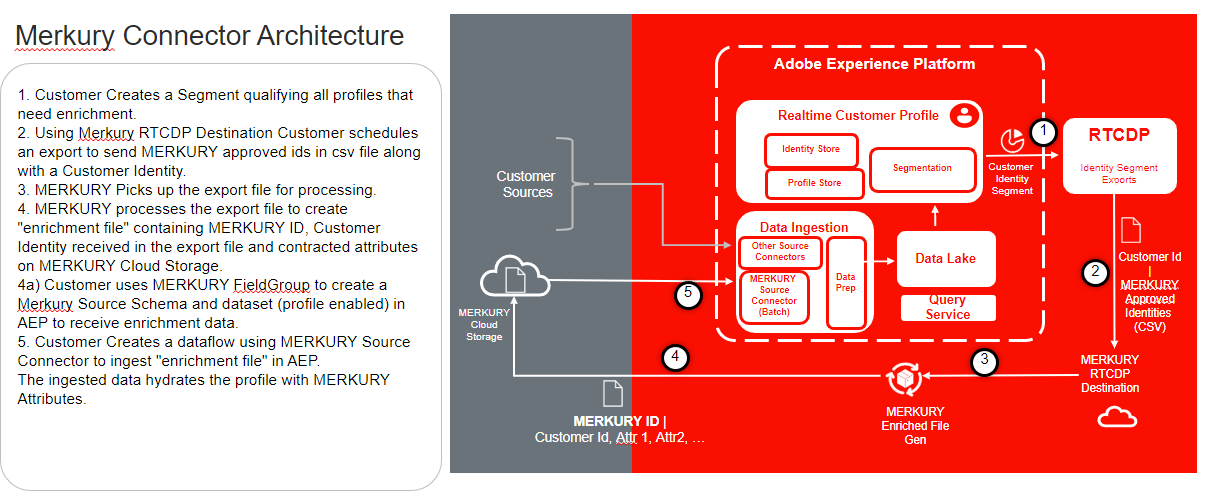
**Adobe Merkury Identity Source Connector**

Merkury, by Merkle, helps you recognize more digital visitors — even without cookies — so you can deliver the relevant, personalized experiences your customer's demand. The **person ID** combines everything your organization knows about an individual – digital behaviors, buying preferences, and identifying information like a name, email address, physical address, or device ID – into one single profile. Ingested data can be formatted as XDM JSON, XDM Parquet, or delimited. Every step of the process is integrated into the Sources workflow.



**IP address allow list**

A list of IP addresses must be added to an allow list prior to working with source connectors. Failing to add your region-specific IP addresses to your allow list may lead to errors or non-performance when using sources. See the [IP address allow list](https://experienceleague.adobe.com/docs/experience-platform/sources/ip-address-allow-list.html?lang=en) page for more information.

**Naming constraints for files and directories**

The following is a list of constraints you must account for when naming your cloud storage file or directory.

* Directory and file component names cannot exceed 255 characters.
* Directory and file names cannot end with a forward slash (/). If provided, it will be automatically removed.
* The following reserved URL characters must be properly escaped: ! ' ( ) ; @ & = + $ , % # [ ]
* The following characters are not allowed: " \ / : | < > \* ?.
* Illegal URL path characters not allowed. Code points like \uE000, while valid in NTFS filenames, are not valid Unicode characters. In addition, some ASCII or Unicode characters, like control characters (0x00 to 0x1F, \u0081, etc.), are also not allowed. For rules governing Unicode strings in HTTP/1.1 see [RFC 2616, Section 2.2: Basic Rules](https://www.ietf.org/rfc/rfc2616.txt) and [RFC 3987](https://www.ietf.org/rfc/rfc3987.txt).
* The following file names are not allowed: LPT1, LPT2, LPT3, LPT4, LPT5, LPT6, LPT7, LPT8, LPT9, COM1, COM2, COM3, COM4, COM5, COM6, COM7, COM8, COM9, PRN, AUX, NUL, CON, CLOCK$, dot character (.), and two dot characters (…).

**Prerequisites**

* Merkury setup completed with your Merkury team.
* Obtained key, secret and bucket name from Merkury team.

A file path like myBucket/folder/subfolder/subsubfolder/abc.csv may lead you to only access subsubfolder/abc.csv. If you want to access the subfolder, you can specify the bucket parameter as myBucket and the folderPath as folder/subfolder to ensure that file exploration starts at subfolder as opposed to subsubfolder/abc.csv.

