

Salesforce.org Easy Datalake User Guide

Prepared for Salesforce by the Envision Engineering Team
May 2022



EnvisionEngineering

“© 2022 Amazon Web Services, Inc. or its affiliates. All Rights Reserved.
This AWS Content is provided subject to the terms of the AWS Customer Agreement available
at <http://aws.amazon.com/agreement> or other written agreement between Customer and
either Amazon Web Services, Inc. or Amazon Web Services EMEA SARL or both.”

Notices

Customers are responsible for making their own independent assessment of the information in this document. This document: (a) is for informational purposes only, (b) represents current AWS product offerings and practices, which are subject to change without notice, and (c) does not create any commitments or assurances from AWS and its affiliates, suppliers or licensors. AWS products or services are provided “as is” without warranties, representations, or conditions of any kind, whether express or implied. The responsibilities and liabilities of AWS to its customers are controlled by AWS agreements, and this document is not part of, nor does it modify, any agreement between AWS and its customers.

© 2022 Amazon Web Services, Inc. or its affiliates. All rights reserved.

Contents

1	Introduction	4
2	Setting up the App	4
3	Datalake Setup in 6 Steps	5
3.1	Step 1	5
3.2	Step 2	7
3.3	Step 3	9
3.4	Step 4	11
3.5	Step 5	12
3.6	Step 6	13
4	Resume Feature	15
4.1	Resume while deployment is in progress.....	15
4.2	Resume after deployment	15
4.3	Resume another datalake	16
5	Data Consistency / Data Deletion.....	17

1 Introduction

This document provides guidance for setting up a datalake with the EZ Datalake App. Refer to the EZ Datalake Handover document for instructions on building and deploying the Easy Datalake PoC code in an AWS account.

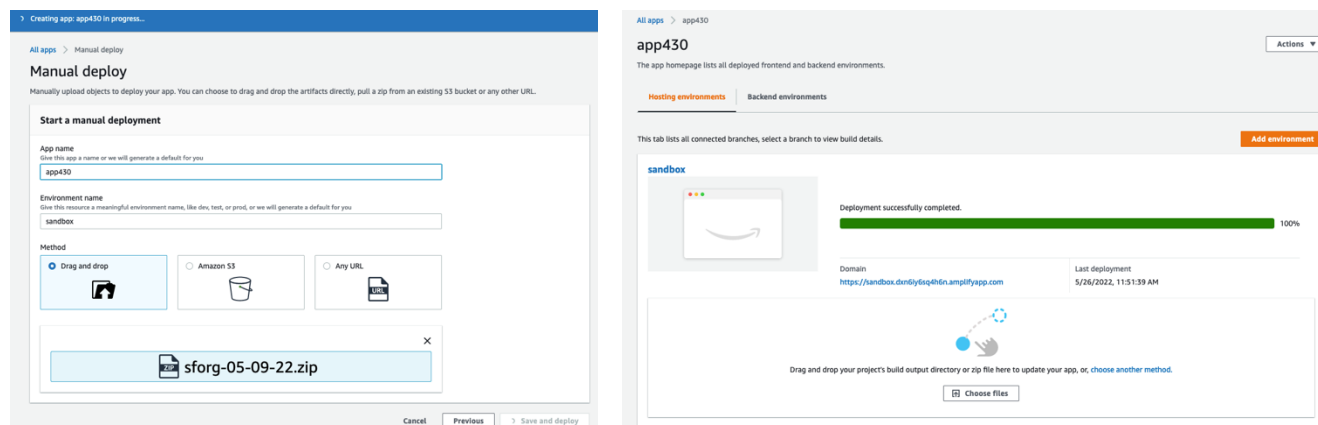
2 Setting up the App

The EZ Datalake is self-hosted within the NPO AWS Account. For this, the app needs to be first deployed as a static website using AWS Amplify.

Sign in to the AWS Management Console and open the [Amplify console](#). Manually deploy the app following the steps here:

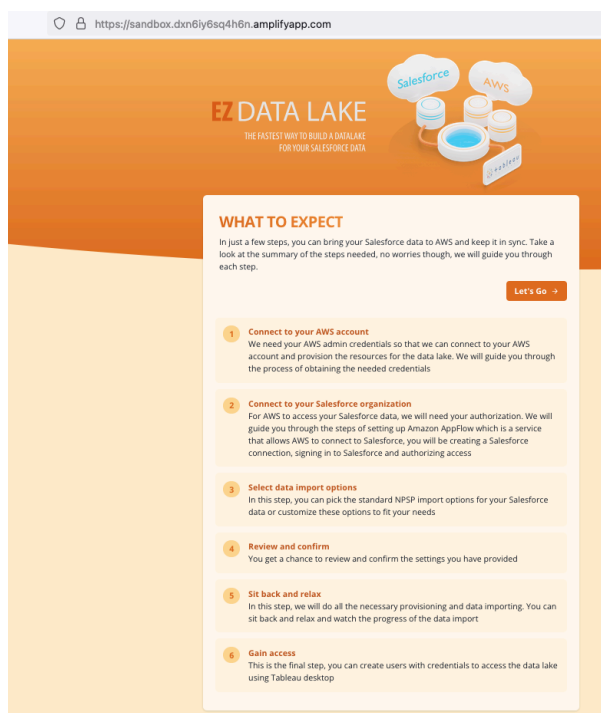
<https://docs.aws.amazon.com/amplify/latest/userguide/manual-deploys.html#drag-and-drop>

The zip file to upload with the manual deployment should be generated with the build process outlined in the EZ Datalake Handover document. Alternatively, you can refer to instructions in the code on generating the build for manual deployment. This build should be made available to users in GitHub to download.



