Informatica Intelligent Cloud Services Accelerator for Snowflake User Guide

Contents

Introduction

The IICS Accelerator for Snowflake guide will help you to navigate the Informatica Intelligent Cloud Services℠ (IICS) interface and to load data into Snowflake quickly.

This guide contains two sections. The first section describes the initial steps for setting up your organization to use the Snowflake Accelerator. It explains the following:

- How to create an IICS organization
- How to install an add-on bundle (optional)
- How to install a Secure Agent to run your tasks
- How to create connections to Snowflake and other data endpoints

The second section of this guide includes examples of commonly used features for loading data into Snowflake. It explains how to perform the following tasks:

- Load Salesforce data to Snowflake
- Load a CSV file to Snowflake
- Load CSV files using mass ingestion
- Use the pushdown optimization option in a mapping

Getting started

To begin, you need to perform the following steps:

1. Create an IICS organization.

   Create an IICS organization from the Snowflake Partner Connect page.

2. Optionally, install the add-on bundle.

   An add-on bundle is a set of mappings that is pre-built and published by an IICS user. The add-on bundles and templates are provided to help you load Salesforce Opportunity and Account data into Snowflake and to load a CSV file into Snowflake without writing a single line of code.

   After you install the bundle, you can start using the mappings. The bundle contains mappings called “Salesforce Oppty_Acct to Snowflake - Parameterized” and “CSV to Snowflake.”

3. Install a Secure Agent to use as the runtime environment for your mappings and tasks.

   The Secure Agent enables IICS to access application, relational database, and file sources and targets in your on-premises network.

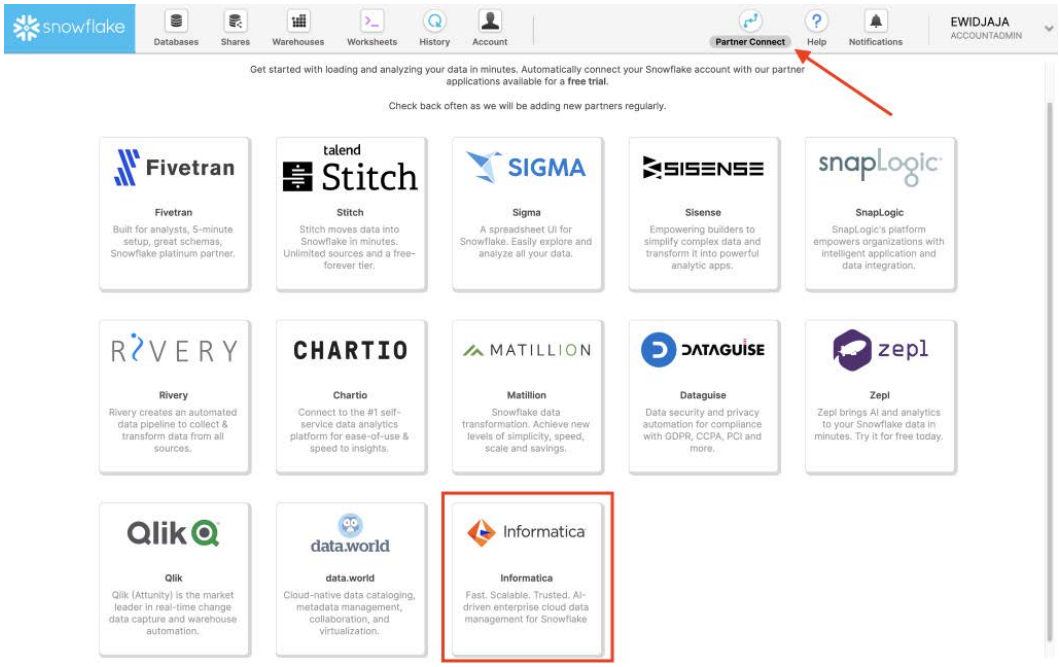
4. Configure connections to Snowflake, Salesforce, and CSV files.

   These connections allow you to connect to your data sources and targets.

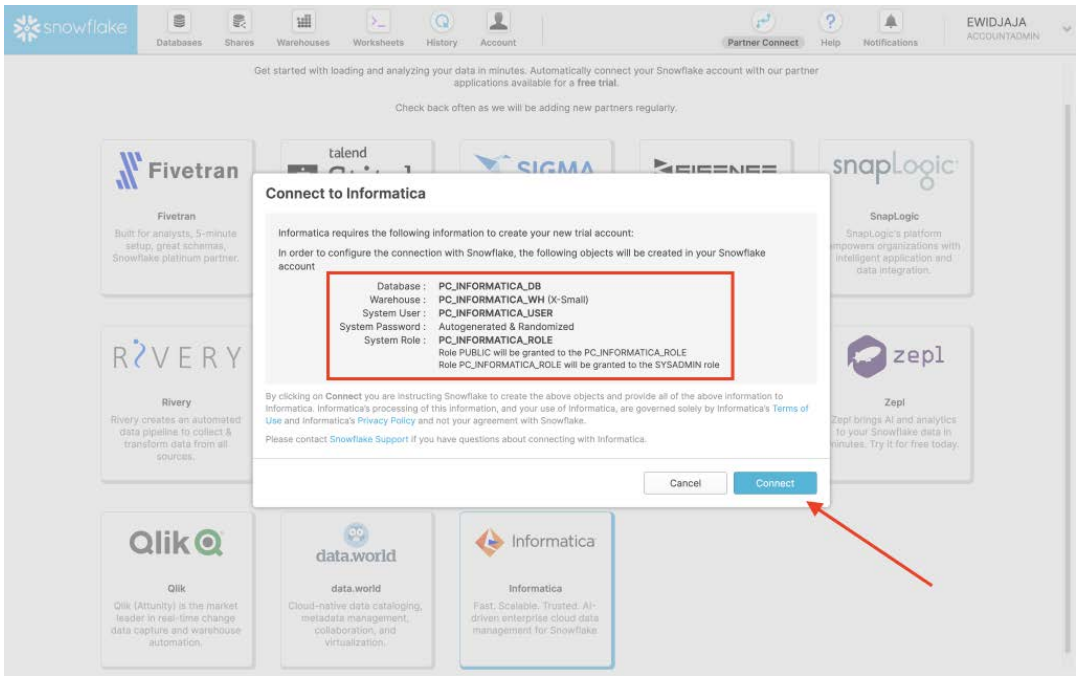
Step 1. Create an IICS organization

Complete the following steps to create your IICS organization and create a Snowflake connection.

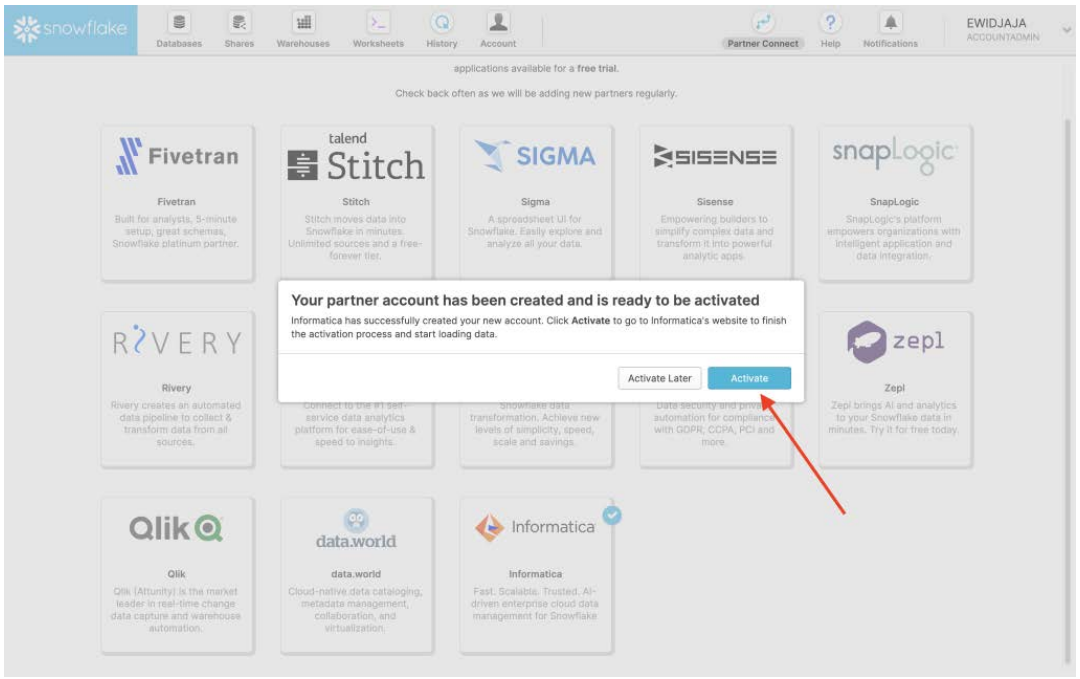
1. Log in to Snowflake and click **Partner Connect**.
2. Click the **Informatica** tile:



Objects that will be created in Snowflake are displayed. These objects will be used to create a Snowflake connection in IICS:

