

# Lab Exercise 4 – Ingestion (Structured + Semi/Unstructured Events)

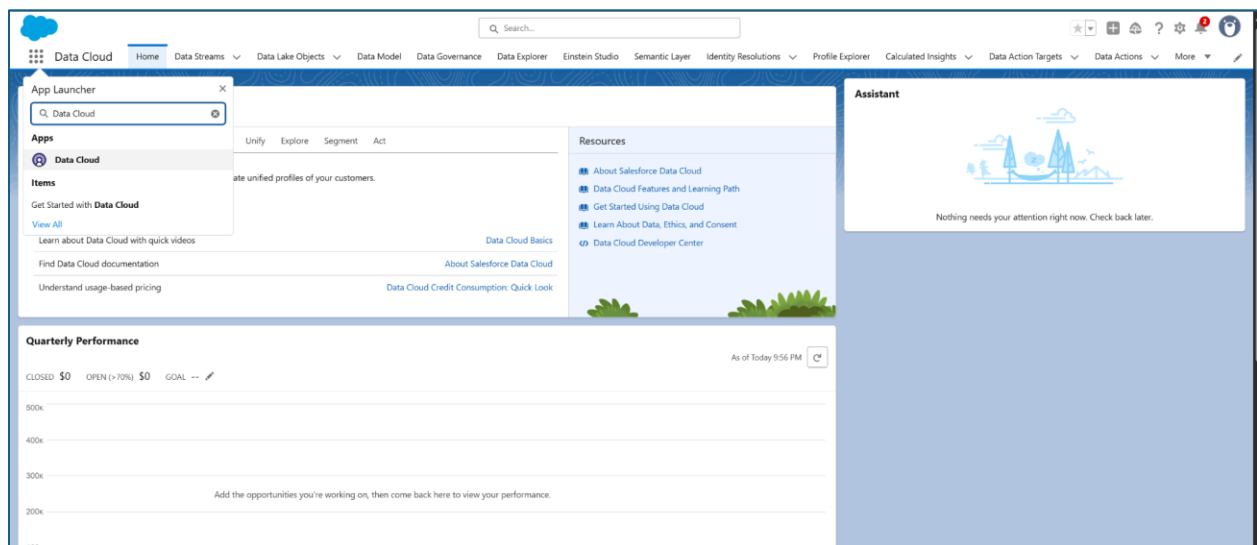
## Objective

By the end of this exercise, you will:

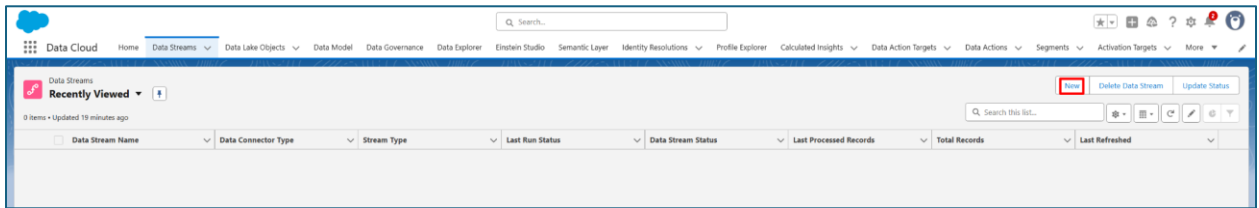
- Create and validate a **Data Stream** (structured data from Salesforce CRM).
- Ingest **Web Engagement events** (semi/unstructured).
- Understand **schema inference, primary keys, and modified date fields**.
- Explore **run history, retries, and error handling**.
- Learn when and why to add **formula fields**.