It typically takes a few minutes to deploy the app after which a URL to the EZ Datalake app is presented. Click on the URL to load the EZ Datalake app and set up the data lake and import data in 6 steps.

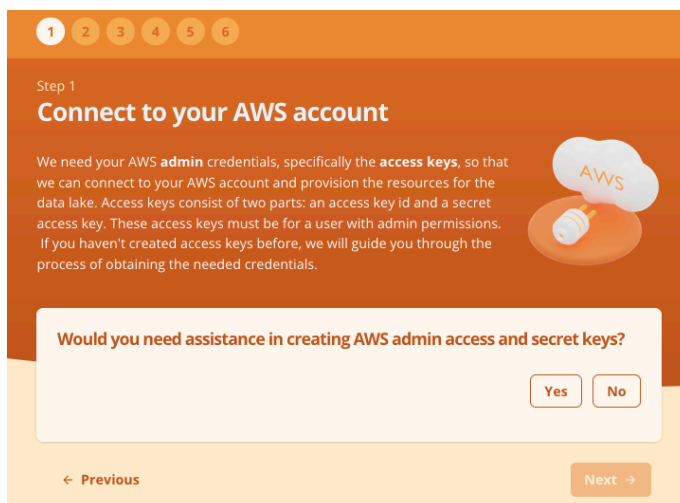
The image below shows the URL that was generated in this instance and the welcome screen showing the user what to expect with an outline of the 6 steps to be completed through the wizard like UI.



3 Datalake Setup in 6 Steps

3.1 Step 1

In Step 1, we connect to the NPO User's AWS Account. The EZ Datalake App requires Access Key ID & Secret Access key for an IAM Admin user.



If the user has never created these, the app shows guidance on how to login to AWS Console and use Identity and Access Management (IAM) to pick the IAM user with admin permissions. The UI then guides the user to create and download the access keys. Once created, the access key information is used to complete Step 1.

1 2 3 4 5 6

Step 1

Connect to your AWS account

We need your AWS admin credentials, specifically the **access keys**, so that we can connect to your AWS account and provision the resources for the data lake. Access keys consist of two parts: an access key id and a secret access key. These access keys must be for a user with admin permissions. If you haven't created access keys before, we will guide you through the process of obtaining the needed credentials.

Would you need assistance in creating AWS admin access and secret keys?

Yes No

Log in to the AWS console using an admin user, not the root user.

Remember to log in using an admin user and not the root user

Then, head over to the Identity and Management section (IAM)

Pick the user name that you used to log in. The user should have admin permissions. Then, select the Security Credentials tab.

Click on the Create access key button to generate the access keys, you will get a chance to view and download the access keys. You can now use the access keys to fill in the information below.

Access Key ID & Secret Access Key

AKIA5NH2BW4GMD7IOEKN

..... Show

Secret access keys are sensitive information, keep them secured and protected, never post them on public platforms or leave them unsecured as this can compromise your account security

AWS Region

AWS has many data centers grouped by geographical regions. Select a region that is closest to your location. We will keep your data in the region you select. If you are not sure, you can select the US East (Ohio) Region

US West (Northern California) Region

At the bottom of Step 1, there is a drop-down box to select the region where the data lake should be provisioned.

Access Key ID & Secret Access Key

Access key id

Secret access key Show

1 Secret access keys are sensitive information, keep them secured and protected, never post them on public platforms or leave them unsecured as this can compromise your account security

AWS Region

US East (Northern Virginia) Region
US West (Oregon) Region
Europe (Ireland) Region
✓ US East (Ohio) Region
US West (Northern California) Region

← Previous Next →

3.2 Step 2

In Step 2, the EZ Datalake app will gather credentials to connect to the NPO Salesforce Organization. For AWS to access the NPO's Salesforce data, we will need user authorization. For this, the app gives guidance to the user to create a dedicated Salesforce user with the necessary profile or permission set that provides:

- Access to the objects and fields in Salesforce that you plan to sync
- System Permission: API Enabled
- System Permission: Manage Connected Apps

This is likely a user with a System Administrator or similar profile. Additionally, the app guides the user to setup Amazon AppFlow which is a service that allows AWS to connect to Salesforce. The user is guided to create a Salesforce connection, sign in to Salesforce and authorize access. The image below shows this in-app guidance.

