



Adobe Experience Platform Data Pipeline

Lab Overview

Ingest Stores data into AEP in batch mode using Data Landing Zone. This Lab will leverage delimited data on Landing Zone and ML Recommendations to map most of the source data. This job will be scheduled to run every 15 minutes.

This lab will introduce you to the basic data ingestion process.

Expected time: 20 minutes

Learning Objectives

What should you walk away with after taking this Lab?

- Understand the basic data ingestion process
- Use Data Landing Zone as a source
- Use Data Prep to Map the non-XDM data to XDM
- Scheduling batch workflows

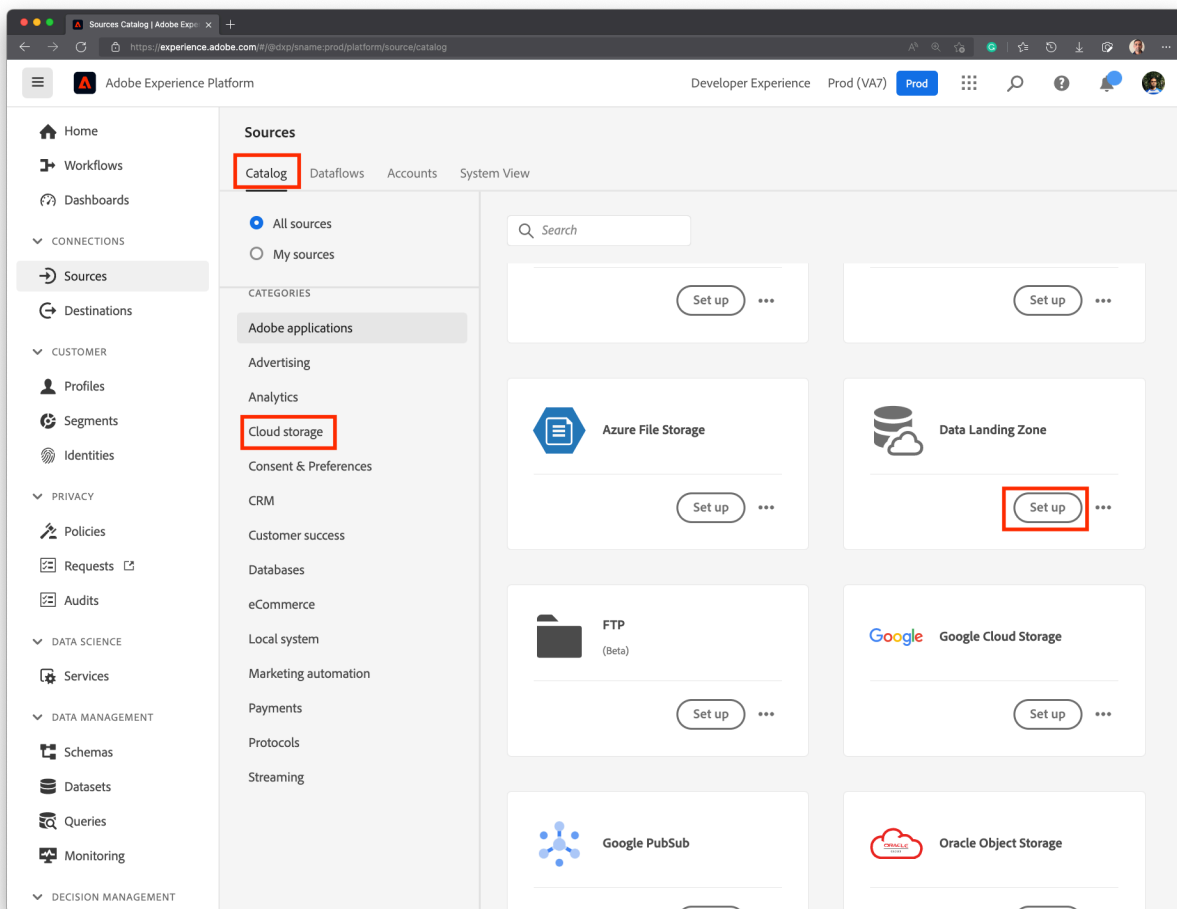
Lab Tasks

Select source data

Go to Adobe Experience Platform → **Sources** → **Catalog**. In the **Cloud Storage** connectors, click **Setup** / **Add Data** in the **Data Landing Zone** card.

