



# HANDS-ON LAB GUIDE FOR HARNESS THE POWER OF SNOWFLAKE WITH INFORMATICA'S INTELLIGENT DATA MANAGEMENT CLOUD

To be used with the Snowflake free 30-day trial at:  
<https://trial.snowflake.com>

Works for any Snowflake edition or cloud provider  
Approximate duration: 90 minutes.

# Table of Contents

## Lab Overview

Module 1: Prepare Your Lab Environment

Module 2: The Snowflake User Interface & Lab “Story”

Module 3: Create IICS Organization

Module 4: Configure an AWS S3 Connection

Module 5: Load Data from AWS S3 into Snowflake

Module 6: Transform Semi-Structured JSON Data

Module 7: Cloning, Time Travel in Snowflake

Summary & Next Steps

# Lab Overview

This entry-level lab introduces you to the user interface and a few basic capabilities of Snowflake, as well as Informatica Intelligent Cloud Services (IICS) Data Integration functionalities that will help your journey to the cloud, specifically with the Snowflake Data Cloud. After completing this lab, you can take advantage of the free Snowflake 30-day trial, and the free IICS organization created through Snowflake Partner Connect to load 1 billion records per month into Snowflake for free. Then take advantage of the free services to learn more about Snowflake and Informatica, and other complementary services and functionalities.

## Target Audience

Data integration architects, data engineers, developers, and analysts.

## What you'll learn

The exercises in this lab will walk you through the steps to:

- Set up an IICS organization through Snowflake Partner Connect.
- Configure an AWS S3 connection to read and load a source dataset
- Perform data transformations.
- Parse and transform semi-structured JSON data
- Clone objects
- Undo user errors

## Prerequisites

- Use of the Snowflake free 30-day trial environment
- Basic knowledge of SQL, and database concepts and objects
- Basic data integration and transformation concepts

## Introduction to Informatica Intelligent Cloud Services

Informatica Intelligent Cloud Services is a next generation iPaaS, which is made up of several data management products. The productivity of the environment is accelerated by a common user experience across all products, the AI/ML-driven intelligence of the [CLAIRES™ engine](#), and a microservices architecture. In this lab, we will be mainly focusing on data integration. However, you may feel free to navigate other available services.

When you log in to Informatica Intelligent Cloud Services, the **My Services** page displays the services that apply to data integration. The **My Services** page might also include other services that you subscribe to and trial versions of other services.

## Working with Cloud Data Integration Services

When you select **Data Integration** from the **My Services** page, the Data Integration **Home** page appears, as follows:

The screenshot shows the Informatica Data Integration Home page with the following sections:

- Overview**: Shows 2 Runtime Environments (Up and Running), 7 Connections, 3 Projects, 3 Folders, and 27 Assets.
- Runtime Environments**: Displays the status of the Informatica Cloud Hosted Agent (Up and Running) and a connection to minfsvc.infa.com (Stopped).
- My Jobs**: Shows "No jobs were run." with a "View All" link.
- Most Recent Project**: Shows a project named "Admin Tasks - Do not delete" with 8 ASSETS, all of which are Valid.
- Recent Assets**: A table listing recent assets modified by the user, including:

| Name              | Location                    | Last Accessed         |
|-------------------|-----------------------------|-----------------------|
| 1_Linear Taskflow | Admin Tasks - Do not delete | Aug 27, 2019, 9:34 AM |
| 2_DS_SUBORG...    | Admin Tasks - Do not delete | Aug 20, 2019, 1:42 AM |
| 3_DS_CREATE...    | Admin Tasks - Do not delete | Aug 4, 2019, 11:44 PM |
| 4_DS_DELETE...    | Admin Tasks - Do not delete | Aug 4, 2019, 11:44 PM |
| 5_DS_IMPORT...    | Admin Tasks - Do not delete | Jul 3, 2019, 4:08 AM  |
| 6_DS_AZURES...    | Admin Tasks - Do not delete | Jul 3, 2019, 4:08 AM  |

The **Home** page displays the following panels:

- **Overview**. A summary of projects, folders, connections, and assets available to your organization.
- **Runtime Environment Status**. The status of all of the organization's runtime environments.
- **My Jobs**. A list of jobs that you ran in the last 24 hours.
- **Most Recent Project**. The last project that you created or that contains the most recently modified asset.
- **Recent Assets**. The assets that you most recently modified.

You can access the following pages from the navigation bar for Data Integration:

- **New**. Create new data integration assets such as mappings, tasks, and taskflows. You can create assets from scratch or use templates as a head start.
- **Home**. Return to the Home page.
- **Explore**. Create and manage data integration projects and assets.
- **Bundles**. Create and manage bundles of assets to share with your organization or other organizations.
- **My Jobs**. View the status of your Data Integration jobs.

When you switch from Data Integration to another service such as Administrator or Monitor, the options in the navigation bar change to suit the service.

When you click on “New” these are the following tasks that you can integrate data with the following tasks:

- **Mass ingestion task**. Transfer files between on-premise and remote servers, and Data Integration.
- **Mapping task**. Process data based on the data flow logic that you define in a mapping.
- **Synchronization task**. Use the synchronization task to synchronize data between a source and a target and use expressions to transform the data according to your business logic or use data filters to filter data before writing it to targets.
- **Masking task**. Replace source data in sensitive columns with realistic test data for non-production environments. Masking rules define the logic to replace the sensitive data. Assign masking rules to the columns you need to mask.
- **Replication task**. Replicate data from Salesforce or database sources to database or file targets. You might replicate data to archive the data, perform offline reporting, or consolidate and manage data.
- **PowerCenter task**. Import a PowerCenter workflow and run it as a Data Integration PowerCenter task.
- **Taskflow**. Taskflow orchestrates the execution sequence of multiple data integration tasks. You can use Taskflow for complex data integration projects.

# Module 1: Prepare Your Lab Environment

## 1.1 Steps to Prepare Your Lab Environment

**1.1.1** If not yet done, register for a Snowflake free 30-day trial at <https://trial.snowflake.com>

- The Snowflake edition (Standard, Premier, Enterprise, e.g.), cloud provider (AWS, Azure, GCP), and Region (US East, EU, e.g.) do \*not\* matter for this lab. But we suggest you select the region which is physically closest to you. And select the Enterprise edition so you can leverage some advanced capabilities that are not available in lower Editions.
- After registering, you will receive an email with an activation link and your Snowflake account URL. Bookmark this URL for easy, future access. After activation, you will create a user name and password. Write down these credentials.

**1.1.2** Resize your browser windows so you can view this lab guide PDF and your web browser side-by-side to more easily follow the lab instructions. If possible, even better is to use a secondary display dedicated to the lab guide.

**1.1.3** Click on

[https://snowflake-corp-se-workshop.s3-us-west-1.amazonaws.com/VHOL\\_Snowflake\\_Informatica/snowflake-informatica-vhol.sql](https://snowflake-corp-se-workshop.s3-us-west-1.amazonaws.com/VHOL_Snowflake_Informatica/snowflake-informatica-vhol.sql) and download the “snowflake-informatica-vhol.sql” file to your local machine. This file contains pre-written SQL commands and we will use this file later in the lab.

## Module 2: The Snowflake User Interface & Lab Data Sets



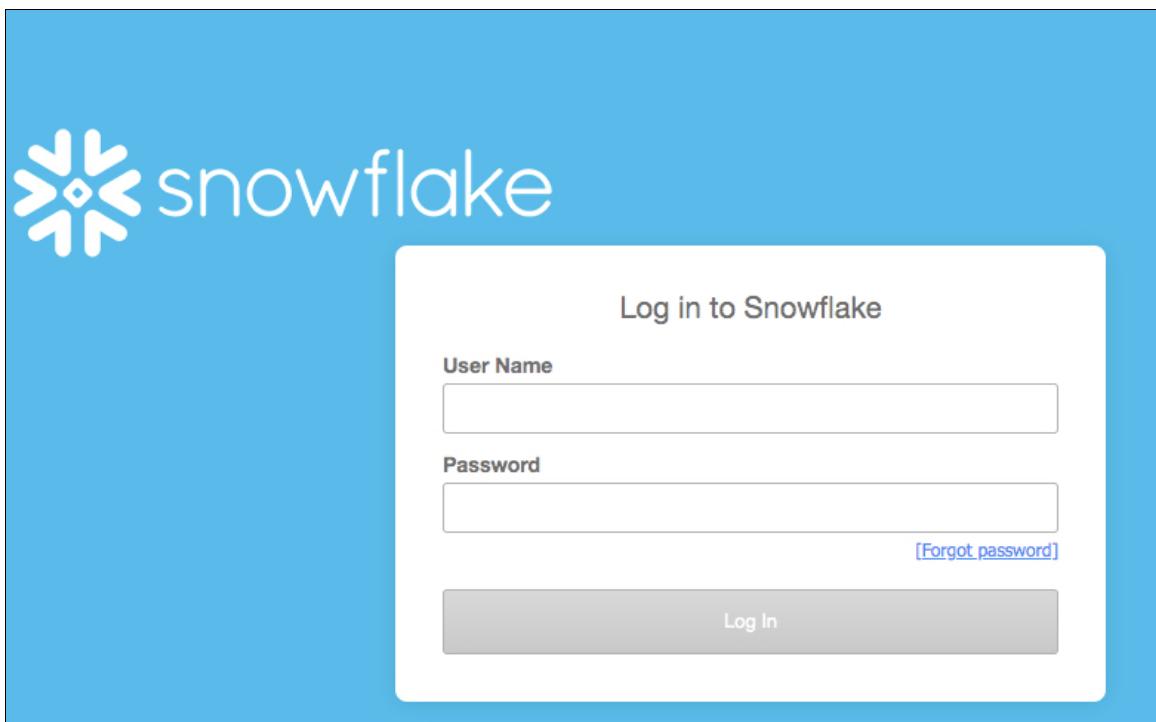
### About the screen captures, sample code, and environment

Screen captures in this lab depict examples and results that may slightly vary from what you may see when you complete the exercises.

### 2.1 Logging Into the Snowflake User Interface (UI)

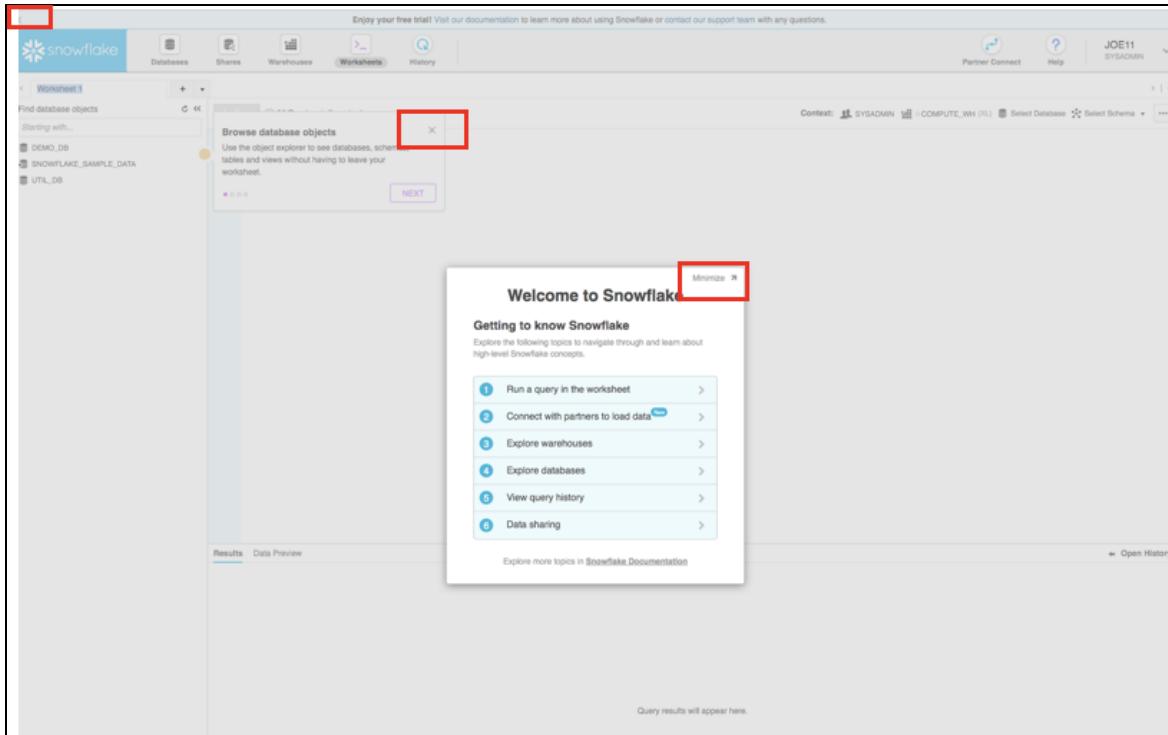
2.1.1 Open a browser window and enter the URL of your the Snowflake 30-day trial environment.

