Step-by-step Guide for Zero Copy and Sharing

Introduction to Zero Copy

Zero Copy serves as a secure feature that enhances your data management and sharing. Zero Copy eliminates the need to move or copy data, bidirectionally sharing your data between Salesforce Data Cloud and your external Data Platforms. Zero Copy ensures you have the most current data, removes data silos, and reduces maintenance costs.

Zero Copy Data Federation enables you to access the data from external data platforms such as Snowflake, Google BigQuery, Databricks and Amazon Redshift in Salesforce Data Cloud. You can think of Data Federation as Data-In. You can use this information to unlock personalized customer experiences and create a comprehensive 360-degree view of the customer.

Data Sharing revolutionizes data collaboration by sharing your Salesforce Data Cloud data to external data platforms. Think of Zero Copy Data Sharing as Data-Out, bringing data from Data Cloud to external data platforms. Once Data Sharing is set up, you will be able to access the Salesforce Data Cloud data as if it is native to the external data platform. Sharing your Data Cloud data prioritizes customer success and allows you to identify strategic insights.

Technical Benefits of Zero Copy

Zero Copy Data Federation offers significant technical benefits by enabling seamless connection between your Data Cloud instance and external data platforms. Through this integration, you can access the data you need without duplicating or moving it. Using Data Streams, you can retrieve only the specific data required, at the desired level of granularity. With Data Sharing (Data Out), you can share your data to external data platforms with clicks not code by creating Data Share Targets. Data Shares allow you to ensure your data analysts and data scientists can use, query or analyze your Data Cloud data in your external data platforms.

Zero Copy in Data Cloud ensures that your data follows the Salesforce security principles ensuring that your data infrastructure is secure.. Zero Copy reduces storage costs and avoids duplication of data. By avoiding multiple copies of data, zero-copy architecture reduces the risk of unauthorized access and data breaches. Zero Copy allows users to access data from their data lakes and warehouses, as well as their SaaS applications in Data Cloud. It also allows users to use their Salesforce data in external systems, like 3rd party analytics or ad software.

Key Technical Features:

- **Ensures data consistency** across platforms by working directly with live data in its native environment.
- **Provides near real-time access to data** from external systems with faster processing time, less complexity, and more scalability than traditional ETL (Extract, Transform, Load) processes.

- Accesses data where it resides (e.g., data lakes, warehouses) with **minimal latency**. Zero Copy Enables real-time use cases like dynamic customer personalization, fraud detection, and instant decision-making.
- **Maintains compliance** with data sovereignty and privacy regulations, as the data stays in its native environment.
- **Leverages external systems' native scalability** to process large datasets without overloading Salesforce infrastructure.
- **Supports growth in data volume** without requiring additional Salesforce storage or compute capacity.
- **Provides out-of-the-box integrations** with external data platforms like Snowflake, Amazon Redshift, Google BigQuery, and others.
- **Reduces complexity in managing data privacy and audit requirements** such as GDPR and CCPA.

Steps to use Zero Copy with Data Cloud

Let's look at our example of Sunshine Trails Hospitality, and how they can use Zero Copy with Data Cloud. Sunshine Trails Hospitality, which possesses valuable POS data, including in-hotel purchases, in Snowflake. By using Zero Copy, we will be able to bring this POS data into Salesforce to create a unified customer profile that combines our Data Cloud and Snowflake data.

To get started, the hotel identifies the tables they wish to bring into Data Cloud from their external data platform,Snowflake. This data can be used to generate rapid insights and provide near real-time information to Customers. Once identified, Sunshine Trails establishes a Connection for Snowflake. Sunshine Trails then creates a Data Stream and enables Acceleration for it. The connected data can be utilized in Data Cloud for segmentation, personalization, analytics, AI & automation.