Step 2
Connect to your Salesforce organization

For AWS to access your Salesforce data, we will need your authorization. We will guide you through the steps of setting up Amazon AppFlow which is a service that allows AWS to connect to Salesforce, you will be creating a Salesforce connection, signing in to Salesforce and authorizing access.

Create a dedicated Salesforce user

To gain access to your Salesforce data, you need a Salesforce user for authentication purposes. It is best security practice to use a dedicated user for this purpose. This helps ensure business continuity and adheres to the principle of least privilege to help secure your data. For detailed instructions, see [Create a secure Salesforce API user](#).

If you do not have an extra full Salesforce user license available, you can skip this step and instead repurpose a user with a profile or permission set that provides:

- Access to the objects and fields in Salesforce that you plan to sync
- System Permission: API Enabled
- System Permission: Manage Connected Apps

This is likely a user with a System Administrator or similar profile.

Create a Salesforce connection using Amazon AppFlow

Amazon AppFlow is a service that allows AWS to connect to Salesforce, for this to work, you need to create a Salesforce connection using Amazon AppFlow.

Go to the AppFlow connections page

- Go to the [AppFlow connections page](#)
- Click on the Create connection button

Choose the Salesforce environment you want to connect to

- Choose the Salesforce environment you want to connect to
- Leave the Private Link selection as 'Disabled'
- Leave the Data Encryption setting as is
- Type a name for the connection, it can be anything you want, then click Continue

Provide the user name and password for the dedicated Salesforce user that you created earlier

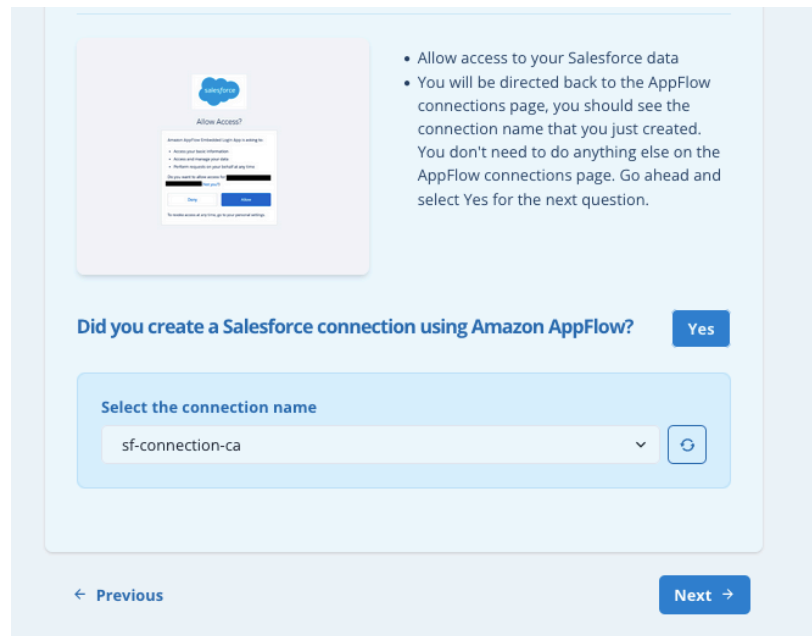
Remember to log in using the newly created dedicated Salesforce user

Allow access to your Salesforce data

- Allow access to your Salesforce data
- You will be directed back to the AppFlow connections page, you should see the connection name that you just created. You don't need to do anything else on the AppFlow connections page. Go ahead and select Yes for the next question.

Did you create a Salesforce connection using Amazon AppFlow?

Once the Salesforce connection is created through Amazon AppFlow, a drop-down box will populate the list of connections that can be selected for use. Connections are region-specific and the list will show only the connections that are created in the selected region. Once a connection is selected from the drop-down box, the NPO User can proceed to Step 3.



The screenshot shows a configuration page for a Salesforce connection in Amazon AppFlow. On the left, there is a preview of the 'Allow Access?' dialog box that will be shown to the user. On the right, there are two bullet points: 'Allow access to your Salesforce data' and 'You will be directed back to the AppFlow connections page, you should see the connection name that you just created. You don't need to do anything else on the AppFlow connections page. Go ahead and select Yes for the next question.' Below this, a question asks 'Did you create a Salesforce connection using Amazon AppFlow?' with a 'Yes' button. Underneath, there is a section 'Select the connection name' with a dropdown menu showing 'sf-connection-ca' and a refresh icon. At the bottom, there are 'Previous' and 'Next' navigation buttons.

- Allow access to your Salesforce data
- You will be directed back to the AppFlow connections page, you should see the connection name that you just created. You don't need to do anything else on the AppFlow connections page. Go ahead and select Yes for the next question.