Tip

If at least one connection exists for that source, you will see “**Add data**” as the default action. If no connections exist for that source, you will see “**Setup**” as the default action.



In the next screen, navigate to **project** → **PIPELINE** and select the **Lab_Lookup_Store.csv**. On the right hand side, choose Data format to be **Delimited**. Preview should load automatically. Click **Next**.

The screenshot shows the Adobe Experience Platform 'Add data' interface. The breadcrumb path is: / > dlz-user-container > project > PIPELINE. The 'Lab_Lookup_Store.csv' file is selected. The 'Data format' is set to 'Delimited'. A preview of the data is shown below.

	STOREID	STORENAME	STOREINAGURATIONDATE	STOREOPEN TIME
0	STORE-1	Cool Place	2021-01	open at 9am / close at 9pm
1	STORE-2	Awesome Town	2021-01	open at 9am / close at 9pm
2	STORE-3	Awesome Town	2022-04	open at 9am / close at 9pm
3	STORE-4	Best Shop	2022-01	open at 9am / close at 9pm
4	STORE-5	Cool Place	2022-01	open at 9am / close at 9pm

Define the target dataset

In the **Dataflow detail** screen, choose the **Existing dataset** [dep: Lookup Store](#). Turn on **Error Diagnostics** (default is ON), and **Enable Partial Ingestion**. Set the Data flow name as [Dataflow - Lookup - Store](#) and click **Next**.

Tip

When Partial Ingestion is enabled, Error diagnostics are automatically enabled and hence the toggle box will disappear.

The screenshot shows the 'Add data' workflow in the Adobe Experience Platform. The 'Dataflow detail' step is active. The 'Target dataset' is set to 'dep: Lookup Store'. The 'Profile dataset' is set to 'dep: Lookup Store'. The 'Partial ingestion' toggle is turned on, and the 'Error threshold %' is set to 5. The 'Dataflow name' is set to 'Dataflow - Lookup - Store'. The 'Description' field is empty. The 'Alerts' section is visible at the bottom, with checkboxes for 'Sources Dataflow Run Start', 'Sources Dataflow Run Success', 'Sources Dataflow Run Failure', and 'Sources Dataflow Lack of Ingestion'.

Data Prep / Transformation

Data Prep (Mapping) step will load. The ML Recommendation Service will pre-populate the mappings for you. Please review and ensure the mappings are as follows. If all mappings look good, click **Next**.

Tip

To change a mapping, click on the arrow next to the source field, or on the target field name respectively.

Adobe Experience Platform

Sources > Data Landing Zone > Add data

Select data Dataflow detail Mapping Scheduling Review

Map desired source fields to target dataset fields

Mapped fields 9 of 9 Required fields 1 of 1 Identity fields 1 of 1 Errors: 0

Search source fields All fields New field type Import mapping

Preview data Clear all

Source: LookUp_Store_CLIENT.csv Mapping schema: dep: Lookup Store

SOURCE DATA	TARGET FIELDS
storeID	_dep.storeID
storeName	_dep.name
storeInagurationDate	_dep.inagurationDate
storeOpenTime	_dep.openTime
storeCloseTime	_dep.closeTime
storeStreet	_dep.address.street1
storeCity	_dep.address.city
storeState	_dep.address.state
storeZipCode	_dep.address.postalCode

End of the mappings.

Schedule

The **Scheduling** screen will load. Change the **Frequency** to **Day** and set the **Interval** as **7**. Leave the **Backfill** toggle turned ON. This will schedule the Dataflow to run once every 7 days.

