

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

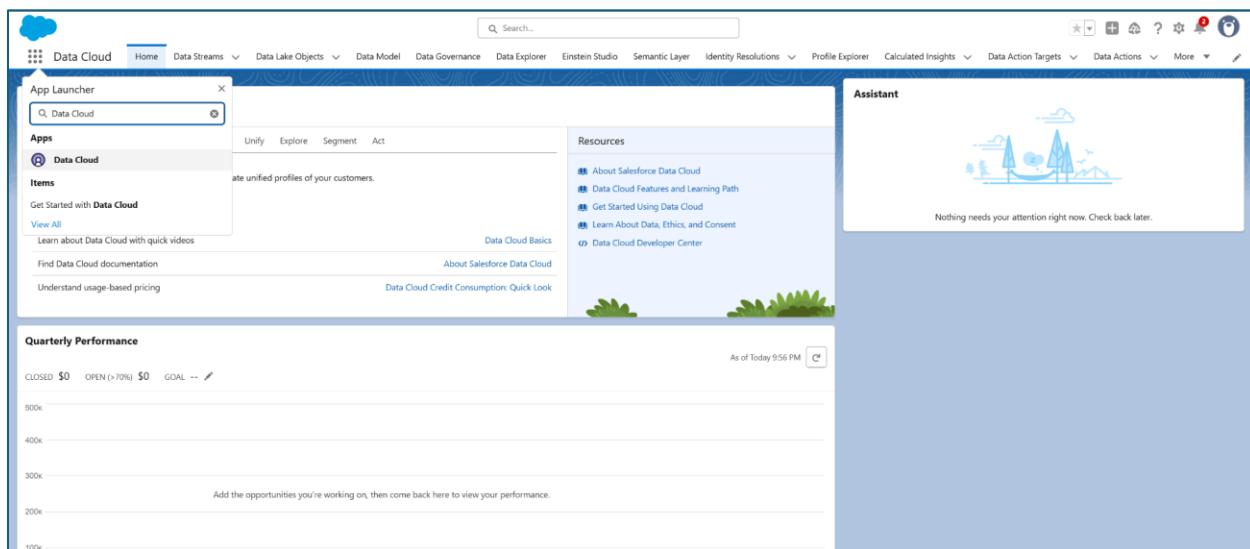
Objective

By the end of this exercise, you will:

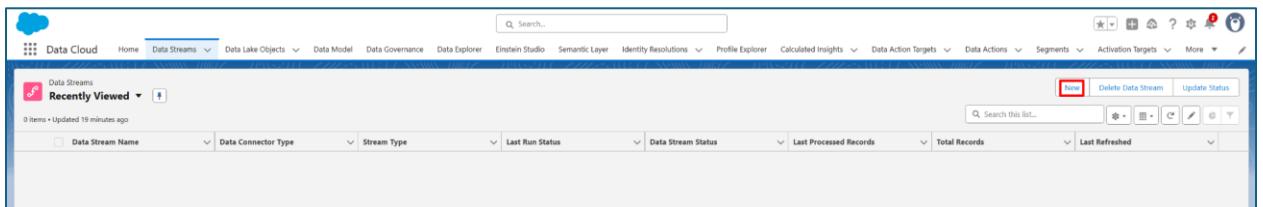
- Create and validate a **Data Stream** (structured data from Salesforce CRM).
- Ingest **Web Engagement events** (semi/unstructured).
- Understand **schema inference, primary keys, and modified date fields**.
- Explore **run history, retries, and error handling**.
- Learn when and why to add **formula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

