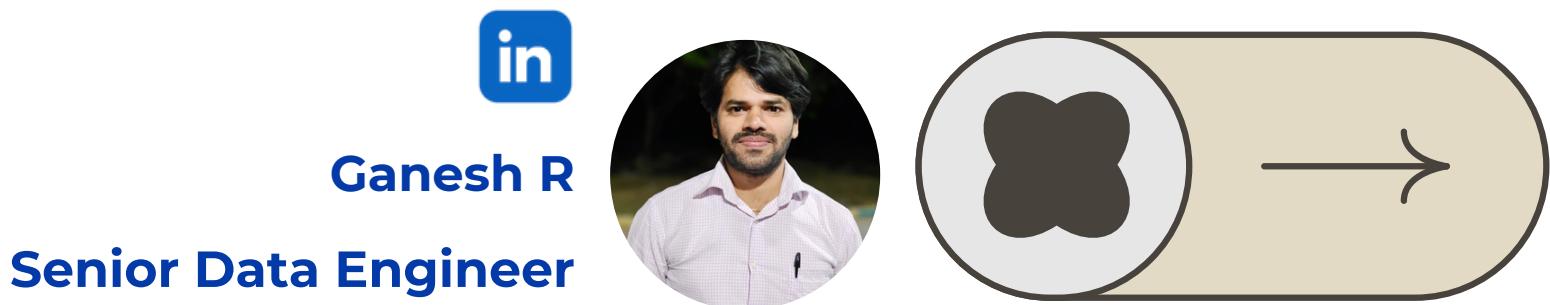
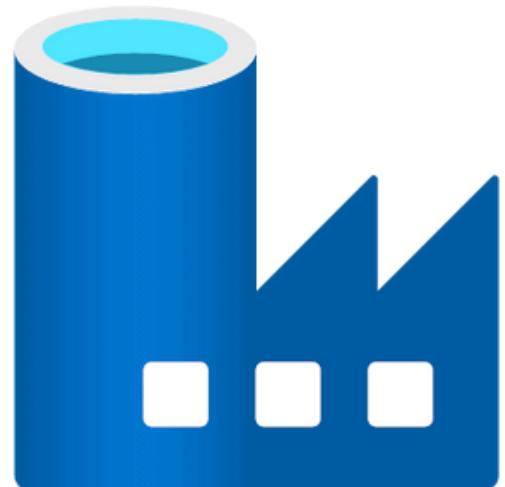


Different Activities in Azure Data Factory



Ganesh R

Senior Data Engineer



Azure Data Factory (ADF) activities are categorized based on their functions within a data pipeline. Here's an overview of the main categories of activities in ADF

Categories:

1

Data Movement Activities

2

Data Transformation Activities

3

Data Control Activities

4

External Control Activities

5

Monitoring and Debugging Activities

6

Custom and Specialized Activities



Data Movement Activities

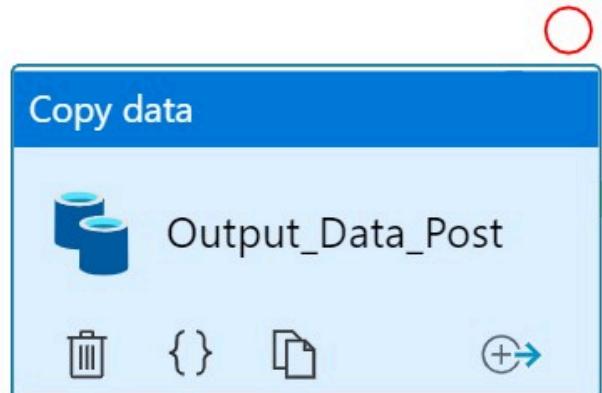


<https://www.linkedin.com/in/rganesh0203/>

Copy Activity:

Copies data between different data stores (e.g. Azure Blob Storage, SQL Server, Amazon S3).

Save Save as template Validate Validate copy runtime Debug Add trigger



General Source Sink Mapping Settings User properties

Source dataset * DS_SQLTABLE Open New Preview data Learn more

Use query Table Query Stored procedure

Query `select top 30 * from anytable`

Query timeout (minutes) 120

Isolation level None

Partition option None Physical partitions of table Dynamic range

i Please preview data to validate the partition settings are correct before you trigger a run or publish the pipeline.