WARNING

Do not set the Frequency to Once. Currently One-time ingestion flows cannot be edited.

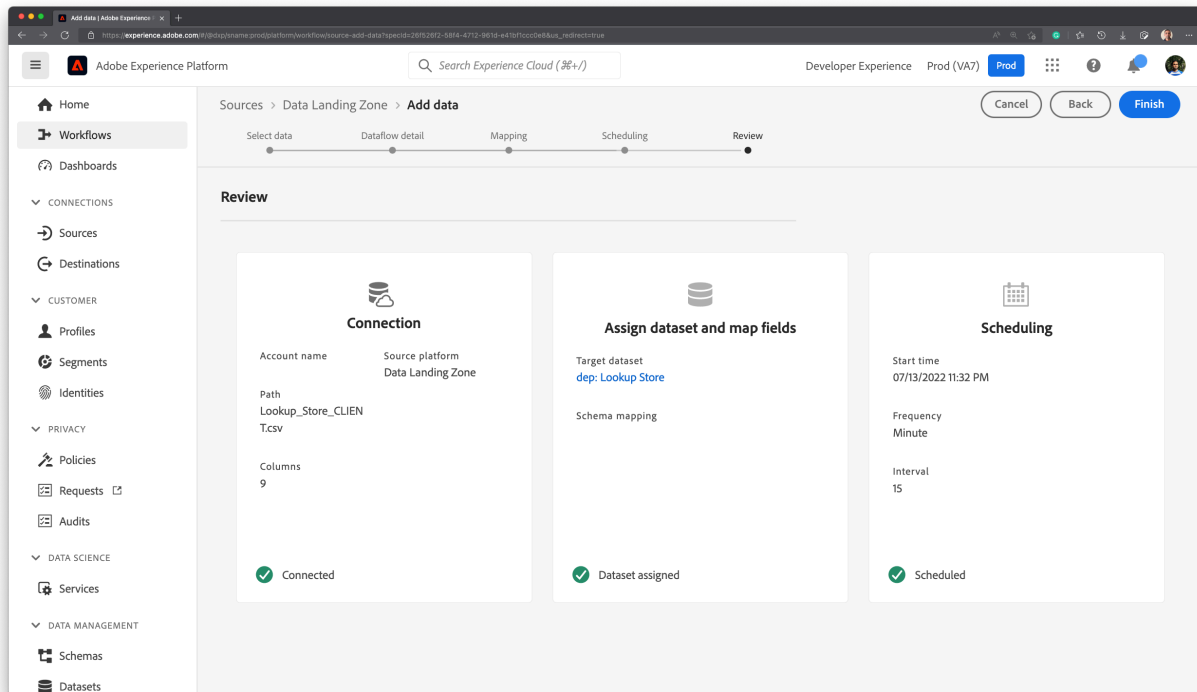
WARNING

Schedule times are in UTC. Not your local time.

The screenshot shows the 'Add data' configuration page in the Adobe Experience Platform. The breadcrumb trail is 'Sources > Data Landing Zone > Add data'. The page has a progress bar with five steps: 'Select data', 'Dataflow detail', 'Mapping', 'Scheduling' (the current step), and 'Review'. On the right, there are 'Cancel', 'Back', and 'Next' buttons. The left sidebar contains navigation links for Home, Workflows, Dashboards, CONNECTIONS (Sources, Destinations), CUSTOMER (Profiles, Segments, Identities), and PRIVACY (Policies). The 'Scheduling' section includes the following fields:

- Frequency ***: Set to 'Day' with a dropdown arrow.
- Interval**: Set to '7' with a green checkmark.
- Start time ***: A date and time picker showing '08/02/2022 6:05 PM'.
- A note: 'All dates and times are in UTC.'
- Backfill**: A toggle switch that is turned ON.

When the **Review** screen appears, click **Finish**.



It takes few minutes to create the Dataflow. Once Dataflow is created, you will see the following screenshot. Notice that Last Dataflow Run Status indicates **No runs**. First run will kick off approximately in 15 minutes.