Step 1: Create a Salesforce CRM Data Stream

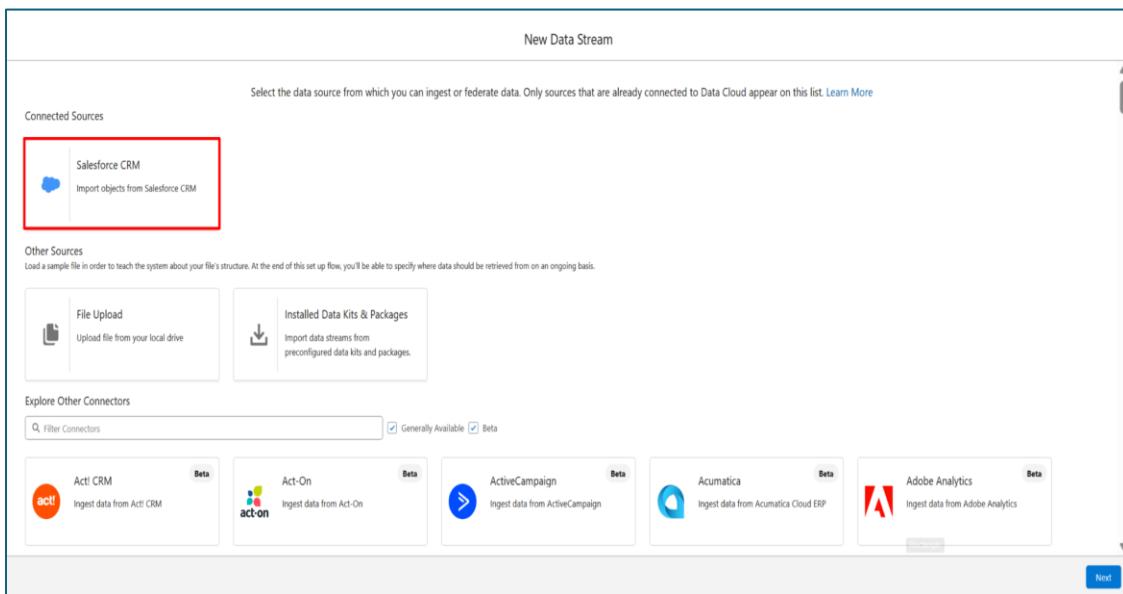
1. Open the **Data Cloud app** (from App Launcher).



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**.

The screenshot shows the 'New Data Stream' configuration page. At the top, there is a note: '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.' A 'Learn More' link is available. Below this, a dropdown menu shows 'Salesforce Org' set to 'Salesforce_Demo'. There are two tabs: 'View Bundles' (disabled) and 'View Objects' (selected). The main area displays a list of 'Salesforce Objects (976)' with columns for 'Object Name', 'Object API Name', 'Data Stream', and 'Is Custom Object'. The 'Data Stream' column contains mostly placeholder names like 'AppointmentCategoryFeed' and 'AppointmentInvitationFeed'. A search bar and a 'Next' button are at the bottom.

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

New Data Stream

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](#)

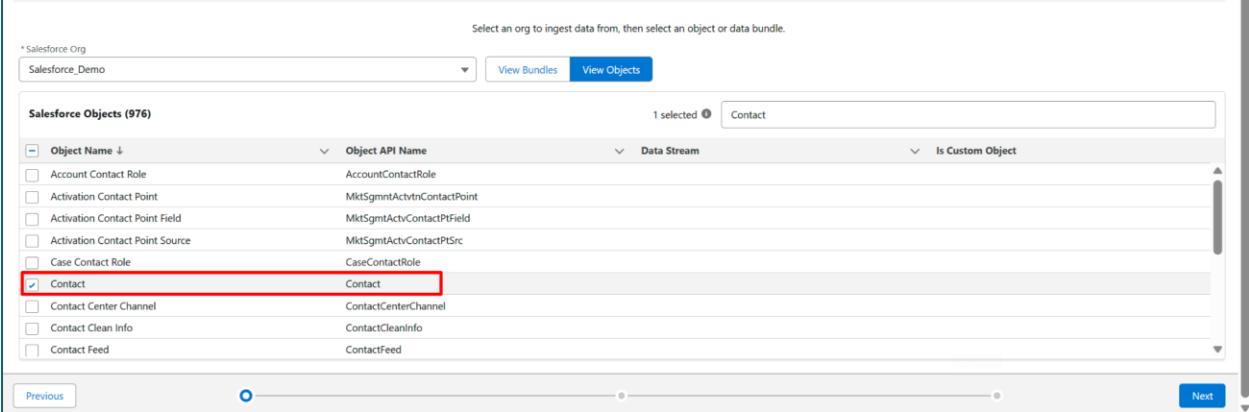
Select an org to ingest data from, then select an object or data bundle.