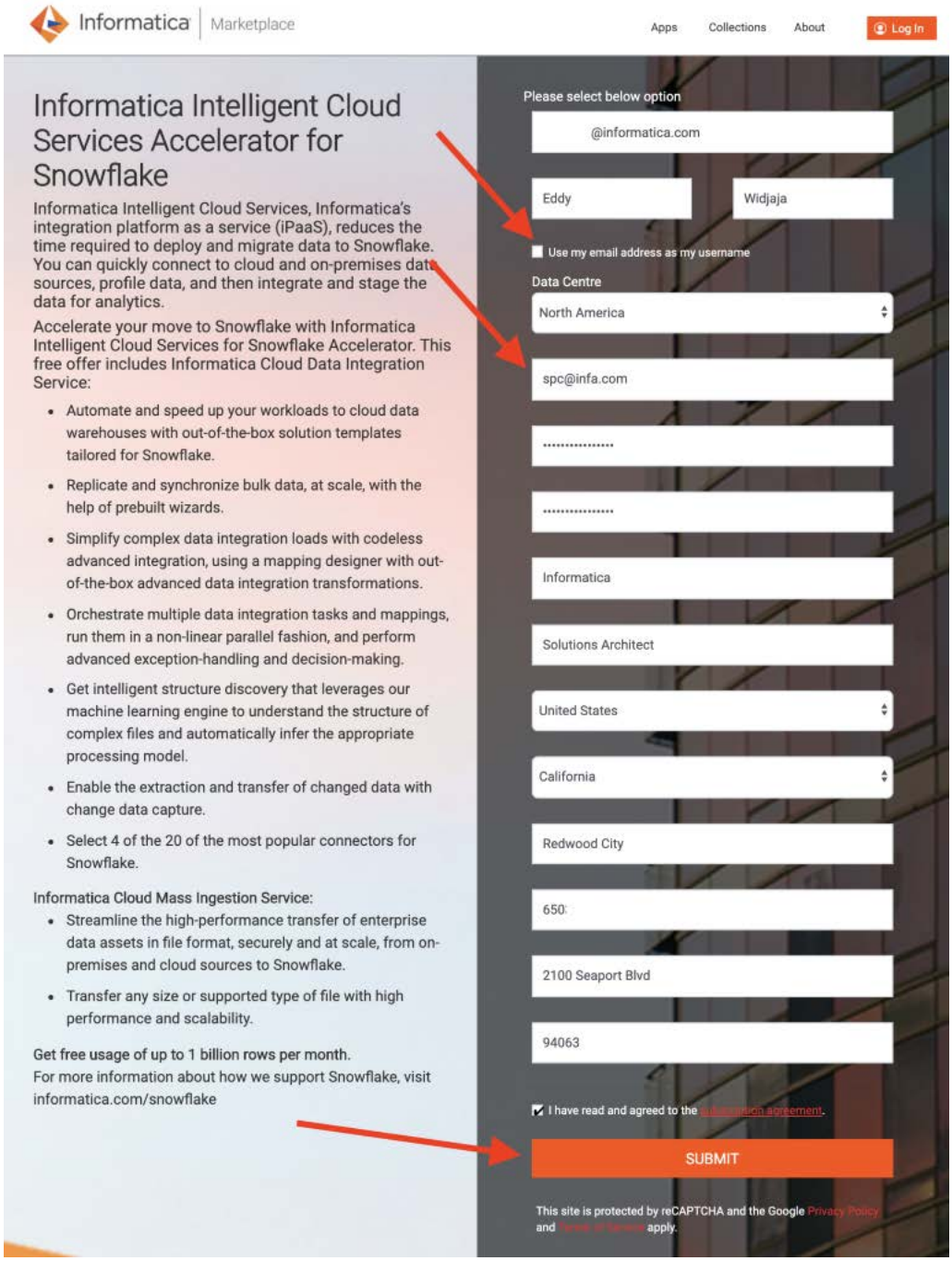


3. Click **Connect**.
4. When you see the message saying that your partner account has been created, click **Activate** to register your IICS organization:

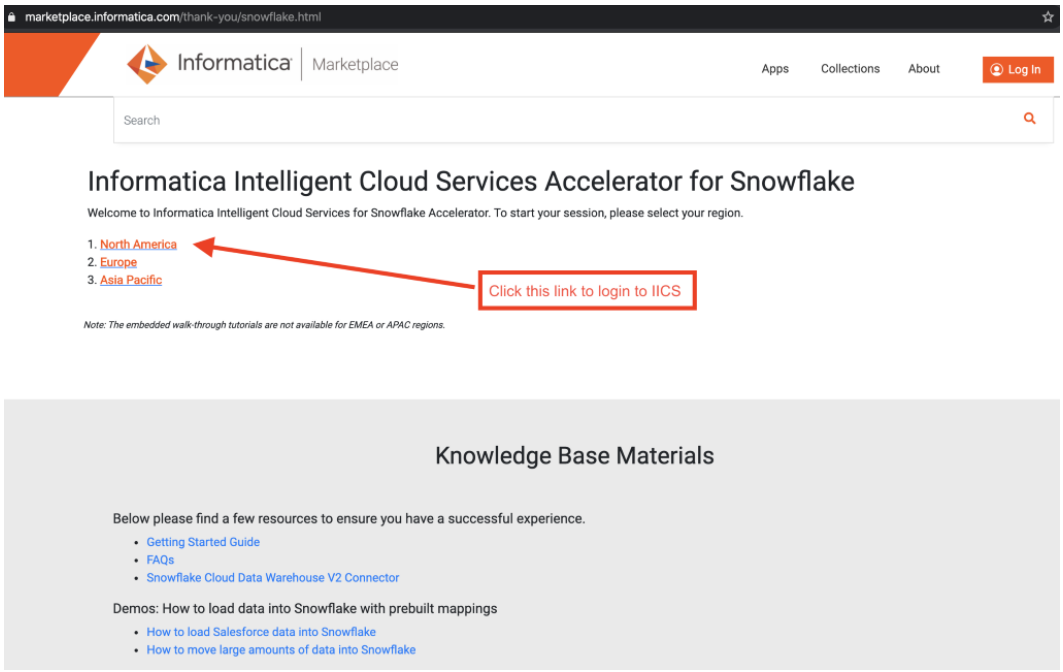


5. Fill in the form and click **Submit**.

   If you have an existing IICS organization with the same email address as your user name, uncheck “Use my email address as my username.”

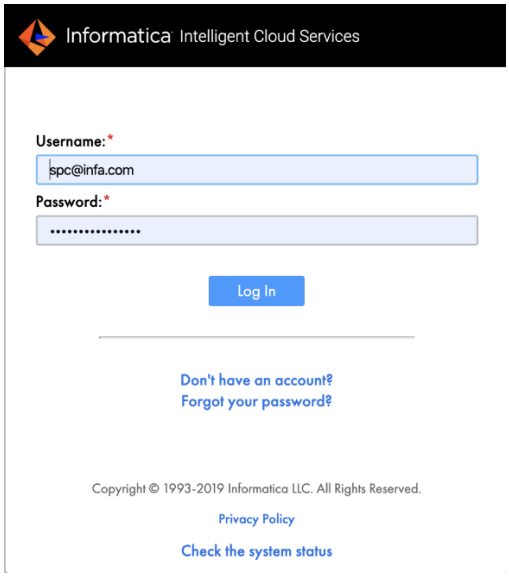


6. Click the appropriate region to log in to IICS. Your region is the Data Center selected in the previous page.

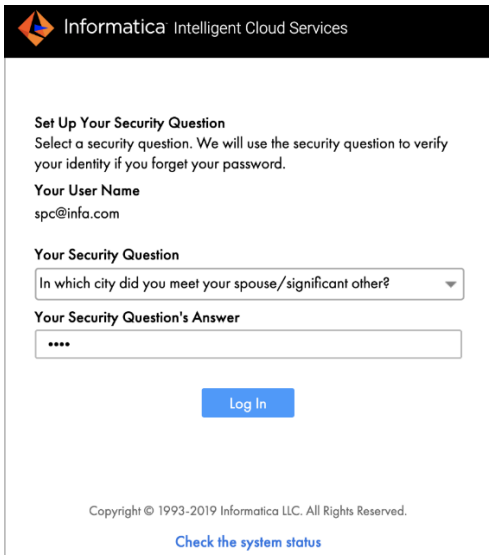


The [Informatica Intelligent Cloud Services Accelerator for Snowflake Marketplace page] also contains links to other materials on the Informatica Knowledge Base. Be sure to check them out!

7. Log in using the IICS user name and password that you provided in the previous form:

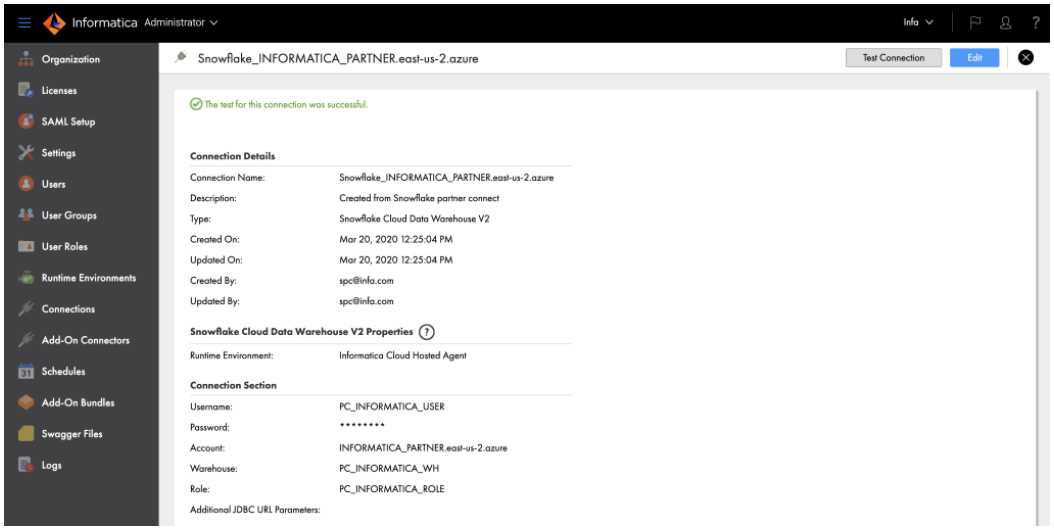


8. Set up the security question and click **Log In**:



A Snowflake connection is automatically created for you.

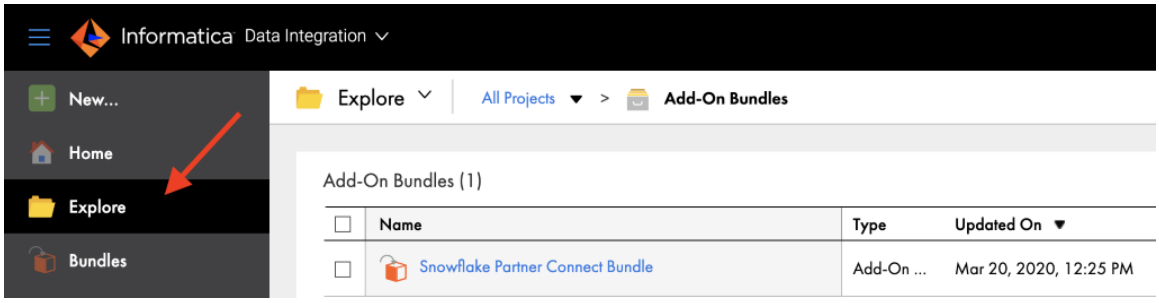
9. Test the connection by opening the Administrator service, clicking **Connections**, and selecting the connection. On the connection details page, click **Test Connection**:



You should see the message, “The test for this connection was successful.”