2.1.2 You should see the login screen below. Enter your unique credentials to log in.



### 2.2 Close any Welcome Boxes and Tutorials

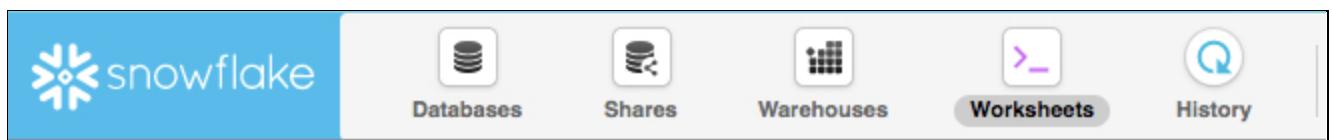
2.2.1 You may see “welcome” and “helper” boxes in the UI when you log in for the first time. Also a “Enjoy your free trial...” ribbon at the top of the UI. Minimize and close them by clicking on the items in the red boxes in the screenshot below.



## 2.3 Navigating the Snowflake UI

First let's get you acquainted with Snowflake! This section covers the basic components of the user interface to help you orient yourself. We will move left to right in the top of the UI.

### 2.3.1 The top menu allows you to switch between the different areas of Snowflake:



2.3.2 The **Databases** tab shows information about the databases you have created or have privileges to access. You can create, clone, drop, or transfer ownership of databases as well as load data (limited) in the UI. Notice several databases already exist in your environment. However, we will not be using these in this lab.

| Database              | Origin            | Creation Time       | Owner        | Comment                    |
|-----------------------|-------------------|---------------------|--------------|----------------------------|
| SNOWFLAKE_SAMPLE_DATA | SFC_SAMPLES.SA... | 6/27/19 11:52:44 PM | ACCOUNTADMIN | TPC-H, OpenWeatherMap, etc |
| DEMO_DB               |                   | 6/27/19 11:52:42 PM | SYSADMIN     | demo database              |
| UTIL_DB               |                   | 6/27/19 11:52:30 PM | SYSADMIN     | utility database           |

2.3.3 The **Shares** tab is where data sharing can be configured to easily and securely share Snowflake table(s) among separate Snowflake accounts or external users, without having to create a second copy of the table data.

2.3.4 The **Warehouses** tab is where you set up and manage compute resources (virtual warehouses) to load or query data in Snowflake. Note a warehouse called “COMPUTE\_WH (X-Small)” already exists in your environment.

| Status    | Warehouse Name | Size    | Clusters       | Scaling Poli... | Runn... | Que... | Auto Suspe... | Auto Resume | Created On | Resumed On | Owner    |
|-----------|----------------|---------|----------------|-----------------|---------|--------|---------------|-------------|------------|------------|----------|
| Suspended | COMPUTE_WH     | X-Small | min: 1, max: 1 | Standard        | 0       | 0      | 10 minutes    | Yes         | 4:18:49 PM | 4:18:49 PM | SYSADMIN |

2.3.5 Let's walk through the settings of this warehouse as there is a lot of functionality here, much of which is unique to Snowflake versus other data warehouses.

NOTE - If you do not have a Snowflake Edition of Enterprise or greater, you will \*NOT\* see the "Maximum Clusters" or "Scaling Policy" configurations from the screenshot below. Multi-clustering is not utilized in this lab, but we will still discuss it as it is a key capability of Snowflake.

- The "Size" drop-down is where the size of the warehouse is selected. For larger data loading operations or more compute-intensive queries, a larger warehouse will be needed. The t-shirt sizes translate to underlying compute nodes, either AWS EC2, Azure Virtual Machines or Google Compute Engine Virtual Machines . The larger the t-shirt size, the more compute resources from the cloud provider are allocated to that warehouse. As an example, the 4-XL option allocates 128 nodes. Also, this sizing can be changed up or down on the fly with a simple click.
- If you have Snowflake Enterprise Edition or greater you will see the Maximum Clusters section. This is where you can set up a single warehouse to be multi-cluster up to 10 clusters. As an example, if the 4-XL warehouse we just mentioned was assigned a maximum cluster size of 10, it could scale up to be 1280 (128 \* 10) AWS EC2, Azure VM or Google VM nodes powering that warehouse...and it can do this in seconds! Multi-cluster is ideal for concurrency scenarios, such as many business analysts simultaneously running different queries using the same warehouse. In this scenario, the various queries can be allocated across the multiple clusters to ensure they run fast.
- The final sections allow you to automatically suspend the warehouse so it suspends (stops) itself when not in use and no credits are consumed. There is also an option to automatically resume (start) a suspended warehouse so when a new workload is assigned to it, it will automatically start back up. This functionality enables Snowflake's fair "pay as you use" compute pricing model which enables customers to minimize their data warehouse costs.

Configure Warehouse

Name COMPUTE\_WH

Size X-Large (16 credits / hour)

Learn more about virtual warehouse sizes [here](#)

Maximum Clusters 1

Multi-cluster warehouses improve the query throughput for high concurrency workloads.

Scaling Policy Standard

The policy used to automatically start up and shut down clusters.

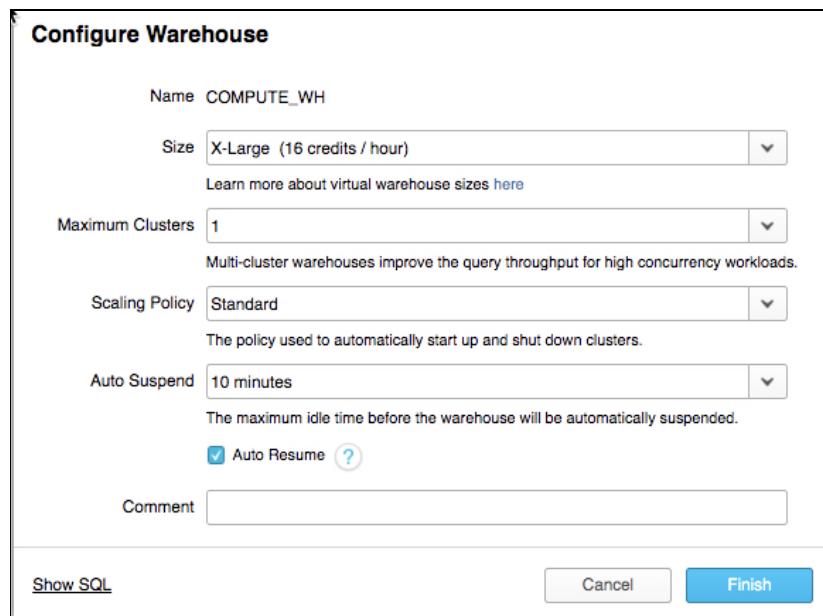
Auto Suspend 10 minutes

The maximum idle time before the warehouse will be automatically suspended.

Auto Resume [?](#)

Comment

Show SQL Cancel Finish





## Snowflake Compute vs Other Warehouses

Many of the warehouse/compute capabilities we just covered, like being able to create, scale up and out, and auto-suspend/resume warehouses are things that are simple in Snowflake and can be done in seconds. Yet for on-premise data warehouses these capabilities are very difficult (or impossible) to do as they require significant physical hardware, over-provisioning of hardware for workload spikes, significant configuration work, and more challenges. Even other cloud data warehouses cannot scale up and out like Snowflake without significantly more configuration work and time.



### Warning - Watch Your Spend!

During or after this lab you should \*NOT\* do the following without good reason or you may burn through your \$400 of free credits more quickly than desired:

- Disable auto-suspend. If auto-suspend is disabled, your warehouses will continue to run and consume credits even when not being utilized.
- Use a warehouse size that is excessive given the workload. The larger the warehouse, the more credits are consumed.

2.3.6 The **History** tab allows you to view the details of all queries executed in the last 14 days in the Snowflake account (click on a Query ID to drill into the query for more detail).

| Status | Query ID     | SQL Text           | User    | Warehouse     | Size | Session ID | Start Time | End Time   | Total Duration |
|--------|--------------|--------------------|---------|---------------|------|------------|------------|------------|----------------|
| ✓      | 018d6a78-... | SHOW GRANTS TO ... | USER123 | COMPUTE...... |      | 225259525  | 1:08:12 PM | 1:08:12 PM | 134ms          |

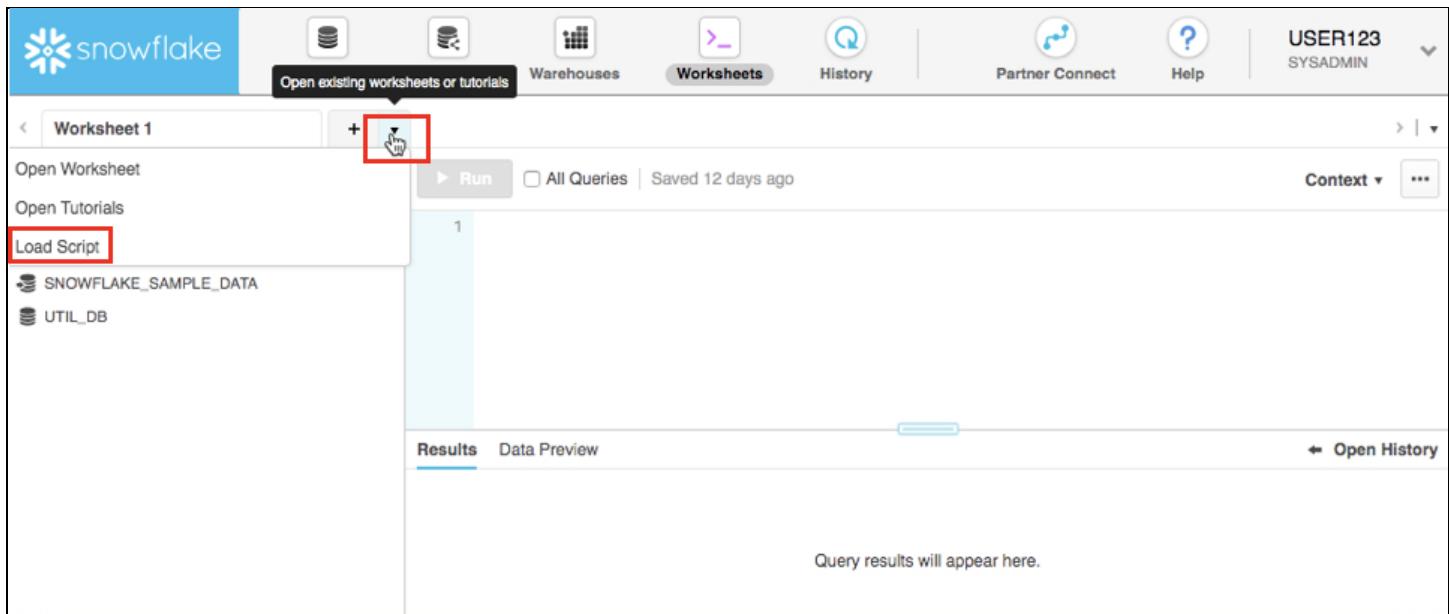
2.3.7 The **Worksheets** tab provides an interface for submitting SQL queries, performing DDL and DML operations and viewing results as your queries/operations complete. The default “Worksheet 1” appears.

In the left pane is the database objects browser which enables users to explore all databases, schemas, tables, and views accessible by the role selected for a worksheet. The bottom pane shows results of queries and operations.

The various windows on this page can be resized by moving the small sliders on them. And if during the lab you need more room to work in the worksheet, collapse the database objects browser in the left pane. Many of the screenshots in this guide will have this database objects browser closed.

The screenshot shows the Snowflake web interface. At the top, there's a navigation bar with icons for Databases, Shares, Warehouses, Worksheets (which is highlighted with a red box), History, Partner Connect, Help, and a user dropdown for 'USER123 SYSADMIN'. Below the navigation bar is a toolbar with buttons for 'Worksheet 1' (with a '+' icon), 'Find database objects' (with a search icon and a red box around it), 'Run' (with a play icon), and other options like 'All Queries' and 'Saved 12 days ago'. To the right of the toolbar is a 'Context' dropdown and a three-dot menu. On the left side, there's a collapsed sidebar titled 'Starting with...' containing database entries: 'DEMO\_DB', 'SNOWFLAKE\_SAMPLE\_DATA', and 'UTIL\_DB'. The main workspace is titled 'Worksheet 1' and contains a single row labeled '1'. At the bottom of the workspace, there are tabs for 'Results' (which is selected) and 'Data Preview', along with a link to 'Open History'. A message at the bottom of the workspace says 'Query results will appear here.'

- 2.3.8 Make sure you are in the Snowflake UI. At the top left of the default “Worksheet 1,” just to the right of the worksheet tab, click on the small, downward facing arrow, select “Load Script”, then browse to the “snowflake-informatica-vhol.sql” file you downloaded in the prior module and select “Open”. All of the SQL commands you need to run for the remainder of this lab will now appear on the new worksheet. Do not run any of the SQL commands yet. We will come back to them later in the lab and execute them one at a time.