* Salesforce Org: **Salesforce_Demo** [View Bundles](#) [View Objects](#)

Salesforce Objects (976) 1 selected Contact

Object Name	Object API Name	Data Stream	Is Custom Object
Account Contact Role	AccountContactRole		
Activation Contact Point	MktSgmnActvtnContactPoint		
Activation Contact Point Field	MktSgmnActvContactPtfld		
Activation Contact Point Source	MktSgmnActvContactPtSrc		
Case Contact Role	CaseContactRole		
<input checked="" type="checkbox"/> Contact	Contact		
Contact Center Channel	ContactCenterChannel		
Contact Clean Info	ContactCleaninfo		
Contact Feed	ContactFeed		

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



New Data Stream

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](#)

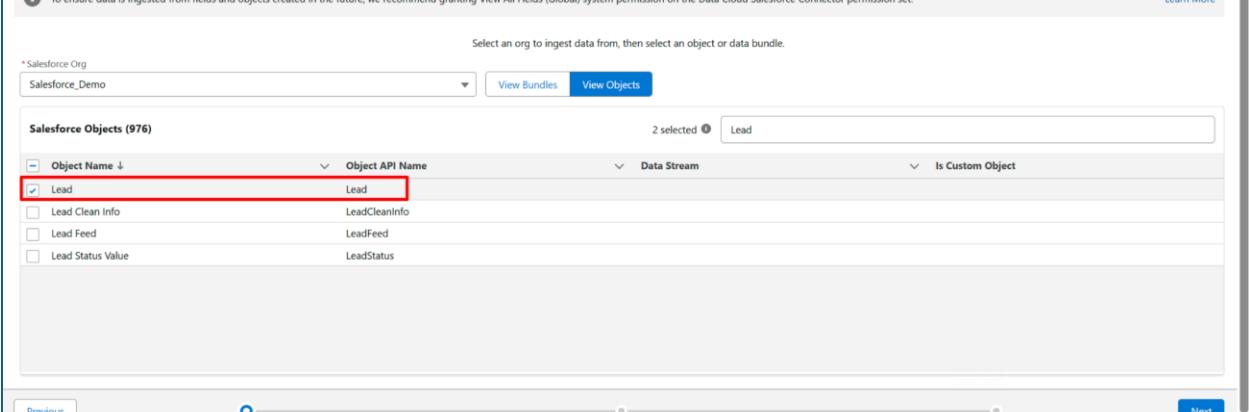
Select an org to ingest data from, then select an object or data bundle.

* 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		
Lead Clean Info	LeadCleanInfo		
Lead Feed	LeadFeed		
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.

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.

Contact Details

* Data Lake Object Label	* Data Lake Object API Name	* Data Stream Name
Contact_Home	Contact_Home	Contact_Home
* Object Category	Primary Key	Contact ID
Profile		

Contact Fields (71) 71 fields selected

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-I)
3 <input checked="" type="checkbox"/>	Salutation	Salutation	Salutation	Text
4 <input checked="" type="checkbox"/>	Reports To ID	Reports To ID	ReportsTold	Text
5 <input checked="" type="checkbox"/>	Pronouns	Pronouns	Pronouns	Text

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-I)
3 <input checked="" type="checkbox"/>	Salutation	Salutation	Salutation	Text
4 <input checked="" type="checkbox"/>	Reports To ID	Reports To ID	ReportsTold	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.

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.

Lead Details

* Data Lake Object Label	* Data Lake Object API Name	* Data Stream Name
Lead_Home	Lead_Home	Lead_Home
* Object Category	Primary Key	Lead ID
Profile		

Lead Fields (67) 67 fields selected

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-I)
4 <input checked="" type="checkbox"/>	Street	Street	Street	Text
5 <input checked="" type="checkbox"/>	Status	Status	Status	Text

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-I)
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 like object(s).

Object Name	Category	Refresh Mode	Data Space Filtering
Lead_Home	Profile	Upsert	Set Filters
Contact_Home	Profile	Upsert	Set Filters

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.