**Connect Merkury to Platform**

The documentation below provides information on how to connect Merkury to Adobe Experience Platform using APIs or the user interface:

**Create Merkury source connection in the UI**

Source connectors in Adobe Experience Platform provide the ability to ingest externally sourced data on a scheduled basis. This tutorial provides steps for creating a Merkury source connector using the Platform user interface.

## Getting started

This tutorial requires a working understanding of the following components of Adobe Experience Platform:

* [Experience Data Model (XDM) System](https://experienceleague.adobe.com/docs/experience-platform/xdm/home.html?lang=en): The standardized framework by which Experience Platform organizes customer experience data.
  + [Basics of schema composition](https://experienceleague.adobe.com/docs/experience-platform/xdm/schema/composition.html?lang=en): Learn about the basic building blocks of XDM schemas, including key principles and best practices in schema composition.
  + [Schema Editor tutorial](https://experienceleague.adobe.com/docs/experience-platform/xdm/tutorials/create-schema-ui.html?lang=en): Learn how to create custom schemas using the Schema Editor UI.
* [Real-Time Customer Profile](https://experienceleague.adobe.com/docs/experience-platform/profile/home.html?lang=en): Provides a unified, real-time consumer profile based on aggregated data from multiple sources.

If you already have a valid S3 connection, you may skip the remainder of this document and proceed to the tutorial on [configuring a dataflow](https://experienceleague.adobe.com/docs/experience-platform/sources/ui-tutorials/dataflow/cloud-storage.html?lang=en).

### Gather required credentials

In order to access your bucket on Platform, you need to provide valid values for the following credentials:

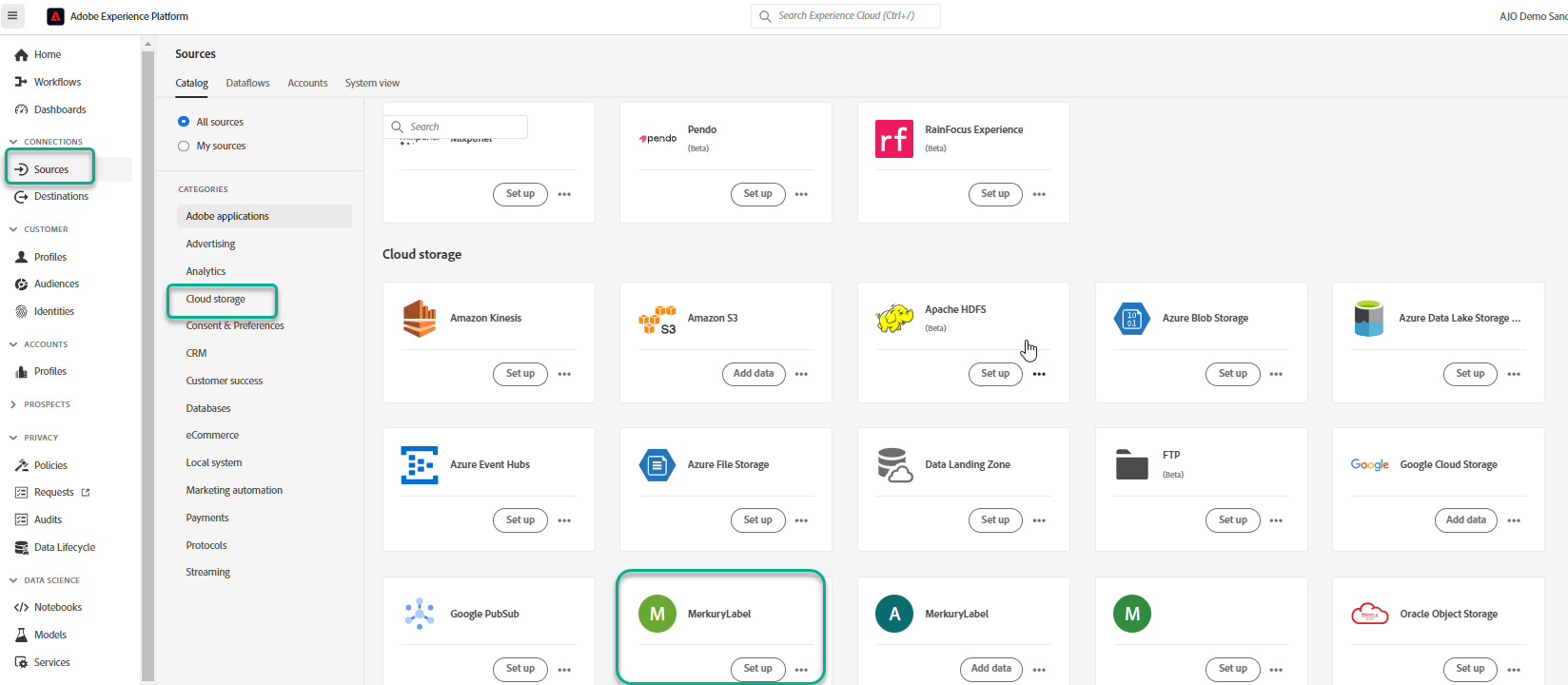
| **Credential** | **Description** |
| --- | --- |
| AccessKey | The access key ID for your bucket. |
| SecretKey | The secret key ID for your bucket. |
| bucketName | The bucket contains your data and its corresponding descriptive metadata. Your S3 bucket name must be between three and 63 characters long and must begin and end with either a letter or a number. The bucket name can only have lowercase letters, numbers, or hyphens (-), and cannot be formatted as an IP address. |