#### **Warning - Do Not Copy/Paste SQL From This PDF to a Worksheet**

Copy-pasting the SQL code from this PDF into a Snowflake worksheet will result in formatting errors and the SQL will not run correctly. Make sure to use the “Load Script” method just covered.



On older or locked-down browsers, this “load script” step may not work as the browser will prevent you from opening the .sql file. If this is the case, open the .sql file with a text editor and then copy/paste all the text from the .sql file to the “Worksheet 1”

#### **Many Options to Run Commands.**



SQL commands can be executed through the UI (limited), via the Worksheets tab, using our SnowSQL command line tool, a SQL editor of your choice via ODBC/JDBC, or through our Python or Spark connectors.

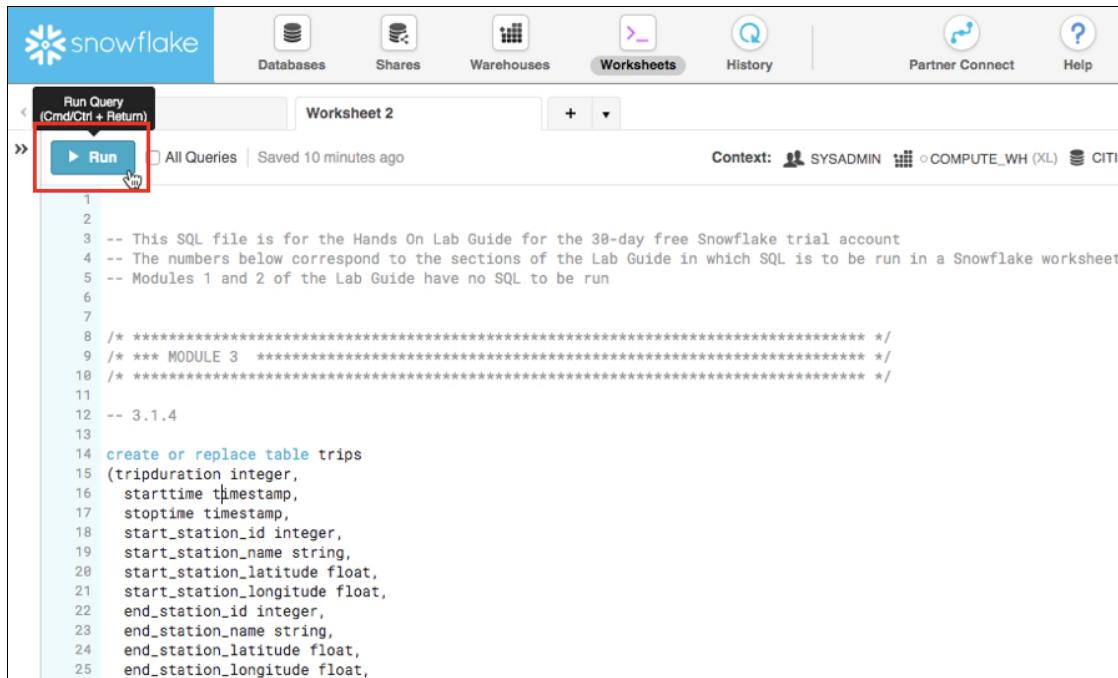
As mentioned earlier, in this lab we will run some operations via pre-written SQL in the worksheet (as opposed to using the UI) to save time.

- 2.3.9 Run the query by placing your cursor anywhere in the command and clicking the blue “Run” button at the top of the page or by hitting Ctrl/Cmd+Enter on your keyboard.



### Warning

In this lab, never check the “All Queries” box at the top of the worksheet. We want to run SQL queries one at a time in a specific order; not all at once.



The screenshot shows the Snowflake Worksheet interface. At the top, there are navigation icons for Databases, Shares, Warehouses, Worksheets (which is selected), History, Partner Connect, and Help. Below the header, a toolbar has a 'Run Query (Cmd/Ctrl + Return)' button, which is highlighted with a red box. The main area is titled 'Worksheet 2'. It displays a SQL script with numbered lines from 1 to 25. Lines 1 through 7 are comments. Lines 8 through 25 define a table named 'trips' with columns: tripduration (integer), starttime (timestamp), stoptime (timestamp), start\_station\_id (integer), start\_station\_name (string), start\_station\_latitude (float), start\_station\_longitude (float), end\_station\_id (integer), end\_station\_name (string), end\_station\_latitude (float), and end\_station\_longitude (float). The context bar at the bottom shows the user is SYSADMIN, connected to COMPUTE\_WH (XL), and the database is CITIB.

```
1
2
3 -- This SQL file is for the Hands On Lab Guide for the 30-day free Snowflake trial account
4 -- The numbers below correspond to the sections of the Lab Guide in which SQL is to be run in a Snowflake worksheet
5 -- Modules 1 and 2 of the Lab Guide have no SQL to be run
6
7
8 /* *****
9  * MODULE 3
10 *****/
11
12 -- 3.1.4
13
14 create or replace table trips
15 (tripduration integer,
16  starttime timestamp,
17  stoptime timestamp,
18  start_station_id integer,
19  start_station_name string,
20  start_station_latitude float,
21  start_station_longitude float,
22  end_station_id integer,
23  end_station_name string,
24  end_station_latitude float,
25  end_station_longitude float,
```

2.3.10 \*If\* you highlighted the entire SQL text of the command (did not just place your cursor in the command) and ran it, a confirmation box should appear asking “Do you want to run the following queries?”. Click the blue “Run” button in the box. In the future you can keep clicking this “Run” button on this confirmation box or check the “Don’t ask me again (All Worksheets)” option in this box.

The screenshot shows the Snowflake web interface. At the top, there's a navigation bar with icons for Databases, Shares, Warehouses, Worksheets (which is selected), History, and Partner Connect. Below the navigation bar, there are two tabs: Worksheet 1 and Worksheet 2. The Worksheet 1 tab has a 'Run' button, an 'All Queries' checkbox, and a note indicating the query was saved 14 minutes ago. To the right, the context is shown as SYSADMIN in COMPUTE\_WH. The main area contains a block of SQL code. A modal dialog box is overlaid on the page, titled 'Do you want to run the following queries?'. It contains the same SQL code as the worksheet. At the bottom of the dialog, there are 'Cancel' and 'Run' buttons, with 'Run' being highlighted with a red box and a cursor icon pointing to it.

```
4 -- The numbers below correspond to the sections of the Lab Guide in which SQL is to be run in a Snowflake
5 -- Modules 1 and 2 of the Lab Guide have no SQL to be run
6
7
8 /* **** MODULE 3 ****
9 */
10 /*
11
12 -- 3.1.4
13
14 create or replace table trips
15 (tripduration integer,
16 starttime timestamp,
17 stoptime timestamp,
18 start_station_id integer,
19 start_station_name string,
20 start_station_latitude float,
21 start_station_longitude float,
22 end_station_id integer,
23 end_station_name string,
24 end_station_latitude float,
25 end_station_longitude float,
26 bikeid integer,
27 membership_type string,
28 usertype string,
29 birth_year integer,
30 gender integer);
```

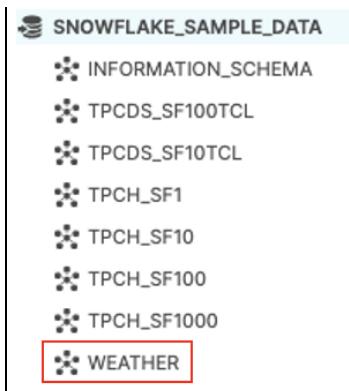
## 2.4 Lab Datasets

For this lab we will be joining the standard TPC-DS ORDERS and LINEITEMS tables, performing some push-down transformations, and loading the data into a new ORDERSLINEITEMS table.

We will also access WEATHER data in its native JSON format and transform it to a structured format.

Note that for most Snowflake deployments, the WEATHER data we'll be using is already available in a **SNOWFLAKE\_SAMPLE\_DATA.WEATHER** schema. **However, for some deployments it is not available, so we'll need to load it first.**

- 2.4.1 From the Worksheets tab, click on the database icon next to SNOWFLAKE\_SAMPLE\_DATA. If you see the WEATHER schema, you already have the data you'll need.



- 2.4.2 If you do NOT see the WEATHER schema, run the following SQL commands to create a schema and load the data into your default DEMO\_DB database.

```
-- Set the context
use role sysadmin;
use warehouse compute_wh;
use database demo_db;
create or replace schema weather;

-- Create the table
create or replace table daily_14_total (
    v variant,
    t timestamp);

-- Allow all users to access the table
grant usage on schema demo_db.weather to public;
grant select on table daily_14_total to public;

-- Define a stage that describes where the data will be loaded from
create or replace stage weather_data_s3
url = 's3://snowflake-corp-se-workshop/VHOL_Snowflake_Informatica/WEATHER/';

-- Re-size the warehouse so we can load the data quicker
alter warehouse compute_wh set warehouse_size = large;

-- Load the data
copy into daily_14_total
from (select $1, to_timestamp($1:time)
      from @weather_data_s3)
file_format = (type=json);
```

Results Data Preview

✓ Query ID SQL 21.52s 68 rows

Filter result...

| Row | file            | status | rows_parsed | rows_loaded | error_limit | errors_seen |
|-----|-----------------|--------|-------------|-------------|-------------|-------------|
| 1   | s3://snowfla... | LOADED | 6433        | 6433        | 1           | 0           |
| 2   | s3://snowfla... | LOADED | 6401        | 6401        | 1           | 0           |
| 3   | s3://snowfla... | LOADED | 9035        | 9035        | 1           | 0           |
| 4   | s3://snowfla... | LOADED | 11774       | 11774       | 1           | 0           |
| 5   | s3://snowfla... | LOADED | 11731       | 11731       | 1           | 0           |
| 6   | parquetfile     | LOADED | 11200       | 11200       | 1           | 0           |

-- IMPORTANT: Set the warehouse back to the original size  
 alter warehouse compute\_wh set warehouse\_size = xsmall;

#### 2.4.3 Let's see what the data looks like.

-- IF YOU ALREADY HAD THE WEATHER SCHEMA

use role sysadmin;

use schema snowflake\_sample\_data.weather;

-- Either way:

select \* from daily\_14\_total limit 10;

Results Data Preview ← Open History

✓ Query ID SQL 783ms 10 rows

Filter result...   Columns ▾

| Row | V   | T                       |
|-----|---|-------------------------|
| 1   | { "city": { "coord": { "lat": 8.8865, "lon": 2.5975 }, "country": "BJ", "id": 239... }  | 2021-02-23 10:33:55.000 |
| 2   | { "city": { "coord": { "lat": 9.25, "lon": 0.7833 }, "country": "TG", "id": 2367... }   | 2021-02-23 10:33:55.000 |
| 3   | { "city": { "coord": { "lat": 6.3193, "lon": 16.3799 }, "country": "CF", "id": 23... }  | 2021-02-23 10:33:55.000 |
| 4   | { "city": { "coord": { "lat": 14.1706, "lon": 120.9236 }, "country": "PH", "id": ... }  | 2021-02-10 10:34:40.000 |
| 5   | { "city": { "coord": { "lat": 15.7333, "lon": 120.5833 }, "country": "PH", "id": ... }  | 2021-02-10 10:34:40.000 |
| 6   | { "city": { "coord": { "lat": 10.747, "lon": 125.0107 }, "country": "PH", "id": 17... } | 2021-02-10 10:34:40.000 |

#### 2.4.4 Click on the first row of data to see the JSON. You can also pull out the individual JSON elements with a query such as:

```
select t, v:city.name::text, v:city.country::text, f1.value:temp.min, f1.value:temp.max,
       f2.value:description::text, f1.value:humidity
  from weather.daily_14_total
 , lateral flatten (input => v:data) f1
 , lateral flatten (input => f1.value:weather) f2
 limit 10;
```

Results Data Preview

✓ Query ID SQL 799ms 10 rows

Filter result...  Columns ▾

| Row | T                    | V:CITY.NAME::TEXT | V:CITY.COUNTRY::TEXT | F1.VALUE:TEMP.MIN | F1.VALUE:TEMP.MAX | F2.VALUE:DESCRIPTION | F1.VALUE:HUMIDITY |
|-----|----------------------|-------------------|----------------------|-------------------|-------------------|----------------------|-------------------|
| 1   | 2021-02-19 22:29:... | Paese             | IT                   | 279.33            | 285.03            | scattered clouds     | 65                |
| 2   | 2021-02-19 22:29:... | Paese             | IT                   | 280.57            | 285.57            | overcast clouds      | 69                |
| 3   | 2021-02-19 22:29:... | Paese             | IT                   | 281.69            | 286.08            | broken clouds        | 64                |
| 4   | 2021-02-19 22:29:... | Paese             | IT                   | 280.06            | 286.7             | broken clouds        | 67                |
| 5   | 2021-02-19 22:29:... | Paese             | IT                   | 280.69            | 289.87            | sky is clear         | 49                |
| 6   | 2021-02-19 22:29:... | Paese             | IT                   | 280.89            | 280.94            | clear skies          | 47                |

We'll work with this data later in the lab.