Step 2. Install the bundle

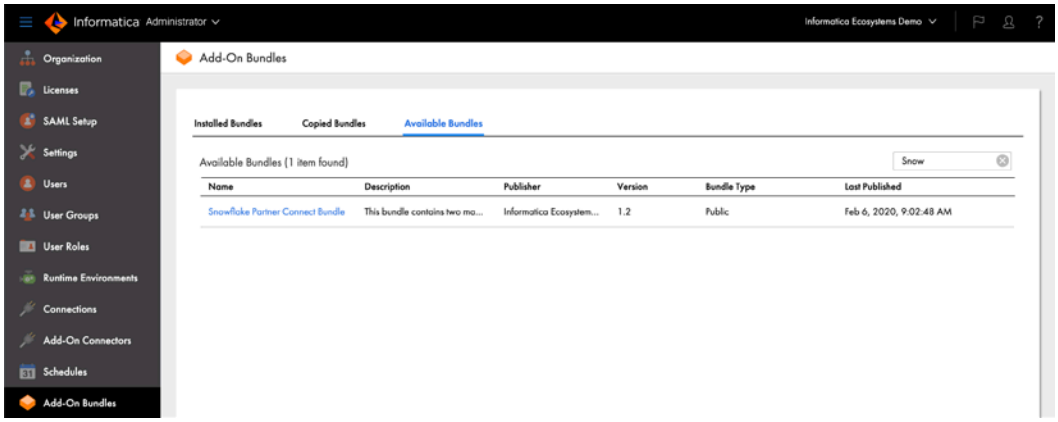
Note: This is an optional step. The add-on bundle is automatically provisioned for you. To view it, open the Data Integration Service in IICS, click **Explore**, and open the **Add-On Bundles** project:



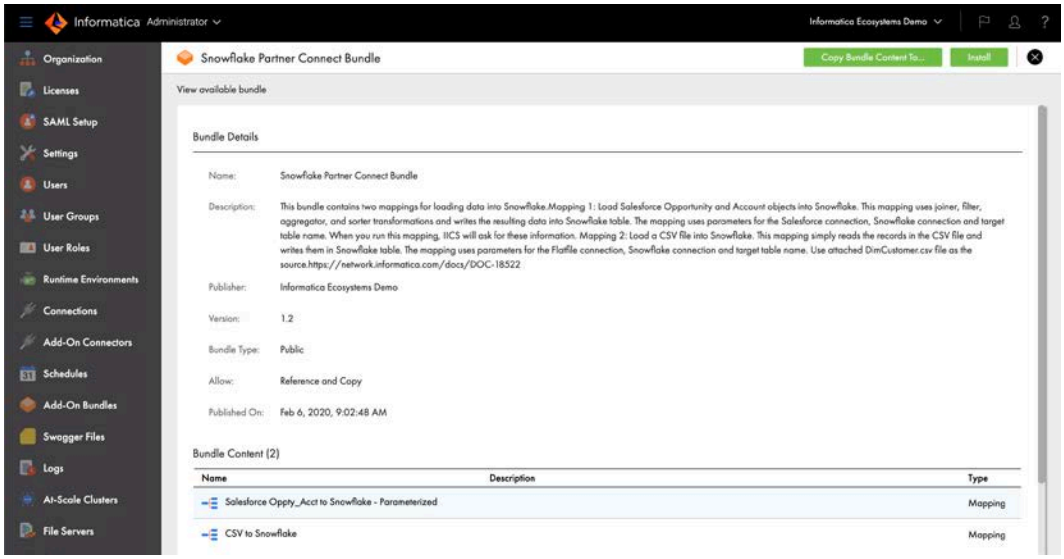
If you don't see the bundle, or you need to reinstall it, the Snowflake Partner Connect Bundle is listed in **Add-On Bundles** page in Administrator.

To install the bundle:

1. In IICS, open the Administrator service.
2. Click **Add-On Bundles** from the menu and select the **Available Bundles** tab.



3. Search for "Snowflake" and click the result.
4. Click **Copy Bundle Content To**:



5. Select the project or folder where you want to copy the mappings.

Step 3. Install the Secure Agent (runtime environment)

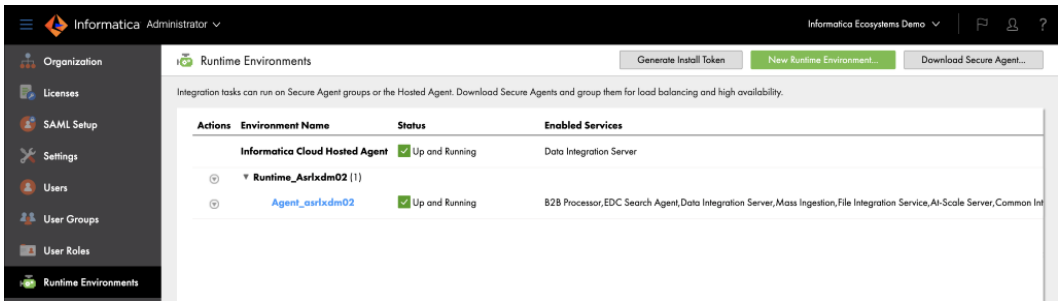
The Secure Agent is a small footprint application that enables secure communication across the firewall between your organization and IICS. It enables IICS to get through the firewall to access application, relational database, and file sources and targets in your on-premises network.

In this step, you install the Secure Agent on a server in your on-premises network or in the cloud service provider of your choice.

1. In the Administrator service, click **Runtime Environments**.
2. Click **Download Secure Agent**.
3. Select the operating system platform where you will install the Secure Agent.
4. Click **Copy** to copy the install token string.

   The install token is needed to associate the Secure Agent with your IICS organization.
5. Click **Download** to start downloading the Secure Agent.
6. Install the secure agent and supply the install token when requested.

   You should see the secure agent listed on the **Runtime Environments** page in Administrator:



Note: Your runtime environment should be listed with the status “Up and Running” before you continue to the next step. If you see the status “Not all the services are running,” wait 5-10 minutes for the agent to start up completely and install updates.

For more information about deploying a Secure Agent, see these instructions in the IICS *Administrator* guide or check out this video.

Step 4. Configure connections

Informatica supports more than 300 connectors. A connector is a pre-built integration that allows you to connect to data sources and targets.

In this step, you configure Snowflake, Salesforce, and flat file connections.

Configure a Snowflake connection

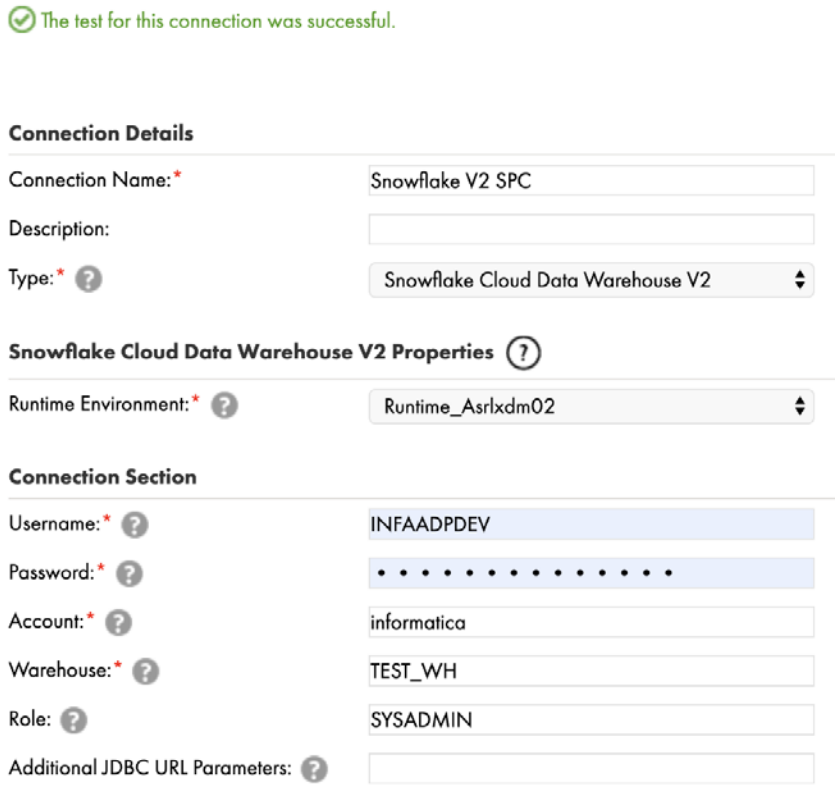
Note: This is an optional step. A Snowflake connection was already created for you when you created your organization through Snowflake Partner Connect.

If you want to configure another Snowflake connection, you can configure a Snowflake connection on the **Connections** page in the Administrator service.

1. In Administrator, click **Connections**.
2. Click **New Connection** in the top right corner of the screen.
3. Enter a name and optional description for the connection.
4. Select **Snowflake Cloud Data Warehouse V2** as the connection type.
5. Configure the following properties:

- Runtime Environment. Select the Secure Agent that you created.
- Username
- Password
- Account
- Warehouse
- Roles

6. To test the connection, click **Test Connection**.
   You should see the message, “The test for this connection was successful”:



7. Click **Save** to save the connection.

For more information about configuring a Snowflake Cloud Data Warehouse V2 connection, see this topic in the IICS *Connections* guide.

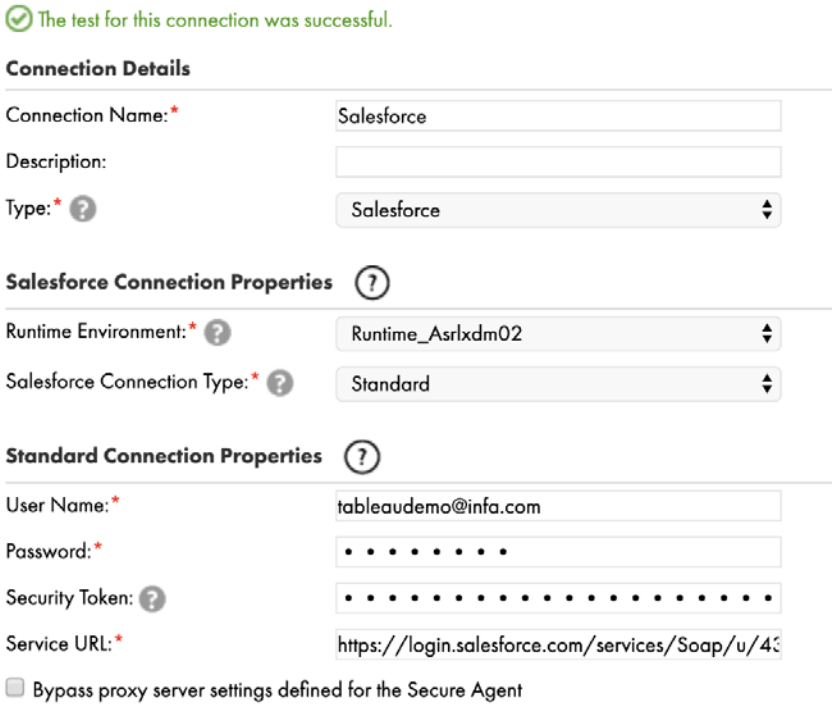
Configure a Salesforce connection

Configure a Salesforce connection on the **Connections** page in the Administrator service.