Previous Deploy

- 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

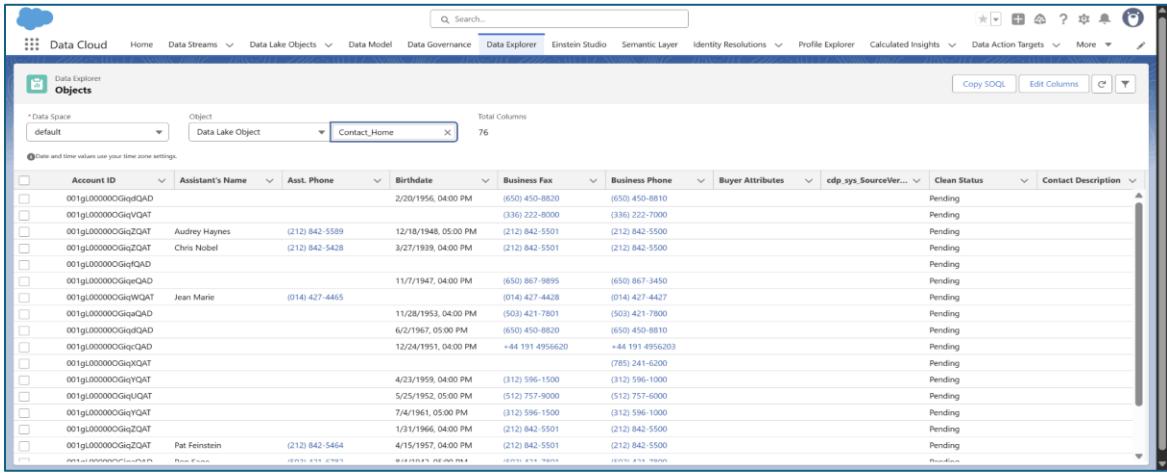
Data Streams Recently Viewed

2 Items • Updated a few seconds ago

Data Stream Name	Data Connector Type	Stream Type	Last Run Status	Data Stream Status	Last Processed Records	Total Records	Last Refreshed
Contact_Home	Salesforce CRM	Ingest	None	Active			
Lead_Home	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.

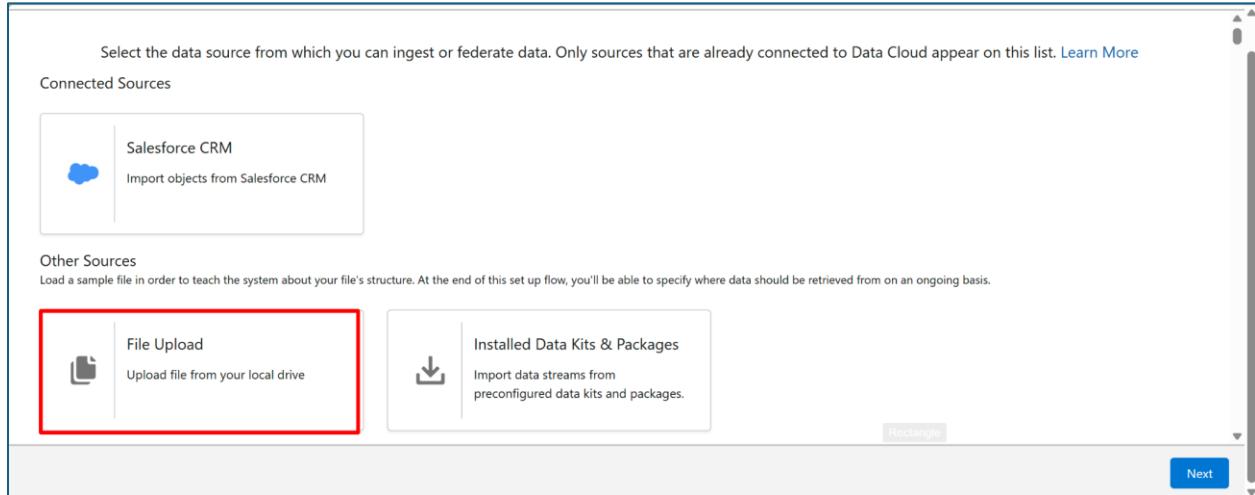


The screenshot shows the Data Explorer interface with the following details:

- Header:** Data Cloud, Home, Data Streams, Data Lake Objects, Data Model, Data Governance, Data Explorer (highlighted), Einstein Studio, Semantic Layer, Identity Resolutions, Profile Explorer, Calculated Insights, Data Action Targets, More.
- Search Bar:** Q Search.
- Filter Bar:** Data Space (default), Object (Data Lake Object), Contact_Home (selected).
- Table Headers:** Account ID, Assistant's Name, Asst. Phone, Birthdate, Business Fax, Business Phone, Buyer Attributes, cdp_sys.SourceVer..., Clean Status, Contact Description.
- Table Data:** A list of contact records with columns including Account ID, Assistant's Name, Asst. Phone, Birthdate, Business Fax, Business Phone, Buyer Attributes, cdp_sys.SourceVer..., Clean Status, and Contact Description. Some rows show placeholder values like '(212) 842-5501' or '(650) 450-8810'.

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)**.

The screenshot shows the 'New Data Stream' setup page. The title 'New Data Stream' is at the top. Below it, it says 'Upload the CSV file with your data. The max file size is 100 MB.' There is a dashed box labeled 'Upload Files' with a plus sign icon and the text 'Or drop files'. At the bottom, there is a navigation bar with 'Previous' and 'Next' buttons, and a progress bar indicating the current step.

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 + New Data Lake Object	* Data Lake Object Label sample_web_events.csv	* Data Lake Object API Name sample_web_eventscsv																																								
Properties Field Defaults <table border="1"> <tr><td>Category</td></tr> <tr><td><input checked="" type="radio"/> Profile</td></tr> <tr><td><input type="radio"/> Engagement</td></tr> <tr><td><input type="radio"/> Other</td></tr> <tr><td>You can't change the Category after saving the data stream. Consider billing and functional implications when making this decision. Learn More in Help</td></tr> <tr><td>* Primary Key</td></tr> <tr><td>Select an Option</td></tr> <tr><td>Record Modified Field</td></tr> </table>			Category	<input checked="" type="radio"/> Profile	<input type="radio"/> Engagement	<input type="radio"/> 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	* Primary Key	Select an Option	Record Modified Field																																
Category																																										
<input checked="" type="radio"/> Profile																																										
<input type="radio"/> Engagement																																										
<input type="radio"/> 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																																										
* Primary Key																																										
Select an Option																																										
Record Modified Field																																										
files (8) <table border="1"> <thead> <tr><th>Sample Data</th><th>Supported Fields (5)</th><th>Unsupported Fields (0)</th><th>Lineage Fields (3)</th><th>Formula Fields (0)</th></tr> <tr><th>EventID</th><th>EventName</th><th>EventDate</th><th>Channel</th><th>UserID</th></tr> </thead> <tbody> <tr><td>1 E001</td><td>VideoPlay</td><td>2025-10-03 00:39:30</td><td>Email</td><td>U100</td></tr> <tr><td>2 E002</td><td>ButtonClick</td><td>2025-10-03 00:34:30</td><td>Web</td><td>U103</td></tr> <tr><td>3 E003</td><td>VideoPlay</td><td>2025-10-03 00:29:30</td><td>Mobile</td><td>U100</td></tr> <tr><td>4 E004</td><td>Purchase</td><td>2025-10-03 00:24:30</td><td>Mobile</td><td>U105</td></tr> <tr><td>5 E005</td><td>AddToCart</td><td>2025-10-03 00:19:30</td><td>Web</td><td>U100</td></tr> <tr><td>6 E006</td><td>VideoPlay</td><td>2025-10-03 00:14:30</td><td>Mobile</td><td>U100</td></tr> </tbody> </table>			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	6 E006	VideoPlay	2025-10-03 00:14:30	Mobile	U100
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																																						
6 E006	VideoPlay	2025-10-03 00:14:30	Mobile	U100																																						
Previous	Next																																									