## 2.5 Get ready to run Snowflake Partner Connect to set up your Informatica Cloud Account

- 2.5.1 If you click on the top right of the UI where your user name appears, you will see that here you can do things like change your password, roles, or preferences. Snowflake has several system defined roles. You are currently in the default role of SYSADMIN. In order to use **Partner Connect** to create an Informatica account, we will need to change our role to the **ACCOUNTADMIN** role.

The screenshot shows the Snowflake interface with the user 'USER123' logged in. The top navigation bar includes links for Databases, Shares, Warehouses, Worksheets, History, Partner Connect, Help, and a user icon. The 'History' tab is selected. A dropdown menu for the user 'USER123' is open, showing options: Change Password, Switch Role, Preferences, and Log Out. The 'Switch Role' option is highlighted with a red box.

- 2.5.2 Hover over Switch Role, and Click on **ACCOUNTADMIN** to change your role.

The screenshot shows the 'Search 14 Roles' dropdown. It lists three roles: 'SYSADMIN (Default)', 'ACCOUNTADMIN', and 'ANALYST'. The 'ACCOUNTADMIN' role is highlighted with a blue box. To the right of the dropdown, a secondary menu is visible with options: Change Password, Switch Role, Preferences, and Log Out. The 'Switch Role' option is also highlighted with a blue box.

## Module 3: Create IICS Organization

In this lab, you will go to Partner Connect in Snowflake UI to create an IICS organization. The free IICS organization created through Snowflake Partner Connect enables you to load 1 billion records per month into Snowflake for free.

### Step 1

1. Login to **Snowflake UI**.
2. Click **Partner Connect**. Switch to **ACCOUNTADMIN** role if you're not already in that role.
3. Click **Informatica** tile.

The screenshot shows the Snowflake Partner Connect interface. At the top, there's a navigation bar with icons for Databases, Shares, Data Marketplace, Warehouses, Worksheets, History, Account, Preview App, Help, Notifications, and a user profile for EWIDJAJA ACCOUNTADMIN. The 'Partner Connect' button is highlighted with a red box. Below the navigation bar, the title 'Snowflake Partner Connect' is displayed, followed by a message: 'Get started with loading and analyzing your data in minutes. Automatically connect your Snowflake account with our partner applications available for a free trial.' A note says 'Check back often as we will be adding new partners regularly.' The main area contains a grid of partner tiles:

- Fivetran**: Built for analysts, 5-minute setup, great schemas, Snowflake platinum partner.
- Stitch**: Stitch moves data into Snowflake in minutes. Unlimited sources and a free-for-ever tier.
- Sigma**: Maximize Snowflake's value. Governed self-service analytics & BI for all. Faster insights.
- Sisense**: Empowering builders to simplify complex data and transform it into powerful analytic apps.
- snapLogic**: SnapLogic's platform empowers organizations with intelligent application and data integration.
- Rivery**: Rivery creates an automated data pipeline to collect & transform data from all sources.
- CHARTIO**: Connect to the #1 self-service data analytics platform for ease-of-use & speed to insights.
- MATILLION**: Snowflake data transformation. Achieve new levels of simplicity, speed, scale and savings.
- DATAGUISE**: Data security and privacy automation for compliance with GDPR, CCPA, PCI and more.
- zepl**: Zepl brings AI and analytics to your Snowflake data in minutes. Try it for free today.
- Qlik**: Qlik (Attunity) is the market leader in real-time change data capture and warehouse automation.
- ThoughtSpot**: Search & AI-driven Analytics on all your data.
- data.world**: Cloud-native data cataloging, metadata management, collaboration, and virtualization.
- Informatica**: Fast, Scalable, Trusted, AI-driven enterprise cloud data management for Snowflake. This tile is highlighted with a red border.
- Hunters.ai**: Autonomous threat hunting - Detect cyberattacks that bypass existing security controls.
- Datalytx**: CI/CD and DataOps for Snowflake. Truly agile data ingestion, modeling and transformation.
- Domo**: High-leverage BI on Snowflake with mobile-first, intelligent apps for business.
- SqlDBM**: Cloud Data Modeling for Snowflake. Develop DB/DW without writing a single line of code.

### Step 2

1. Note the objects that will be created in your Snowflake account.
2. Click **Connect**.

## Connect to Informatica

Informatica requires the following information to create your new trial account:

In order to configure the connection with Snowflake, the following objects will be created in your Snowflake account

Database : PC\_INFORMATICA\_DB  
Warehouse : PC\_INFORMATICA\_WH (X-Small)  
System User : PC\_INFORMATICA\_USER  
System Password : Autogenerated & Randomized  
System Role : PC\_INFORMATICA\_ROLE  
Role PUBLIC will be granted to the PC\_INFORMATICA\_ROLE  
Role PC\_INFORMATICA\_ROLE will be granted to the SYSADMIN role

By clicking on **Connect** you are instructing Snowflake to create the above objects and provide all of the above information to Informatica. Informatica's processing of this information, and your use of Informatica, are governed solely by Informatica's [Terms of Use](#) and Informatica's [Privacy Policy](#) and not your agreement with Snowflake.

Please contact [Snowflake Support](#) if you have questions about connecting with Informatica.

Cancel

Connect

## Step 3

1. Click **Activate** to proceed to the activation process.

### Your partner account has been created and is ready to be activated

Informatica has successfully created your new account. Click **Activate** to go to Informatica's website to finish the activation process and start loading data.

Activate Later

Activate

## Step 4

1. Fill in the form.
2. Select **Data Center** in your region.
3. Click **Submit**.

## Informatica Intelligent Cloud Services Accelerator for Snowflake

Informatica Intelligent Cloud Services, Informatica's Integration platform as a service (PaaS), reduces the time required to deploy and migrate data to Snowflake. You can quickly connect to cloud and on-premises data sources, profile data, and then integrate and stage the data for analytics.

Accelerate your move to Snowflake with Informatica Intelligent Cloud Services for Snowflake Accelerator. This free offer includes Informatica Cloud Data Integration Service:

- Automate and speed up your workloads to cloud data warehouses with out-of-the-box solution templates tailored for Snowflake.
- Replicate and synchronize bulk data, at scale, with the help of prebuilt wizards.
- Simplify complex data integration loads with codeless advanced integration, using a mapping designer with out-of-the-box advanced data integration transformations.
- Orchestrate multiple data integration tasks and mappings, run them in a non-linear parallel fashion, and perform advanced exception-handling and decision-making.
- Get intelligent structure discovery that leverages our machine learning engine to understand the structure of complex files and automatically infer the appropriate processing model.
- Enable the extraction and transfer of changed data with change data capture.
- Select 4 of the 20 of the most popular connectors for Snowflake.

### Informatica Cloud Mass Ingestion Service:

- Streamline the high-performance transfer of enterprise data assets in file format, securely and at scale, from on-premises and cloud sources to Snowflake.
- Transfer any size or supported type of file with high performance and scalability.

Get free usage of up to 1 billion rows per month.

For more information about how we support Snowflake, visit [informatica.com/snowflake](https://informatica.com/snowflake)

Please select below regions

Email

First Name  Last Name

Use my email address as my username

Data Centre

Password

Confirm Password

Company

Job Title

Select Country

State/Province

City

Phone

Address

Postal Code

I have read and agreed to the [Terms and Conditions](#).

**SUBMIT**

This site is protected by reCAPTCHA and the Google [Privacy Policy](#) and [Terms of Service](#) apply.

### Step 5

1. This page will automatically open up in your browser <https://marketplace.informatica.com/thank-you/snowflake.html>. Bookmark this page for future reference. Please also read through **Knowledge Base** materials and demo recording for more information.
2. Click the **region** you selected in the previous step to log in to IICS.

Search



## Informatica Intelligent Cloud Services Accelerator for Snowflake

Welcome to Informatica Intelligent Cloud Services for Snowflake Accelerator. To start your session, please select your region.

1. [North America](#)
2. [Europe](#)
3. [Asia Pacific](#)

*Note: The embedded walk-through tutorials are not available for EMEA or APAC regions.*

### Knowledge Base Materials

Below please find a few resources to ensure you have a successful experience.

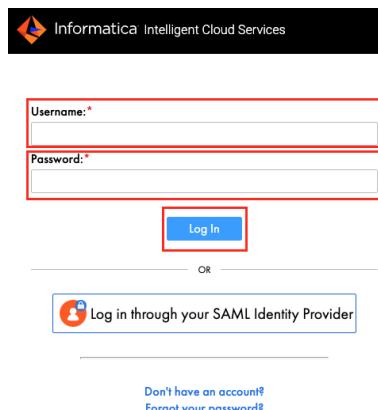
- [Getting Started Guide](#)
- [FAQs](#)
- [Snowflake Cloud Data Warehouse V2 Connector](#)

Demos: How to load data into Snowflake with prebuilt mappings

- [How to load Salesforce data into Snowflake](#)
- [How to move large amounts of data into Snowflake](#)

### Step 6

1. Enter your **username** and **password**.
2. Click **Log In**.



The screenshot shows a login interface for Informatica Intelligent Cloud Services. It features a black header bar with the Informatica logo and the text "Informatica Intelligent Cloud Services". Below this is a white form area with two red-outlined input fields: "Username:" and "Password:". A blue "Log In" button is centered below the fields. Below the button is a horizontal line with the word "OR" in the center. At the bottom of the form are two links: "Don't have an account?" and "Forgot your password?".

Copyright © 1993-2020 Informatica LLC. All Rights Reserved.  
[Privacy Policy](#)  
[Check the system status](#)

### Step 7

1. The first time logging in, you will be prompted to enter a security **question** and **answer**. Fill them in.
2. Click **Log In**.



**Set Up Your Security Question**

Select a security question. We will use the security question to verify your identity if you forget your password.

**Your User Name**

handsonlab1@infa.com

**Your Security Question**

In which city did you meet your spouse/significant other? ▾

**Your Security Question's Answer**

\*\*\*\*

**Log In**

Copyright © 1993-2020 Informatica LLC. All Rights Reserved.

[Check the system status](#)

----- Congratulations! You have successfully completed this lab -----

## Module 4: Configure an AWS S3 Connection

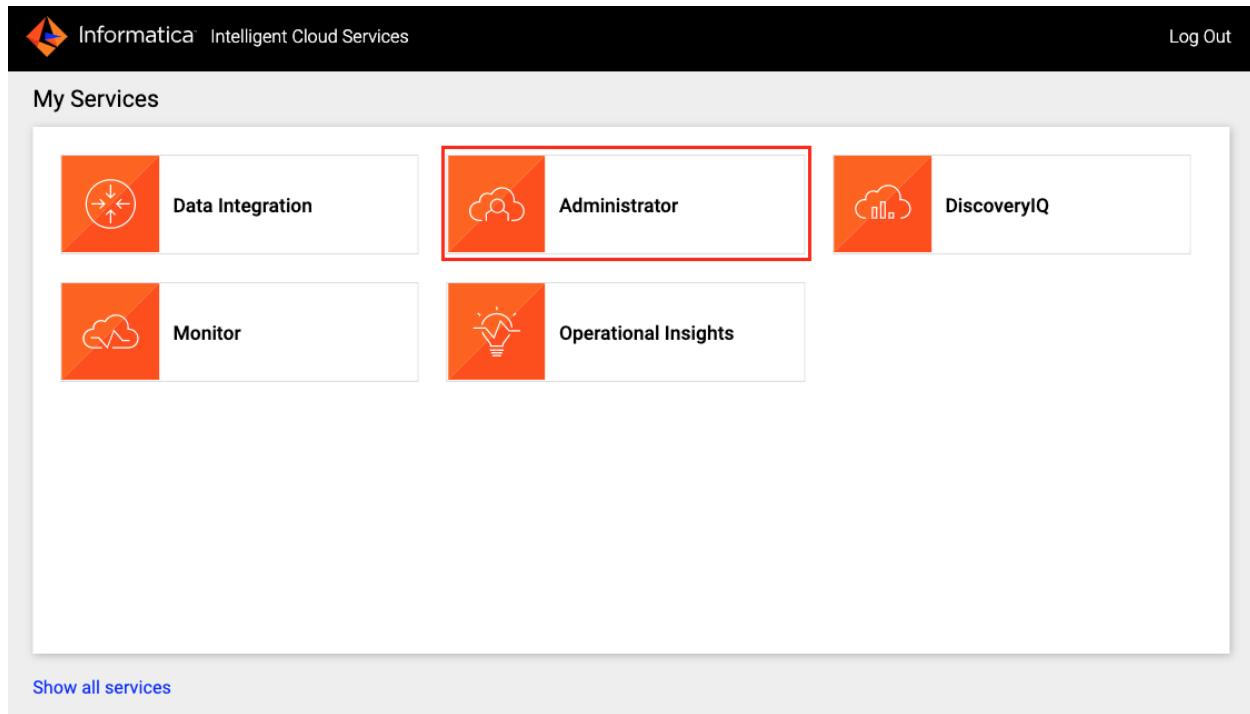
In this lab, you will create an AWS S3 connection. Connection is used to connect to data source and target. In the following lab, we will use this connection as a data source and write to a Snowflake table using a Snowflake connection.