Additional columns New



<https://www.linkedin.com/in/r Ganesh0203/>

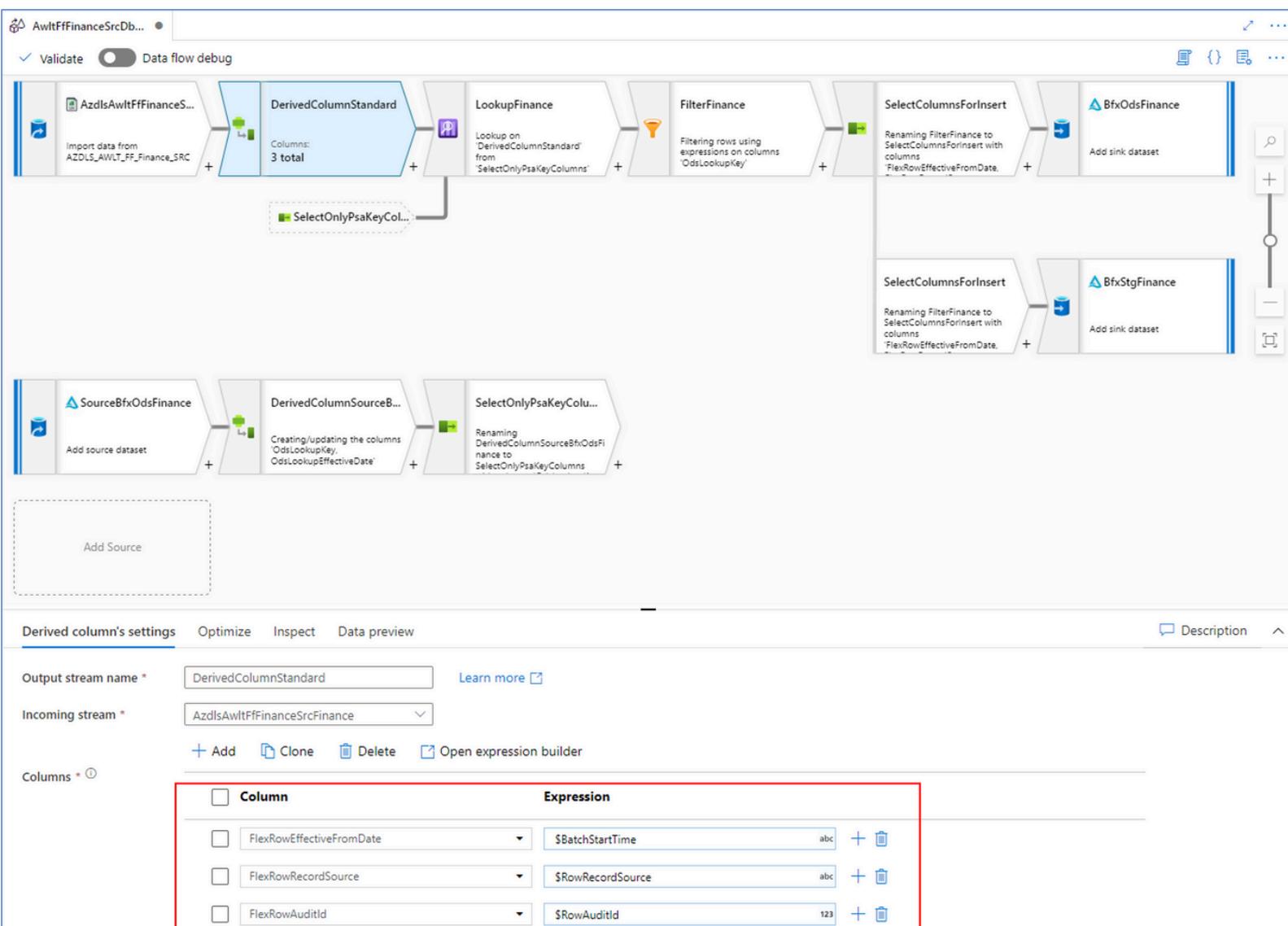
Data Transformation Activities



<https://www.linkedin.com/in/rganesh0203/>

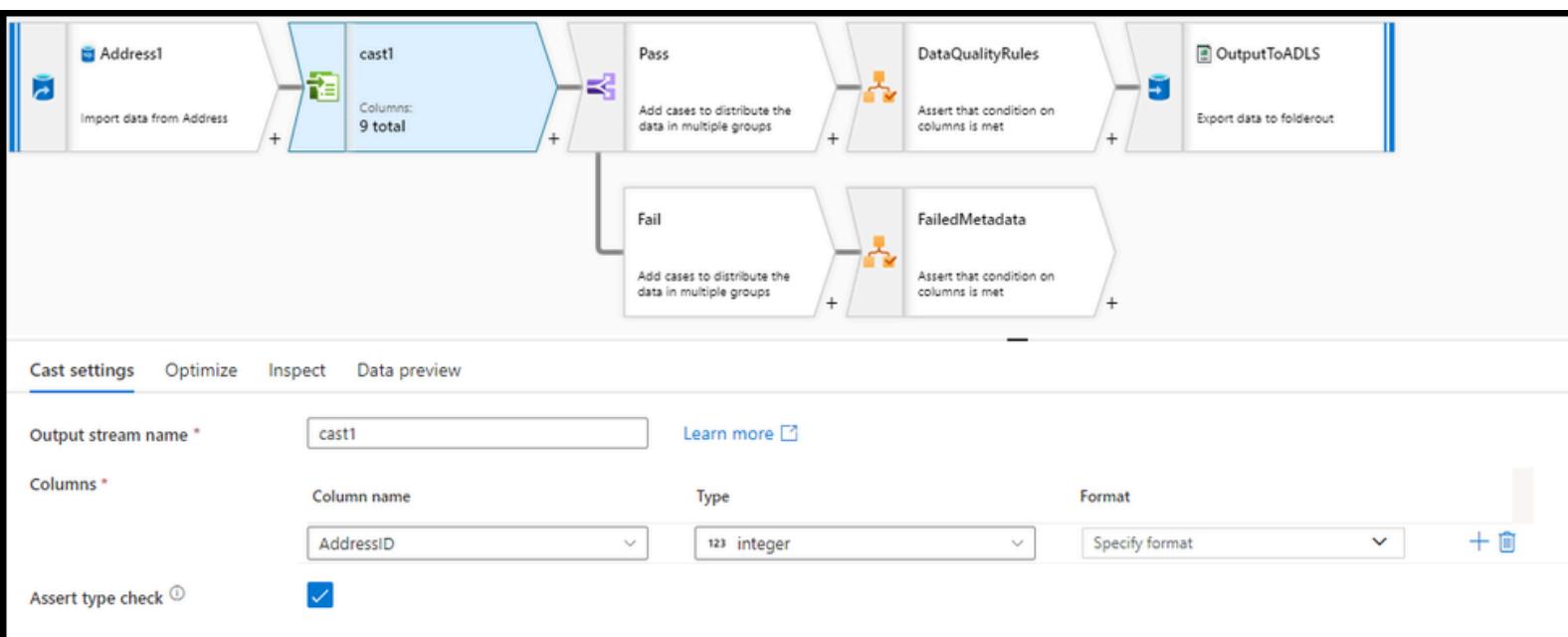
Mapping Data Flow:

Performs data transformations using a graphical interface.



Wrangling Data Flow:

Cleans and shapes data using Power Query.



<https://www.linkedin.com/in/r Ganesh0203/>

Stored Procedure Activity:

Calls stored procedures in a database.

