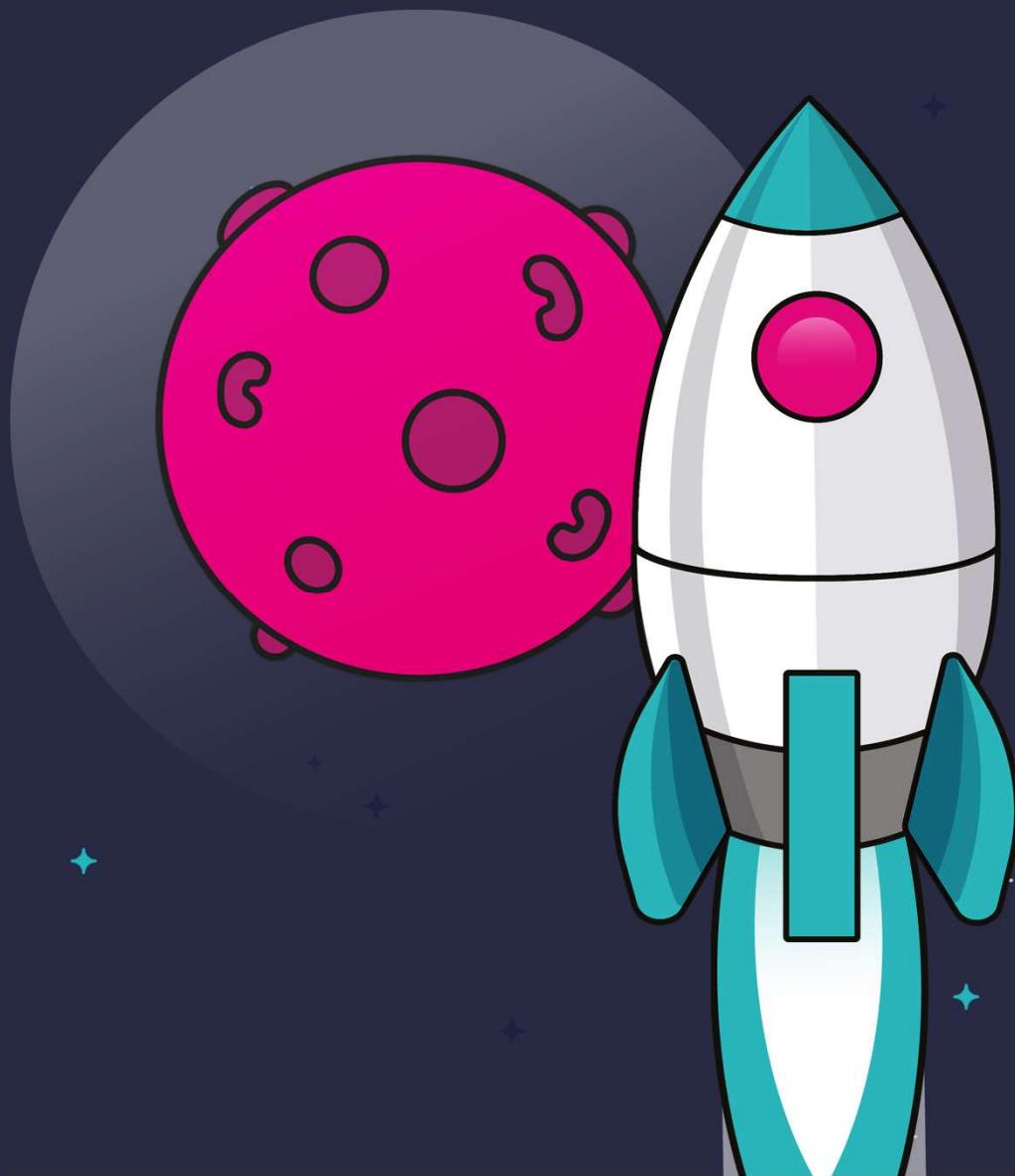
## Pipeline\_ExecutionLog

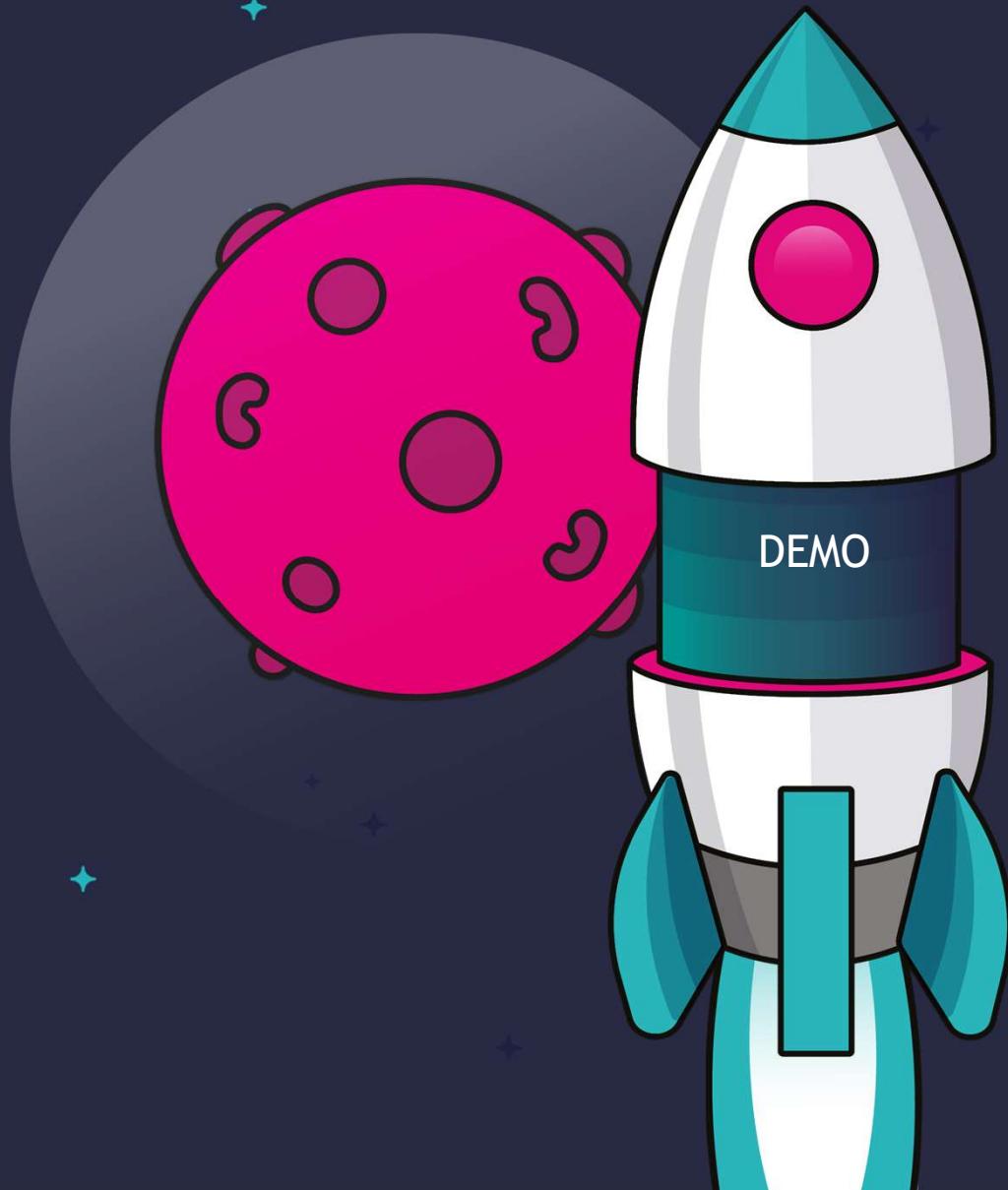
Results		Messages	
LogId	StartTime	EndTime	
1	5	2021-02-08 19:55:35.517	2021-02-08 19:55:57.687
2	4	2021-02-08 19:55:04.670	2021-02-08 19:55:28.327
3	3	2021-02-08 19:54:35.920	2021-02-08 19:54:58.937
4	2	2021-02-08 19:54:09.117	2021-02-08 19:54:29.653
5	1	2021-02-08 19:53:57.763	NULL

LogId	StartTime	En...	Status	FailureMessage
1	15	2021-02-0...	20...	1
2	14	2021-02-0...	20...	1
3	13	2021-02-0...	20...	2
4	12	2021-02-0...	20...	1
5	11	2021-02-0...	20...	2

Failure happened on 'Source' side. ErrorCode=SqlOperationFailed,'Type=Microsoft.DataTransfer.Common.Shared.HybridDeliveryException,Message=A database operation failed with the following error: '.