The Snowflake connection is already configured by default to use the Snowflake objects when you signed up via the Informatica tile in Partner Connect.

Informatica develops and maintains connector code so our customers do not have to. Go to **Administrator > Add-on Connectors** to view available connectors.

### Step 1

1. Log in to IICS and navigate to the **Administrator** service.



### Step 2

1. Click **Connections** on the left panel
2. Click **New Connection** to create a connection.

The screenshot shows the Informatica Administrator interface. On the left sidebar, under the 'Connections' section, the 'Connections' button is highlighted with a red box. The main area is titled 'Connections' and contains a table with columns: Actions, Name, Type, and Runtime Environment. One row in the table is highlighted in blue and shows the connection 'Snowflake\_INFORMATICAPARTNER'.

| Actions | Name                         | Type  | Runtime Environment            |
|---------|------------------------------|---|--------------------------------|
|         | Snowflake_INFORMATICAPARTNER | Snowflake Cloud Data Warehouse V2 (Informatica Cloud) | Informatica Cloud Hosted Agent |

### Step 3

1. Enter **S3** in the Connection Name field.
2. Select **Amazon S3 v2** from the Type dropdown field.
3. Select **Informatica Cloud Hosted Agent** from the Runtime Environment dropdown field.
4. Enter **AKIA4LKFNB2T5EBYGVS5** in the Access Key field.
5. Enter **SuND78yc1oYYquDnAMLvZSKZ5WEk4xrI9bthPlqj** in the Secret Key field.
6. Enter **dataforingestion** in the Folder Path field.
7. Select **US West(Oregon)** from the Region Name dropdown field.
8. Click **Test Connection**. If the configuration is correct, the page should display "**The test for this connection was successful.**"
9. Click **Save**.

S3

Test Connection Save

✓ The test for this connection was successful.

**Connection Details**

|   |              |
|---|--------------|
| Connection Name: <sup>*</sup>                               | S3           |
| Description:  |              |
| Type: <sup>*</sup> <span style="font-size: small;">?</span> | Amazon S3 v2 |

**Amazon S3 v2 Properties** (?)

|  |                                |
|--|--------------------------------|
| Runtime Environment: <sup>*</sup> <span style="font-size: small;">?</span> | Informatica Cloud Hosted Agent |
|--|--------------------------------|

**Connection Section**

|  |                          |
|--|--------------------------|
| Access Key:  | *****                    |
| Secret Key:  | *****                    |
| IAM Role ARN:  |                          |
| External Id:   |                          |
| Use EC2 Role to Assume Role:                                       | <input type="checkbox"/> |
| Folder Path: <sup>*</sup> <span style="font-size: small;">?</span> | dataforingestion         |
| Master Symmetric Key:  |                          |
| Customer Master Key ID:  |                          |
| Region Name: <sup>*</sup> <span style="font-size: small;">?</span> | US West(Oregon)          |
| Federated SSO IdP:   | NONE                     |
| Other Authentication Type :  | NONE                     |

#### Step 4

1. You should now have an **Amazon S3** and **Snowflake** connections. The name of your Snowflake connection will be **Snowflake\_<account name>**.

Connections New Connection

Configure connections to work with applications, databases, and files.

| Actions | Name <span style="font-size: small;">▲</span> | Type                             | Runtime Environment            | Service URL |
|---------|---|----------------------------------|--------------------------------|-------------|
|         | <b>S3</b>                                     | Amazon S3 v2 (Informatica Cloud) | Informatica Cloud Hosted Agent |             |

----- Congratulations! You have successfully completed this lab -----

## Module 5: Load Data from AWS S3 into Snowflake

IICS Data Integration allows you to load source data from databases, applications, and data files in the cloud or on-premises into Snowflake. Data Integration supports many transformations that can be used to transform and enrich the source data. In addition, pushdown optimization (PDO) can be utilized for some transformations and functions to take advantage of Snowflake warehouse for data processing.

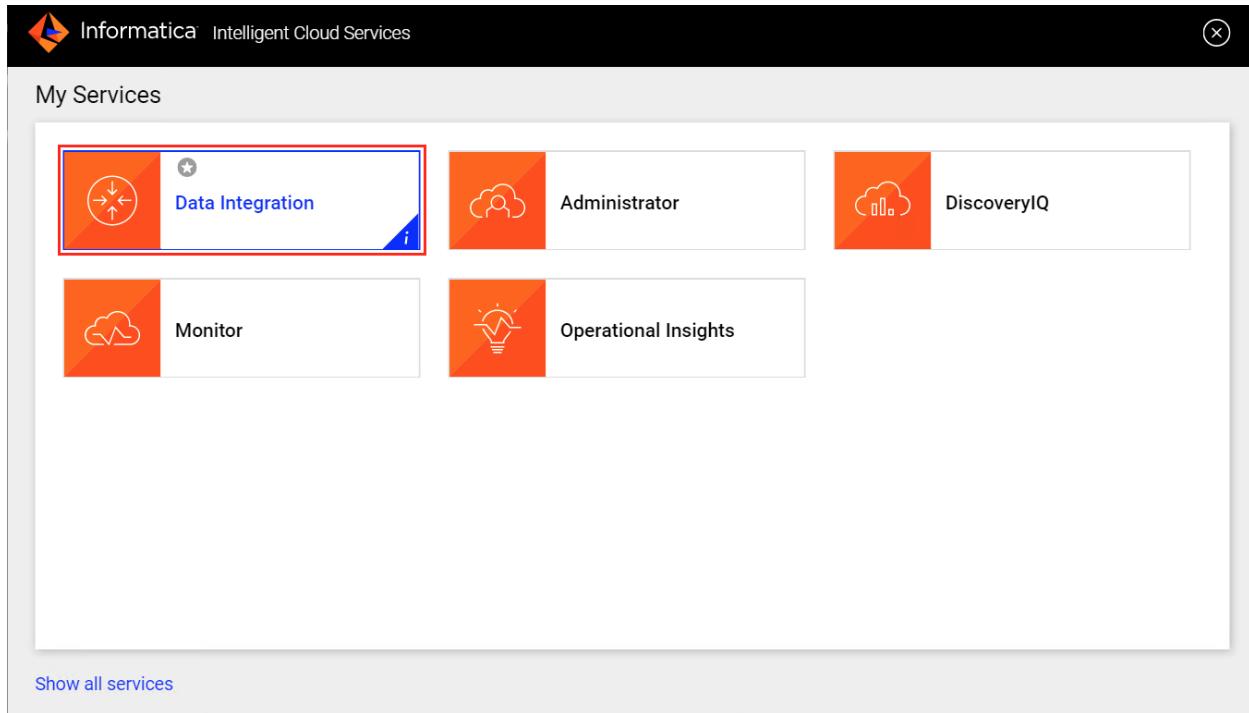
In this lab, you will create a mapping to read two delimited files, Orders and Lineitem, from S3, join the files, and aggregate to create a count and total, and write the results into a new table in Snowflake. Then in the mapping task, turn on pushdown optimization to enable Snowflake warehouse for processing.

### Prerequisite:

AWS S3 connection is created in Module 4.

### Step 1

1. Log in to Informatica Intelligent Cloud Services and navigate to the “**Data Integration**” Service.



### Step 2

Let's create a project/folder to store our new mapping.

1. Click option "**Explore**" on the left panel.
2. Click **New Project** to create a new project.

The screenshot shows the Informatica Data Integration application. The left sidebar has icons for 'New...', 'Home', 'Explore' (which is selected and highlighted with a red box), 'Bundles', 'My Jobs', and 'My Import/Export Logs'. The main area is titled 'Explore' and 'All Projects'. It shows two projects: 'Add-On Bund...' and 'Default'. A red box highlights the 'New Project' button at the top right of the project list.

3. Enter **Hands-on Lab** in Name field.
4. Click **Save**.

The screenshot shows the 'New Project Properties' dialog box. It has fields for 'Name:' (containing 'Hands-on Lab') and 'Description:' (empty). At the bottom are 'Save' and 'Cancel' buttons, with 'Save' highlighted with a red box.

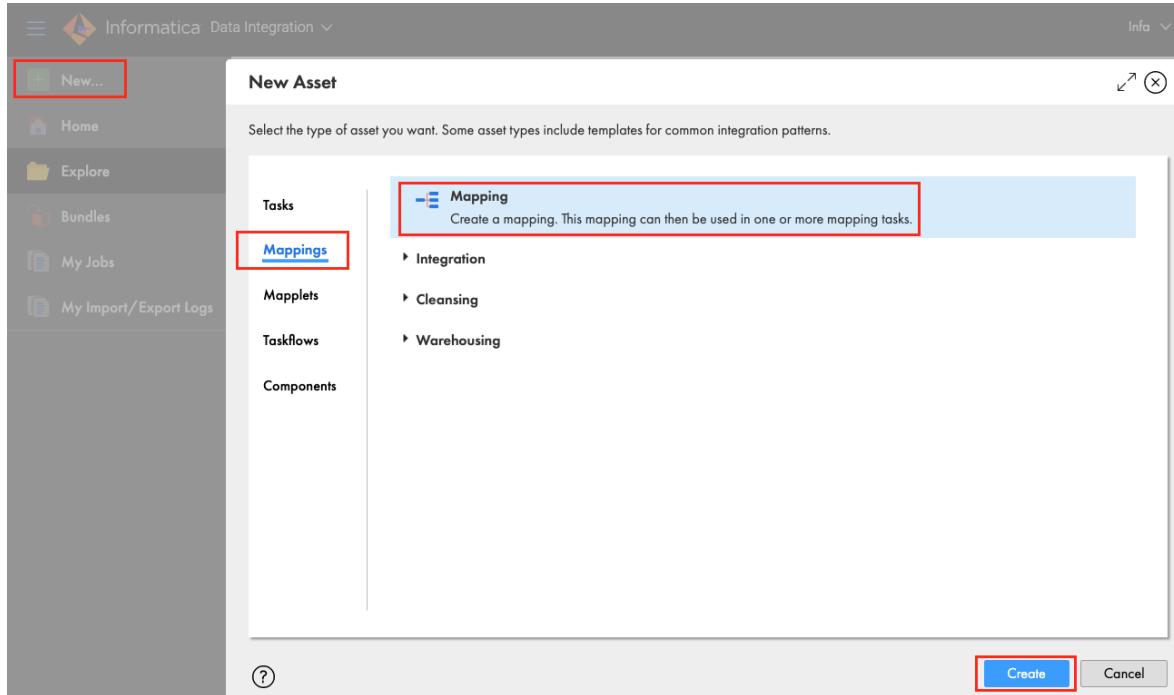
5. Click Hands-on Lab project.

The screenshot shows the Informatica Data Integration application. The left sidebar has icons for 'New...', 'Home', 'Explore' (selected and highlighted with a red box), 'Bundles', 'My Jobs', and 'My Import/Export Logs'. The main area is titled 'Explore' and 'All Projects'. It shows three projects: 'Add-On Bundles', 'Default', and 'Hands-on Lab' (which is highlighted with a red box).

### Step 3

Create a new mapping.

1. Click **New...**
2. Click **Mappings**.
3. Select **Mapping**.
4. Click **Create**.



5. Under properties, enter **m\_S3\_Orders\_Lineitem\_into\_Snowflake** in Name field.
6. Ensure Location is **Hands-on Lab**. If not, click **Browse** and select it.

**m\_S3\_Orders\_Lineitem\_to\_Snowflake\_PDO**

|           |   |
|-----------|---|
| Name:*    | <input type="text" value="m_S3_Orders_Lineitem_to_Snowflake_PDO"/>              |
| Location: | <input type="text" value="Hands-on Lab"/> <input type="button" value="Browse"/> |

#### Step 4

Let's configure the Orders data source from S3.

1. Click **Source** transform in the mapping canvas to assign its properties.
2. In General tab, enter **src\_S3\_Orders** in the Name field.

**Properties** Preview |  src\_S3\_Orders

- General** Name: \* **src\_S3\_Orders**
- Source Description:
- Fields
- Partitions

3. In Source tab, select **S3** in the Connection dropdown field.
4. Click **Select** to select a source file.

**Properties** Preview |  src\_S3\_Orders

- General**
- Source** Connection: **S3 (Amazon S3 v2)** **Select...**
- Fields
- Partitions

**Details**

5. Click on **dataforingestion** S3 bucket.
6. From the results on the right, select **orders.tbl** file.
7. Click **OK**.