The screenshot shows the Azure Data Factory pipeline editor interface. At the top, there is a title bar with the pipeline name "pipeline1". Below the title bar, the main workspace displays a single activity: a "Stored procedure" activity named "Stored procedure1". The activity icon is blue with white horizontal lines, and it has a red circle icon above it. To the left of the workspace, there is a sidebar titled "Activities" with a search bar containing the text "store". Under the search bar, there is a section for "General" activities, with "Stored procedure" selected. The bottom half of the screen shows the "Settings" tab for the "Stored procedure1" activity. The "General" tab is selected, while the "Settings" tab is highlighted with a red box. The "User properties" tab is also visible. In the "Settings" tab, there are fields for "Linked service" (with a dropdown menu and a "New" button) and "Stored procedure name" (with a dropdown menu, a "Refresh" button, and an "Edit" checkbox). Below these fields is a collapsed section for "Stored procedure parameters". On the right side of the screen, there is a "Properties" panel with tabs for "General" and "Related". The "General" tab is selected, showing fields for "Name" (set to "pipeline1") and "Description". There are also sections for "Annotations" and a "+ New" button.



Databricks Notebook Activity

Runs a Databricks notebook for advanced transformations.

The screenshot shows the Azure Data Factory designer interface. On the left, the 'Activities' pane is open, with the 'Databricks' category selected (indicated by a red box). A red arrow points from the 'Notebook' icon in this list to the 'Notebook' activity component in the main canvas area. The main canvas contains a single 'Notebook' activity named 'Notebook1'. The properties pane for this activity is visible, showing the following settings:

| General | Azure Databricks | Settings | User properties |
|------------------|--------------------------|------------|-----------------|
| Name * | Notebook1 | Learn more | |
| Description | | | |
| Timeout ⓘ | 7.00:00:00 | | |
| Retry ⓘ | 0 | | |
| Retry interval ⓘ | 30 | | |
| Secure output ⓘ | <input type="checkbox"/> | | |
| Secure input ⓘ | <input type="checkbox"/> | | |

Below the main canvas, there is a second, partially visible row of settings for another activity, showing identical timeout and retry configurations.



<https://www.linkedin.com/in/r Ganesh0203/>

Azure HDInsight Activities:

Hive

Processes data using Hive scripts.

Pig

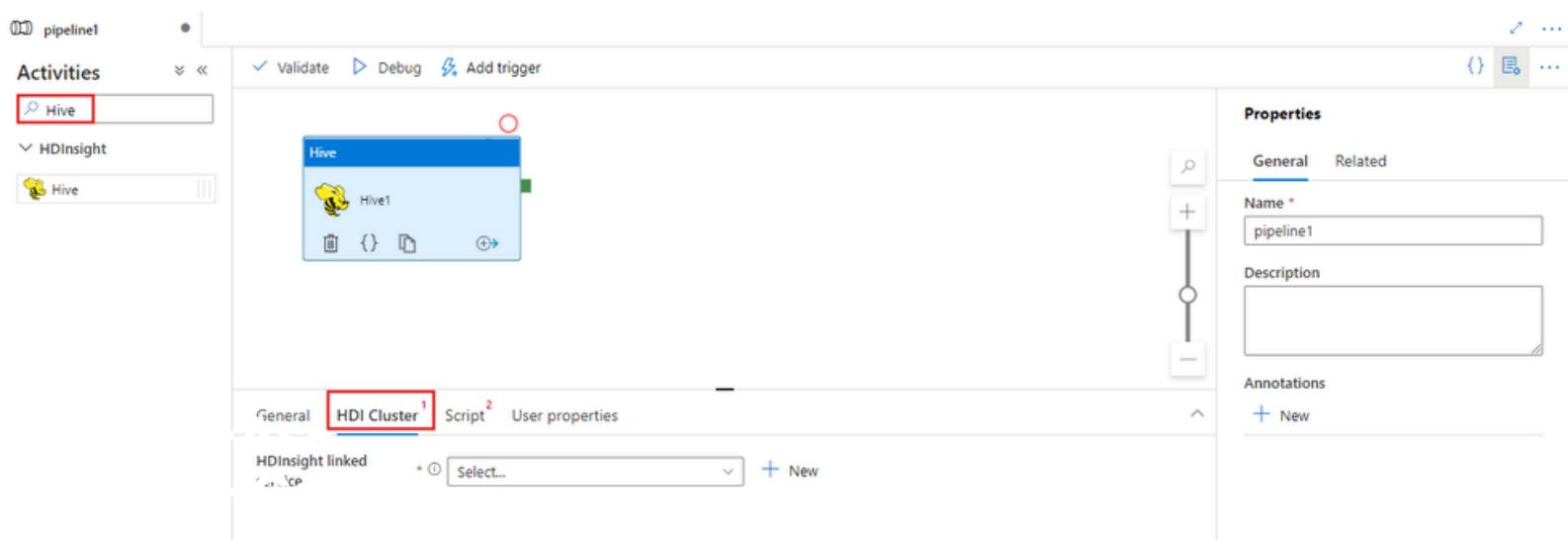
Processes data using Pig scripts.

Spark

Executes Spark jobs.

MapReduce

Runs MapReduce programs.



<https://www.linkedin.com/in/rganesh0203/>

Data Control Activities

Filter

Filters data based on a condition.