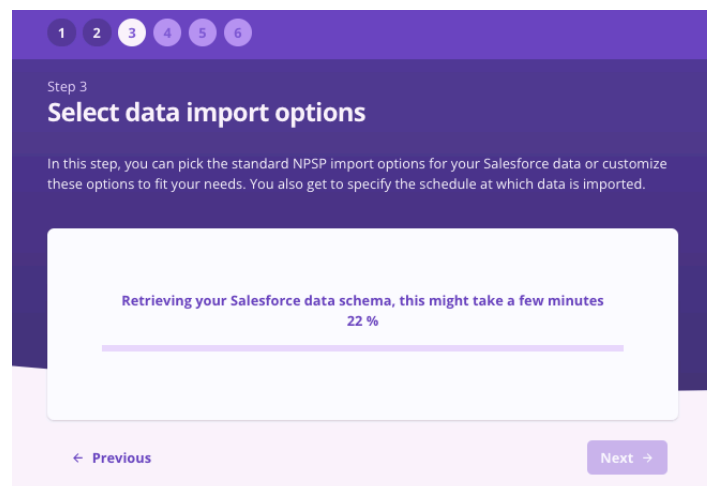
Did you create a Salesforce connection using Amazon AppFlow? **Yes**

Select the connection name

sf-connection-ca

← Previous **Next** →

3.3 Step 3



The screenshot shows 'Step 3: Select data import options' in a wizard. At the top, there is a progress bar with six steps, where step 3 is currently selected. Below the title, there is a description: 'In this step, you can pick the standard NPSP import options for your Salesforce data or customize these options to fit your needs. You also get to specify the schedule at which data is imported.' A large white box in the center contains a progress indicator: 'Retrieving your Salesforce data schema, this might take a few minutes' followed by a 22% progress bar. At the bottom, there are 'Previous' and 'Next' navigation buttons.

Step 3

Select data import options

In this step, you can pick the standard NPSP import options for your Salesforce data or customize these options to fit your needs. You also get to specify the schedule at which data is imported.

Retrieving your Salesforce data schema, this might take a few minutes

22 %

← Previous **Next** →

During step 3, Salesforce data schema is retrieved to show the NPSP data model. An NPO organization can have standard objects, core NPSP objects as well as custom objects. The retrieved schema is used to show the list of objects to the NPO user to be included or excluded as desired.

Step 3
Select data import options

In this step, you can pick the standard NPSP import options for your Salesforce data or customize these options to fit your needs. You also get to specify the schedule at which data is imported.

How often should we import the data?

☒ **Monthly** RECOMMENDED

☐ Weekly

☐ Daily

We will import the data every month, on the last day of the month at 23:00 UTC

Day

☒ Last day of the month

☐ Last weekday of the month

☐ The 1st day of the month

Time (24-hour format)

23 : 00 UTC

Salesforce NPSP data model

Objects to import 23

If needed, you can include or exclude **objects** by clicking on the object name. You can also exclude **fields** by clicking on the 'Customize Fields' button.

Show ☒ All Objects 2490 ☐ Included Objects 23 ☐ Excluded Objects 2467

Search Page 1 of 100 >

OBJECTS	FIELDS
<input checked="" type="checkbox"/> Account <small>NPSP</small> Account	520 Customize Fields
<input checked="" type="checkbox"/> Account Soft Credit <small>NPSP</small> npnp_Account_Soft_Credit__c	16 Customize Fields
<input checked="" type="checkbox"/> Address <small>NPSP</small> npnp_Address__c	52 Customize Fields
<input checked="" type="checkbox"/> Affiliation <small>NPSP</small> npa5_Affiliation__c	26 Customize Fields

<input checked="" type="checkbox"/> Campaign <small>NPSP</small> Campaign	97	Customize Fields
<input checked="" type="checkbox"/> Campaign Member <small>NPSP</small> CampaignMember	40	Customize Fields
<input checked="" type="checkbox"/> Contact <small>NPSP</small> Contact	315	Customize Fields
<input checked="" type="checkbox"/> Deliverable <small>NPSP</small> npnp_Grant_Deliverable__c	18	Customize Fields
<input checked="" type="checkbox"/> Donation <small>NPSP</small> Opportunity	215	Customize Fields
<input checked="" type="checkbox"/> Donation Contact Role <small>NPSP</small> OpportunityContactRole	11	Customize Fields
<input checked="" type="checkbox"/> Engagement Plan <small>NPSP</small> npnp_Engagement_Plan__c	32	Customize Fields
<input checked="" type="checkbox"/> Engagement Plan Task <small>NPSP</small> npnp_Engagement_Plan_Task__c	26	Customize Fields
<input checked="" type="checkbox"/> Event <small>NPSP</small> Event	69	Customize Fields
<input checked="" type="checkbox"/> GAU Allocation <small>NPSP</small> npnp_Allocation__c	27	Customize Fields
<input checked="" type="checkbox"/> General Accounting Unit <small>NPSP</small> npnp_General_Accounting_Unit__c	45	Customize Fields
<input checked="" type="checkbox"/> Lead <small>NPSP</small> Lead	170	Customize Fields
<input checked="" type="checkbox"/> Level <small>NPSP</small> npnp_Level__c	25	Customize Fields
<input checked="" type="checkbox"/> Opportunity Stage <small>NPSP</small> OpportunityStage	16	Customize Fields
<input checked="" type="checkbox"/> Partial Soft Credit <small>NPSP</small> npnp_Partial_Soft_Credit__c	18	Customize Fields
<input checked="" type="checkbox"/> Payment <small>NPSP</small> npa01_OppPayment__c	59	Customize Fields
<input checked="" type="checkbox"/> Recurring Donation <small>NPSP</small> npa03_Recurring_Donation__c	53	Customize Fields
<input checked="" type="checkbox"/> Relationship <small>NPSP</small> npa4_Relationship__c	23	Customize Fields
<input checked="" type="checkbox"/> Task <small>NPSP</small> Task	62	Customize Fields
<input type="checkbox"/> *Case History Case_History__c		
<input type="checkbox"/> *Chat History Chat_History__c		