1. In Administrator, click **Connections**.
2. Click **New Connection** in the top right corner of the screen.
3. Enter a name and optional description for the connection.
4. Select **Salesforce** as the connection type.
5. Configure the following properties:
   - Runtime Environment. Select the Secure Agent that you created.
   - User Name
   - Password

- Security Token
- Service URL

6. To test the connection, click **Test Connection**.
7. You should see the message, “The test for this connection was successful”:



8. Click **Save** to save the connection.

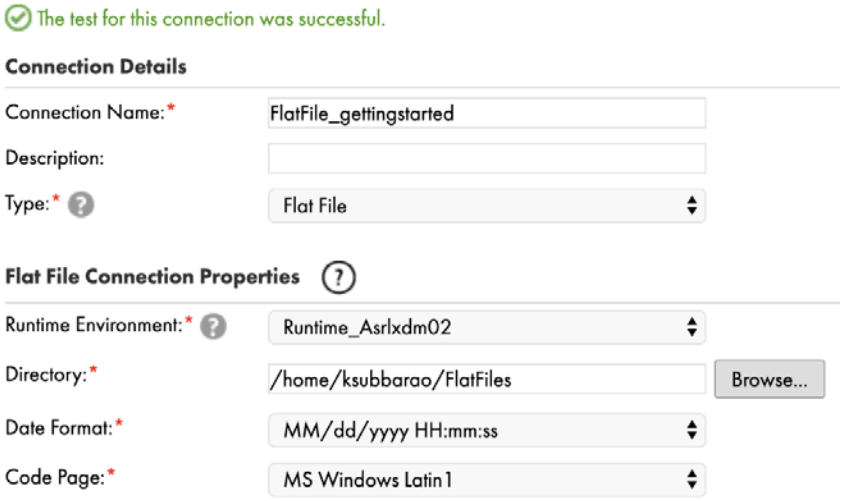
For more information about configuring a Salesforce connection, see this topic in the IICS *Connections* guide.

Configure a flat file connection

Configure a flat file connection on the **Connections** page in the Administrator service.

1. In Administrator, click **Connections**.
2. Click **New Connection** in the top right corner of the screen.
3. Enter a name and optional description for the connection.
4. Select **Flat File** as the connection type.
5. Configure the following properties:
   - Runtime Environment. Select the Secure Agent that you created.
   - Directory
   - Date Format
   - Code Page
6. To test the connection, click **Test Connection**.

You should see the message, “The test for this connection was successful:”



7. Click **Save** to save the connection.

For more information about configuring a flat file connection, see this topic in the IICS *Connections* guide.

Using the IICS Accelerator for Snowflake

This section uses mappings and tasks included in the bundle that you installed to perform the following tasks:

- Load Salesforce data to Snowflake
- Load a CSV file to Snowflake using a mapping
- Load CSV files to Snowflake using a mass ingestion task
- Perform ELT in Snowflake using pushdown optimization

Loading Salesforce data to Snowflake

The Data Integration Mapping Designer allows you to define data integration flow in a visual designer interface. It starts from defining the data sources to read data from, applying transformations based on your requirements on that data set, and then writing the resulting data set to the target system.

Note: If you still have the Administrator service open, click **Administrator** next to the Informatica logo in the top left corner of the screen. Then select **Data Integration** from the menu.

The first mapping in the bundle is called “Salesforce Oppty_Acct to Snowflake - Parameterized.”

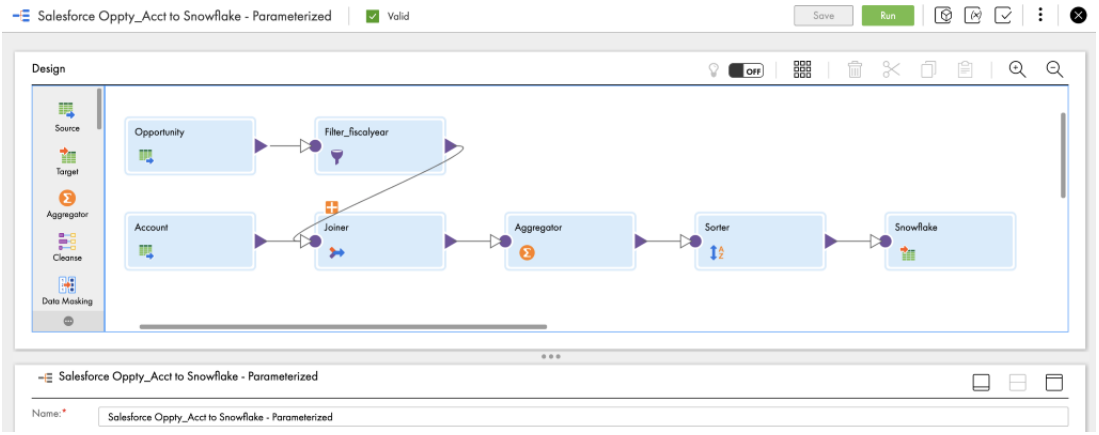
This mapping can:

1. Read the Opportunity object from Salesforce.
2. Filter records prior to 2018.
3. Read the Account object from Salesforce.
4. Join the Opportunity object with the Account object.
5. Aggregate the Amount column in the Opportunity object and group by AccountId, ForecastCategory, and FiscalYear.
6. Sort the records by AccountId, FiscalYear, FiscalQuarter, and ForecastCategory.
7. Write the resulting records to the Snowflake table.

The source and target connections in this mapping are parameterized using input parameters. This means that you have the option to choose the sources and the target Snowflake table at run time.

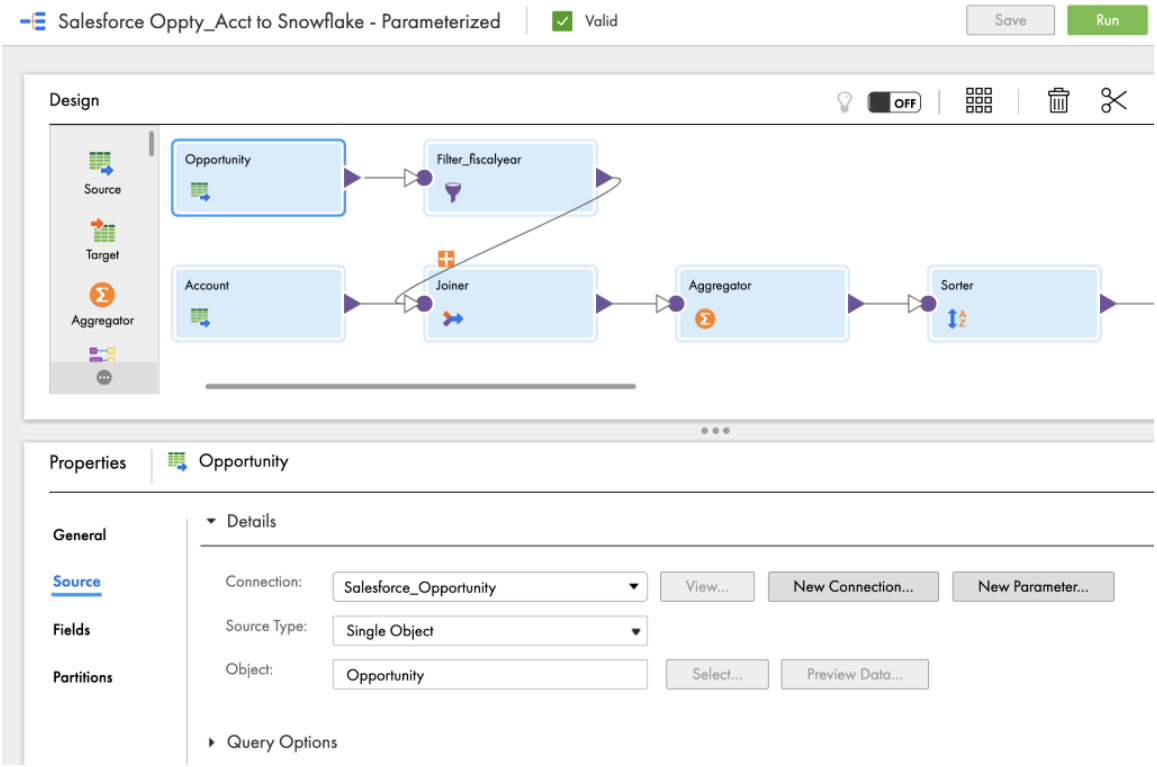
For more information about input parameters, see this topic in the Data Integration *Mappings* guide.

The following image shows the mapping in the Mapping Designer:



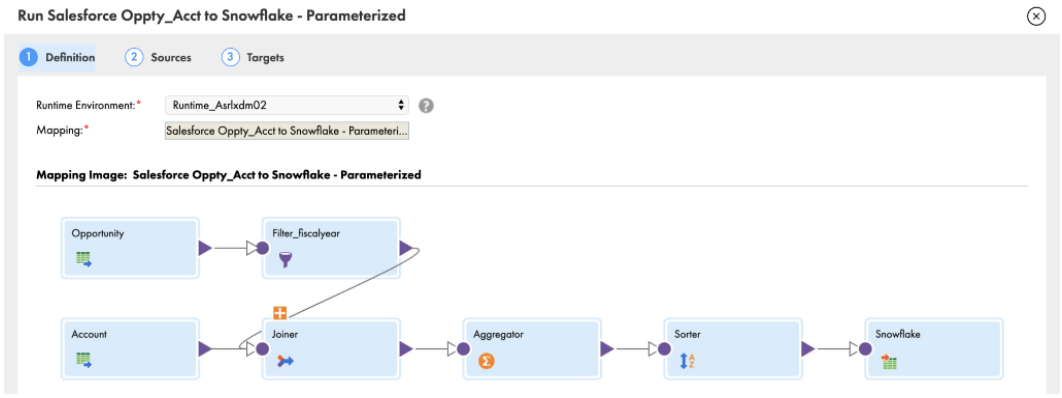
Click each transformation and view the transformation details on the **Properties** panel.