Innovate to accelerate

# DEMO

# *Can we get answers on the following questions?*



Can we build Synapse Pipelines dynamically?



Can we load the active(current) or historical records to a DataStore?



Can we extract data from my sources based on MetaData?



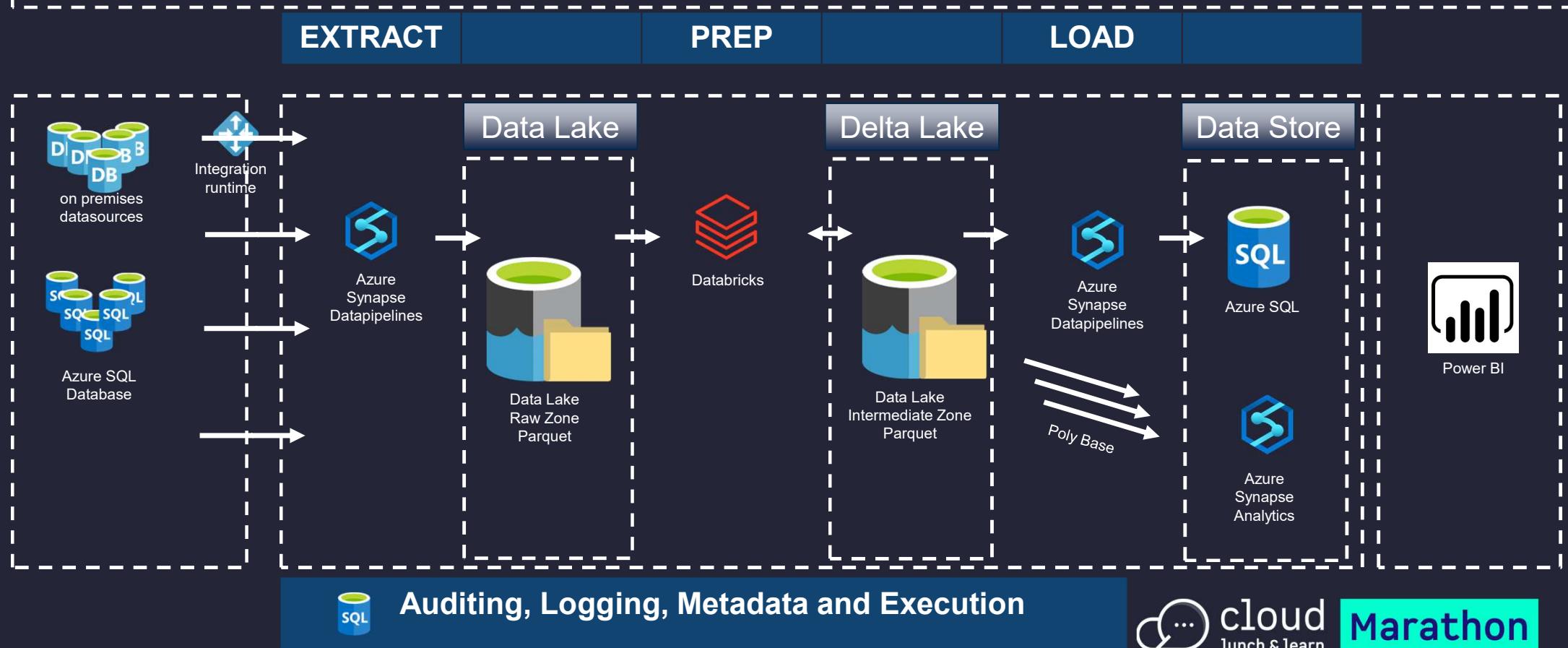
Can we build history from extracted data based on MetaData?