Page 1 of 100 >

[← Previous](#) [Next →](#)

In Step 3, the NPO User can click on default options to setup the datalake and periodic data import. A typical NPO User is expected to click on all default options to proceed to Step 4. However, an advanced NPO User can customize the objects and fields to be included or excluded in the data import. The image on the left below shows a custom object being added for import and the image on the right, shows fields being excluded from the Account object.

Salesforce NPSP data model

Objects to import 24

If needed, you can include or exclude **objects** by clicking on the object name.

You can also exclude **fields** by clicking on the 'Customize Fields' button.

Show ☒ All Objects 2490 ☐ Included Objects 24 ☐ Excluded Objects 2466

campaign

Found 21 matches

OBJECTS	FIELDS
<input checked="" type="checkbox"/> Campaign <small>NPSP</small>	97 Customize Fields
<input checked="" type="checkbox"/> Campaign Member <small>NPSP</small>	40 Customize Fields
<input type="checkbox"/> Campaign Feed	
<input type="checkbox"/> Campaign Field History	
<input type="checkbox"/> Campaign Influence Model	
<input type="checkbox"/> Campaign Member Configuration <small>enriched__Campaign_Member_Configuration__c</small>	
<input type="checkbox"/> Campaign Member Sort Field <small>mdm__TS__CampaignMemberSortField__c</small>	
<input type="checkbox"/> Campaign Member Status	
<input checked="" type="checkbox"/> Campaign Share	8 Collapse Fields Table

Show ☒ All Fields 8 ☐ Included Fields 8 ☐ Excluded Fields 0

☒ Campaign Access CampaignAccessLevel picklist

☒ Campaign ID CampaignId reference

☒ Deleted IsDeleted boolean

☒ Row Cause RowCause picklist

☒ User/Group ID UserOrGroupId reference

The following 3 fields cannot be excluded. They will always be imported.

☒ Campaign Share ID id

☒ Last Modified By ID LastModifiedById reference

☒ Last Modified Date LastModifiedDate datetime

Step 3

Select data import options

In this step, you can pick the standard NPSP import options for your Salesforce data or customize these options to fit your needs. You also get to specify the schedule at which data is imported.

How often should we import the data?

☒ Monthly RECOMMENDED

☐ Weekly

☐ Daily

We will import the data every month, on the last day of the month at 23:00 UTC

Day

☒ Last day of the month

☐ Last weekday of the month

☐ The 1st day of the month

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31

Time (24-hour format)

23 : 00 UTC

hour minute

Salesforce NPSP data model

Objects to import 24

If needed, you can include or exclude **objects** by clicking on the object name.

You can also exclude **fields** by clicking on the 'Customize Fields' button.

Show ☒ All Objects 2490 ☐ Included Objects 24 ☐ Excluded Objects 2466

Search

Page 1 of 100

OBJECTS	FIELDS
<input checked="" type="checkbox"/> Account <small>NPSP</small>	516 Collapse Fields Table

Show ☒ All Fields 520 ☐ Included Fields 516 ☐ Excluded Fields 4

Search

Page 1 of 26

☒ # of Children in Household #ChildrenInHousehold__gc number

☒ % Goal Attainment %GoalAttainment__gc percent

☒ % of Reps Trained %RepsTrained__gc percent

☒ Account Address AccountAddress__gc string

☐ Account Description Description textarea

☒ Account Fax Fax phone

☒ Account Group acc__g__AccountGroup__gc reference

☒ Account Hierarchy H1B1s acc__hierarchyH1B1s__gc string

☐ Account Location AccountLocation__gc string

☐ Account Name Name string

☒ Account Number AccountNumber string

☒ Account Phone Phone phone

☒ Account Priority AccountPriority string

☐ Account Rating Rating percent

☒ Account Site Site string

☒ Account Source AccountSource string

☒ Account Source Custom AccountSourceCustom__gc string

☒ Account Status AccountStatus__gc string

A default monthly sync is recommended. However, the user can select a weekly or even a daily sync and choose the date and the time of the sync. Note that we don't recommend a daily sync option for a large NPO Organization as this could incur additional costs.