## Step 1: Create a Salesforce CRM Data Stream

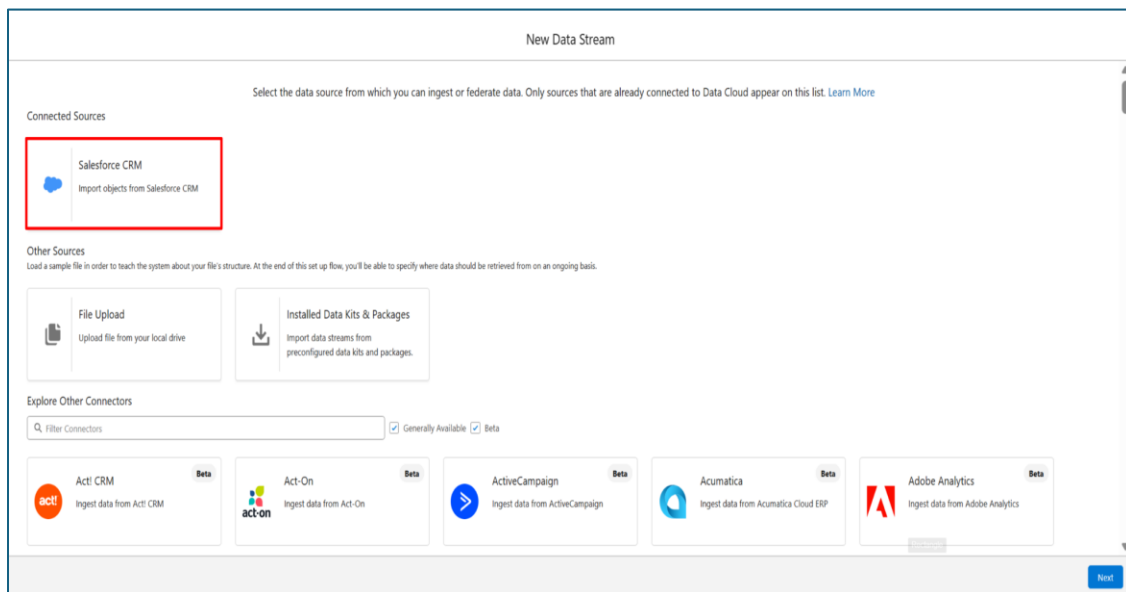
1. Open the **Data Cloud** app (from App Launcher).



2. Navigate to the **Data Streams** tab → click **New Data Stream**.



3. Select **Salesforce CRM** as the source.



4. You'll see two possible cases:

#### **Case A: CRM Bundles Available**

- Choose the **Contacts & Leads bundle** or **Sales bundle**.
- Review objects included (e.g., Account, Contact, Lead).
- Assign **Category = Profile**.
- Select **Primary Key = Id**.
- Select **Record Modified Date = LastModifiedDate**.

## Case B: Bundles Not Available (DE Limitation)

Instead of **View Bundles**, click **View Objects**.

New Data Stream

**i** To ensure data is ingested from fields and objects created in the future, we recommend granting View All Fields (Global) system permission on the Data Cloud Salesforce Connector permission set. [Learn More](#)

Select an org to ingest data from, then select an object or data bundle.

\*Salesforce Org

Salesforce, Demo [View Bundles](#) **View Objects**

Salesforce Objects (976) 0 selected

<input type="checkbox"/> Object Name ↓	<input type="checkbox"/> Object API Name	<input type="checkbox"/> Data Stream	<input type="checkbox"/> Is Custom Object
<input type="checkbox"/> Data Action Job Summary	DataActionJobSummary		
<input type="checkbox"/> _MISSING LABEL_ PropertyFile - val AppointmentCategory n...	AppointmentCategoryFeed		
<input type="checkbox"/> _MISSING LABEL_ PropertyFile - val AppointmentInvitation n...	AppointmentInvitationFeed		
<input type="checkbox"/> _MISSING LABEL_ PropertyFile - val AppointmentTopicTimeS...	AppointmentTopicTimeSlotFeed		
<input type="checkbox"/> _MISSING LABEL_ PropertyFile - val CreditMemoInvApplicati...	CreditMemoInvApplicationFeed		
<input type="checkbox"/> _MISSING LABEL_ PropertyFile - val DataCommCapActvTarg...	DataCommCapActvTargetFeed		
<input type="checkbox"/> _MISSING LABEL_ PropertyFile - val DataObjSecondaryIndex...	DataObjSecondaryIndexFeed		
<input type="checkbox"/> _MISSING LABEL_ PropertyFile - val DataSemanticSearch not...	DataSemanticSearchFeed		
<input type="checkbox"/> _MISSING LABEL_ PropertyFile - val DataSpace not found in ...	DataSpaceFeed		

[Previous](#) [Next](#)

Select **Contact** (and optionally **Lead**).

New Data Stream

**i** To ensure data is ingested from fields and objects created in the future, we recommend granting View All Fields (Global) system permission on the Data Cloud Salesforce Connector permission set. [Learn More](#)

Select an org to ingest data from, then select an object or data bundle.