## Connect your Merkury Account

In the Platform UI, select **Sources** from the left navigation bar to access the Sources workspace. The Catalog screen displays a variety of sources for which you can create an account with.

You can select the Enrichment category from the catalog on the left-hand side of your screen. Alternatively, you can find the specific source you wish to work with using the search option.

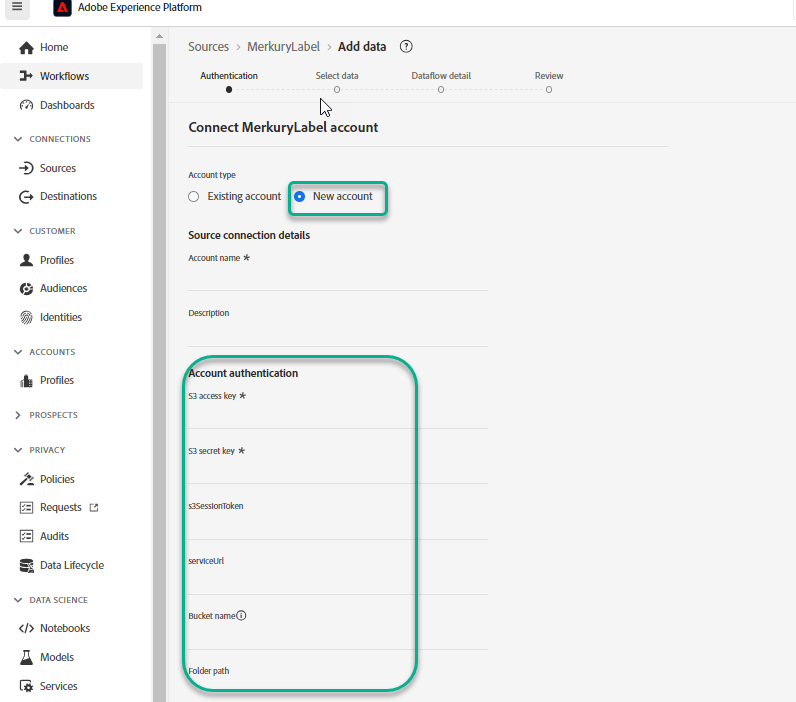
Under the **Enrichment** category, select **Merkury** and then select **Add data**.



The **Connect to Merkury** page appears. On this page, you can either use new credentials or existing credentials.