3. Assign Category = Engagement.

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

* Data Lake Object + New Data Lake Object	* Data Lake Object Label sample_web_events.csv	* Data Lake Object API Name sample_web_eventscsv																																								
Properties Field Defaults <table border="1"> <tr><td>Category</td></tr> <tr><td><input type="radio"/> Profile</td></tr> <tr><td><input checked="" type="radio"/> Engagement</td></tr> <tr><td><input type="radio"/> Other</td></tr> <tr><td>You can't change the Category after saving the data stream. Consider billing and functional implications when making this decision. Learn More in Help</td></tr> <tr><td>* Event Time Field</td></tr> <tr><td>Select an Option</td></tr> <tr><td>* Primary Key</td></tr> </table>			Category	<input type="radio"/> Profile	<input checked="" type="radio"/> Engagement	<input type="radio"/> 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	Select an Option	* Primary Key																																
Category																																										
<input type="radio"/> Profile																																										
<input checked="" type="radio"/> Engagement																																										
<input type="radio"/> 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																																										
Select an Option																																										
* Primary Key																																										
files (8) <table border="1"> <thead> <tr><th>Sample Data</th><th>Supported Fields (5)</th><th>Unsupported Fields (0)</th><th>Lineage Fields (3)</th><th>Formula Fields (0)</th></tr> <tr><th>EventID</th><th>EventName</th><th>EventDate</th><th>Channel</th><th>UserID</th></tr> </thead> <tbody> <tr><td>1 E001</td><td>VideoPlay</td><td>2025-10-03 00:39:30</td><td>Email</td><td>U100</td></tr> <tr><td>2 E002</td><td>ButtonClick</td><td>2025-10-03 00:34:30</td><td>Web</td><td>U103</td></tr> <tr><td>3 E003</td><td>VideoPlay</td><td>2025-10-03 00:29:30</td><td>Mobile</td><td>U100</td></tr> <tr><td>4 E004</td><td>Purchase</td><td>2025-10-03 00:24:30</td><td>Mobile</td><td>U105</td></tr> <tr><td>5 E005</td><td>AddToCart</td><td>2025-10-03 00:19:30</td><td>Web</td><td>U100</td></tr> <tr><td>6 E006</td><td>VideoPlay</td><td>2025-10-03 00:14:30</td><td>Mobile</td><td>U100</td></tr> </tbody> </table>			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	6 E006	VideoPlay	2025-10-03 00:14:30	Mobile	U100
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																																						
6 E006	VideoPlay	2025-10-03 00:14:30	Mobile	U100																																						
Previous	Next																																									

4. Required setup:

- Primary Key = EventID

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

The screenshot shows the Data Cloud interface for creating a new data stream. On the left, there's a sidebar with a 'Category' dropdown set to 'Engagement'. Below it, a note says 'You can't change the Category after saving the data stream. Consider billing and functional implications when making this decision.' Two fields are highlighted with red boxes: 'Event Date Field' (set to 'EventDate') and 'Primary Key' (set to 'EventID'). On the right, a 'Sample Data' table displays 8 rows of event 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

At the bottom, there are 'Previous' and 'Next' buttons.

5. Finish and Deploy.

The screenshot shows the final step of deploying a new data stream. The top bar says 'New Data Stream'. The form includes fields for 'Data Stream Name' (set to 'sample_web_events.csv') and 'Data Space' (set to 'default'). A 'Set Filters' button is also visible. At the bottom, there are 'Previous' and 'Next' buttons, and a large blue 'Deploy' button which is highlighted with a red box.

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 shows the Data Stream details page for "sample_web_events.csv". The "Refresh History" tab is selected. A red box highlights the "Refresh History" button. The table below shows one refresh history entry:

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

Data Mapping section on the right:

- 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
- Buttons: READY (orange), Start (blue)

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

The screenshot shows the Refresh History table from the previous step. It contains one entry:

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

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

Data Stream
sample_web_events.csv

+ Follow Update File Delete Data Stream Update Status

Stream Type Ingest	Data Stream Status Active	Last Run Status Success	Last Refreshed 10/2/2025, 5:53 PM	Last Processed Records 10	Total Records 10
-----------------------	------------------------------	----------------------------	--------------------------------------	------------------------------	---------------------