\* Salesforce Org  
 Salesforce\_Demo View Bundles View Objects

**Salesforce Objects (976)** 1 selected ⓘ Contact

Object Name ↓	Object API Name	Data Stream	Is Custom Object
<input type="checkbox"/> Account Contact Role	AccountContactRole		
<input type="checkbox"/> Activation Contact Point	MktSgmtActvtnContactPoint		
<input type="checkbox"/> Activation Contact Point Field	MktSgmtActvContactPtField		
<input type="checkbox"/> Activation Contact Point Source	MktSgmtActvContactPtSrc		
<input type="checkbox"/> Case Contact Role	CaseContactRole		
<input checked="" type="checkbox"/> Contact	Contact		
<input type="checkbox"/> Contact Center Channel	ContactCenterChannel		
<input type="checkbox"/> Contact Clean Info	ContactCleanInfo		
<input type="checkbox"/> Contact Feed	ContactFeed		

Previous ○ Next

New Data Stream

**i** To ensure data is ingested from fields and objects created in the future, we recommend granting View All Fields (Global) system permission on the Data Cloud Salesforce Connector permission set. [Learn More](#)

Select an org to ingest data from, then select an object or data bundle.

\* Salesforce Org  
 Salesforce\_Demo View Bundles View Objects

**Salesforce Objects (976)** 2 selected ⓘ Lead

Object Name ↓	Object API Name	Data Stream	Is Custom Object
<input checked="" type="checkbox"/> Lead	Lead		
<input type="checkbox"/> Lead Clean Info	LeadCleanInfo		
<input type="checkbox"/> Lead Feed	LeadFeed		
<input type="checkbox"/> Lead Status Value	LeadStatus		

Previous ○ Next

- Assign Category = **Profile** in both **Contact** and **Lead**.

**New Data Stream**

**⚠️** The object's category impacts how you can map it to data model objects. After saving the data stream, you can't change the category. [Learn More](#)

Data Cloud creates a data lake object (DLO) to store data from each Salesforce object in the data stream. Select an object to configure details for its related DLO. After saving the data stream, you can't change these configurations.

Search...

**Objects**

**Contact**

Lead

### Contact Details

\* Data Lake Object Label <sup>?</sup>

\* Data Lake Object API Name <sup>?</sup>

\* Data Stream Name

\* Object Category <sup>?</sup> Profile

\* Primary Key <sup>?</sup>

**Contact Fields (71)** 71 fields selected  [New Formula Field](#)

Standard Fields (69) Custom Fields (2) Formula Fields (0)

	Required	Field Name	Field Label	Field API Name	Data Type
1	<input checked="" type="checkbox"/>	Title	Title	Title	Text
2	<input checked="" type="checkbox"/>	System Modstamp	System Modstamp	SystemModstamp	DateTime (yyyy-MM-dd HH:mm:ss)
3	<input checked="" type="checkbox"/>	Salutation	Salutation	Salutation	Text
4	<input checked="" type="checkbox"/>	Reports To ID	Reports To ID	ReportsToId	Text
5	<input checked="" type="checkbox"/>	Pronouns	Pronouns	Pronouns	Text

[Previous](#)
●
○
[Next](#)

**New Data Stream**

**⚠️** The object's category impacts how you can map it to data model objects. After saving the data stream, you can't change the category. [Learn More](#)

Data Cloud creates a data lake object (DLO) to store data from each Salesforce object in the data stream. Select an object to configure details for its related DLO. After saving the data stream, you can't change these configurations.

Search...

**Objects**

Contact

**Lead**

### Lead Details

\* Data Lake Object Label <sup>?</sup>

\* Data Lake Object API Name <sup>?</sup>

\* Data Stream Name

\* Object Category <sup>?</sup> Profile

\* Primary Key <sup>?</sup>

**Lead Fields (67)** 67 fields selected  [New Formula Field](#)

Standard Fields (62) Custom Fields (5) Formula Fields (0)

	Required	Field Name	Field Label	Field API Name	Data Type
1	<input checked="" type="checkbox"/>	Website	Website	Website	URL
2	<input checked="" type="checkbox"/>	Title	Title	Title	Text
3	<input checked="" type="checkbox"/>	System Modstamp	System Modstamp	SystemModstamp	DateTime (yyyy-MM-dd HH:mm:ss)
4	<input checked="" type="checkbox"/>	Street	Street	Street	Text
5	<input checked="" type="checkbox"/>	Status	Status	Status	Text

[Previous](#)
●
○
[Next](#)

- Click **Next** → **Deploy**.