**Select Source Object**

Select a source object, then click OK. You can also search for a source object.

| S3       |                  | orders.tbl{dataforingestion}     |              |              |
|----------|------------------|----------------------------------|--------------|--------------|
| Packages | Search...        | Select                           | Name         | Label        |
| /        | dataforingestion | <input type="radio"/>            | lineitem.tbl | lineitem.tbl |
|          |                  | <input checked="" type="radio"/> | orders.tbl   | orders.lbl   |

Displaying 2 matching object(s).

OK Cancel

8. Click Format dropdown field and select **Delimited**.
9. Click **Formatting Options**.

Properties Preview |  src\_S3\_Orders

General ▾ Details

**Source**

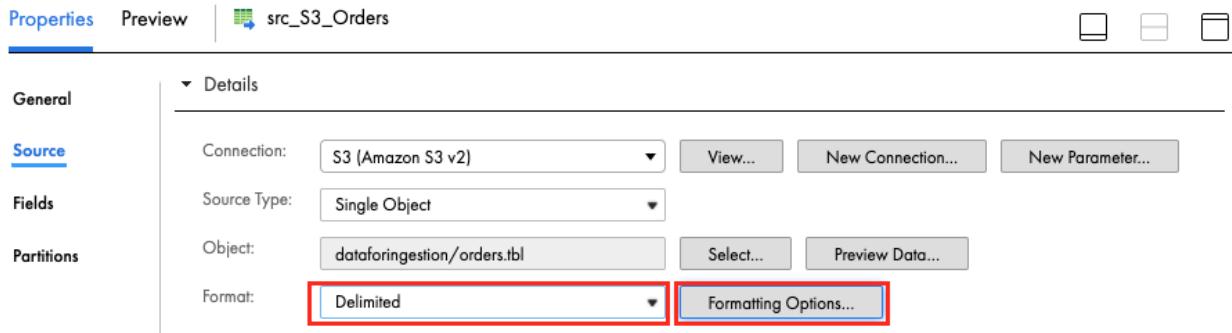
Connection: S3 (Amazon S3 v2) View... New Connection... New Parameter...

Fields

Object: dataforingestion/orders.tbl Select... Preview Data...

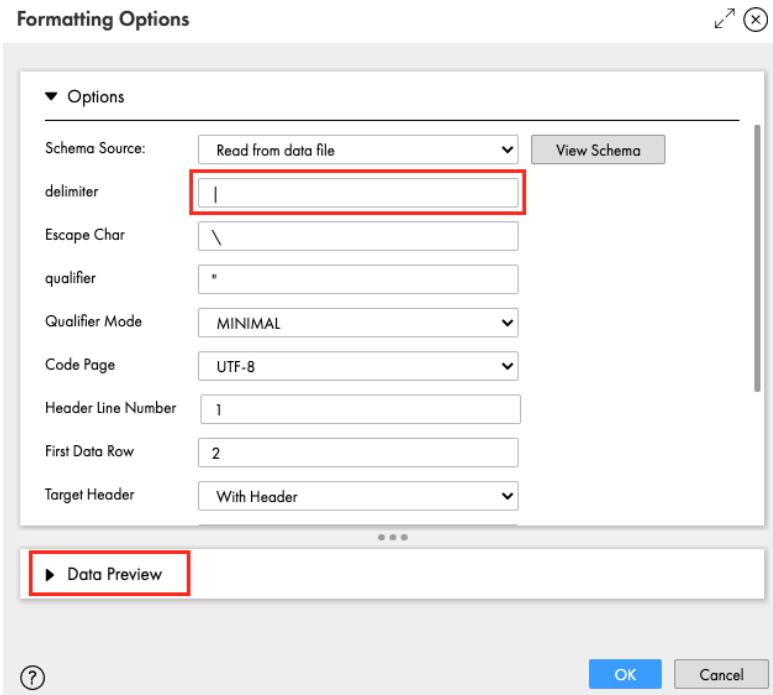
Partitions

Format: Delimited **Formatting Options...**

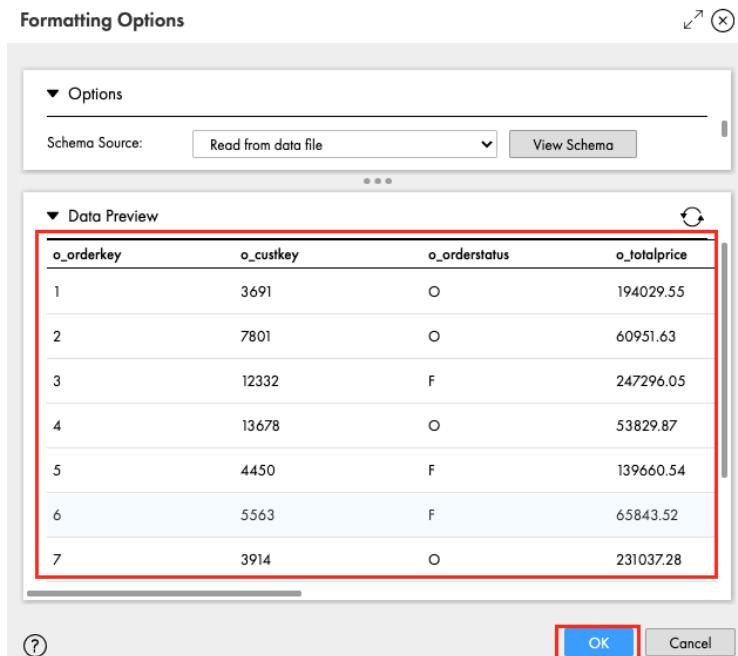


10. Enter a **vertical bar (|)** character in the delimiter field.

11. Click **Data Preview** to view the first 10 records.



12. Records should be separated by fields.



13. Click **OK**.
14. In Fields tab, select fields **7, 8, and 9**. Then click **trash** to remove those fields.
15. Click **Yes**.

|   | Name   | Type   | Precision | Scale | Origin                   |
|---|--|--------|-----------|-------|--------------------------|
| 1 | <input type="checkbox"/> o_orderkey                | string | 256       | 0     | dataforingestion/orde... |
| 2 | <input type="checkbox"/> o_custkey                 | string | 256       | 0     | dataforingestion/orde... |
| 3 | <input type="checkbox"/> o_orderstatus             | string | 256       | 0     | dataforingestion/orde... |
| 4 | <input type="checkbox"/> o_totalprice              | string | 256       | 0     | dataforingestion/orde... |
| 5 | <input type="checkbox"/> o_orderdate               | string | 256       | 0     | dataforingestion/orde... |
| 6 | <input type="checkbox"/> o_orderpriority           | string | 256       | 0     | dataforingestion/orde... |
| 7 | <input checked="" type="checkbox"/> o_clerk        | string | 256       | 0     | dataforingestion/orde... |
| 8 | <input checked="" type="checkbox"/> o_shippriority | string | 256       | 0     | dataforingestion/orde... |
| 9 | <input checked="" type="checkbox"/> o_comment      | string | 256       | 0     | dataforingestion/orde... |

16. Let's edit the **o\_totalprice** metadata so that it is a decimal field.
17. Click Options dropdown, select Edit Metadata.
18. Click flat\_string Native Type field for o\_totalprice and select **flat\_number**.
19. Change the Type to **decimal**.
20. Change the Native Precision and Precision to **38**.
21. Change the Native Scale and Scale to **2**.

Properties Preview | src\_S3\_Orders

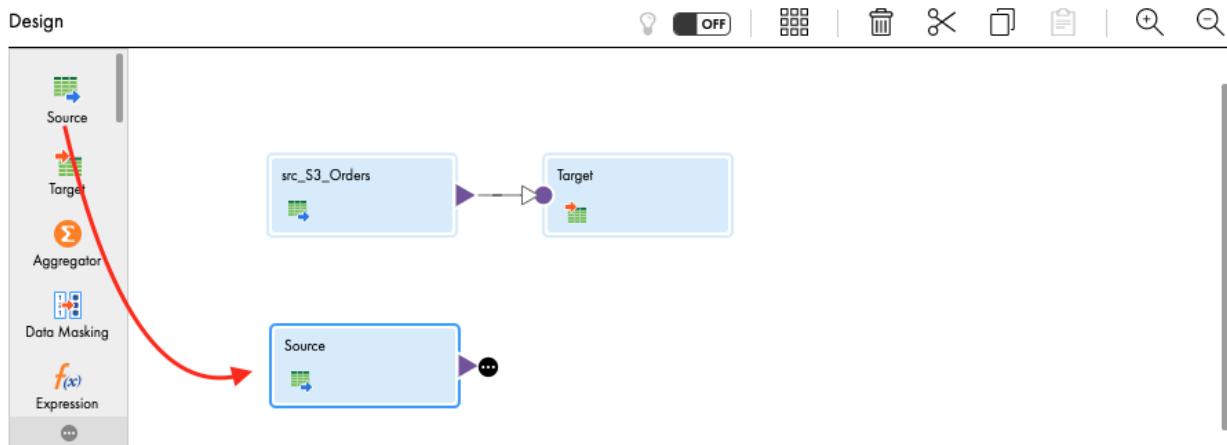
| General    | Fields      |               |               |             |                  |           |                 | Options |
|------------|-------------|---------------|---------------|-------------|------------------|-----------|-----------------|---------|
| Source     | Native Name | Name          | Native Type   | Type        | Native Precision | Precision | Native Scale    |         |
| Fields     | 1           | o_orderkey    | o_orderkey    | flat_string | string           | 256       | 256             | 0       |
| Partitions | 2           | o_custkey     | o_custkey     | flat_string | string           | 256       | 256             | 0       |
|            | 3           | o_ordersta... | o_ordersta... | flat_string | string           | 256       | 256             | 0       |
|            | 4           | o_totalprice  | o_totalprice  | flat_number | decimal          | 38        | 38              | 2       |
|            | 5           | o_orderdate   | o_orderdate   | flat_string | string           | 256       | 256             | 0       |
|            | 6           | o_orderpri... | o_orderpri... | flat_string | string           | 256       | 256             | 0       |
|            |             |               |               |             |                  | 0         | dataforinges... |         |
|            |             |               |               |             |                  | 0         | dataforinges... |         |

22. Click **Save** to periodically save work in progress.

### Step 5

Now we will add the Lineitem file as another data source. The steps are the same as the above Orders data source.

- From the transformation palette, drag **Source** transform and drop in the mapping canvas.



- Let's assign its properties.
- In General tab, enter **src\_S3\_Lineitem** in the Name field.
- In Source tab, select **S3** in the Connection dropdown field.
- Click **Select** to select a source file.
- Click on **dataforingestion** S3 bucket.
- From the results on the right, select **lineitem.tbl** file.
- Click **OK**.
- Click Format dropdown field and select **Delimited**.
- Click **Formatting Options**.
- Enter a **vertical bar** character in the delimiter field.
- Click **Data Preview** to view the first 10 records.
- Records should be separated by fields.
- Click **OK**.
- In Fields tab, remove all fields except **I\_orderkey, I\_extendedprice, I\_discount, I\_tax**.
- Click **Yes**.