Fields Details Refresh History

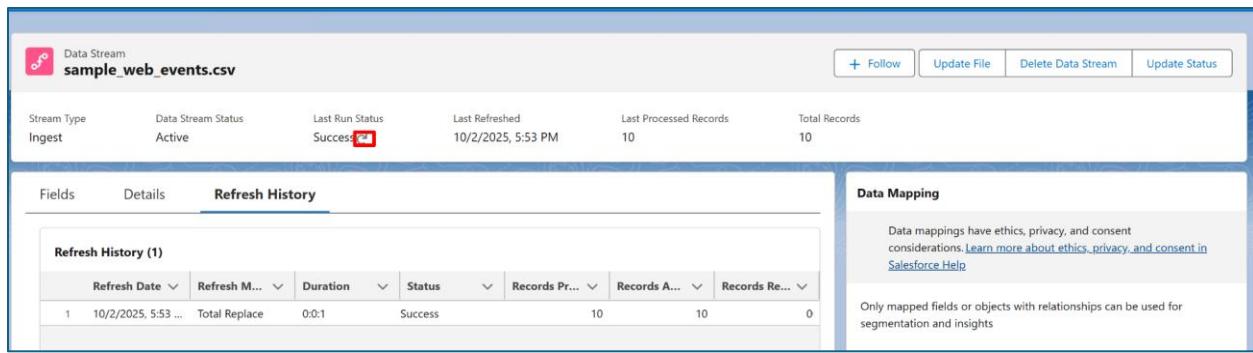
Refresh History (1)

Refresh Date	Refresh M...	Duration	Status	Records Pr...	Records A...	Records R...
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 Stream
sample_web_events.csv

✓ Data stream status was updated.

+ Follow Update File Delete Data Stream Update Status

Stream Type Ingest	Data Stream Status Active	Last Run Status Success	Last Refreshed 10/2/2025, 5:53 PM	Last Processed Records 10	Total Records 10
-----------------------	------------------------------	----------------------------	--------------------------------------	------------------------------	---------------------

Fields Details Refresh History

Refresh History (1)

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

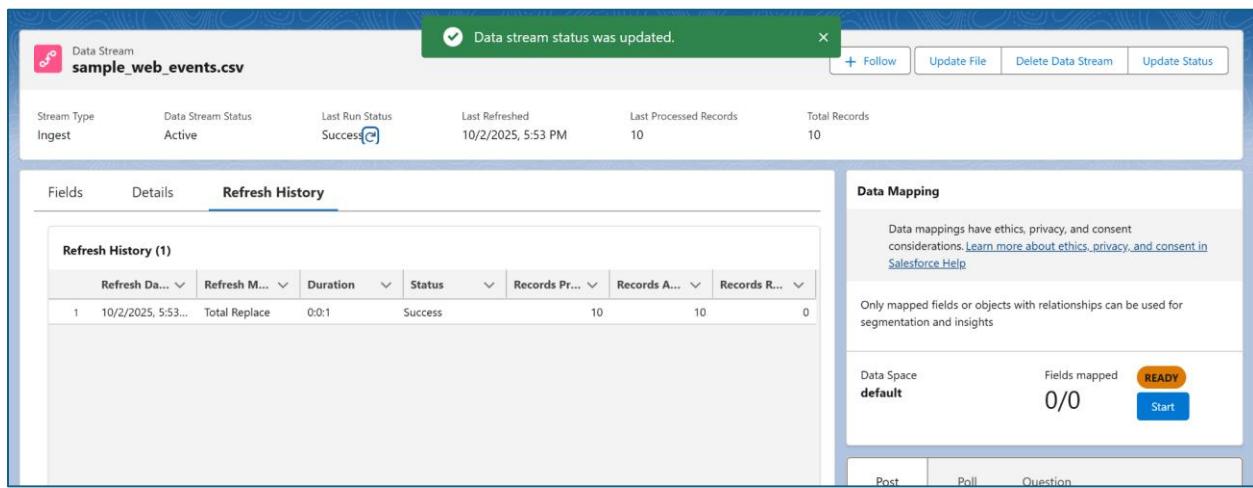
Data Mapping

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 READY
Start

Post Poll Question



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