New Data Stream

Put the finishing touches on your data stream(s).

\*Data Space   
 default

This bundle creates 2 data stream(s) and data lake object(s).

Object Name	Category	Refresh Mode	Data Space Filtering
Lead_Home	Profile	Upsert	<a href="#">Set Filters</a>
Contact_Home	Profile	Upsert	<a href="#">Set Filters</a>

Previous

Deploy

#### Frequently Asked Questions

**What are data space filters?**  
Data space filters let you determine which records from the data lake object are available in the context of a data space.

**What is a refresh mode?**  
After the initial data ingestion, you can opt to replace only the fields for which new data was received (partial refresh) or to replace the entire record with the data received (incremental refresh). When refresh mode is incremental, existing values can be replaced by blank values.

- Check the status → it should move from *Processing* to *Active*.

Data Cloud

Home
 Data Streams
 Data Lake Objects
 Data Model
 Data Governance
 Data Explorer
 Einstein Studio
 Semantic Layer
 Identity Resolutions
 Profile Explorer
 Calculated Insights
 Data Action Targets
 Data Actions
 Segments
 Activation Targets
 More

Data Streams
 Recently Viewed

2 Items • Updated a few seconds ago

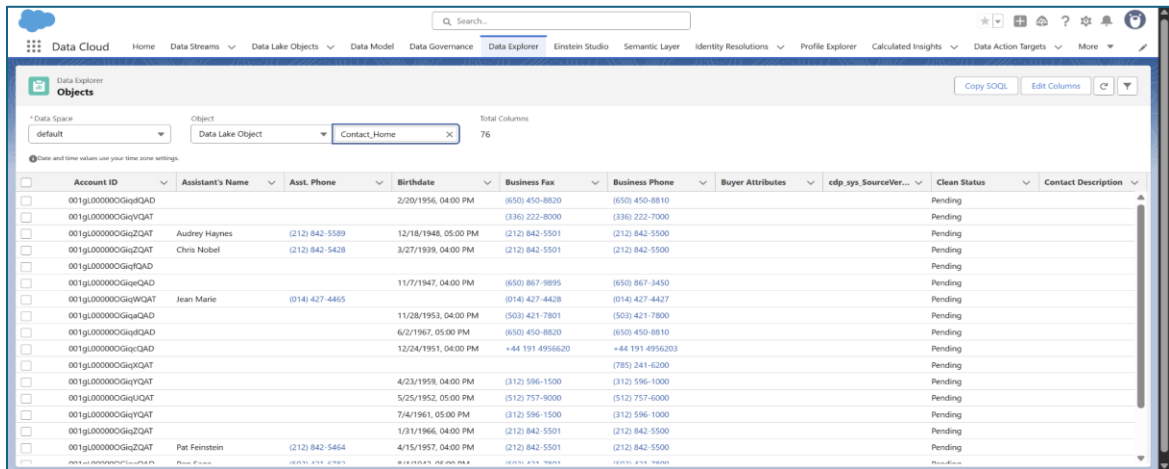
New
 Delete Data Stream
 Update Status

Search this list...

	Data Stream Name	Data Connector Type	Stream Type	Last Run Status	Data Stream Status	Last Processed Records	Total Records	Last Refreshed
1	<a href="#">Contact_Home</a>	Salesforce CRM	Ingest	None	Active			
2	<a href="#">Lead_Home</a>	Salesforce CRM	Ingest	None	Active			

## Step 2: Validate Data

1. Open **Data Explorer**.
2. Select the ingested object (e.g., Contact).
3. Spot-check a few records for values like Assistant's Name, Account ID, and Phone.