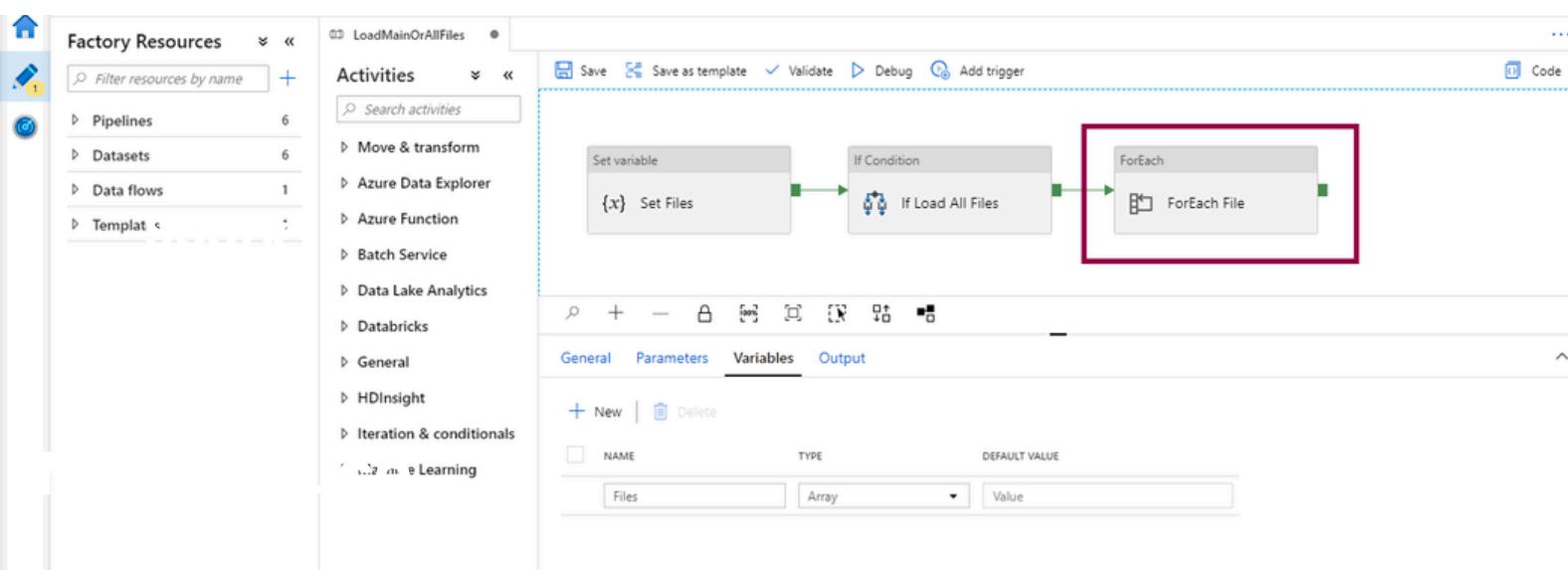
The screenshot shows the Microsoft Power Automate designer interface. On the left, there's a sidebar with a search bar for 'Activities' and a section for 'Iteration & conditionals'. A 'Filter' activity is selected, highlighted with a red border. The main workspace contains a single 'Filter' activity named 'Filter1'. To the right is the 'Properties' panel, which is currently displaying the 'General' tab. The 'Name' field is set to 'pipeline1'. Below it is a large red box highlighting the 'Settings' tab and its contents. The 'Settings' tab shows two fields: 'Items' set to '@variables('AnimalsArray')' and 'Condition' set to '@not>equals(item(), 'Bear'))'. There are also tabs for 'User properties' and a collapse button '^'.



<https://www.linkedin.com/in/r Ganesh0203/>

ForEach

Iterates over a collection.



If Condition:

Executes activities based on a true/false condition.

The screenshot shows the Microsoft Power Automate designer interface. A central modal window titled "If Condition" is open, showing two branches: "True" and "False", both labeled "No activities". The "True" branch has a red box around its edit icon. The "False" branch also has a red box around its edit icon. Below the modal, the main canvas shows a single "If Condition" step. On the left, the "Activities" pane shows an "If" step selected. On the right, the "Properties" pane shows the pipeline is named "pipeline1". The "Activities" tab is selected in the pipeline properties. A tooltip "This property should be parameterized." is shown over the "Expression" field, which is also highlighted with a red box. Another red box highlights the "Add dynamic content [Alt+Shift+D]" link below the expression field. The bottom section shows a table for mapping cases to activities, with rows for "True" and "False", both mapped to "No activities".



Switch

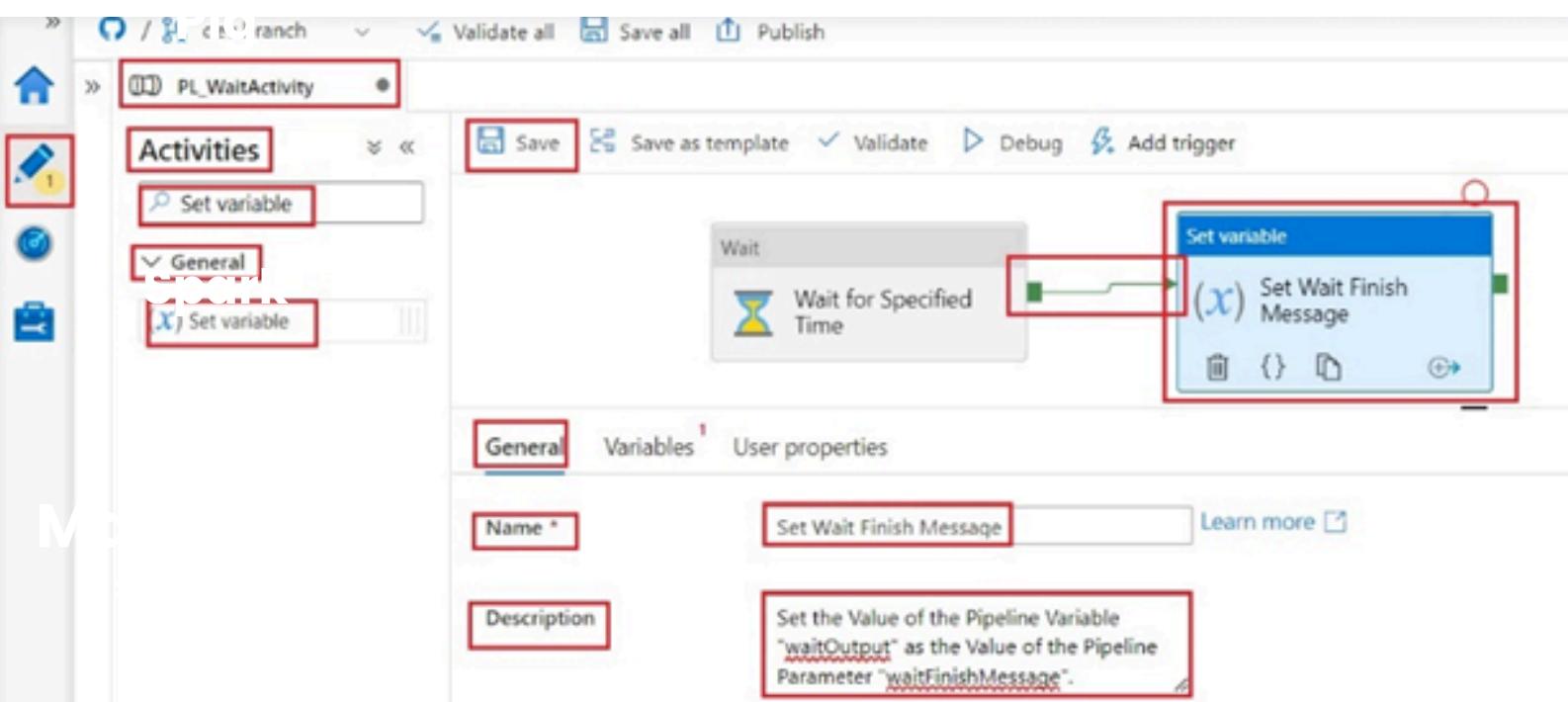
Routes data flow based on matching conditions.