The following image shows the properties on the **Source** tab of the Opportunity source:

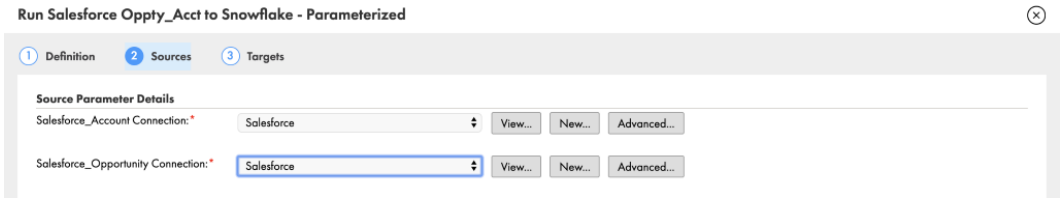


To run the mapping, click **Run** in the top right corner of the canvas, and then do the following:

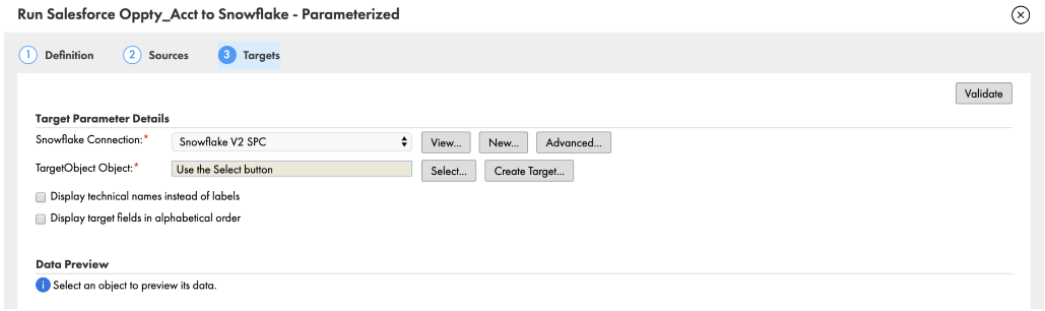
1. Select your runtime environment and then click **Next**:



2. In the **Salesforce_Account Connection** and **Salesforce_Opportunity Connection** source parameter fields, select the Salesforce connections you created and then click **Next**:



3. In the **Snowflake Connection** target parameter field, select the Snowflake connection you created:

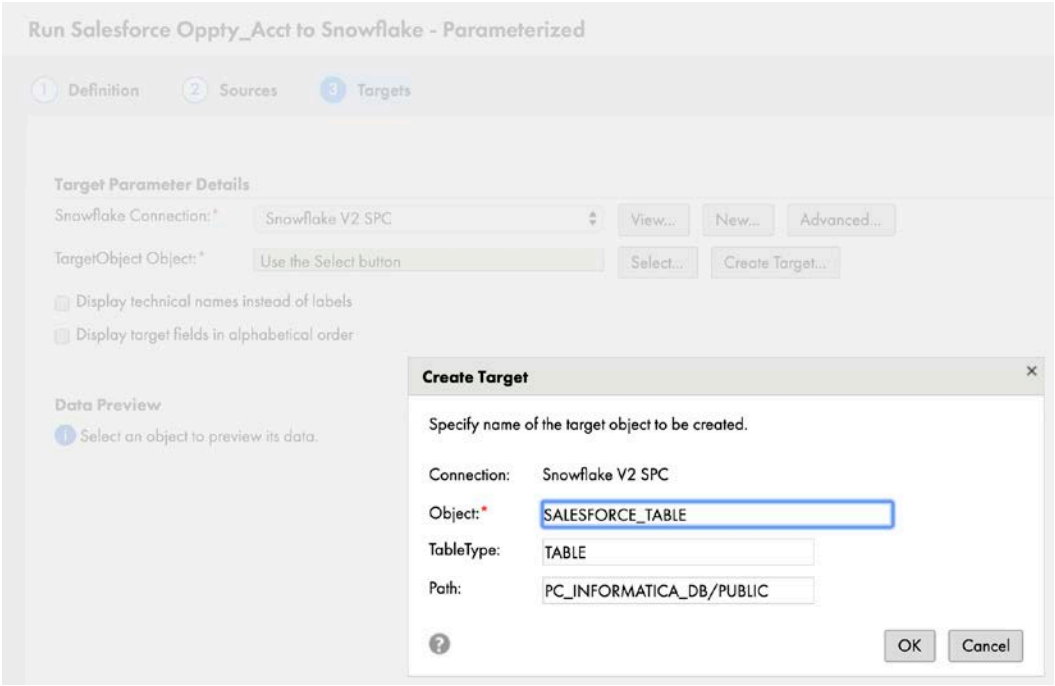


4. Click **Create Target** to create a new table.

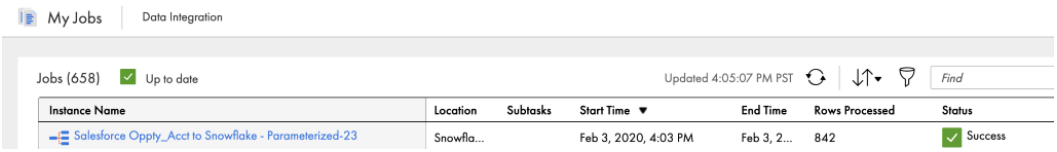
   Alternatively, you can click **Select** to use an existing table.

5. In the **Create Target** dialog box, enter the following information and then click **OK**:

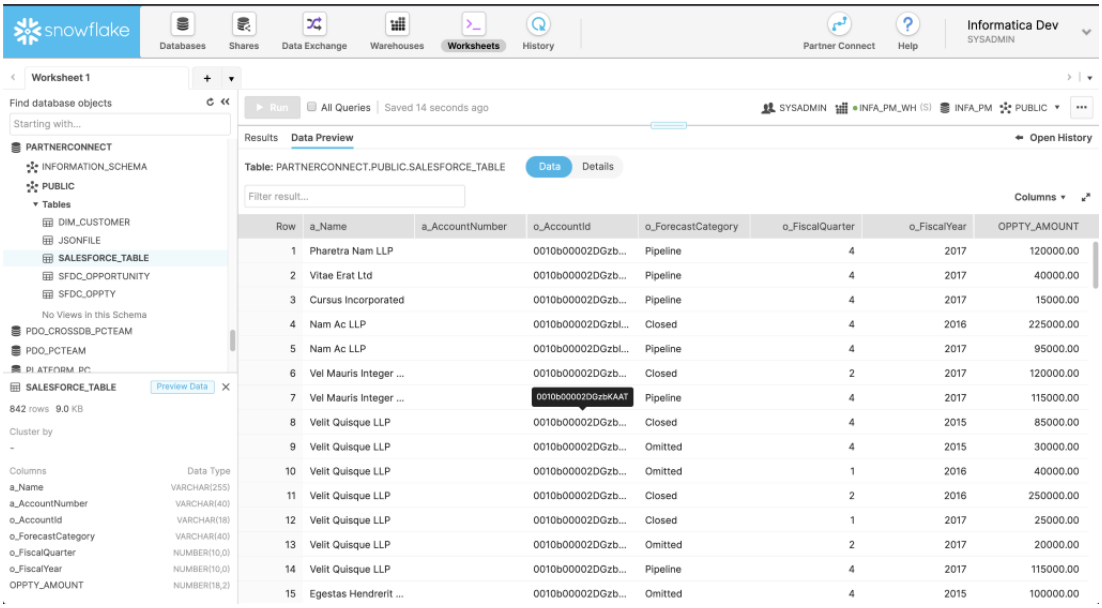
Property	Description
Object	Enter the target table name. This table will be created when you run the mapping.
TableType	Enter TABLE.
Path	Enter the database/schema.



6. Click **Run**.
7. Click **My Jobs** to open the job activity page:



The target table is created in Snowflake:



Loading a CSV file to Snowflake

In this step, you load a CSV file from the flat file directory that you configured in the flat file connection and write it to a Snowflake table. Please download the CSV file and place it in the flat file directory.

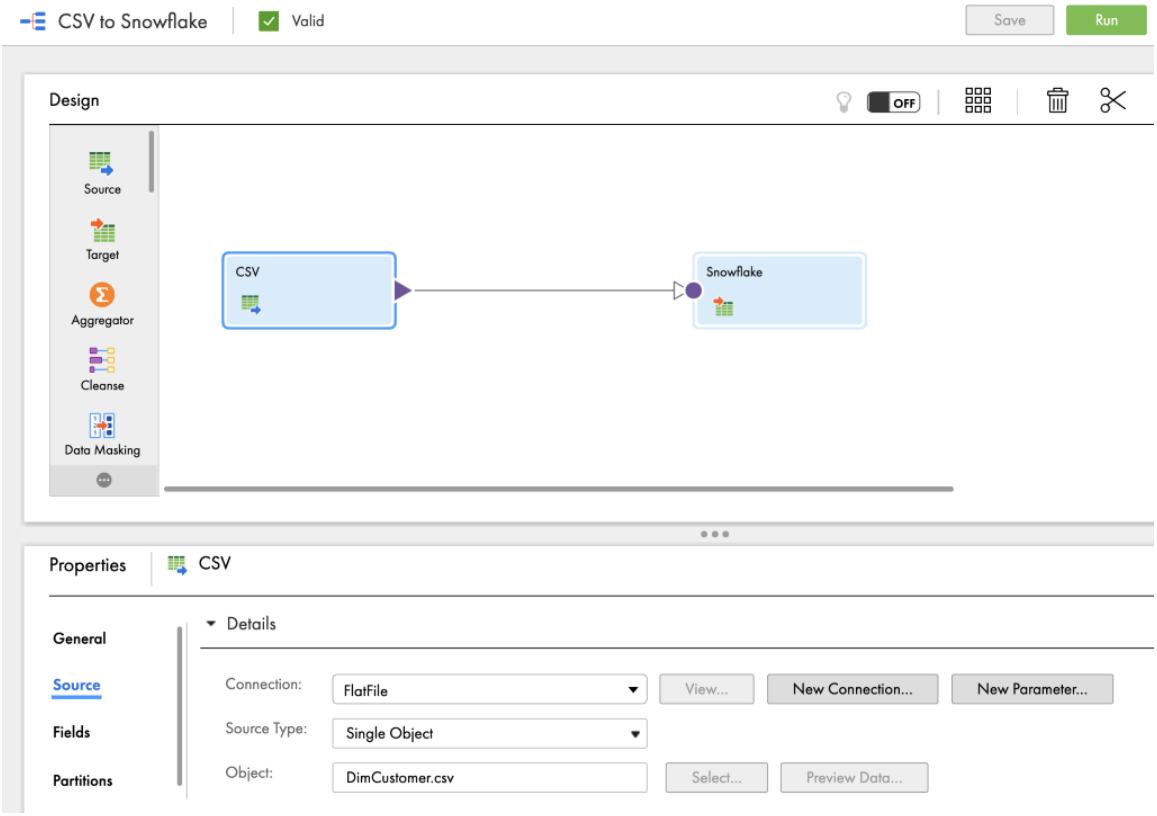
The second mapping in the bundle is called “CSV to Snowflake.” This mapping does the following:

- Reads the DimCustomer.csv file.
- Writes the records into Snowflake table.

The target connection in this mapping is parameterized using input parameters. This means that you have the option to choose the Snowflake target table at run time.