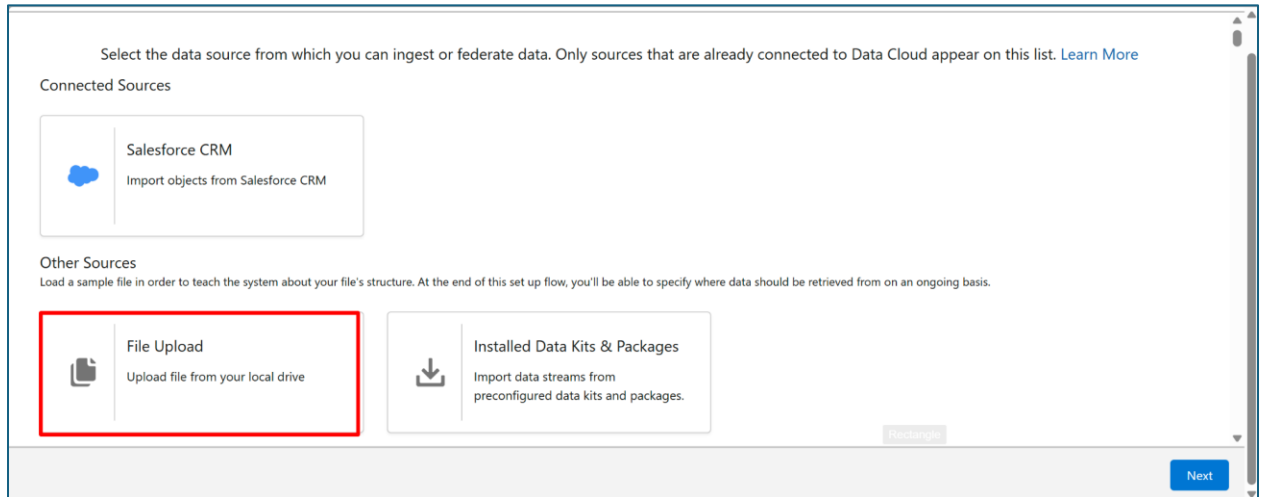


The screenshot shows the Data Cloud Data Explorer interface. The top navigation bar includes links to Home, Data Streams, Data Lake Objects, Data Model, Data Governance, Data Explorer (selected), Einstein Studio, Semantic Layer, Identity Resolutions, Profile Explorer, Calculated Insights, Data Action Targets, and More. A search bar is located in the top right. Below the navigation bar, the 'Data Explorer' section is active, showing a list of objects. The 'Contact\_Home' object is selected, displaying a table of 76 records. The table has columns for Account ID, Assistant's Name, Asst. Phone, Birthdate, Business Fax, Business Phone, Buyer Attributes, cdp\_sys.SourceVer..., Clean Status, and Contact Description. The first few rows of data are visible, showing records for Audrey Haynes, Chris Nobel, Jean Marie, and Pat Feinstein.

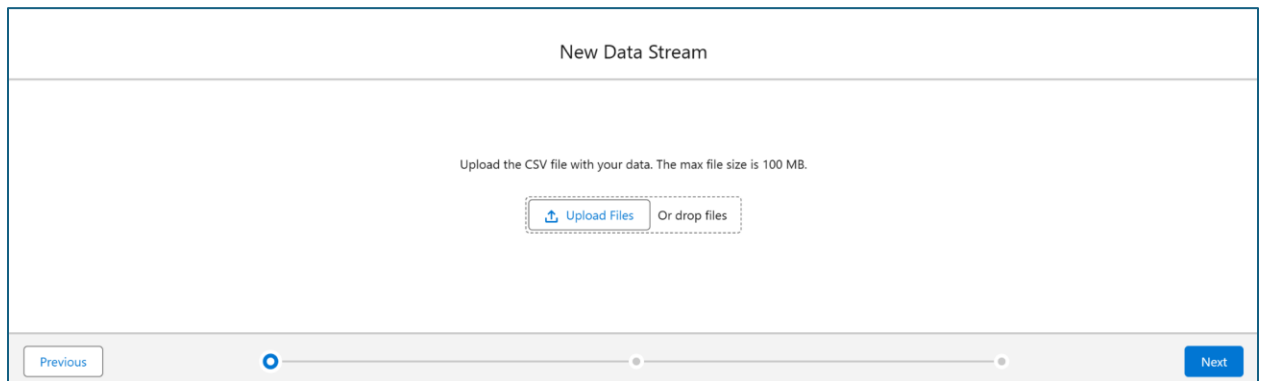
Account ID	Assistant's Name	Asst. Phone	Birthdate	Business Fax	Business Phone	Buyer Attributes	cdp_sys.SourceVer...	Clean Status	Contact Description
001g000000GiqpQAD			2/25/1956, 04:00 PM	(850) 450-8820	(650) 450-8810			Pending	
001g000000GiqvQAT				(336) 222-8000	(336) 222-7000			Pending	
001g000000GiqzQAT	Audrey Haynes	(212) 842-5589	12/18/1948, 05:00 PM	(212) 842-5501	(212) 842-5500			Pending	
001g000000GiqzQAT	Chris Nobel	(212) 842-5428	3/27/1939, 04:00 PM	(212) 842-5501	(212) 842-5500			Pending	
001g000000GiqpQAD								Pending	
001g000000GiqpQAD			11/7/1947, 04:00 PM	(850) 867-9895	(650) 867-3450			Pending	
001g000000GiqpQAT	Jean Marie	(014) 427-4465		(014) 427-4428	(014) 427-4427			Pending	
001g000000GiqpQAD			11/28/1953, 04:00 PM	(503) 421-7801	(503) 421-7800			Pending	
001g000000GiqpQAD			6/2/1967, 05:00 PM	(650) 450-8820	(650) 450-8810			Pending	
001g000000GiqpQAD			12/24/1951, 04:00 PM	+44 191 4956620	+44 191 4956203			Pending	
001g000000GiqpQAT					(785) 241-6200			Pending	
001g000000GiqpQAT			4/23/1959, 04:00 PM	(312) 596-1500	(312) 596-1000			Pending	
001g000000GiqpQAT			5/25/1952, 05:00 PM	(512) 757-9000	(512) 757-6000			Pending	
001g000000GiqpQAT			7/4/1961, 05:00 PM	(312) 596-1500	(312) 596-1000			Pending	
001g000000GiqpQAT			1/31/1966, 04:00 PM	(212) 842-5501	(212) 842-5500			Pending	
001g000000GiqzQAT	Pat Feinstein	(212) 842-5464	4/15/1957, 04:00 PM	(212) 842-5501	(212) 842-5500			Pending	