3.4 Step 4

In Step 4, the user can review the data lake setup and data import options and confirm the changes they've done (if any). This step is where the NPO User is allowed to go back to previous step to make any other needed changes before confirming the setup of resources for the data lake.

1 2 3 4 5 6

Step 4

Installation ID
uby9q0ccbum

Review and confirm

Did we get everything right? Take a look at all the answers you provided and confirm your selections.

AWS Region

US West (Northern California) Region

Connection Name

sf-connection-ca

Import Schedule

We will import the data every month, on the last day of the month at 23:00 UTC

Data Model

We will import 24 objects

TYPE	# FIELDS
> Account (Account)	516/520
> Account Soft Credit (npsp__Account_Soft_Credit__c)	16/16
> Address (npsp__Address__c)	52/52
> Affiliation (npe5__Affiliation__c)	26/26
> Campaign (Campaign)	97/97
> Campaign Member (CampaignMember)	40/40
> Contact (Contact)	315/315
> Deliverable (npsp__Grant_Deadline__c)	18/18
> Donation (Opportunity)	215/215
> Donation Contact Role (OpportunityContactRole)	11/11
> Engagement Plan (npsp__Engagement_Plan__c)	32/32
> Engagement Plan Task (npsp__Engagement_Plan_Task__c)	26/26
> Event (Event)	69/69
> GAU Allocation (npsp__Allocation__c)	27/27
> General Accounting Unit (npsp__General_Accounting_Unit__c)	45/45
> Lead (Lead)	170/170

> Level (npsp__Level__c)

25/25

> Opportunity Stage (OpportunityStage)

13/16

Included Fields 13

Api Name ApiName Closed IsClosed Created By ID CreatedById

Created Date CreatedDate Description Description

Forecast Category Name ForecastCategoryName Is Active IsActive

Last Modified By ID LastModifiedById Last Modified Date LastModifiedDate

Opportunity Stage ID Id Sort Order SortOrder System Modstamp SystemModstamp

Won IsWon

Excluded Fields 3

Forecast Category ForecastCategory

Master Label MasterLabel

Probability (%) DefaultProbability

> Partial Soft Credit (npsp__Partial_Soft_Credit__c)

18/18

> Payment (npe01__OppPayment__c)

59/59

> Recurring Donation (npe03__Recurring_Donation__c)

53/53

> Relationship (npe4__Relationship__c)

23/23

> Task (Task)

62/62

> Campaign Share (CampaignShare)

8/8

Included Fields 8

Campaign Access CampaignAccessLevel Campaign ID CampaignId Campaign Share ID Id

Deleted IsDeleted Last Modified By ID LastModifiedById

Last Modified Date LastModifiedDate Row Cause RowCause User/Group ID UserOrGroupId

1,936

What's next?

We will start the process of setting up the datalake at AWS and importing your Salesforce records. Depending on the number of records to import, the next step might take 15 minutes to a few hours to complete.

< Previous

Next >

Once review is complete, the NPO User can click “Next” to start provisioning the resources in Step 5.

3.5 Step 5

Step 5 is where the data lake is provisioned and the data is imported at the end. This can take several minutes to several hours depending on the size of the data in the NPO Organization. We caution the user to leave the EZ Datalake app running in the browser while the data lake is being provisioned. However, there is a resume feature described in Section 4.1 where a user can resume the data lake provisioning if the browser is accidentally closed.

Step 5

Installation ID
ubyl9q0ccbum

Sit back and relax

We are provisioning the data lake and importing the data. You can sit back and relax and watch the progress of the data import. Don't close the browser tab, otherwise the provisioning will pause before it's fully completed.

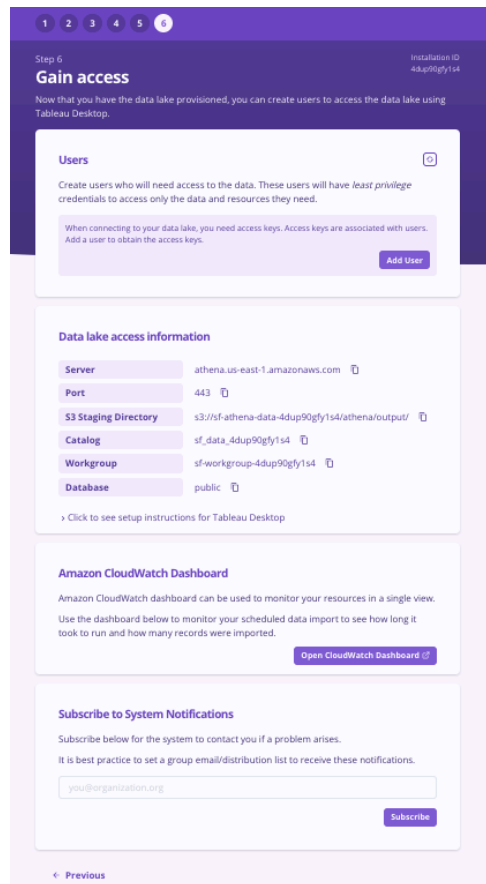
Don't close the browser tab, otherwise the provisioning will pause before it's fully completed.