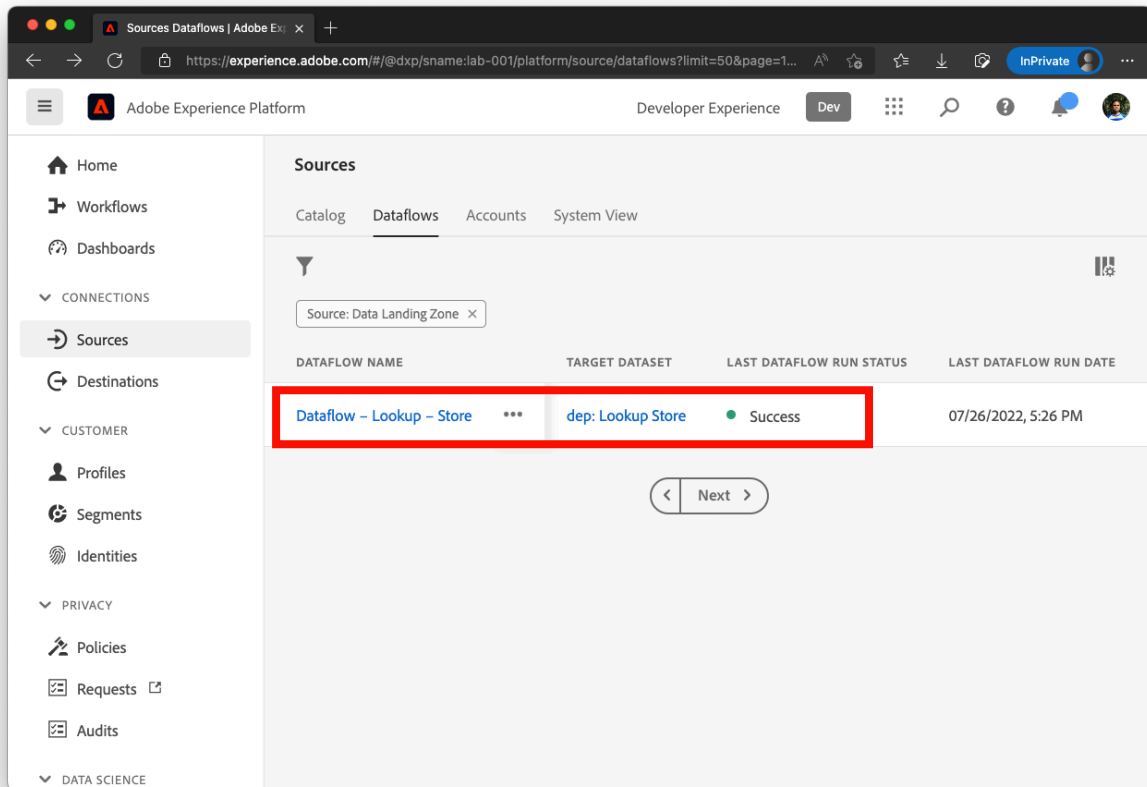
The screenshot shows the Adobe Experience Platform interface. The left sidebar contains navigation links for Home, Workflows, Dashboards, CONNECTIONS (Sources, Destinations), CUSTOMER (Profiles, Segments, Identities), PRIVACY (Policies, Requests, Audits), and DATA SCIENCE. The main content area is titled 'Sources' and has tabs for Catalog, Dataflows, Accounts, and System View. The 'Dataflows' tab is active, showing a table with the following data:

DATAFLOW NAME	TARGET DATASET	LAST DATAFLOW RUN STATUS	LAST DATAFLOW RUN DATE	SOURCE DATA	ACCOUNT NAME
Dataflow - Lookup - Store	dep: Lookup Store	No runs	-	dlz-user-container/project/CSV/Lookup_St...	-

Below the table is a 'Next' button. A blue information banner at the bottom states: 'This temporary Data Landing Zone is provided to you for the sole purpose of the ingress and egress of data in and out of Adobe Experience Platform. All data in this Data Landing Zone has a Time-To-Live of 7 days, which means any file ingested will be deleted after 7 days. For more information about Data Landing Zone, please see the documentation.'