When creating Data Shares, we will identify the desired Data Cloud objects and add them to a Data Share. Sunshine Trails will link its Data Share to the Data Share targets, enabling you to perform queries in Snowflake. Now that both Data Cloud and external data systems are able to access the data there is now a comprehensive view. In Data Cloud, this means that this data can be used in identity resolution, segmentation, queries, CRM enrichment, and creating calculated insights. For your external data platforms, you will be able to create meaningful insights without copying data.

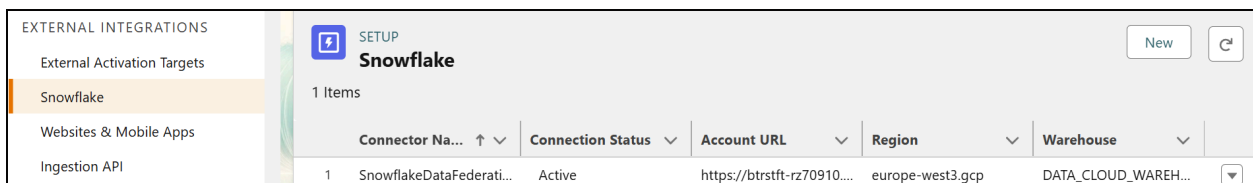| Source System	Entity	Description
Snowflake	Customer Details	Customer demographic attributes and their contact details.
	In-hotel Purchases	Customer purchase information for hotel restaurants, bars, and activities including the total amount, and any applied discounts.

Steps for Data Federation

Remember that Data Federation (Data-In) is the process of bringing your data from external data platforms to Data Cloud. With Data Federation, you are removing any data silos allowing you to analyze all of your information in one place, and enhance your company's ability to deliver personalized customer experiences. To begin, we will want to create connections to our external data platforms and define our Data Streams.

Step 1: Creating Connections

Initiating Zero Copy requires setting up external integrations, a process that can be quickly and easily done in your Data Cloud Setup. To establish an External Integration, Administrative Access to the desired external system is necessary. From there, you will log in with your credentials and follow the steps to create a new connection. Let's look at these steps in more detail:

1. From Data Cloud Setup, select Snowflake under External Integrations > click **New**.
2. Click **Snowflake** > click **Next.**
3. Enter a connection name, connection API name, Account URL, Snowflake username, and private key details > click **Next.**
4. Select the desired database > click **Save.**

Step 2: Creating Data Streams

Data Streams are continuous flows of near real-time data from your external data platforms. This will allow Sunshine Trails Hospitality to have the most up-to-date in-hotel purchase information, meaning they will be able to tell when a customer orders a dish multiple times. Salesforce can then use this information to make recommendations to increase loyalty.

Sunshine Trails Hospitality will need a Data Stream connection for Snowflake. When you select the Data Source, you will be able to make a connection to your specific source to identify the desired databases and schemas as needed.

Ensure you are selecting the desired fields that you want to bring into Data Cloud. For more information on the data types that are supported in Data Cloud, review the Help Article: Data Type Compatibility in Data Federation. For Snowflake, we will pull in the following Fields related to in-hotel purchases:

Source Header	Field Label	Field API Name	Source Data Type	Inferred Data Type
EMAIL	EMAIL	email	VARCHAR	Text
FIRST_NAME	FIRST_NAME	first_name	VARCHAR	Text
GUEST_ID	GUEST_ID	guest_id	Number	Number
SURNAME	SURNAME	surname	VARCHAR	Text
PURCHASE_LINE_ITEM	PURCHASE_LINE_ITEM	purchase_line_item	VARCHAR	Text
TOTAL_COST	TOTAL_COST	total_cost	Number	Number
DISCOUNT	DISCOUNT	discount	Number	Number
DATE_OF_PURCHASE	DATE_OF_PURCHASE	date_of_purchase	DATE	Date

You will bring your external data by creating Data Streams and map it to the Salesforce Data Model Objects. The fields that you have selected to import will be imported to Salesforce as external Data Lake Objects (DLO) in your Salesforce Data Stream.