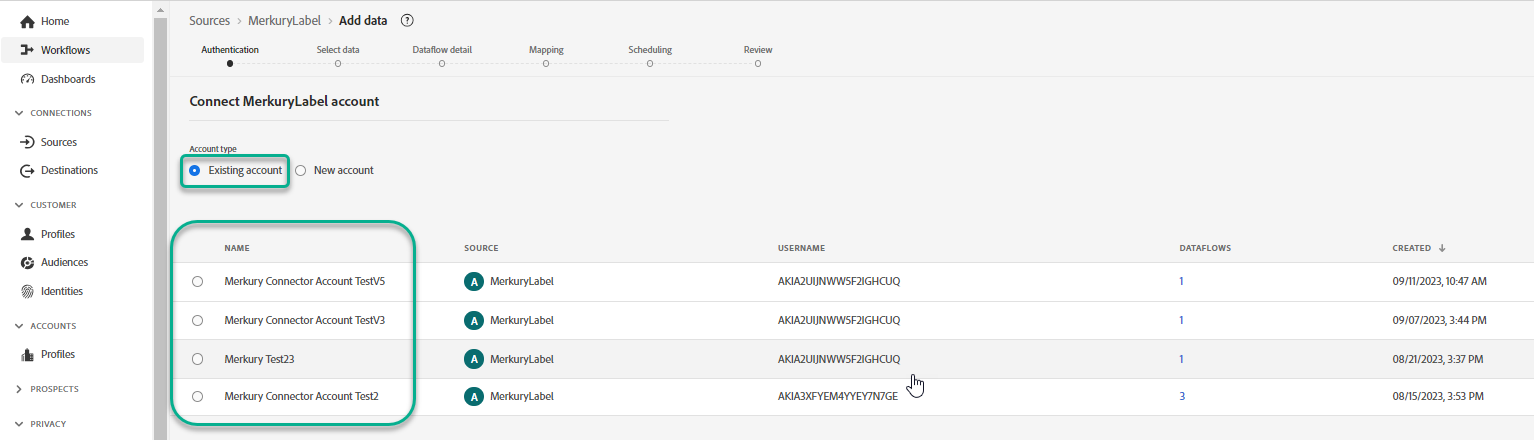
New account

If you are using new credentials, select **new account**. On the input form that appears, provide a name, an optional description, and your Merkury credentials. When finished, select **Connect to source** and then allow some time for the new connection to establish.



### Existing account

To connect an existing account, select Merkury account you want to connect with, then select **Next** to proceed.



**Configure a dataflow to ingest batch data from a cloud storage source in the UI**

This tutorial provides steps on how to configure a dataflow to bring batch data from your source connector to Adobe Experience Platform.

## Getting started

**NOTE**

In order to create a dataflow to bring batch data from a cloud storage, you must already have access to an authenticated cloud storage source. If you do not have access, go to the [sources overview](https://experienceleague.adobe.com/docs/experience-platform/sources/home.html?lang=en#cloud-storage) for a list of sources that you can create an account with.

This tutorial requires a working understanding of the following components of Experience Platform:

* [Experience Data Model (XDM) System](https://experienceleague.adobe.com/docs/experience-platform/xdm/home.html?lang=en): The standardized framework by which Experience Platform organizes customer experience data.
  + [Basics of schema composition](https://experienceleague.adobe.com/docs/experience-platform/xdm/schema/composition.html?lang=en): Learn about the basic building blocks of XDM schemas, including key principles and best practices in schema composition.
  + [Schema Editor tutorial](https://experienceleague.adobe.com/docs/experience-platform/xdm/tutorials/create-schema-ui.html?lang=en): Learn how to create custom schemas using the Schema Editor UI.
* [Real-Time Customer Profile](https://experienceleague.adobe.com/docs/experience-platform/profile/home.html?lang=en): Provides a unified, real-time consumer profile based on aggregated data from multiple sources.

### Supported file formats

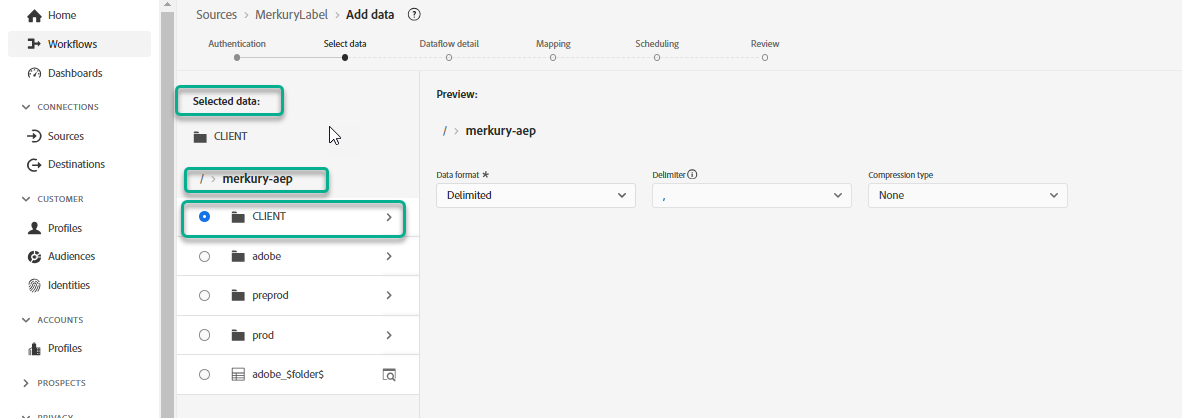
Cloud storage sources for batch data supports the following file formats for ingestion:

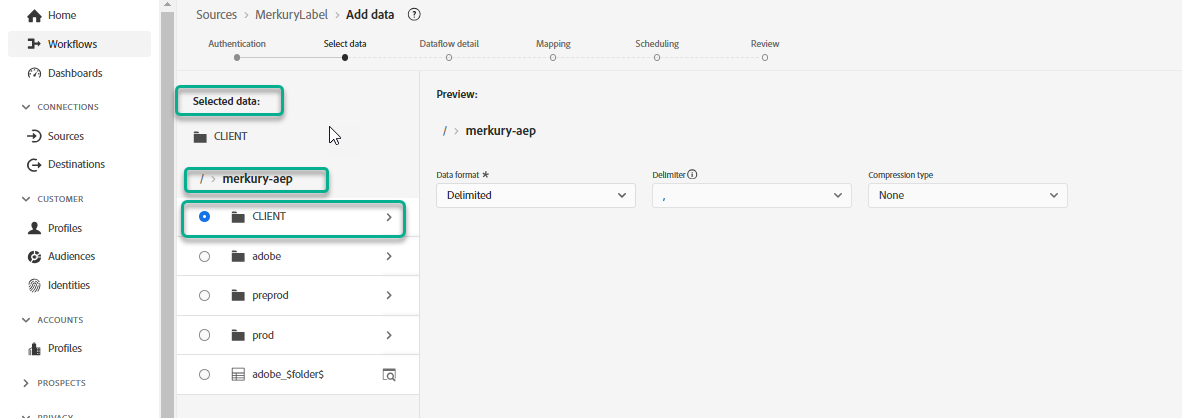
* Delimiter-separated values (DSV): Any single-character value can be used as a delimiter for DSV-formatted data files.
* JavaScript Object Notation (JSON): JSON-formatted data files must be XDM-compliant.
* Apache Parquet: Parquet-formatted data files must be XDM-compliant.
* Compressed files: JSON and delimited files can be compressed as: bzip2, gzip, deflate, zipDeflate, tarGzip, and tar.

## Add data

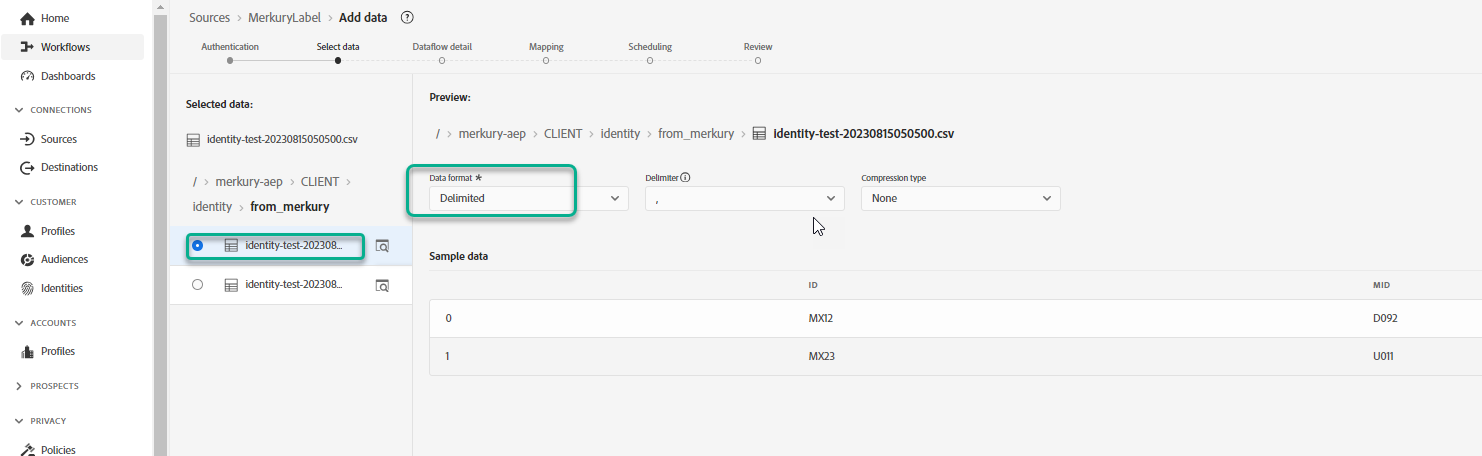
After creating your cloud storage account, the **Add data** step appears, providing an interface for you to explore your cloud storage file hierarchy and select the folder or specific file that you want to bring to Platform.

* The left part of the interface is a directory browser, displaying your cloud storage file hierarchy.
* The right part of the interface lets you preview up to 100 rows of data from a compatible folder or file.



Select the root folder to access your folder hierarchy. From here, you can select a single folder to ingest all files in the folder recursively. When ingesting an entire folder, you must ensure that all files in that folder share the same data format and schema. 

Once you have selected a folder, the right interface updates to a preview of the contents and structure of the first file in the selected folder.



During this step, you can make several configurations to your data, before proceeding. First, select **Data format** and then select the appropriate data format for your file in the dropdown panel that appears.