The screenshot shows the Azure Data Factory pipeline editor interface. On the left, there's a sidebar with 'Activities' and a 'Switch' item highlighted with a red box and circled in blue. Below it are sections for 'Iteration & conditionals' and another 'Switch' item. The main workspace has a 'Validate', 'Debug', and 'Add trigger' toolbar at the top. A 'Properties' panel on the right shows 'General' tab selected with 'Name * pipeline1'. The central area displays a 'Switch' activity named 'Switch1'. It has a 'Default' case with 'No activities' and a single '1' case also with 'No activities'. The 'Activities (0)' tab is selected in the ribbon. An 'Expression' field contains '@string(rand(0, 1))'. Below it, a '+ Add case' button is highlighted with a red box and circled in blue. A table lists cases: 'Default' and '1'. The '1' row is highlighted with a red box and circled in blue. A small edit icon is also circled in blue next to the '1' row. The bottom right corner features a large blue arrow pointing right.



Wait

Introduces a delay.



Until

Loops until a condition is met.

The screenshot shows the Microsoft Power Automate designer interface. A central activity card for an 'Until' loop is displayed, with its 'Activities' section currently empty. The 'Properties' pane on the right shows the pipeline is named 'pipeline1'. The 'Settings' tab is selected, revealing an 'Expression' field which contains the message 'This property should be parameterized.' and a 'Timeout' field set to '7.00:00:00'. A red box highlights the 'Expression' field, and another red box highlights the 'Add dynamic content [Alt+Shift+D]' link below it.