## General

## Source

## Fields

## Partitions

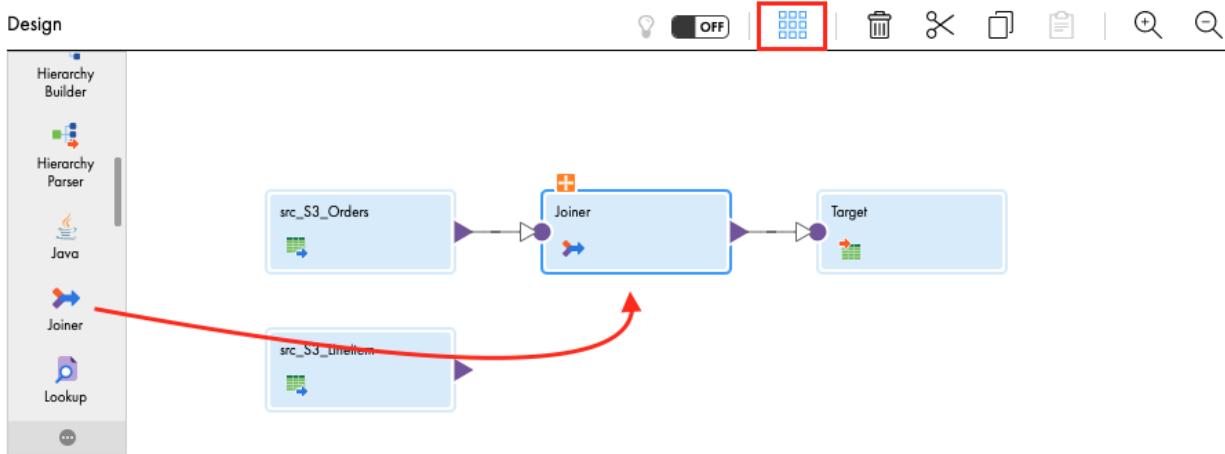
| Fields |  |        |           |       | <input style="border: none; background-color: transparent; font-size: small; margin-right: 10px;" type="button" value="+"/> | <input style="border: none; background-color: transparent; font-size: small; margin-right: 10px;" type="button" value="Delete"/> | <input style="border: none; background-color: transparent; font-size: small; margin-right: 10px;" type="button" value="Up"/> | <input style="border: none; background-color: transparent; font-size: small; margin-right: 10px;" type="button" value="Find"/> | <input style="border: none; background-color: transparent; font-size: small; margin-right: 10px;" type="button" value="Options"/> | <input style="border: none; background-color: transparent; font-size: small; margin-right: 10px;" type="button" value="Refresh"/> |
|--------|--|--------|-----------|-------|---|--|--|--|---|---|
|        | <input checked="" type="checkbox"/> Name | Type   | Precision | Scale | Origin  |  |  |  |   |   |
| 1      | <input type="checkbox"/> l_orderkey      | string | 256       | 0     | dataforingestion/lineit...  |  |  |  |   |   |
| 2      | <input type="checkbox"/> l_extendedprice | string | 256       | 0     | dataforingestion/lineit...  |  |  |  |   |   |
| 3      | <input type="checkbox"/> l_discount      | string | 256       | 0     | dataforingestion/lineit...  |  |  |  |   |   |
| 4      | <input type="checkbox"/> l_tax           | string | 256       | 0     | dataforingestion/lineit...  |  |  |  |   |   |

24. Click **Save** to periodically save work in progress.

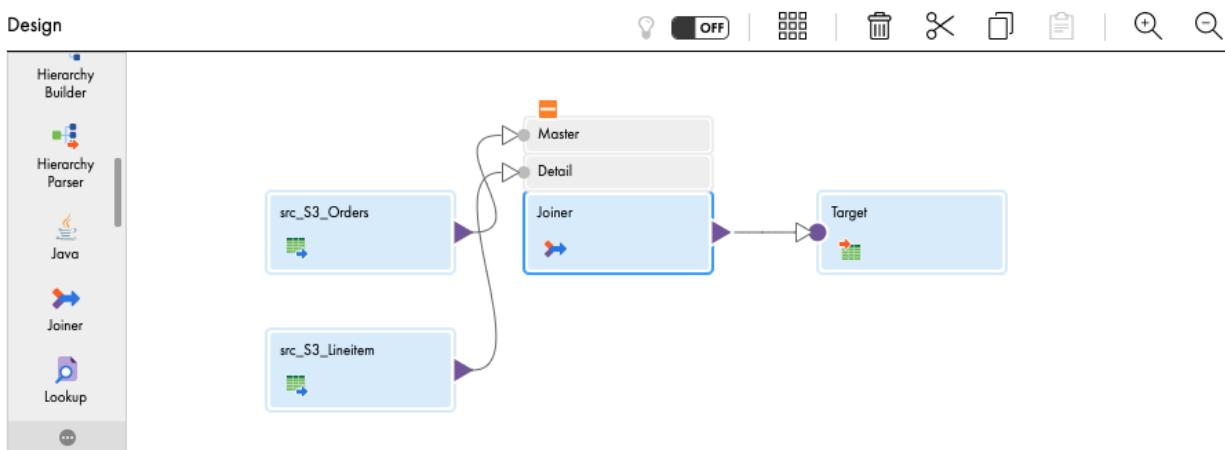
**Step 6**

Let's join the two data sources.

- From the transformation palette, drag **Joiner** transform and drop it over the line between the src\_S3\_Orders source and target transforms. The Joiner should now be linked to the Orders and target. If not, manually link them.
- Click align icon  to align transformations in the mapping canvas.



- Click the plus icon  above Joiner to expand.
- Link src\_S3\_Lineitem to the **Detail** of Joiner transform.



- Let's assign the Joiner properties.
- In General tab, enter **jnr\_orders\_lineitem** in the Name field.
- In Join Condition tab, click the plus icon  $\oplus$  to add a new condition.
- Select **o\_orderkey** for Master and **l\_orderkey** for Detail.

**Properties** Preview | jnr\_orders\_lineitem

**General** Join Type: Normal Master: src\_S3\_Orders Detail: src\_S3\_Lineitem

**Incoming Fields** Join Condition: Simple

**Join Condition**

**Advanced**

| Master              | Operator | Detail              |
|---------------------|----------|---------------------|
| o_orderkey (string) | =        | l_orderkey (string) |

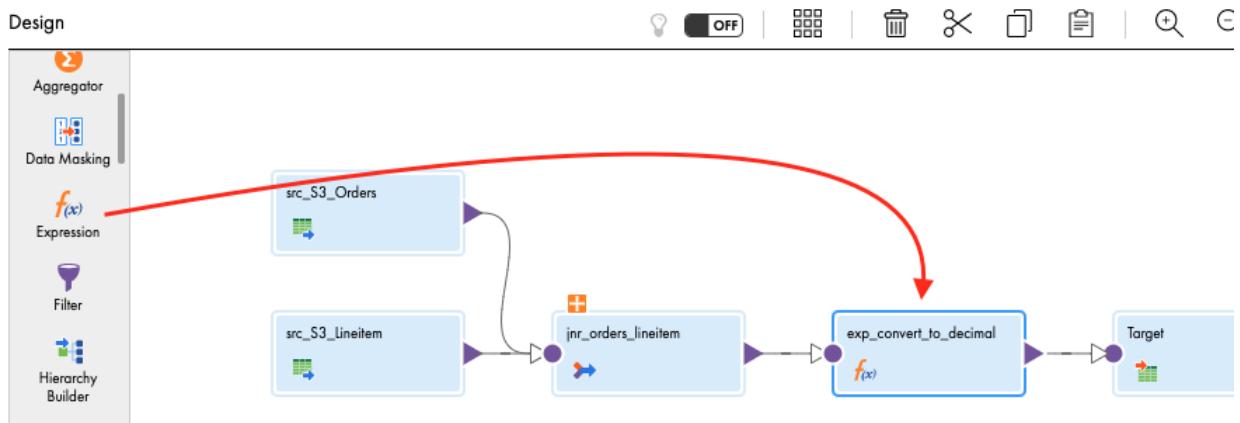
( $\oplus$ )

- Click **Save** to periodically save work in progress.

### Step 7

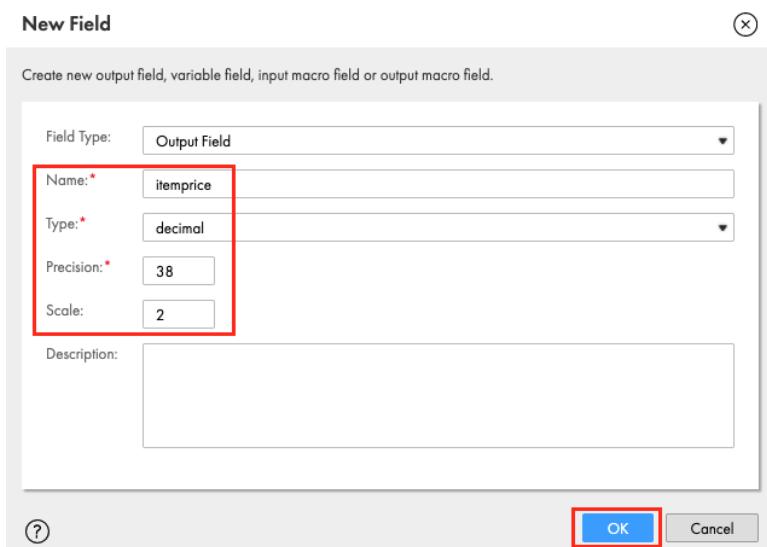
Let's add an Expression to convert input fields from string to decimal data type.

- From the transformation palette, drag **Expression** transform and drop it over the line between the **jnr\_orders\_lineitem** and target transforms. The expression should now be linked to the Joiner and Target. If not, manually link them.
- Click align icon to align transformations in the mapping canvas.



- Let's assign the properties.
- In General tab, enter **exp\_convert\_to\_decimal** in the Name field.
- In Expression tab, click the plus icon  $\oplus$  to add a new field.
- Enter **itemprice** in the Name field.
- Select **decimal** in the Type dropdown field.
- Enter **38** in the Precision field.
- Enter **2** in the Scale field.

10. Click **OK**.



11. Click **Configure** to configure the expression.

The screenshot shows the 'Properties' tab for the 'exp\_convert\_to\_decimal' field. The 'Expression' tab is selected. On the left, there are tabs for 'Properties', 'Preview', and 'fx' (expression). Below the tabs, there are sections for 'General', 'Incoming Fields', 'Expression' (which is selected), 'Window', and 'Advanced'. In the 'Expression' section, there is a note: 'Create simple expressions. You can also use expression macros to create complex expressions.' Below this is a checkbox for 'Allow additional fields and expressions during task creation'. Under 'Expressions', there is a table:

| Field Name | Expression   | Field Description |
|------------|--------------|-------------------|
| itemprice  | Configure... |                   |

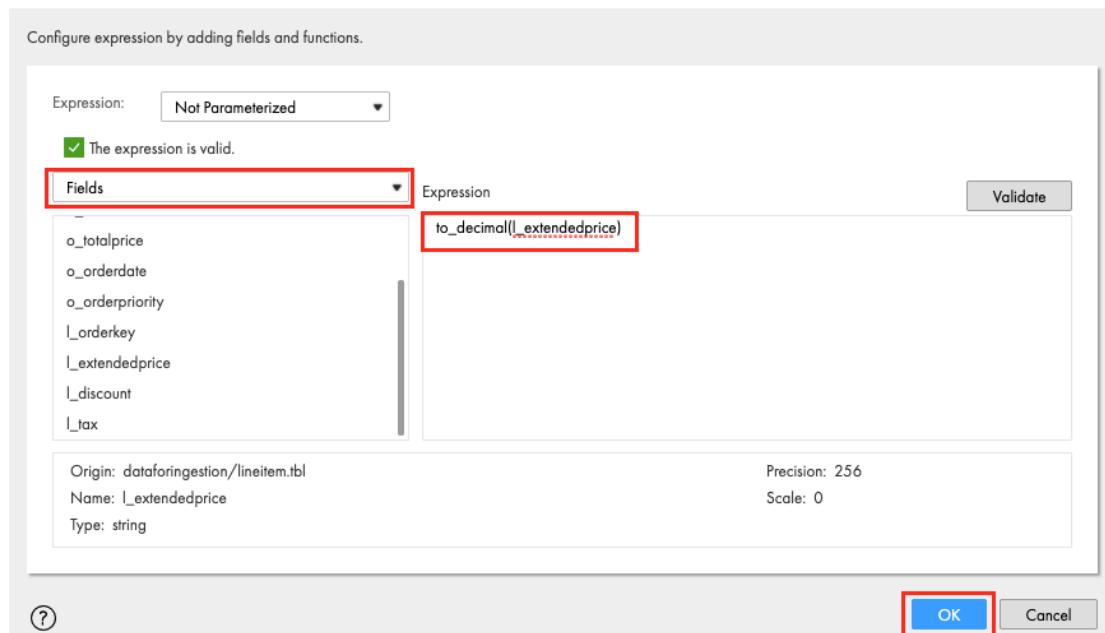
The 'Configure...' button in the 'Expression' column is highlighted with a red box.

12. Enter **to\_decimal(l\_extendedprice)** in the Expression field. Click Fields dropdown list to view available System Variables, Parameters, and Functions.