Check scheduled execution

After 15 minutes, the following will appear. Notice the **Last Dataflow Run Status** and **Last Dataflow Run Date**.



The screenshot shows the Adobe Experience Platform interface. The left sidebar contains navigation links: Home, Workflows, Dashboards, CONNECTIONS (Sources, Destinations), CUSTOMER (Profiles, Segments, Identities), PRIVACY (Policies, Requests, Audits), and DATA SCIENCE. The main content area is titled 'Sources' and has tabs for Catalog, Dataflows, Accounts, and System View. The 'Dataflows' tab is active, showing a filter for 'Source: Data Landing Zone'. Below the filter is a table with the following data:

DATAFLOW NAME	TARGET DATASET	LAST DATAFLOW RUN STATUS	LAST DATAFLOW RUN DATE
Dataflow - Lookup - Store	dep: Lookup Store	Success	07/26/2022, 5:26 PM

A red rectangle highlights the row for 'Dataflow - Lookup - Store'. Below the table is a 'Next' button.

Click on the Dataflow name to get a list of Dataflow Runs. 100 Records should be ingested.

Adobe Experience Platform

Sources > Dataflow - Lookup - Store

Dataflow activity

Records ingested: 100, Records failed: 0

Properties: Update dataflow, Edit schedule, Enabled, Delete

Dataflow name: Dataflow - Lookup - Store

Description:

Source data: dlz-user-container/project/CSV/Lookup_Store_CLIENT.csv

DATAFLOW RUN START	PROCESSING TIME	RECORDS INGESTED	RECORDS FAILED	STATUS
07/26/2022, 5:26 PM	2 minutes	100	0	Success

Click on the Dataflow Run Start time to see error diagnostic details.

Adobe Experience Platform

Sources > Dataflow - Lookup - Store > Dataflow Run 7b900fa1-b01c-4dcf-b255-82b79b7415b1

Dataflow run overview

Records ingested: 100, Records failed: 0, Total files: 1, Size of data: 11.57 kB

Status: Success

Dataflow run start: 07/26/2022, 5:26 PM

Dataflow run end: 07/26/2022, 5:28 PM

Partial ingestion: Enabled: 5% Error threshold

Error diagnostics: Enabled

Error summary: CONNECTOR-2001-500: Error in processing (parsing, validation or transformation) the copied data.

Dataflow run ID: 7b900fa1-b01c-4dcf-b255-82b79b7415b1

IMS org ID: 37E0399C61687C4E0A495E06@AdobeOrg

Dataset: dep: Lookup Store

Dataflow run errors

Records failed: 0, Records skipped: 0

ERROR CODE	RECORD COUNT	DESCRIPTION
MAPPER-3700-199	0	Field transformation warning

NOTE

Ignore the Error code being generated. Error handling and debugging will be discussed in a subsequent lab

Verify the data

In the Left Nav bar, Go to **Datasets** in Platform and click on **dep: Lookup Store**

The screenshot shows the Adobe Experience Platform interface. The left sidebar contains a navigation menu with categories like PRIVACY, DATA SCIENCE, DATA MANAGEMENT, DECISION MANAGEMENT, and ADMINISTRATION. The 'Datasets' option is selected. The main area displays a table of datasets with columns: NAME, CREATED, SOURCE, SCHEMA, LAST BATCH STATUS, and LAST UPDATED. The 'dep: Lookup Store' dataset is highlighted with a red box. The right sidebar shows a summary of 28 datasets and a list of most recently updated datasets.