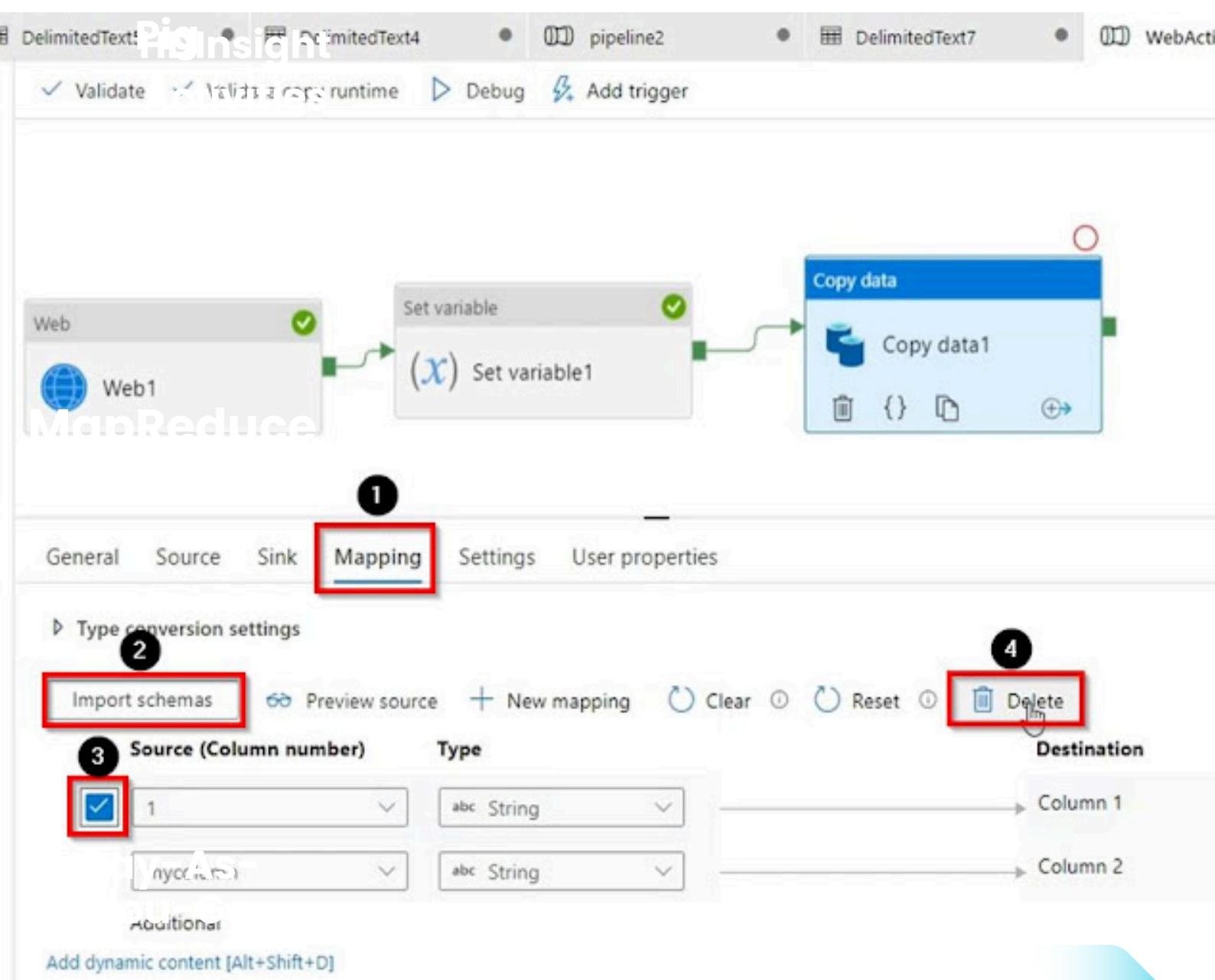


<https://www.linkedin.com/in/r Ganesh0203/>

External Control Activities

Web

Calls REST APIs or web services.



<https://www.linkedin.com/in/r Ganesh0203/>

WebHook

Triggers activities based on an external event.

The screenshot shows the Microsoft Power Automate pipeline editor interface. A central workspace contains a single "WebHook" activity named "WebHook1". To the left, a sidebar lists "Activities" under "General", with "WebHook" selected. At the top, there are tabs for "Validate", "Debug", and "Add trigger". On the right, a "Properties" panel is open for the selected "WebHook1" activity. The "General" tab is selected, showing the following settings:

- Name:** pipeline1
- Description:** (empty)
- Annotations:** (empty)

The "Settings" tab is currently active, displaying the configuration for the WebHook:

- URL ***: (empty input field)
- Method ***: POST
- Headers**: + New
- Body ***: (empty input field)
- Timeout**: 00:10:00
- Disable certificate validation**:
- Report status on callback**:
- Authentication**: None



Executes Azure Functions.

Azure Function:

The screenshot shows the Azure Logic Apps designer interface. On the left, the 'Activities' sidebar is open, with the 'Azure Functions' section expanded. A red box highlights the 'Azure Function' activity icon. In the main workspace, there is one 'Azure Function' activity named 'Azure Function1'. To the right, the 'Properties' pane is visible, showing the 'General' tab with the name 'pipeline1'. The 'Settings' tab is selected, showing the 'Azure Function linked service' dropdown set to 'AzureFunction1' (also highlighted with a red box). Other settings include 'Function name' (YourAzureFunction), 'Method' (GET), and 'Headers'. A large blue arrow graphic is positioned in the bottom right corner of the slide.



<https://www.linkedin.com/in/r Ganesh0203/>

Execute Pipeline:

Invokes another pipeline.

The screenshot shows the Azure Data Factory pipeline editor interface. On the left, there's a sidebar with a search bar labeled 'Pipeline' and a list of activities: General (selected), Machine Learning, and Machine Learning Execution. The main area displays a single 'Execute Pipeline' activity named 'Execute Pipeline1'. The 'Settings' tab is selected, showing the following configuration:

- Invoked pipeline ***: A dropdown menu is open, with 'SendEmail' highlighted and selected.
- Wait on completion**: A checked checkbox.
- Parameters**: A table with one row:

| Name | Type | Value | Default value |
|----------|--------|-----------------------|-----------------------|
| receiver | string | youremail@contoso.com | youremail@contoso.com |

On the right, the 'Properties' panel shows the pipeline's name as 'pipeline1' and other basic details like description and annotations.



Execute SSIS Package

Executes an SSIS package in Azure-SSIS Integration Runtime.

- ▶ Databricks
- ▶ Move & Transform
- ▶ Data Lake Analytics
- ▶ General
- ▶ HDInsight
- ▶ Iteration & Conditionals
- ▶ Machine Learning

The screenshot shows the 'Execute SSIS Package' blade in the Azure portal. At the top, there's a preview of the package 'Execute_SSIS_AC'. Below it, the 'Settings' tab is selected. The configuration includes:

- Azure-SSIS IR ***: Set to 'SsisIR' (highlighted with a red box).
- Windows authentication**: Unchecked.
- 32-Bit runtime**: Unchecked.
- Folder ***: Set to 'ADF' (highlighted with a red box).
- Project ***: Set to 'SSIS_DEMO' (highlighted with a red box).
- Package ***: Set to 'Package.dtsx' (highlighted with a red box).
- Environment**: Set to 'Select...'.
- Logging level ***: Set to 'Basic' (highlighted with a red box). A checkbox for 'Customized' is also present.
- Manual entries**: An empty input field (highlighted with a red box).

A blue arrow points from the bottom right towards the bottom left corner of the slide.



<https://www.linkedin.com/in/rganesh0203/>

Lookup:

Retrieves data from a data source for use in subsequent activities.

The screenshot shows the Azure Data Factory pipeline editor interface. A 'Lookup' activity named 'Lookup1' is selected in the main workspace. The 'Properties' pane on the right shows the pipeline is named 'pipeline1'. The 'Settings' tab is selected in the 'Lookup' activity configuration pane. The 'Source dataset' dropdown is set to 'DelimitedText1', which is highlighted with a red box. Below it, the 'File path type' is set to 'File path in dataset'. Other settings include 'Recursively' checked, 'First row only' checked, and 'Skip line count' empty. The 'Annotations' section is empty.



Monitoring and Debugging Activities



<https://www.linkedin.com/in/r Ganesh0203/>

Get Metadata:

Retrieves metadata of a file or folder.