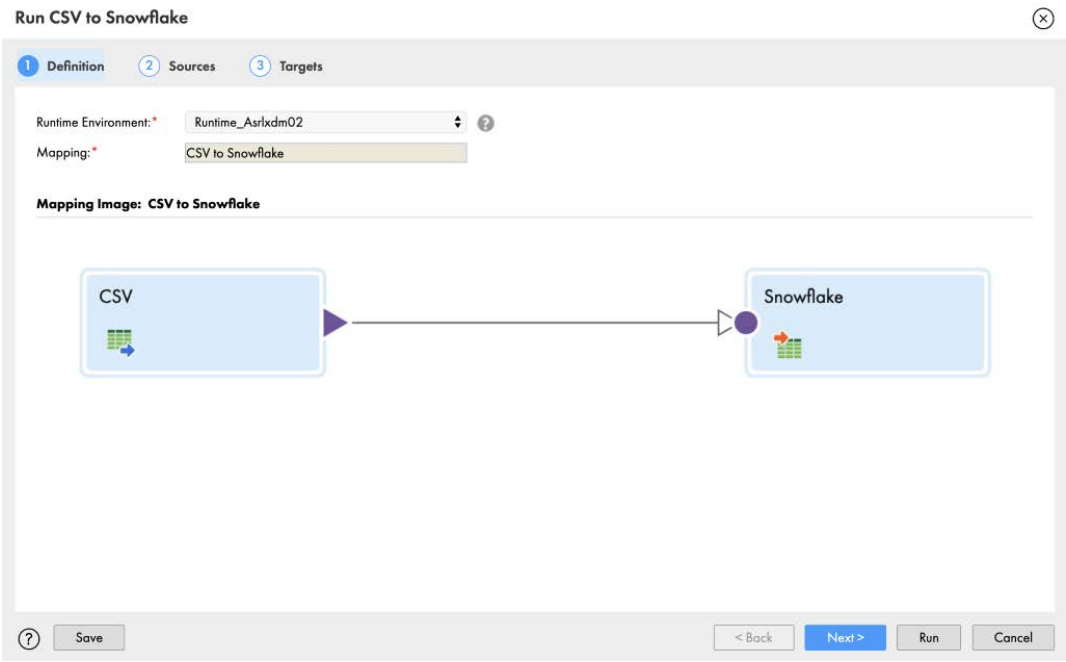
Click each transformation and view the transformation details on the **Properties** panel.

The following image shows the properties on the **Source** tab of the CSV source:



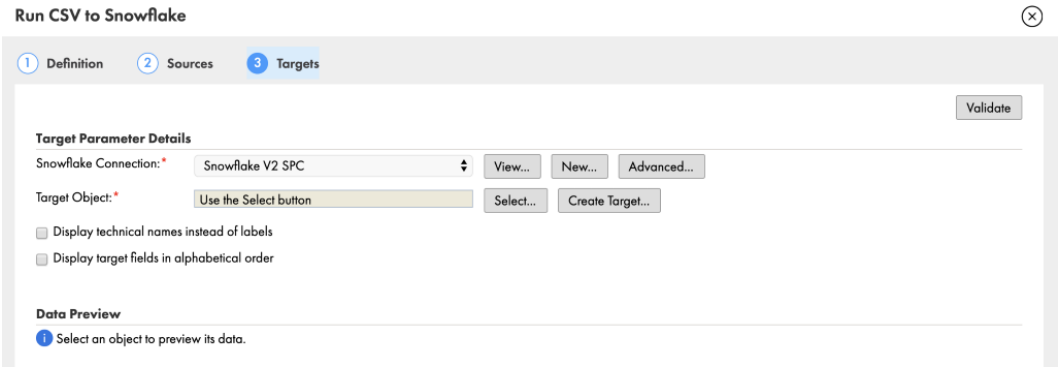
To run the mapping, click **Run** in the top right corner of the canvas, and then do the following:

1. Select your runtime environment and then click **Next**:



2. In the **FlatFile Connection** source parameter field, select the flat file connection you created and then click **Next**:

3. In the **Snowflake Connection** target parameter field, select the Snowflake connection you created:

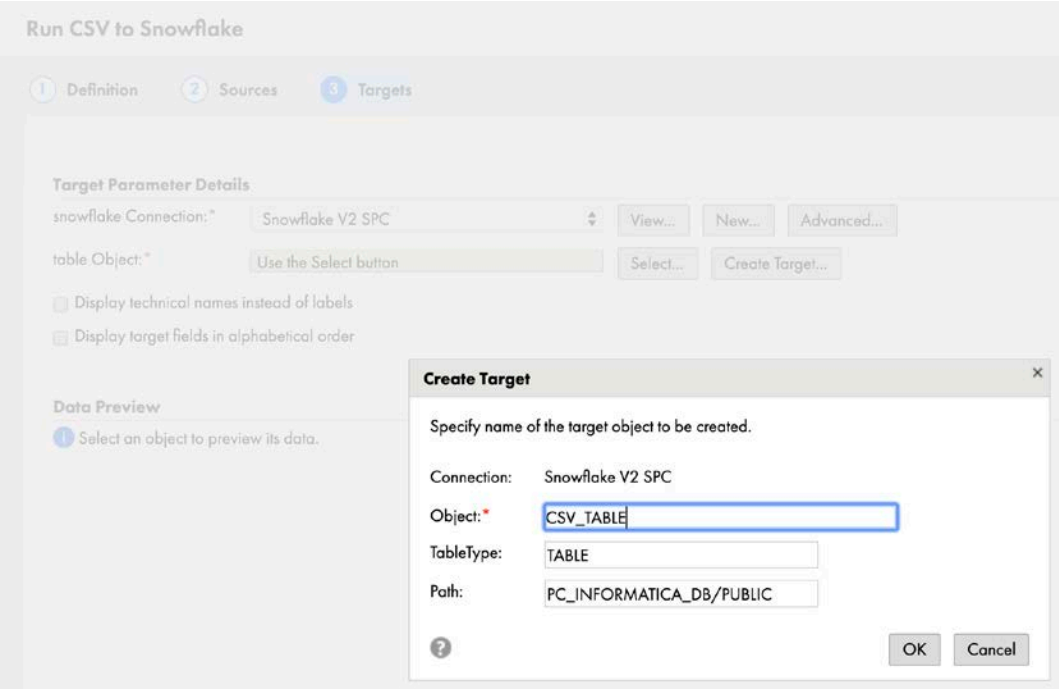


4. Click **Create Target** to create a new table.

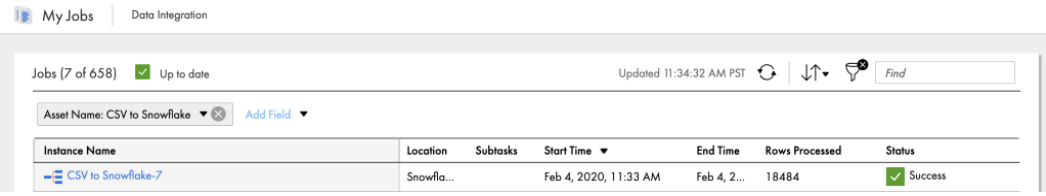
   Alternatively, you can click **Select** to use an existing table.

5. In the **Create Target** dialog box, enter the following information and then click **OK**:

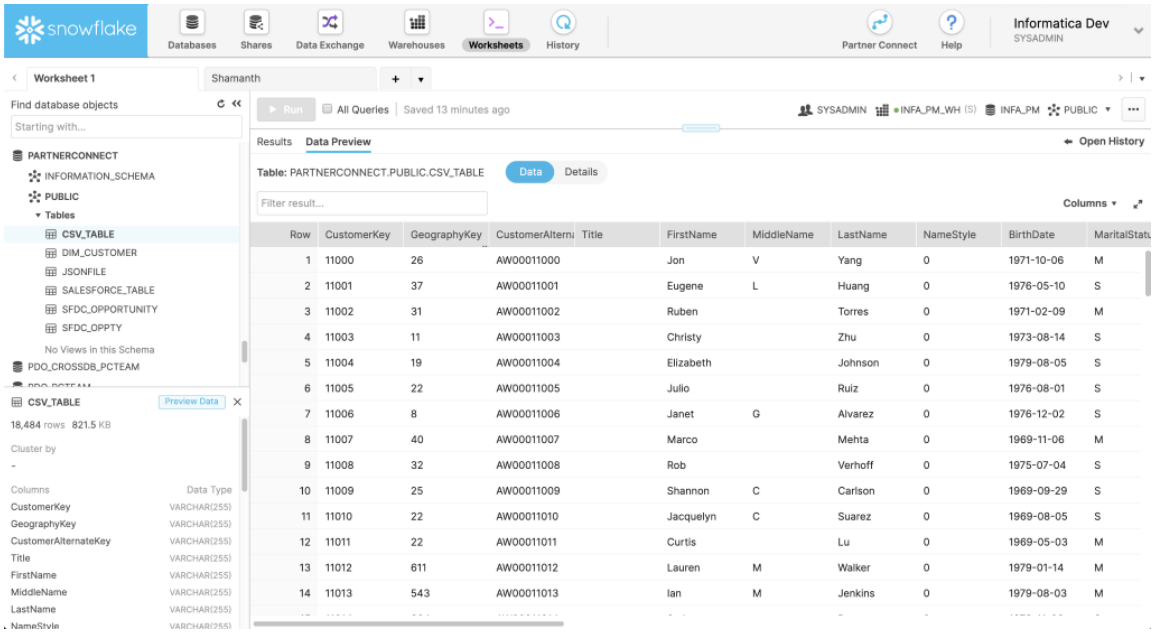
Property	Description
Object	Enter the target table name. This table will be created when you run the mapping.
TableType	Enter TABLE.
Path	Enter the database/schema.



6. Click **Run**.
7. Click **My Jobs** to open the job activity page:



The target table is created in Snowflake:



For more information about creating mappings, see the Data Integration *Mappings* guide.

Loading CSV files using a mass ingestion task

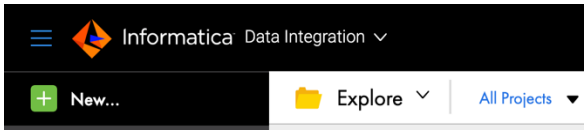
Loading many CSV files into Snowflake is easy to do using a mass ingestion task. A mass ingestion task can transfer a large number of files of any file type between on-premises and cloud repositories and can track and monitor file transfers. This section describes the step-by-step instructions to create and run a mass ingestion task.

Before you create the task, complete the following prerequisites:

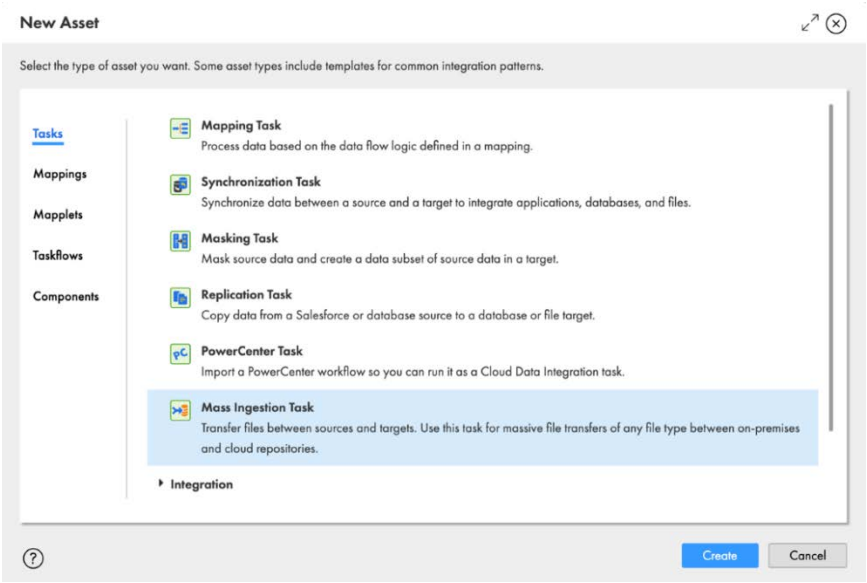
- Place the CSV files into the directory in which your flat file connection is configured to read. You can download the sample CSV files from here. The sample files are extracts from the Salesforce Opportunity object.
- Create a table in Snowflake with the same columns that are in the CSV files. File mass ingestion requires a target table to be available. You can download an SQL script to create the table here.

