Microsoft Fabric in a Day Lab Manual – Lab 3

Prepared by: Will Crayger

# Working with Pipelines – Extracting Source Data

## Introduction:

For this lab, we will create two pipelines. Copying data from source to target is one of the most commonly used patterns for pipelines. As such, our first pipeline will be created to extract a single table from an Azure SQL Database. Our second pipeline will also be used to extract data, but we’ll create a pattern that loops through many tables using a lookup activity. Leveraging lookups and parameters are incredibly powerful and efficient, as you’ll see during the exercise.

## Part 1: Single Object Copy

**Creating a Data Pipeline:** To begin creating your data pipeline, navigate to the **Data Engineering Fabric Landing** page. From here, choose **Data Pipeline** from the available options. Name your pipeline **single\_object\_copy** and click **Create** to move to the next step.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Setting Up the Copy Data Activity:** Once your pipeline is created, you'll land on the pipeline page. Select **Copy data** **assistant** to start the process. You will notice a variety of available data sources listed for data extraction. Choose **Azure SQL Database** from the list and click **Next**.

A screenshot of a computer

Description automatically generated

With the **Create new connection** option is selected, reauthenticate with the read-only credentials from lab 2 and click **Next**. In the table selection screen, check the box next to **Sales.Customers** and click **Next**.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Selecting the Destination:** In the **Destination** menu, switch to the **OneLake data hub** tab. Choose the **bronze\_lakehouse** created in Lab 1 as the **Lakehouse** destination.

On the next page, review the settings to familiarize yourself, but leave everything as default. Change the table name to **sales\_customers.** Once you've reviewed the configuration, click **Next**. You’ll be presented with the **Pipeline Summary**. Uncheck the **Start data transfer immediately** box and click **OK** to save the configuration.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Configuring and Running the Pipeline:** Return to the **Canvas** where the **Copy Data** activity is visible. Select it and rename the activity to **single\_object\_copy** for clarity. Open the **Source** tab within the activity to review the pre-populated configuration details that were set using the **Copy Data Tool**. Next, open the **Destination** tab and review the settings there as well.

If needed, expand the **Advanced** section. Here you can change the writer behavior to either **Append** or **Overwrite**, depending on your requirements.

Once you've reviewed and finalized the configurations, return to the **Home** tab of the pipeline and click the **Validate** button. If all steps were done correctly, no errors should appear, and you can safely close the **Pipeline Validation Output**.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Executing and Monitoring the Pipeline:** Now that your pipeline is validated, click **Run** from the activity bar. Select **Save and Run** to begin the process. The **Output** of the pipeline will automatically display, allowing you to track the execution in real-time. You can also click on the **Activity Name** to launch the **Copy Data details blade**, where you can monitor progress and review various metrics related to the run.

Once the pipeline completes, navigate back to your **Lakehouse**. There, you should see the newly created table for **Sales\_Customers**. You have now successfully completed **Part 1** of the lab.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

## Part 2: Dynamic Object Copy

**Creating the Dynamic Object Copy Pipeline:** To begin, navigate to the **Data Engineering Fabric Landing** page. From there, choose **Data Pipeline**. Name the pipeline **dynamic\_object\_copy** and click **Create** to proceed. Once the pipeline is created, click **Pipeline activity** from the landing page. We will start by executing a **Lookup activity** to retrieve a list of schema and table combinations. From the list of activities, choose **Lookup** and add it to the canvas. Select the **Lookup activity** and rename it to **lookup\_table\_and\_schema**.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Configuring the Lookup Activity:** Navigate to the **Settings** tab of the Lookup activity and select the connection created in part 1 of this lab. Change the connection type to **Azure SQL Database** and click the **Query** radial button. Now, locate the **Dynamic Load Source Query** file shared as part of the lab materials. Copy the query from the file and paste it into the **Query** box of the Lookup activity. Click **Preview Data** to see the output of the Lookup query, and be sure to uncheck the box labeled **First row only**.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Setting Up the ForEach Loop:** Next, open the **Activities** tab and add a **ForEach** activity to the pipeline canvas. Drag from the green checkmark on the **Lookup Schema and Table** activity to the **ForEach** activity to establish a connection between them. Select the ForEach activity and rename it **extraction\_loop**. Go to the **Settings** tab of the ForEach activity, click inside the **Items** box, and select **Add dynamic content** to specify the items that will be looped through. Select **Lookup Schema and Table value array** from the list of activity outputs to populate the field, then click **OK**.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Configuring the Copy Data Activity:** Once the ForEach activity is configured, click the **Pencil** icon to edit the loop’s contents. In the **Activities** tab, add a **Copy Data** activity to the canvas inside the loop. Select the **Copy Data** task and rename it **dynamic\_object\_copy**. In the **Source** tab, set the connection to connection created in part 1, choose **Azure SQL Database** as the connection type, and check the **Enter Manually** box for the table query. For the **Schema Name** field, click **Add content**, then select **Extraction Loop** from the ForEach iterator options. Specify the field from the Lookup activity by choosing **schema\_name** and click **OK**. Repeat this process for the **Table Name** field.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Setting the Destination:** Navigate to the **Destination** tab of the **dynamic\_object\_copy** activity and select the **Lakehouse** created in **Lab 1**. In the **Table Name** field, click **Add content** to specify the destination. Open the **Advanced** menu and change the table action to **Overwrite**.

A screenshot of a computer

Description automatically generated

**Validating and Running the Pipeline:** Return to the **Home** tab of the pipeline and click **Validate** to ensure that there are no errors in the setup. Once the validation is complete, click **Run** from the activity bar to execute the pipeline. You can monitor the pipeline run in the **Output** tab, where you’ll notice multiple **Copy** activities running simultaneously.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Confirming the Results:** After the pipeline run completes, navigate back to your **Lakehouse** to confirm that the delta tables have been created. If the tables aren’t immediately visible, refresh your browser or right-click on **Tables** and select **Refresh**.

A screenshot of a computer

Description automatically generated

You have now successfully completed **Part 2** of the lab.