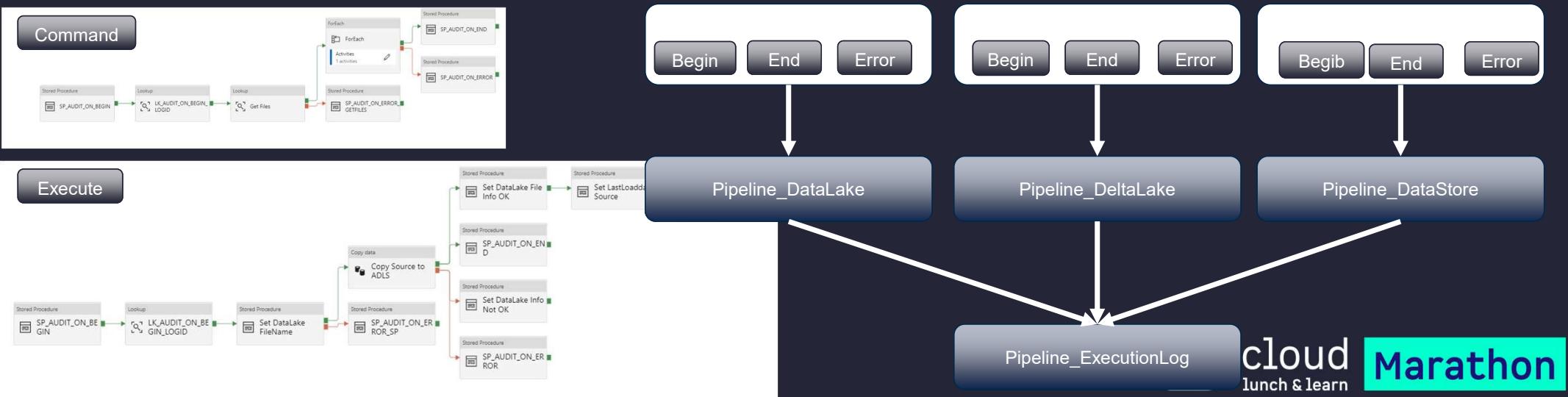
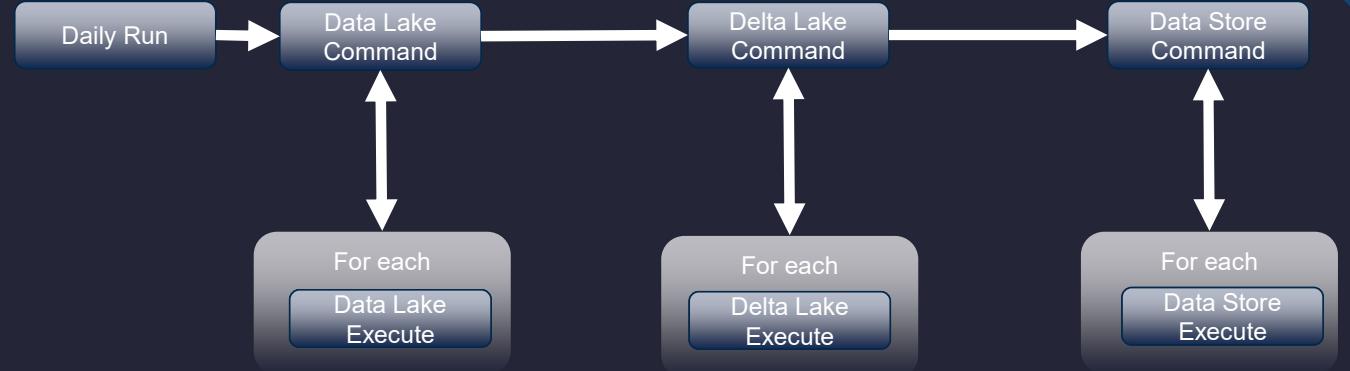