Provisioning the S3 buckets stack. This might take 4 minutes or more.	10%	0:00:17
Provision the VPC stack. This might take 3 minutes or more.	0%	0:00:00
Provision the Datastore stack. This might take 14 minutes or more.	0%	0:00:00
Provision the Athena stack. This might take 6 minutes or more.	0%	0:00:00
Provision the Step Functions stack. This might take 4 minutes or more.	0%	0:00:00
Upload data model schema. This might take a few seconds.	0%	0:00:00
Import Salesforce objects. This might take 15 minutes to a few hours.	0%	0:00:00
Update the Step Functions stack. This might take 2 minutes or more.	0%	0:00:00
Create Dashboard. This might take a few seconds.	0%	0:00:00
Overall Progress	1%	0:00:17

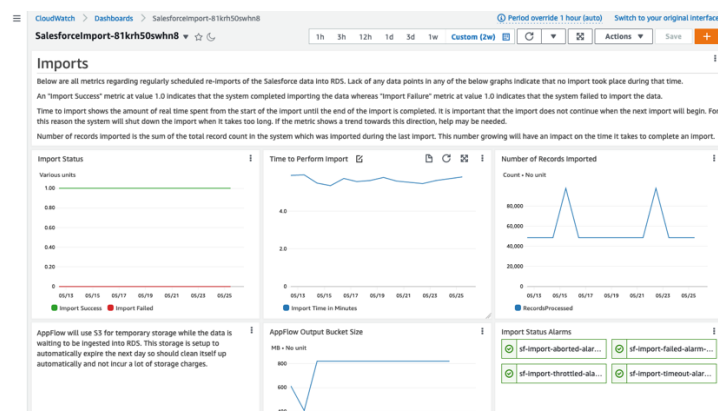
Next →

3.6 Step 6

Once the data lake is setup, in Step 6, the NPO Users can create and manage IAM users to grant them access to Tableau Desktop. These users will have *least privilege* credentials to access only the data and resources they need. In Step 6, the user can also view the Data Lake access information and instructions for Tableau Desktop installation and access for Windows and Mac, access CloudWatch Dashboard and subscribe to systems notifications.



Amazon CloudWatch dashboard can be used to monitor your resources in a single view. The screenshot below shows metrics for the user to monitor the scheduled data import to see how long it took to run and how many records were imported.



Additionally, users can subscribe to system notifications to be alerted if a problem arises.

4 Resume Feature

The resume feature allows an NPO System Administrator to continue deployment if the app was closed during any of the five initial steps. Additionally, the resume feature can also be used to retrieve the Data Lake provisioning information after deployment and also a start a new datalake deployment if that is desired.

4.1 Resume while deployment is in progress

During any of the six steps with the EZ Datalake App, it is recommended that the NPO User keep the EZ Datalake app in the browser running till the datalake is deployed. However, if the user does accidentally close the browser, the “resume” feature can help pick up where the deployment was left off. Note that depending on which step the user is in when the browser is closed, the app will resume in the same step.

4.2 Resume after deployment

Once the datalake is deployed, the “resume” feature can be used for the NPO User to retrieve details on the datalake that was setup earlier. Additionally, the NPO System Administrator can use the app to create and manage IAM users to grant them access to Tableau Desktop. These users will have *least privilege* credentials to access only the data and resources they need. The screenshots below show how the NPO System Administrator can enter their IAM Admin Key ID and Secret Access Key to retrieve the details of the datalake from Step 6 and manage users for Tableau Desktop access, view Data Lake access information and retrieve instructions for Tableau Desktop, access CloudWatch Dashboard and subscribe to systems notifications in Step 6.

EZ DATA LAKE
THE FASTEST WAY TO BUILD A DATALAKE FOR YOUR SALESFORCE DATA

DATA LAKE
2 weeks ago • by rk-demo

Installation Id: 4dup90gfy1s4

AWS Account Id	921792788236
AWS Region	US East (Northern Virginia) Region
Connection Name	sf-prod-east1
Import Schedule	We will import the data every month, on the last day of the month at 23:00 UTC

Click **Resume** to continue the steps of provisioning the data lake, we will ask you for your AWS admin credentials so that we can connect to your AWS account your . You can also choose to start a new data lake.

[Start New](#) [Resume →](#)

EZ DATA LAKE
THE FASTEST WAY TO BUILD A DATALAKE FOR YOUR SALESFORCE DATA

DATA LAKE
2 weeks ago • by rk-demo

Installation Id: 4dup90gfy1s4

AWS Account Id	921792788236
AWS Region	US East (Northern Virginia) Region
Connection Name	sf-prod-east1
Import Schedule	We will import the data every month, on the last day of the month at 23:00 UTC

Access Key ID & Secret Access Key

AKIA5NHZBW4GMD7IOEKN

..... [Show](#)

Secret access keys are sensitive information, keep them secured and protected, never post them on public platforms or leave them unsecured as this can compromise your account security

[Cancel](#) [Resume →](#)

Step 5
Sit back and relax

We are provisioning the data lake and importing the data. You can sit back and relax and watch the progress of the data import. Don't close the browser tab, otherwise the provisioning will pause before it's fully completed.