The following table displays the appropriate data formats for the supported file types:

| **File type** | **Data format** |
| --- | --- |
| CSV | Delimited |
| JSON | JSON |
| Parquet | XDM Parquet |

### Select a column delimiter

After configuring your data format, you can set a column delimiter when ingesting delimited files. Select the **Delimiter** option and then select a delimiter from the dropdown menu. The menu displays the most frequently used options for delimiters, including a comma (,), a tab (\t), and a pipe (|).If you prefer to use a custom delimiter, select **Custom** and enter a single-character delimiter of your choice in the pop-up input bar.

### Ingest compressed files

You can also ingest compressed JSON or delimited files by specifying their compression type.

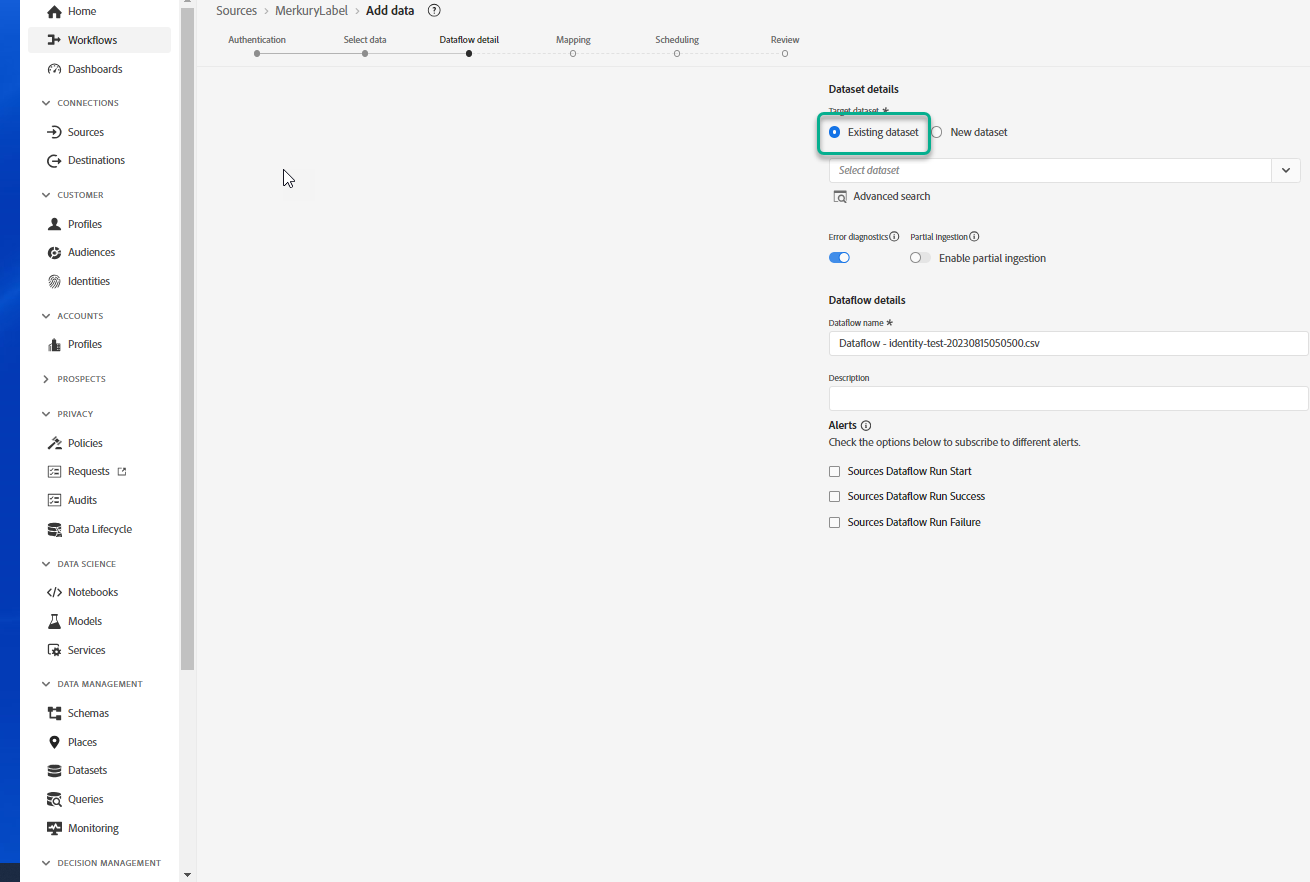
In the Select data step, select a compressed file for ingestion and then select its appropriate file type and whether it’s XDM-compliant or not. Next, select **Compression type** and then select the appropriate compressed file type for your source data.

To bring a specific file to Platform, select a folder, and then select the file that you want to ingest. During this step, you can also preview file contents of other files within a given folder by using the preview icon beside a file name.