### Step 3: Ingest Web Engagement Data (Events)

1. In **Data Streams** → **New Data Stream**, select **File Upload** or **Web Engagement Events** (if available).



- For DE orgs, the quickest option is **Upload File (CSV)**.



2. Upload a sample events file (you can create one in Excel/CSV with columns like EventID, EventName, Timestamp, Channel, UserID).



Select the data lake object your data will land in, configure object details, and manage your fields.

\* Data Lake Object:

\* Data Lake Object Label:

\* Data Lake Object API Name:

**Properties** Field Defaults

\* Category

☒ Profile <sup>?</sup>

☐ Engagement <sup>?</sup>

☐ Other <sup>?</sup>

**You can't change the Category after saving the data stream. Consider billing and functional implications when making this decision. [Learn More in Help](#)**

\* Primary Key

Record Modified Field <sup>?</sup>

files (8)

**Sample Data** Supported Fields (5) Unsupported Fields (0) Lineage Fields (3) Formula Fields (0)

	EventID	EventName	EventDate	Channel	UserID
1	E001	VideoPlay	2025-10-03 00:39:30	Email	U100
2	E002	ButtonClick	2025-10-03 00:34:30	Web	U103
3	E003	VideoPlay	2025-10-03 00:29:30	Mobile	U100
4	E004	Purchase	2025-10-03 00:24:30	Mobile	U105
5	E005	AddToCart	2025-10-03 00:19:30	Web	U100

☒ ☐ ☐

### 3. Assign **Category = Engagement**.

Select the data lake object your data will land in, configure object details, and manage your fields.

\* Data Lake Object:

\* Data Lake Object Label:

\* Data Lake Object API Name:

**Properties** Field Defaults

\* Category

☐ Profile <sup>?</sup>

☒ **Engagement** <sup>?</sup>

☐ Other <sup>?</sup>

**You can't change the Category after saving the data stream. Consider billing and functional implications when making this decision. [Learn More in Help](#)**

\* Event Time Field

\* Primary Key

files (8)

**Sample Data** Supported Fields (5) Unsupported Fields (0) Lineage Fields (3) Formula Fields (0)

	EventID	EventName	EventDate	Channel	UserID
1	E001	VideoPlay	2025-10-03 00:39:30	Email	U100
2	E002	ButtonClick	2025-10-03 00:34:30	Web	U103
3	E003	VideoPlay	2025-10-03 00:29:30	Mobile	U100
4	E004	Purchase	2025-10-03 00:24:30	Mobile	U105
5	E005	AddToCart	2025-10-03 00:19:30	Web	U100

☒ ☐ ☐

### 4. Required setup:

- **Primary Key = EventID**

- **Timestamp = EventDate** (or your event timestamp column).