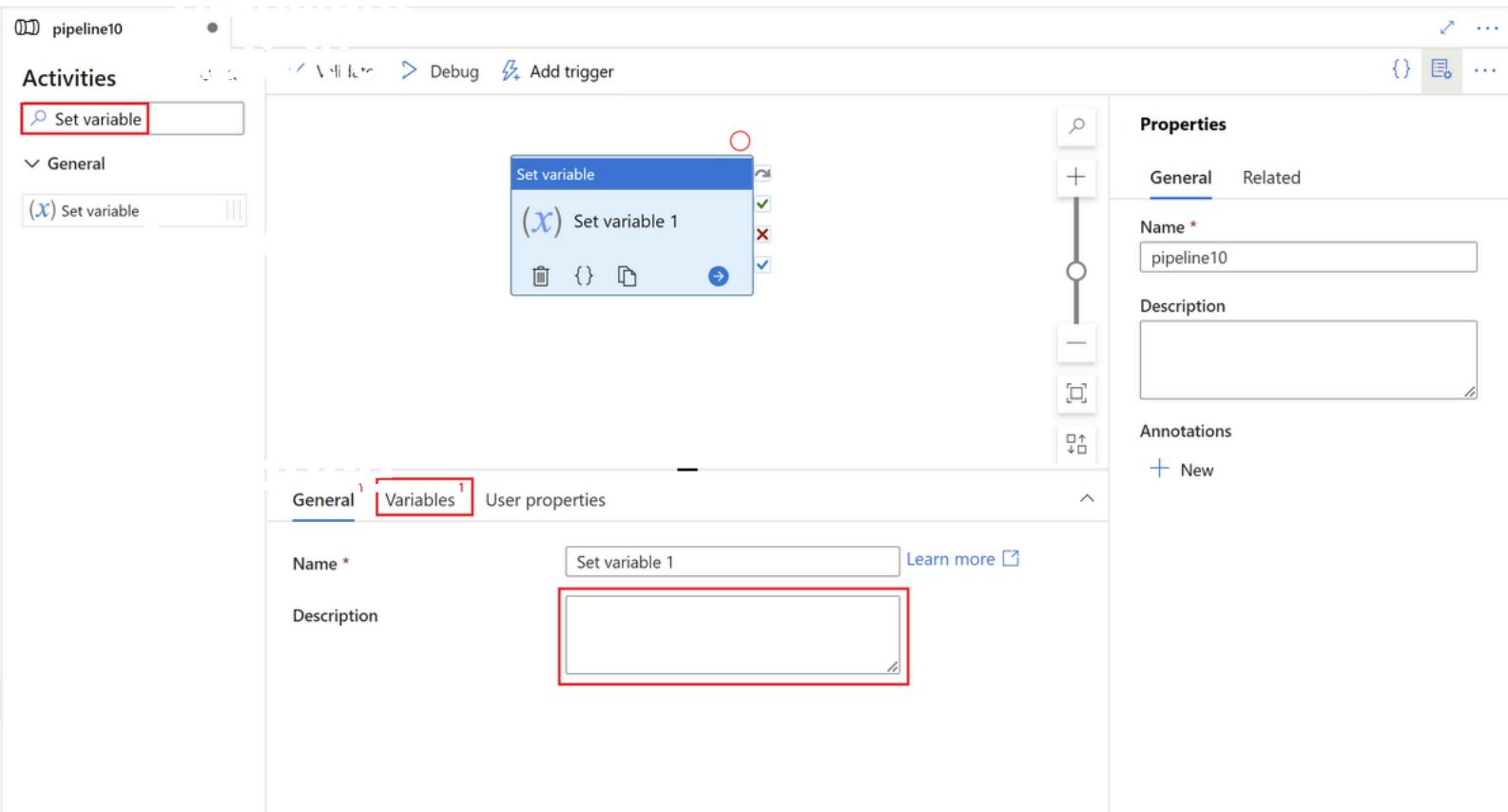
The screenshot shows the Azure Data Factory pipeline editor interface. On the left, there's a sidebar with 'Activities' and a search bar. A 'Get Metadata' activity is selected. The main area displays the 'Get Metadata' activity configuration. The 'Settings' tab is active. Under 'Dataset *', 'DelimitedText1' is selected. Below it, there's a section for 'View list' with fields for 'Argument', 'Item name', and 'Item type'. This entire section is highlighted with a red box. At the bottom, there are filters for 'Start time (UTC)' and 'End time (UTC)', and a 'Skip line count' input field. To the right, the 'Properties' panel shows the pipeline name as 'pipeline1'.



<https://www.linkedin.com/in/rganesh0203/>

Set Variable:

Sets a variable value.



<https://www.linkedin.com/in/r Ganesh0203/>

Append Variable

Appends a value to an array variable.

The screenshot shows the Azure Logic Apps designer interface. On the left, the 'Activities' pane is open, with 'Append variable' selected and highlighted with a red box. Below it, there is another 'Append variable' activity with a delete icon next to it. The main workspace contains a single 'Append variable' activity named 'Append variable1'. To the right of the workspace is the 'Properties' panel, which is currently on the 'General' tab. It shows the pipeline name as 'pipeline1' and a description field. The 'Variables' tab is selected and highlighted with a blue box. Under 'Variables', the 'Name' field is set to 'TestVariable' and the 'Value' field is empty. A note at the bottom of the 'Variables' section states: 'Disclaimer: Append Variables Only Supports Adding To 'Array' Type Variables'. At the bottom of the Properties panel, there is a link to 'Add dynamic content [Alt+Shift+D]'. The top of the screen has tabs for 'Validate', 'Debug', and 'Add trigger', along with standard window controls.



<https://www.linkedin.com/in/r Ganesh0203/>

Validation Activity:

Validates the existence of data before processing.

The screenshot shows the Azure Data Factory pipeline editor interface. On the left, there's a navigation pane with 'Activities' and a search bar set to 'Validation'. Below it, under 'General', is a 'Validation' activity. In the center, a validation dialog box is open, showing 'Validation1' with a green checkmark icon. On the right, the 'Properties' panel is displayed for the 'pipeline10' pipeline. The 'General' tab is selected, showing the 'Name' field set to 'pipeline10'. The 'Settings' tab is also visible. Under 'Settings', the 'Dataset' dropdown is set to 'DelimitedText5', which is highlighted with a red box. Below that, the 'Child items' section has three radio button options: 'Ignore' (selected), 'True', and 'False', all of which are also highlighted with a red box. Other settings like 'Timeout' (0.12:00:00) and 'Sleep' (10) are shown but not highlighted.