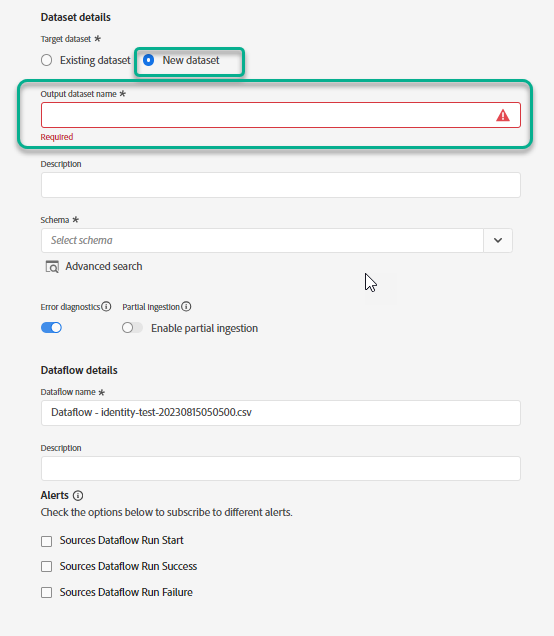
When finished, select **Next**.

### Use an existing dataset

To ingest data into an existing dataset, select **Existing dataset**. You can either retrieve an existing dataset using the Advanced search option or by scrolling through the list of existing datasets in the dropdown menu. Once you have selected a dataset, provide a name and a description for your dataflow.



### Use a new dataset

To ingest into a new dataset, select **New dataset** and then provide an output dataset name and an optional description. Next, select a schema to map to using the Advanced search option or by scrolling through the list of existing schemas in the dropdown menu. Once you have selected a schema, provide a name and a description for your dataflow 

### Enable Profile and error diagnostics

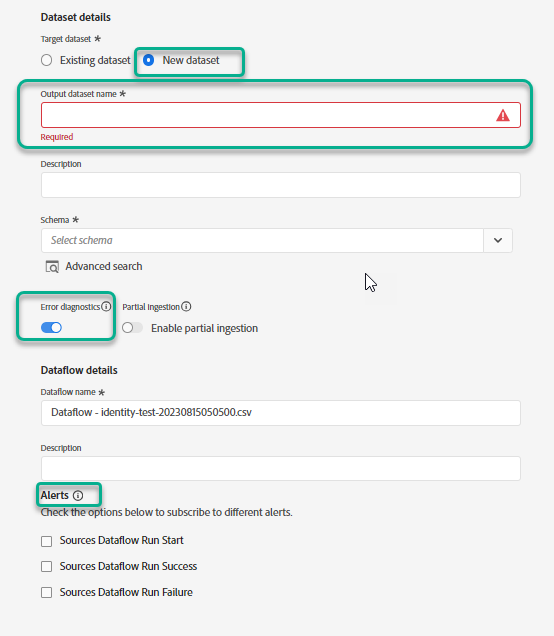
Next, select the **Profile dataset** toggle to enable your dataset for Profile. This allows you to create a holistic view of an entity’s attributes and behaviors. Data from all Profile-enabled datasets will be included in Profile and changes are applied when you save your dataflow.

Error diagnostics enables detailed error message generation for any erroneous records that occur in your dataflow, while Partial ingestion allows you to ingest data containing errors, up to a certain threshold that you manually define. See the [partial batch ingestion overview](https://experienceleague.adobe.com/docs/experience-platform/ingestion/batch/partial.html?lang=en) for more information.

### Enable alerts

You can enable alerts to receive notifications on the status of your dataflow. Select an alert from the list to subscribe to receive notifications on the status of your dataflow. For more information on alerts, see the guide on [subscribing to sources alerts using the UI](https://experienceleague.adobe.com/docs/experience-platform/sources/ui-tutorials/alerts.html?lang=en).