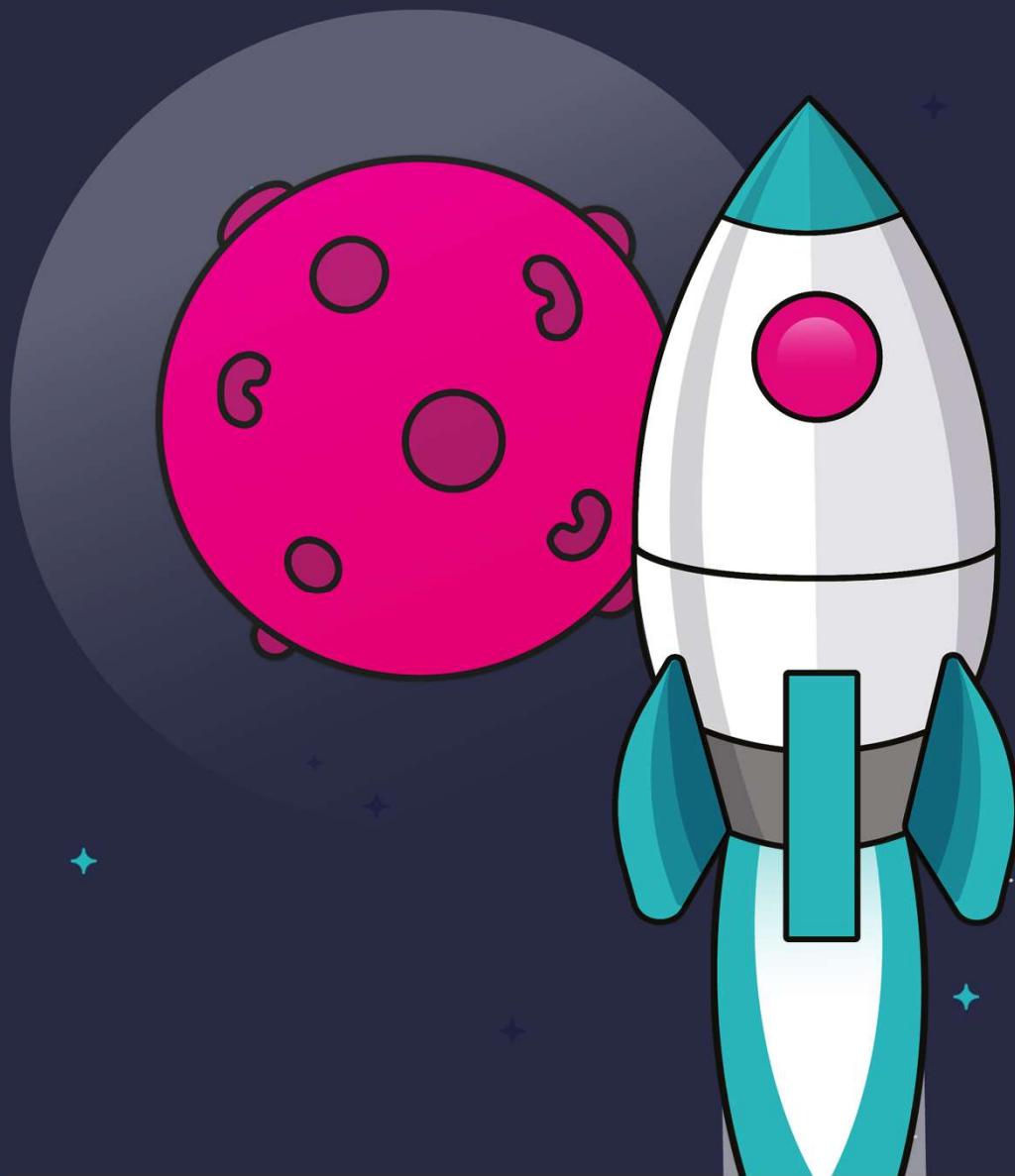
Can we log the execution of the Pipelines?