Step 3: Defining Data Streams

Acceleration

Enabling acceleration uses caching to improve performance and makes this data lake object available for some features that can't use external data lake objects as source data.
Using acceleration impacts the consumption of credits used for billing. Learn More

☑ Enable acceleration

Incremental Column ⓘ

| Select an Option ▾ |

*Frequency

| Daily ▾ |

* Enter a Time (in UTC)

| 6:00 ▾ |

☐ Refresh initial file immediately

After selecting your external objects, you determine whether to perform a live query with your data stream or if you can cache your data. In the case of Live Queries, you are pull the most up to date information each time you perform a query, and this can decrease your data storage costs. Cache Acceleration enables Salesforce to only import in the new and changed records to your Data Stream. It proves particularly useful for datasets which are frequently accessed but rarely change.

💡 For fast-moving data, where real-time updates are critical, use Live Query to prioritize data accuracy. If there is a change in your external Data Lake, it will not automatically trigger the associated Data Action. For such scenarios, you should set up an incremental cache to query the external Data Lake for any changes every 15 minutes.

To activate Cache Acceleration in your Data Stream, you will first need to create your Data Stream. You can click the **Edit Data Stream** button and ensure the Enable acceleration checkbox is checked. You can define the acceleration schedule for the individual Data Stream, with incremental accelerations as frequent as 15 minutes.You are now ready to deploy your data stream!

💡 By enabling Cache Acceleration, you can improve performance, and Cache Acceleration enables you to access additional Data Cloud features.

Once you create your Data Stream, the information is available for you in Data Cloud and you can start making decisions. You will then be able to map the DLOs to the Salesforce Data Model Object (DMO), these mapping objects will allow this data to become a unified source of data in Data Cloud. Once mapped this begins to harmonize your external data so you can begin to create your customer 360 unified profile in Data Cloud.

The team at Sunshine Trails Hospitality can now see what the customer has purchased in the past and provide them with a potential discount to reward their loyalty. Data Federation ensure that you don't have to bring large amounts of data into Data Cloud as the external systems will provide the needed information.

Steps for Data Sharing

Data Sharing (Data-Out) allows you to share data from Data Cloud to external data platforms. By creating a Data Share in Salesforce, you can view Salesforce Data Cloud data in your external data platform and it is natively presented as tables or views. You are not only sharing the Data Cloud data but also have the opportunity to share the unified profile and the rich insights that Data Cloud has produced. Let's explore how to create Data Shares and Data Share Targets.

Step 1: Adding Objects to your Data Share

Before you share your data, it is important to identify which objects are going to be sent out of Data Cloud. In our example, we have identified that we want to share our booking and reservation data with Snowflake. To understand what Objects are available for you, you need to understand the object category:

- Data Lake Objects (DLO) – Storage containers within the data lake for the data ingested into all data streams within Data Cloud.
- Data Model Objects (DMO) – A grouping of data (made up of attributes) created from data streams, insights, and other sources.
- Calculate Insight Objects – A data model object created after a calculated insight is processed.

For our example, we will select the booking object and the reservation object. When you have determined what objects will be sent to Snowflake, you will create your Data Share. Here is what you'll see once you create your Data Share:

Step 2: Creating a Data Share Target

Data Share Targets facilitate you sharing information with your partners such as Snowflake. For Sunshine Trails Hospitality, we have one destination where we want to share data from Data Cloud: Snowflake. If you have multiple accounts, you will need to create a Data Share Target for each of these accounts. To create a Data Share Target follow the steps below:

1. On the Data Share Target tab > click New
2. Select your connection type.
3. Enter the required details including your Account URL.
4. Gather your user credentials, you will first need to create a Data Cloud admin users, create a security integration, gather the OAUTH_AUTHORIZATION_ENDPOINT details, and get your client credentials.
   a. For more information refer to the Help Article: Get User Credentials from Snowflake.
5. Enter the Client ID and Client Secret generated in Snowflake.
6. Enter the description and click Save.
7. Authenticate with your Snowflake credentials.

Your Data Share Target will now be active status.