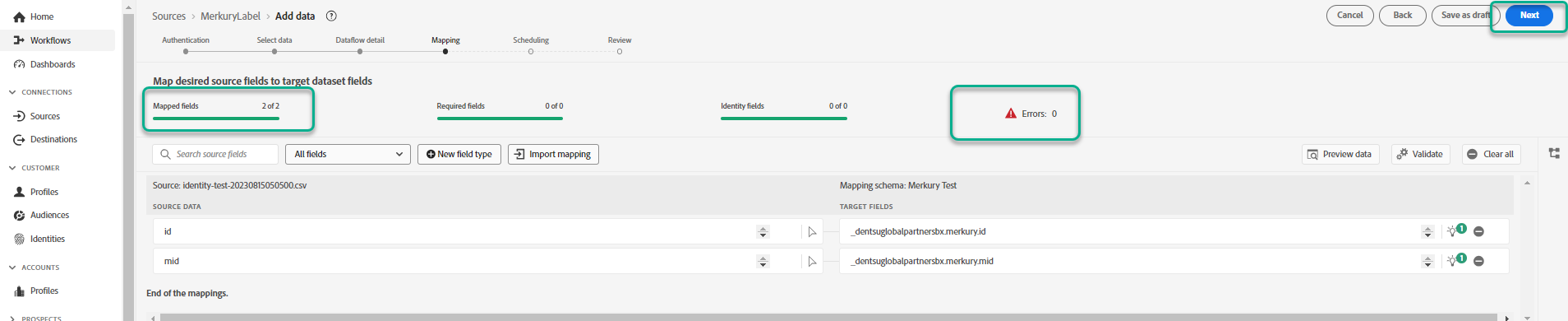
When you are finished providing details to your dataflow, select **Next**.



## Map data fields to an XDM schema

The Mapping step appears, providing you with an interface to map the source fields from your source schema to their appropriate target XDM fields in the target schema.

Platform provides intelligent recommendations for auto-mapped fields based on the target schema or dataset that you selected. You can manually adjust mapping rules to suit your use cases. Based on your needs, you can choose to map fields directly, or use data prep functions to transform source data to derive computed or calculated values. For comprehensive steps on using the mapper interface and calculated fields, see the [Data Prep UI guide](https://experienceleague.adobe.com/docs/experience-platform/data-prep/ui/mapping.html?lang=en).

Once your source data is successfully mapped, select **Next**. 

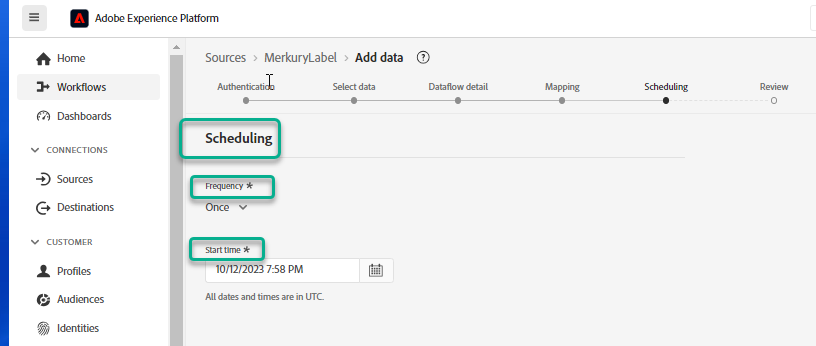
**Schedule ingestion runs**

**IMPORTANT** It is strongly recommended to schedule your dataflow for one-time ingestion when using the [FTP source](https://experienceleague.adobe.com/docs/experience-platform/sources/connectors/cloud-storage/ftp.html?lang=en).

The Scheduling step appears, allowing you to configure an ingestion schedule to automatically ingest the selected source data using the configured mappings. By default, scheduling is set to Once. To adjust your ingestion frequency, select **Frequency** and then select an option from the dropdown menu.

**TIP**

Interval and backfill are not visible during a one-time ingestion.



If you set your ingestion frequency to Minute, Hour, Day, or Week, then you must set an interval to establish a set time frame between every ingestion. For example, an ingestion frequency set to Day and an interval set to 15 means that your dataflow is scheduled to ingest data every 15 days.

During this step, you can also enable **backfill** and define a column for the incremental ingestion of data. Backfill is used to ingest historical data, while the column you define for incremental ingestion allows new data to be differentiated from existing data.

A screenshot of a computer

Description automatically generated

**NOTE**

For batch ingestion, every ensuing dataflow selects files to be ingested from your source based on their **last modified** timestamp. This means that batch dataflows select files from the source that are either new or have been modified since the last flow run. Furthermore, you must ensure that there’s a sufficient time span between file upload and a scheduled flow run because files that are not entirely uploaded to your cloud storage account before the scheduled flow run time may not be picked up for ingestion.

When finished configuring your ingestion schedule, select **Next**.

**Review your dataflow**

The **Review** step appears, allowing you to review your new dataflow before it is created. Details are grouped within the following categories:

* **Connection**: Shows the source type, the relevant path of the chosen source file, and the amount of columns within that source file.
* **Assign dataset & map fields**: Shows which dataset the source data is being ingested into, including the schema that the dataset adheres to.
* **Scheduling**: Shows the active period, frequency, and interval of the ingestion schedule.

Once you have reviewed your dataflow, click **Finish** and allow some time for the dataflow to be created.