NAME	CREATED	SOURCE	SCHEMA	LAST BATCH STATUS	LAST UPDATED
dep: Customer Account	07/26/2022, 5:11 PM	Schema	dep: Customer Account	No batches	-
dep: Lookup Store	07/26/2022, 4:45 PM	Schema	dep: Lookup Store	Success	07/26/2022, 5:28 PM
dep: Orders	07/26/2022, 4:40 PM	Schema	dep: Orders	No batches	-
dep: Lookup Plan	07/19/2022, 1:38 PM	Schema	dep: Lookup Plan	No batches	-
dep: Lookup Product	07/19/2022, 1:37 PM	Schema	dep: Lookup Product	No batches	-
dep: Web	07/19/2022, 1:37 PM	Schema	dep: Web	No batches	-
dep: Ecommerce	07/19/2022, 1:37 PM	Schema	dep: Ecommerce	No batches	-
dep: Billing	07/19/2022, 1:37 PM	Schema	dep: Billing	No batches	-
dep: Customer Aggregates	07/19/2022, 1:36 PM	Schema	dep: Customer Aggregates	No batches	-
dep: Customer Active Lines	07/19/2022, 1:36 PM	Schema	dep: Customer Active Lines	No batches	-

Click on the **Preview Dataset**

The screenshot shows the Adobe Experience Platform interface for the 'dep: Lookup Store' dataset. The left sidebar contains a navigation menu with categories like PRIVACY, DATA SCIENCE, DATA MANAGEMENT, DECISION MANAGEMENT, and ADMINISTRATION. The 'Datasets' option is selected. The main area displays the 'Dataset activity' section with metrics for total records, ingested records, and failed batches. A chart shows ingested records over time. The right sidebar shows dataset details.

Metric	Value
Total records in previous month	100
Ingested records in the last 7 days	0
Ingested batches today	1
Ingested batches in the last 7 days	0
Size of data in previous month	7.81 kB
Failed batches today	0
Failed batches in the last 7 days	0

The chart shows ingested records over time. The x-axis represents dates from July 20 to July 26, 2022. The y-axis represents the number of records, ranging from 0 to 3. The legend indicates that blue bars represent 'Ingested batches' and red bars represent 'Failed batches'.

The right sidebar shows dataset details for 'dep: Lookup Store':

- Name: dep: Lookup Store
- Description: dep: Lookup Store
- Dataset ID: 62e07ca90db3df1c073ceb32
- Table name: dep_lookup_store
- Profile: ☐
- Schema: dep: Lookup Store
- Source: Schema
- Created: 07/26/2022, 4:45 PM

Data is loaded and displayed. Expand the XDM tree on the right to see child attributes.

TIP

When you click on a node such as **_dxp** → **address**, the right hand side will update to show only attributes within the selected parent node (**address** in this example). To see all attributes click on the master node (**dep: Lookup Store** in this example).

The screenshot shows the Adobe Experience Platform interface. On the left, the 'Datasets' menu is expanded, showing 'dep: Lookup Store' selected. The XDM tree on the right is expanded to show the 'address' node. The main panel displays a table of data with columns for city, close time, postal code, street, and open time.

_DXP.ADDRESS.CITY	_DXP.CLOSETIME	_DXP.ADDRESS.POSTALCODE	_DXP.ADDRESS.STREET	_DXP.D
Hisai	open at 9am / close at 9pm	515-2605	57403 Arizona Alley	open z
Palopo	open at 9am / close at 9pm	-	20415 Messerschmidt Park	open z
Covas da Raposa	open at 9am / close at 9pm	2970-124	6 Norway Maple Road	open z
Jaša Tomić	open at 9am / close at 9pm	-	770 Granby Parkway	open z
Jinshan	open at 9am / close at 9pm	-	473 Milwaukee Hill	open z
Palaiópyrgos	open at 9am / close at 9pm	-	729 Maple Lane	open z
Dmytrivka	open at 9am / close at 9pm	-	5 Tony Point	open z
Itô	open at 9am / close at 9pm	414-0055	7155 Truax Center	open z
Plesná	open at 9am / close at 9pm	725 27	92 Parkside Circle	open z