To create and run the mass ingestion task:

1. In Data Integration, click **New** in the navigation menu on the left:

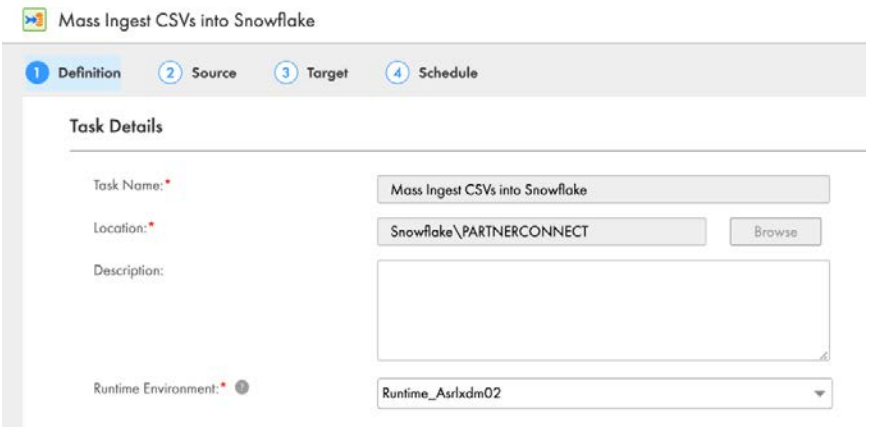


2. In the **New Asset** dialog box, select **Mass Ingestion Task**:



3. On the **Definition** page, enter the following information:

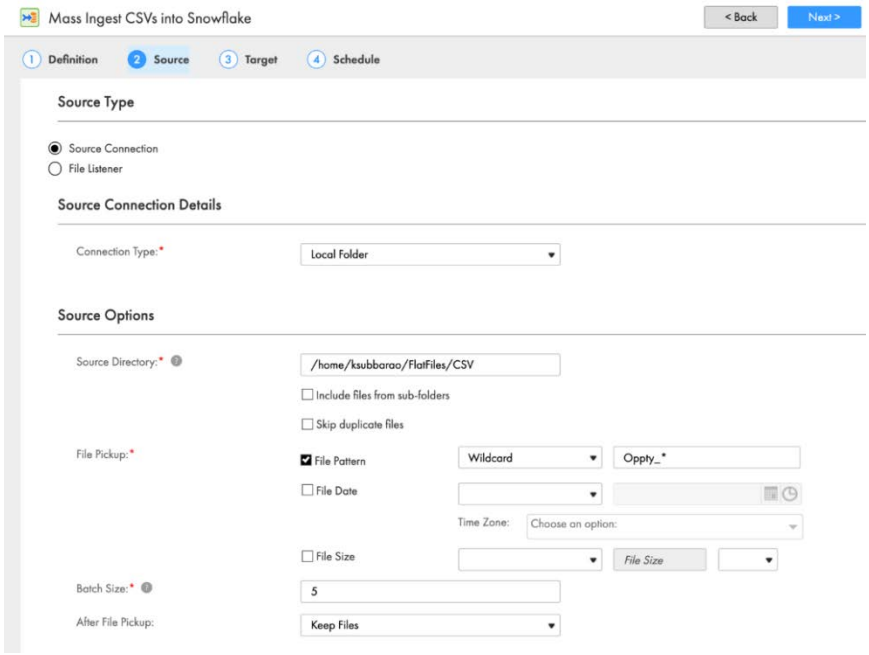
Property	Description
Task Name	Enter a task name.
Location	Select the project and folder where you want to save this task.
Runtime Environment	Select the runtime environment to run the task.



4. Click **Next**.
5. On the **Source** page, enter the following information:

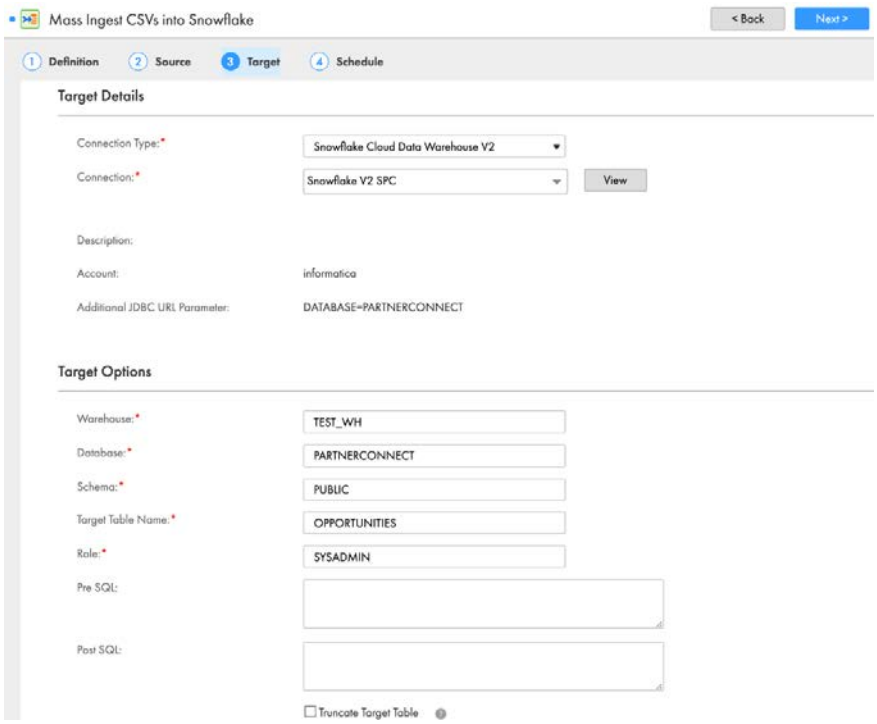
Property	Description
Connection Type	Select **Local Folder**.
Source Directory	Select the directory where you placed the CSV files.

Property	Description
File Pickup	Select **File Pattern**, leave the default **Wildcard**, and enter `Oppty_*`**.**

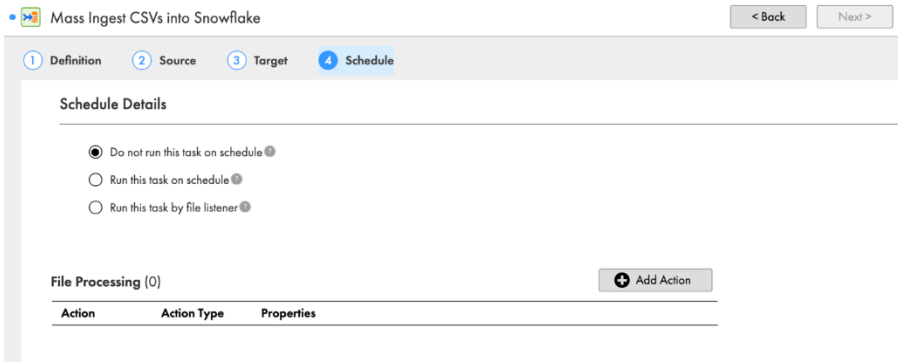


6. Click **Next**.
7. On the **Target** page, enter the following information:

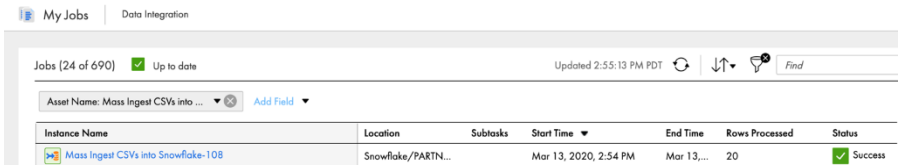
Property	Description
Connection Type	Select Snowflake Cloud Data Warehouse V2.
Connection	Select the Snowflake connection from the list.
Warehouse	Enter your warehouse name.
Database	Enter the database name where you created the opportunity table.
Schema	Enter the schema name.
Target Table Name	Enter `OPPORTUNITIES`.
Role	Enter a role name that has access to the target table.



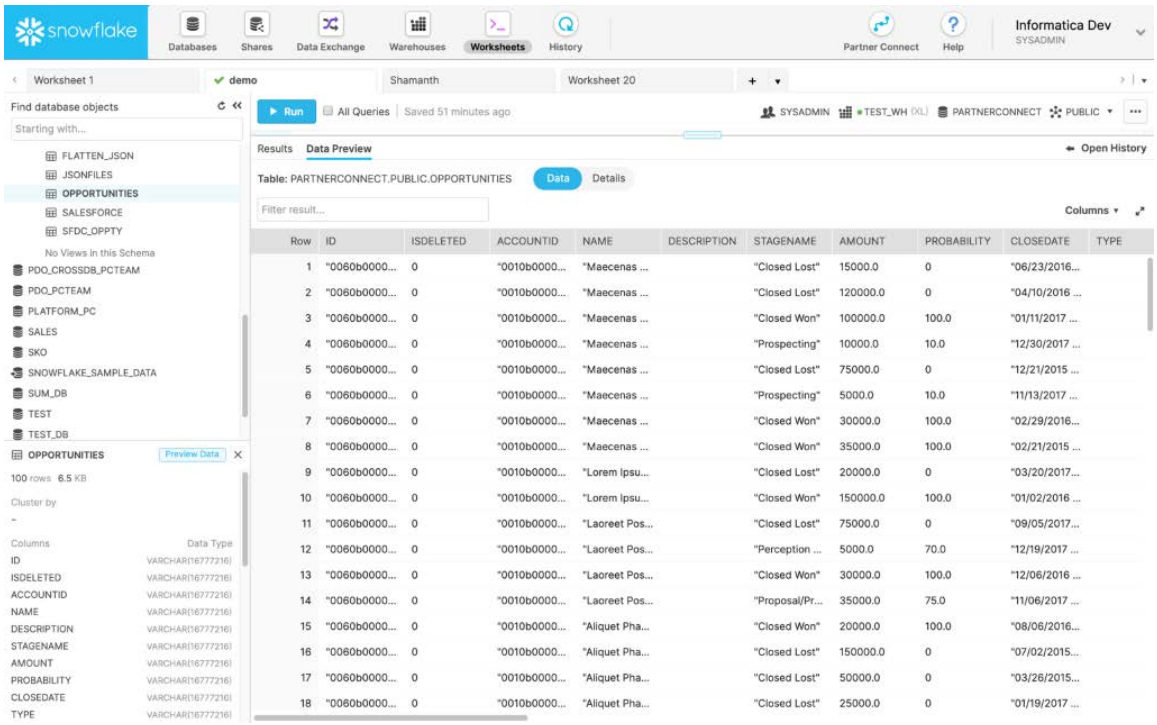
8. Click **Next**.
9. On the **Schedule** page, select **Do not run this task on a schedule**:



10. Click **Save** to save the task.
11. Click **Run**.
12. Open the **My Jobs** page to check the job status:



The target table is populated in Snowflake:



For more information about mass ingestion tasks, see "Mass ingestion tasks" in the Data Integration *Tasks* guide.