Data Lake is ready!
Great news! we got everything in place and you are ready to proceed to the next step

Task	Progress	Time
Successfully provisioned the S3 buckets stack	100%	0:02:37
Successfully provisioned the VPC stack	100%	0:02:58
Successfully provisioned the Datastore stack	100%	0:09:23
Successfully provisioned the Athena stack	100%	0:04:21
Successfully provisioned the Step Functions stack	100%	0:06:02
Successfully uploaded the data model schema	100%	0:00:00
Successfully imported 22 Salesforce objects	100%	0:04:39
Successfully updated the Step Functions stack	100%	0:01:12
Successfully created the dashboard	100%	0:00:00
Overall Progress	100%	0:31:12

Latest Data Import 22 2 weeks ago

Objects	Records	Fields	Size	Time
22	48,778	1,848	201MB	277 sec

[see more](#)

[Next](#)

Step 6
Gain access

Now that you have the data lake provisioned, you can create users to access the data lake using Tableau Desktop.

Users
Create users who will need access to the data. These users will have *least privilege* credentials to access only the data and resources they need.
When connecting to your data lake, you need access keys. Access keys are associated with users. Add a user to obtain the access keys.
[Add User](#)

Data lake access information

Server	athena.us-east-1.amazonaws.com
Port	443
S3 Staging Directory	s3://sf-athena-data-4dup90gfy1s4/athena/output/
Catalog	sf_data_4dup90gfy1s4
Workgroup	sf-workgroup-4dup90gfy1s4
Database	public

[Click to see setup instructions for Tableau Desktop](#)

Amazon CloudWatch Dashboard
Amazon CloudWatch dashboard can be used to monitor your resources in a single view. Use the dashboard below to monitor your scheduled data import to see how long it took to run and how many records were imported.
[Open CloudWatch Dashboard](#)

Subscribe to System Notifications
Subscribe below for the system to contact you if a problem arises. It is best practice to set a group email/distribution list to receive these notifications.

[Subscribe](#)

[Previous](#)

4.3 Resume another datalake

Typically, an NPO User is expected to setup a single datalake. If for any reason, a new datalake needs to be setup, the same app can be utilized to deploy another datalake. This can be useful perhaps if a datalake for a sandbox needs to be setup first before moving to production organization. The "resume" feature can then be used to view and manage the multiple datalake as show below in the screenshots.

EZ DATA LAKE
THE FASTEST WAY TO BUILD A DATA LAKE FOR YOUR SALESFORCE DATA

WHAT TO EXPECT
In just a few steps, you can bring your Salesforce data to AWS and keep it in sync. Take a look at the summary of the steps needed, no worries though, we will guide you through each step.

Let's Go →

- 1 Connect to your AWS account**
We need your AWS admin credentials so that we can connect to your AWS account and provision the resources for the data lake. We will guide you through the process of obtaining the needed credentials
- 2 Connect to your Salesforce organization**
For AWS to access your Salesforce data, we will need your authorization. We will guide you through the steps of setting up Amazon AppFlow which is a service that allows AWS to connect to Salesforce, you will be creating a Salesforce connection, signing in to Salesforce and authorizing access
- 3 Select data import options**
In this step, you can pick the standard NPSP import options for your Salesforce data or customize these options to fit your needs
- 4 Review and confirm**
You get a chance to review and confirm the settings you have provided
- 5 Sit back and relax**
In this step, we will do all the necessary provisioning and data importing. You can sit back and relax and watch the progress of the data import
- 6 Gain access**
This is the final step, you can create users with credentials to access the data lake using Tableau desktop

Resume
If you have already created a datalake using the EZ Data Lake app and would like to view or resume its setup, we got you covered. Click the button to reconnect to an existing datalake, we will ask you for your AWS admin credentials so that we can detect the existing datalake setup.

Resume *

Available data lakes 2

<input type="radio"/> Data Lake - 4dup90gy1s4	US East (Northern Virginia) Region
2 weeks ago • by r1-demo Connection name: sf-prod-east1	
<input type="radio"/> Data Lake - uby19qccbum	US West (Northern California) Region
2 hours ago • by r1-demo Connection name: sf-connection-ca	

Cancel Resume

5 Data Consistency / Data Deletion

Data can be deleted in Salesforce in multiple ways triggering soft deletes and hard deletes. The EZ Datalake app ensures data consistency with the data in the Salesforce by performing full data import on each periodic sync operation. Refer to [EZ_Datalake_Testing_Deletion_Scenarios.doc](#) where different deletion scenarios are detailed along with testing outcomes.