Step 3: Link your Data Share to a Data Share Target

Once you have created both the Data Share and Data Share Target, you will want to link the two. This linkage allows you to create a virtual database or table where all of your information is shared between Data Cloud and Snowflake.

To link your data, simply navigate to the desired Data Share, and link your Data Share Target button to select the Data Share Target.

Step 4: Access Data in External System

Now your data is connected to Snowflake. When you go to your external system, you can now see your Salesforce tables that have been shared. To view this information in Snowflake, you will follow these steps:

1. Navigate to the **Data** section and select **Private Sharing**.
2. Refresh your direct share by clicking the **Get Data** button.
   a. You will have the ability to rename and grant access to the database as needed.

3. Your database will now appear in the **Databases** section,
4. You can view your data under the **Data Preview** tab.

Native Snowflake users can now incorporate Salesforce data into their work. You can combine your customer data with your operational data in Snowflake, allowing you to understand your business as a whole.

Best Practices for Data Federation

Common Discovery Questions	Design Considerations
● Understand the volume and types of data from external systems	● Filter data at the source if the data is not needed in Data Cloud
● Details about systems where customer data is stored and profile information	● Ensure you choose the right object type during ingestion.
● What use cases drive access to data in real-time?	● Evaluate each object being ingested. Only bring in the objects you require
● Establish your data model prior to bringing in data.	● Only bring in the engagement data you need
● What levels of latency are permissible for near real-time use cases?	● For large data sets use acceleration and set the right period for refresh.
● How often does the underlying data change?	● Start with a small set of 50M rows, and grow from there
● How many segments are going to use the same data?	● Use a WHERE clause when setting up query federation

Best Practices for Data Sharing

Design Considerations
● For data sharing, only share the objects you are going to use.

Use Case for File Federation and Query Federation
File Federation enables direct access to your external data platforms, eliminating the need for additional external compute resources and is best when you are working with large datasets. On the other hand, Query Federation is ideal for real-time queries as you distribute the required compute power between Data Cloud and your external data platform.

Query Federation is best used when there is a greater need for compute resources, while File Federation is preferred when you work with large datasets from storage without involving external compute.

Network and Bandwidth Considerations

You will also want to understand your Network and Bandwidth when designing. Cross-System network latency can impact the speed at which data is retrieved. It is crucial to minimize latency to maintain optimal performance, as queries are executed across external data platforms and involve moving data across networks.

High-bandwidth networks are critical, as Data Cloud directly accesses large data sets stored externally. Slow networks may cause delays in data transfers, which can impact real-time applications. Ensuring Data Cloud and the external data platforms are located in the same region or availability zone can effectively reduce latency.

> When data platforms are not located in the same region, enabling Acceleration serves as an alternative solution to mitigate latency issues.

Securing your Data with Zero Copy

Data security is a key concern when sharing data between platforms, with Zero Copy, you do not need to take any additional steps to ensure your data is secure. Zero Copy with TLS1.2 security handles the end to end management of your data for data federation and data sharing. Your responsibility is in ensuring that only the necessary data is shared with Data Cloud and external data platforms. On the external data platforms, it is crucial to establish who can view the information transferred via Data Shares by setting appropriate access policies.

Conclusion

As you can see, Zero Copy capability facilitates efficient data management and informed decision-making by eliminating data silos and offering a unified view of Salesforce records, including external data sources like Snowflake. Zero Copy can be deployed immediately with Data Cloud and offers customization options.

By following this comprehensive guide, users can effectively set up Zero Copy with Snowflake, ensuring a seamless data-sharing experience. The integration of Zero Copy with Salesforce Data Cloud streamlines data accessibility and management, enhancing the capacity for strategic business decisions.

Resources

Help Article: Data Streams in Data Cloud
Help Article: Manage Data Shares
Help Article: Data Type Compatibility in Data Federation.
Trailhead: Ingestion and Modeling in Data Cloud
Trailhead: BYOL Data Shares in Data Cloud: Quick Look