Using pushdown optimization in a mapping

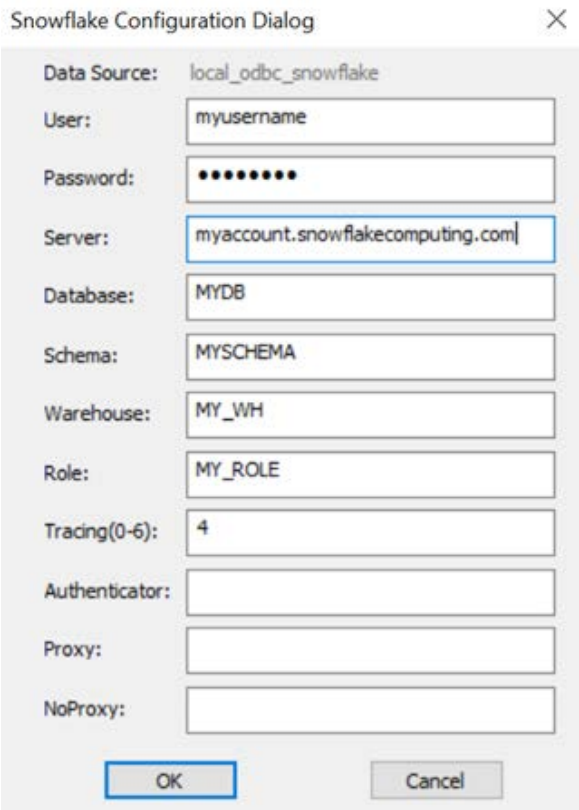
This section gives an example of a pushdown optimization scenario in which you develop a mapping in Data Integration and configure it to run entirely on Snowflake.

The steps involved are:

1. Create an ODBC DSN.
2. Create one or more ODBC connections in IICS.
3. Create a mapping.
4. Create a mapping task and run it.
5. Validate the pushdown implementation.

Step 1. Create an ODBC DSN

On the machine where you downloaded and installed the Secure Agent, create an ODBC DSN entry using a 64-bit ODBC client. Use the SnowflakeDSI driver that you download from the Snowflake website:



Be sure to specify the attributes **Database**, **Schema**, **Warehouse**, and **Role** in addition to the other credentials, as shown above.

Step 2. Create a connection

1. In Administrator, click **Connections**.
2. Click **New connection** in the top right corner of the screen.
3. Create an ODBC connection using the following values:

Property	Description
Connection Name	Name of the connection. The connection will appear in mappings and tasks with this name.
Description	Optional description for the connection.
Type	Select **OBDC**.
Runtime Environment	Runtime environment you have configured with the IICS organization.
Username	Your Snowflake account user name.
Password	Password for your Snowflake account.
Data Source Name	Name of the ODBC data source that you created.

Property	Description
Schema	Name of the schema in your Snowflake account that contains your tables. If you use more than one schema, you can create a connection for each schema.
Code Page	Select the code page that you want to use.
OBDC Subtype	Select **Snowflake**.
Driver Manager for Linux	Select **unixODBC 2.3.0** (the default value).

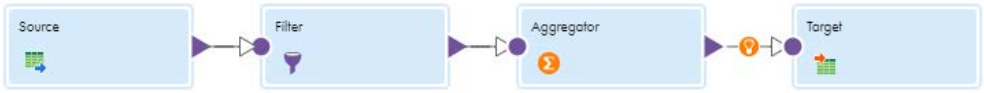
4. To test the connection, click **Test Connection**.

   You should see the message, “The test for this connection was successful.”

5. Click **Save** to save the connection.

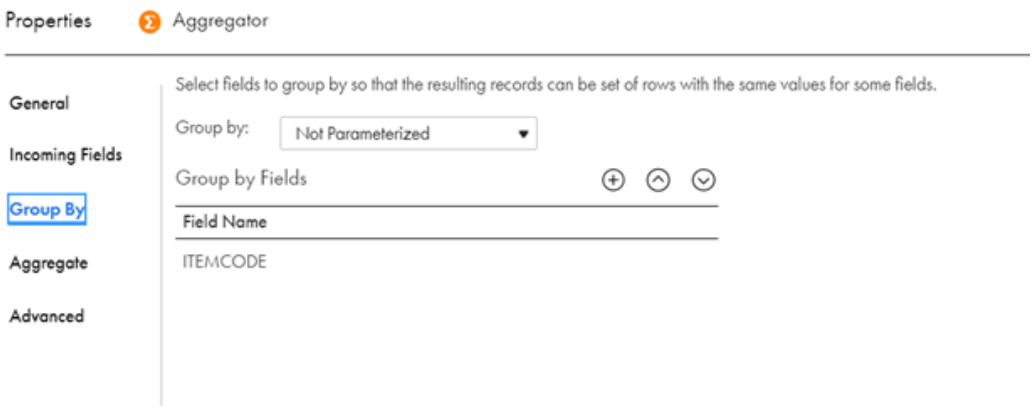
Step 3. Create a mapping

1. Click **Administrator** next to the Informatica logo in the top left corner of the screen, and then select **Data Integration** from the menu.
2. Click **New** in the navigation menu on the left.
3. In the **New Asset** dialog box, select **Mapping**.
4. In the Mapping Designer, create a mapping with the following transformations:



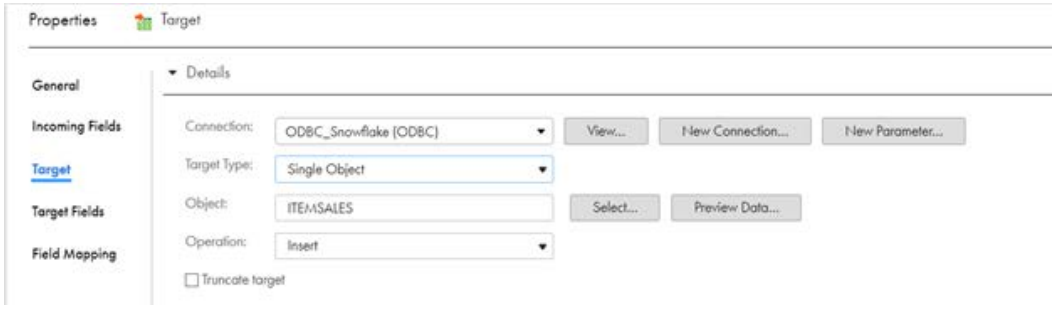
Transformation	Description	Configuration
Source	The Snowflake object that you read data from.	Browse and select the object.
Filter	Filter the rows that you don’t want to process or load.	Configure the filter condition.
Aggregator	Group by rows and apply aggregate functions such as SUM.	Configure the **Group By** fields and use the **Aggregate** tab to apply aggregate functions for the rest of the fields using the **Field Expression** dialog. An example is shown below.
Target	The Snowflake object that you want to write data to.	Browse and select the object or create a new target object based on your field projection. Configure the write action (insert, upsert etc.). If you use an existing Snowflake object, configure the field mapping on the **Field Mapping** tab.

Aggregator transformation configuration – **Group By** tab:



Aggregator transformation configuration – **Aggregate** tab:

Target transformation – **Target** tab:



5. In the top right corner of the Mapping Designer, click  to validate the mapping.

   If the mapping has validation errors, check the Data Integration online help for more information.

   The most common validation errors are:

   - No source or target object selected
   - Transformations not linked
   - In the Aggregator transformation, no fields are created, or no fields are specified for grouping
   - No fields are mapped for an existing target

6. Click **Save** to save the mapping.

Step 4. Create and run a mapping task

1. In the top right corner of the Mapping Designer, open the **Actions** menu and select **New Mapping Task**.
2. On the **Definition** page of the New Mapping Task wizard, enter a name for the mapping task and select the runtime environment you created.
3. Click **Next**.
4. On the **Schedule** page, enable pushdown optimization:
   a. Scroll down to the **Advanced Session Properties** and click **Add**.
   b. In the **Session Property Name**, select **Pushdown Optimization** in the Performance Settings.
   c. In the **Session Property Value**, select **Full**.
   d. Keep the **Enable cross-schema Pushdown Optimization** option selected, regardless of whether your source and target objects are in the same schema and database.
5. Click **Finish** to save and create the task.
6. Click **Run** to run the task.
7. Click **My Jobs** to open the job activity page.

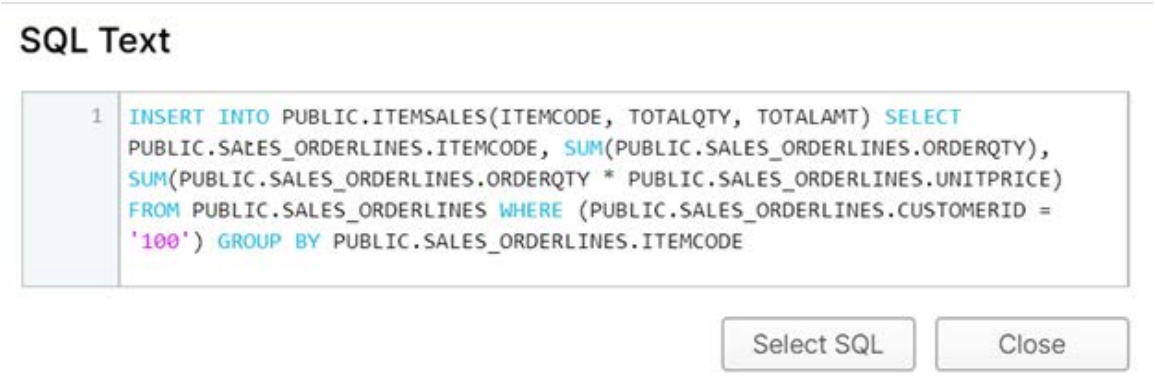
   It might take a few seconds for the new job to show up on the **My Jobs** page. Click refresh if you don't see it after a few seconds.

Step 5. Monitor the pushdown job

When the job completes and shows the status as "Success," you can go to the Snowflake Console and check the command history. Refer to the Frequently Asked Questions or the forum if the task fails and you need more information about the errors.

When the task shows either **Success** or **Warning**, the Snowflake command history should show a command that contains an INSERT INTO the target object followed by a SELECT from the source and a GROUP BY applied to the group by field.

An example appears below:



This completes the tryout of the pushdown optimization feature to implement an ELT pattern on Snowflake. Refer to the Snowflake Connector Guide to find the transformations and functions that support pushdown and any limitations that apply.

Next steps

Congratulations on completing the tasks in this guide!

There is much more that you can do with IICS. Please check out the Cloud Data Integration documentation on Informatica Network to find out more.