Log:

Logs information for monitoring purposes.

The screenshot shows the Azure Data Factory pipeline editor interface. On the left, the 'Activities' pane is open, displaying various data movement and transformation options. A 'Copy data' activity named 'Copy data1' is selected in the main workspace. The 'Settings' tab is currently active in the configuration pane. A red box highlights the 'Logging' section, which includes fields for 'Storage connection name' (set to 'linkedService1'), 'Logging level' (set to 'Warning'), and 'Logging mode' (set to 'Best effort'). The 'Enable logging' checkbox is checked. Other tabs like 'General', 'Source', 'Sink', 'Mapping', and 'User properties' are also visible.



<https://www.linkedin.com/in/r Ganesh0203/>

Custom and Specialized Activities



<https://www.linkedin.com/in/r Ganesh0203/>

Azure Batch

Runs parallel batch processing jobs

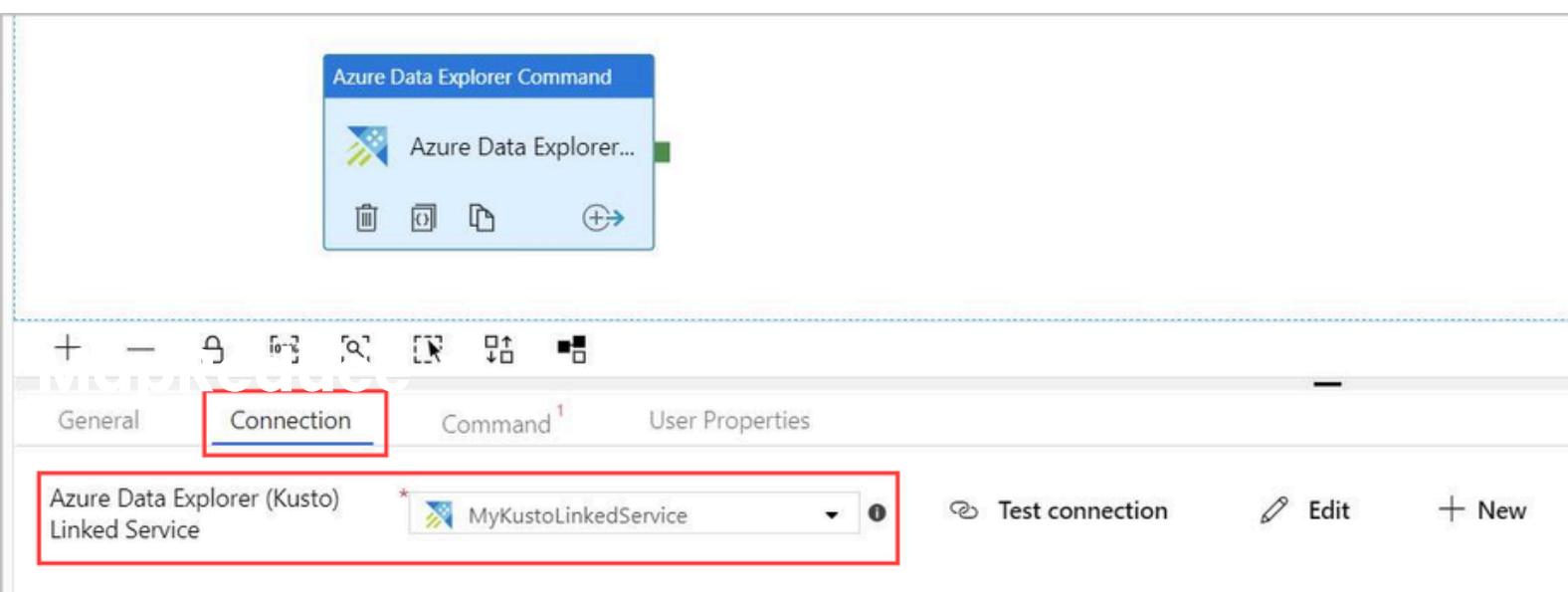
The screenshot shows the Azure Data Factory pipeline editor interface. On the left, there's a sidebar with 'Activities' listed under 'Batch Service'. A 'Custom' activity is selected, indicated by a red box around its name. In the main workspace, a 'Custom' activity is inserted into the pipeline, labeled 'Custom1'. Below it, a 'attività' placeholder is shown. To the right, the 'Properties' pane is open, displaying the 'General' tab for the pipeline. The pipeline is named 'pipeline1'. Other tabs in the properties pane include 'Related', 'Description', and 'Annotations'. At the bottom of the properties pane, there are buttons for 'Test connection', 'Edit', and 'New'.



<https://www.linkedin.com/in/r Ganesh0203/>

Data Explorer Command:

Executes KQL commands in Azure Data Explorer.



Key Benefits of Azure Data Factory (ADF) Activities:

- **Data Integration:** Supports diverse sources and handles large-scale data.
- **Transformation:** Enables visual and big data processing.
- **Dynamic Workflows:** Provides conditional logic, loops, and runtime flexibility.
- **Azure Integration:** Seamlessly connects with Azure services.
- **Monitoring:** Offers runtime insights, custom logging, and error handling.
- **Automation:** Simplifies orchestration with reusable, scalable pipelines.
- **Cost-Effective:** Reduces infrastructure and manual effort.
- **Secure:** Ensures data security with encryption and RBAC.
- **Error Recovery:** Supports automatic retries and error triggers.
- **Hybrid Support:** Integrates on-premises and multi-cloud environments.

Follow for more content like this



Ganesh R

Senior Data Engineer



Azure Cloud for Data



<https://www.linkedin.com/in/rganesh0203/>