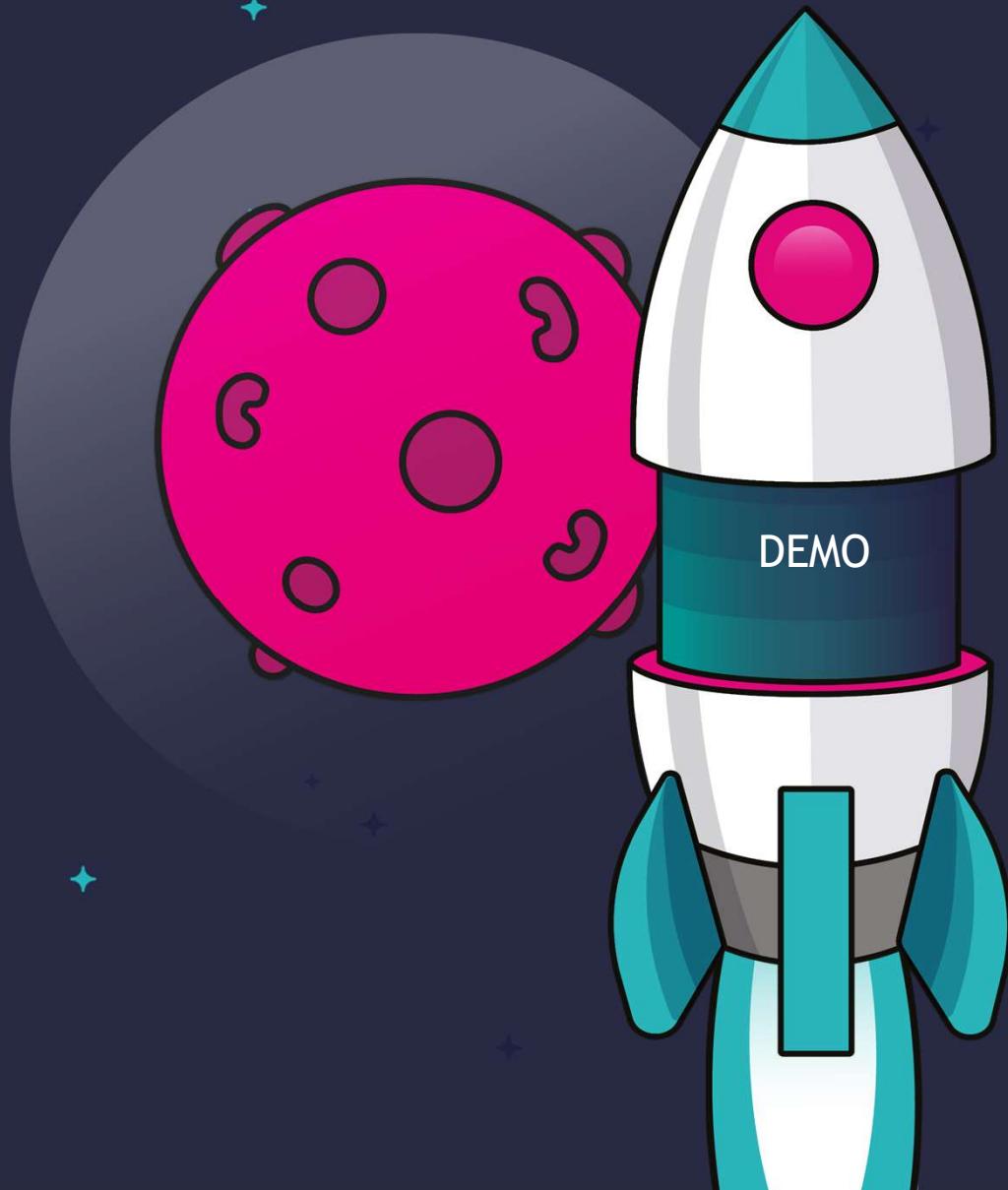


# HIGH OVERVIEW ARCHITECTURE NITROGEN Data Integration









Innovate to accelerate

# DEMO

# *Can we get answers on the following questions?*



Can we build Synapse Pipelines dynamically?



Can we load the active(current) or historical records to a DataSt...



Can we extract data from my sources based on MetaData?



Can we build history from extracted data based on MetaData?



Can we log the execution of the Pipelines?



# Questions?



<https://erwindekreuk.com>



@erwindekreuk



<https://www.linkedin.com/in/erwindekreuk/>

Slides will be available on my blog

cloud  
lunch & learn **Marathon**



cloud  
lunch & learn

Marathon

# Thank you!

Share your Love with the speaker and team on social, using the hashtag `#cllmarathon` and `@CloudLunchLearn`

And follow us 😊



[youtube.com/c/CloudLunchandLearn](https://youtube.com/c/CloudLunchandLearn)



[@CloudLunchLearn](https://twitter.com/CloudLunchLearn)

In collaboration with