**Category**

- ☐ Profile
- ☒ Engagement
- ☐ Other

You can't change the Category after saving the data stream. Consider billing and functional implications when making this decision. [Learn More in Help](#)

**Event Time Field**

EventDate

**Primary Key**

EventID

**Record Modified Field**

Select an Option

**Organization Unit Identifier**

Select an Option

**Sample Data**

	EventID	EventName	EventDate	Channel	UserID
1	E001	VideoPlay	2025-10-03 00:39:30	Email	U100
2	E002	ButtonClick	2025-10-03 00:34:30	Web	U103
3	E003	VideoPlay	2025-10-03 00:29:30	Mobile	U100
4	E004	Purchase	2025-10-03 00:24:30	Mobile	U105
5	E005	AddToCart	2025-10-03 00:19:30	Web	U100
6	E006	VideoPlay	2025-10-03 00:14:30	Mobile	U100
7	E007	VideoPlay	2025-10-03 00:09:30	Web	U101
8	E008	ButtonClick	2025-10-03 00:04:30	Mobile	U101

Previous Next

## 5. Finish and Deploy.

**New Data Stream**

Upsert may cause slower ingestion. [Learn more in Help](#)

**Data Stream Name**

sample\_web\_events.csv

**Data Space**

default [Set Filters](#)

Previous Deploy

**Note:** Engagement category requires a timestamp field because Data Cloud needs event ordering for analytics and insights.

## Step 4: Review Run History, Errors & Retries

1. Open the **Data Streams** tab.
2. Click on your Data Stream name [eg. sample\_web\_events]. This takes you to the **Data Stream details page → Refresh History**.

The screenshot displays the Salesforce Data Cloud interface for a data stream named 'sample\_web\_events.csv'. The 'Refresh History' tab is highlighted with a red box. The history table shows one successful refresh on 10/2/2025 at 5:53 PM, processing 10 records. The 'Data Mapping' section on the right shows 0/0 fields mapped and a 'READY' status.

Refresh Date	Refresh M...	Duration	Status	Records Pr...	Records A...	Records Re...	
1	10/2/2025, 5:53 ...	Total Replace	0:0:1	Success	10	10	0

3. Check for Status:
  - Successful ingestion.
  - Errors (e.g., missing fields, bad schema).

This is a close-up view of the 'Refresh History' table. It shows a single entry with a status of 'Success' and 10 records processed.

Refresh Date	Refresh M...	Duration	Status	Records Pr...	Records A...	Records Re...	
1	10/2/2025, 5:53 ...	Total Replace	0:0:1	Success	10	10	0

4. Try **Manual Refresh** to re-run ingestion.

DataStream

sample\_web\_events.csv

+ Follow

Update File

Delete Data Stream

Update Status

Stream Type

Data Stream Status

Last Run Status

Last Refreshed

Last Processed Records

Total Records

Ingest

Active

Success

10/2/2025, 5:53 PM

10

10

Fields

Details

Refresh History

Refresh History (1)

	Refresh Date	Refresh M...	Duration	Status	Records Pr...	Records A...	Records Re...
1	10/2/2025, 5:53 ...	Total Replace	0:0:1	Success	10	10	0

Data Mapping

Data mappings have ethics, privacy, and consent considerations. [Learn more about ethics, privacy, and consent in Salesforce Help](#)

Only mapped fields or objects with relationships can be used for segmentation and insights

DataStream

sample\_web\_events.csv

+ Follow

Update File

Delete Data Stream

Update Status

Stream Type

Data Stream Status

Last Run Status

Last Refreshed

Last Processed Records

Total Records

Ingest

Active

Success

10/2/2025, 5:53 PM

10

10

Fields

Details

Refresh History

Refresh History (1)

	Refresh Da...	Refresh M...	Duration	Status	Records Pr...	Records A...	Records R...
1	10/2/2025, 5:53...	Total Replace	0:0:1	Success	10	10	0

Data Mapping

Data mappings have ethics, privacy, and consent considerations. [Learn more about ethics, privacy, and consent in Salesforce Help](#)

Only mapped fields or objects with relationships can be used for segmentation and insights

Data Space

default

Fields mapped

0/0

READY

Start

Post

Poll

Question

If errors exist, use **Retry** to fix and rerun.