13. Click **Validate**.

14. Click **OK**.

### Field Expression: itemprice( decimal, 38, 2 )



15. Click **OK**.

16. Repeat no. 5-14 to add the following fields and expressions:

- Field name: **discountpercentage**; Expression: **to\_decimal(l\_discount)**
- Field name: **taxpercentage**; Expression: **to\_decimal(l\_tax)**

17. When completed, your Expression tab properties should look like this:

Properties Preview | exp\_convert\_to\_decimal

**General** Create simple expressions. You can also use expression macros  
 Allow additional fields and expressions during task creation

**Incoming Fields**

**Expression**

| Field Name         | Expression                               |
|--------------------|--|
| itemprice          | <code>to_decimal(l_extendedprice)</code> |
| discountpercentage | <code>to_decimal(l_discount)</code>      |
| taxpercentage      | <code>to_decimal(l_tax)</code>           |

**Window**

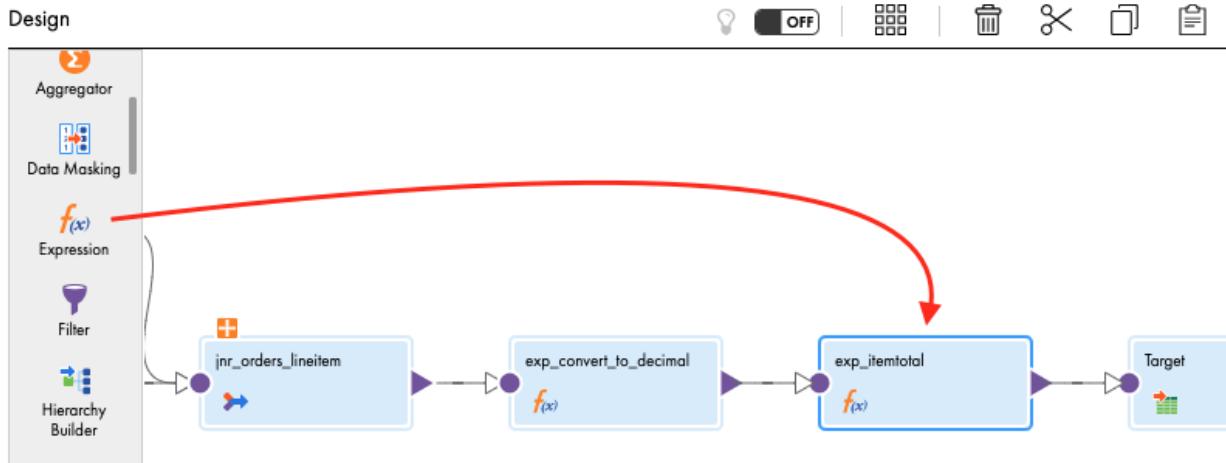
**Advanced**

18. Click **Save** to periodically save work in progress.

### Step 8

Let's add another Expression transform to calculate the item price.

- From the transformation palette, drag **Expression** transform and drop it over the line between the **exp\_convert\_to\_decimal** and target transforms. The expression should now be linked to the expression and Target. If not, manually link them.
- Click align icon to align transformations in the mapping canvas.



3. Let's assign the properties.
4. In General tab, enter **exp\_itemtotal** in the Name field.
5. In Expression tab, click the plus icon  $\oplus$  to add a new field.
6. Enter **itemtotal** in the Name field.
7. Select **decimal** in the Type dropdown field.
8. Enter **38** in the Precision field.
9. Enter **2** in the Scale field.
10. Click **OK**.
11. Click **Configure** to configure the expression.
12. Enter **itemprice \* (1-discountpercentage) \* (1+taxpercentage)** in the Expression field.
13. Click **Validate**.
14. Click **OK**.
15. When completed, your Expression tab properties should look like this:

**Properties** Preview | **f(x)** **exp\_itemtotal**

---

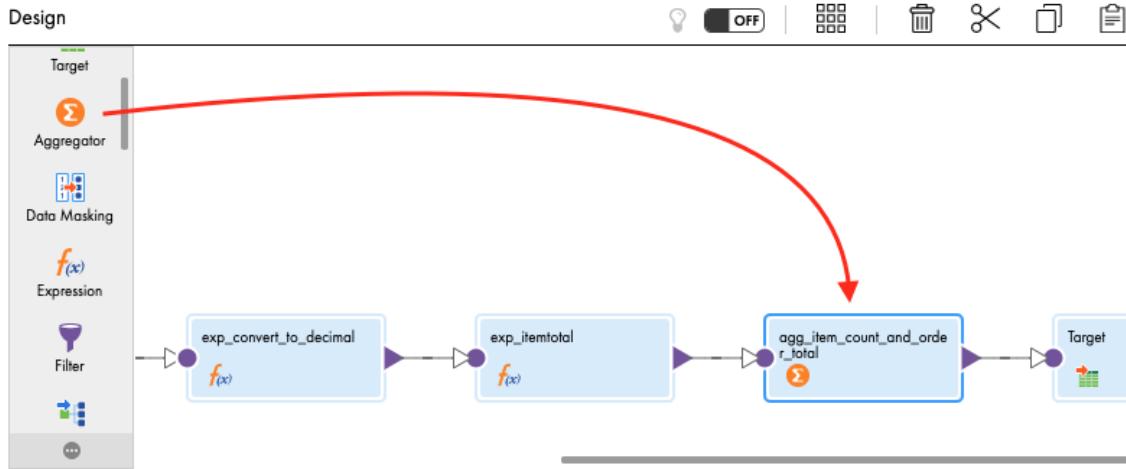
| <b>General</b>         | Create simple expressions. You can also use expression macros to create complex expressions.<br><input type="checkbox"/> Allow additional fields and expressions during task creation  |            |            |          |                  |   |  |
|------------------------|--|------------|------------|----------|------------------|---|--|
| <b>Incoming Fields</b> |  |            |            |          |                  |   |  |
| <b>Expression</b>      | <b>Expressions</b>   |            |            |          |                  |   |  |
| <b>Window</b>          | <table border="1"> <thead> <tr> <th>Field Name</th> <th>Expression</th> <th>Field De</th> </tr> </thead> <tbody> <tr> <td><b>itemtotal</b></td> <td><b>itemprice * (1-discountpercentage) * (1+taxpercentage)</b></td> <td></td> </tr> </tbody> </table> | Field Name | Expression | Field De | <b>itemtotal</b> | <b>itemprice * (1-discountpercentage) * (1+taxpercentage)</b> |  |
| Field Name             | Expression   | Field De   |            |          |                  |   |  |
| <b>itemtotal</b>       | <b>itemprice * (1-discountpercentage) * (1+taxpercentage)</b>  |            |            |          |                  |   |  |
| <b>Advanced</b>        |  |            |            |          |                  |   |  |

16. Click **Save** to periodically save work in progress.

### Step 9

Now we will add an Aggregator transformation in the mapping to calculate the number of items for an order and the total of all items.

1. From transformation palette, select **Aggregator** transformation, drag and drop between the **exp\_itemtotal** and **Target** in mapping canvas window.
2. Click align icon to align transformations in the mapping canvas.



3. Let's assign the properties.
4. In General tab, enter **agg\_item\_count\_and\_order\_total** in the Name field.
5. In Group By tab, click the plus icon  $\oplus$  to add new fields.
6. Add the following fields:
  - a. **o\_orderkey**
  - b. **o\_custkey**
  - c. **o\_orderstatus**
  - d. **o\_totalprice**
  - e. **o\_orderdate**
  - f. **o\_orderpriority**
7. When completed, the Group By tab properties should look like this:

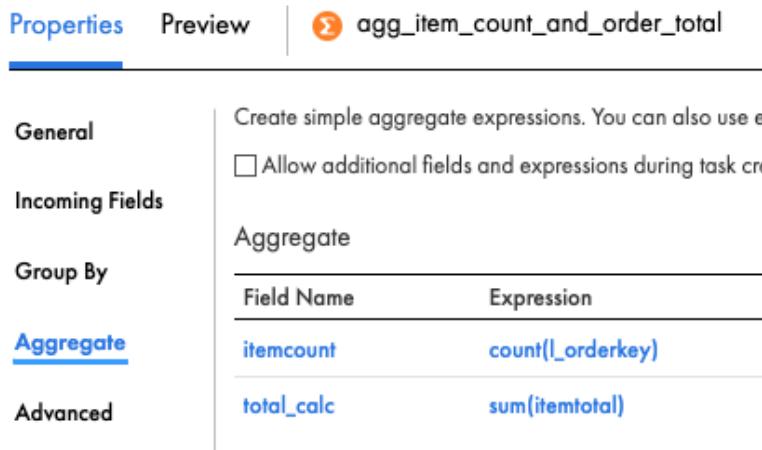
**Properties** Preview | **agg\_item\_count\_and\_order\_total**

---

|                        |   |
|------------------------|---|
| <b>General</b>         | Select fields to group by so that the resulting records can be set of rows with the : |
| <b>Incoming Fields</b> | Group by: Not Parameterized   |
| <b>Group By</b>        | Group by Fields $\oplus \ominus \odot$  |
| <b>Aggregate</b>       | Field Name<br>o_orderkey  |
| <b>Advanced</b>        | o_custkey<br>o_orderstatus<br>o_totalprice<br>o_orderdate<br>o_orderpriority          |

8. In Aggregate tab, click the plus icon  $\oplus$  to add a new field.
9. Enter **itemcount** in the Name field.
10. Select **integer** in the Type dropdown field.
11. Enter **10** in the Precision field.
12. Enter **0** in the Scale field.
13. Click **OK**.
14. Click **Configure** to configure the expression.

15. Enter **count(l\_orderkey)** in the Expression field. This function will result in the total number of items in an order.
16. Click **Validate**.
17. Click **OK**.
18. Click the plus icon  to add another new field.
19. Enter **total\_calc** in the Name field.
20. Select **decimal** in the Type dropdown field.
21. Enter **38** in the Precision field.
22. Enter **2** in the Scale field.
23. Click **OK**.
24. Click **Configure** to configure the expression.
25. Enter **sum(itemtotal)** in the Expression field. This function will add the total of all items in an order.
26. Click **Validate**.
27. Click **OK**.
28. When completed, your Expression tab properties should look like this:



**Properties**   Preview |  agg\_item\_count\_and\_order\_total

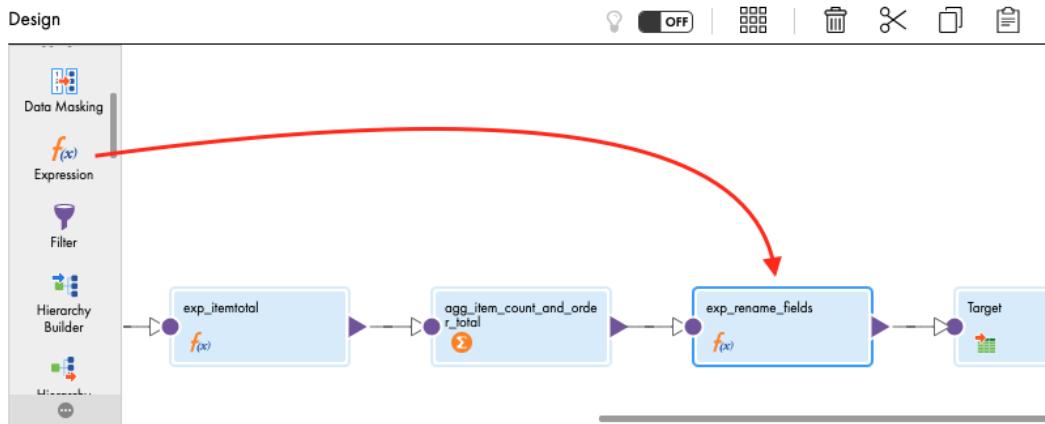
|                  | Field Name   | Expression |            |            |           |                   |            |                |
|------------------|--|------------|------------|------------|-----------|-------------------|------------|----------------|
| General          | Create simple aggregate expressions. You can also use e<br><input type="checkbox"/> Allow additional fields and expressions during task creation   |            |            |            |           |                   |            |                |
| Incoming Fields  |  |            |            |            |           |                   |            |                |
| Group By         |  |            |            |            |           |                   |            |                |
| <u>Aggregate</u> | <table border="1"> <thead> <tr> <th>Field Name</th> <th>Expression</th> </tr> </thead> <tbody> <tr> <td>itemcount</td> <td>count(l_orderkey)</td> </tr> <tr> <td>total_calc</td> <td>sum(itemtotal)</td> </tr> </tbody> </table> |            | Field Name | Expression | itemcount | count(l_orderkey) | total_calc | sum(itemtotal) |
| Field Name       | Expression   |            |            |            |           |                   |            |                |
| itemcount        | count(l_orderkey)  |            |            |            |           |                   |            |                |
| total_calc       | sum(itemtotal)   |            |            |            |           |                   |            |                |
| Advanced         |  |            |            |            |           |                   |            |                |

29. Click **Save** to periodically save work in progress.

### Step 10 - Optional

Now we will add another expression to rename the fields so that they look better and are in the order we want in the Snowflake table. This is an optional transformation.

1. From the transformation palette, drag **Expression** transform and drop it over the line between the **agg\_item\_count\_and\_order\_total** and target transforms. The expression should now be linked to the aggregator and Target transforms. If not, manually link them.
2. Click align icon  to align transformations in the mapping canvas.



3. Let's assign the properties.
4. In General tab, enter **exp\_rename\_fields** in the Name field.
5. In Expression tab, click the plus icon  $\oplus$  to add the following:

| Field Name    | Type    | Precision | Scale | Expression      |
|---------------|---------|-----------|-------|-----------------|
| orderkey      | string  | 256       | 0     | o_orderkey      |
| custkey       | string  | 256       | 0     | o_custkey       |
| orderdate     | string  | 256       | 0     | o_orderdate     |
| orderpriority | string  | 256       | 0     | o_orderpriority |
| orderstatus   | string  | 256       | 0     | o_orderstatus   |
| totalprice    | decimal | 38        | 2     | o_totalprice    |

6. When completed, your Expression tab properties should look like this:

The screenshot shows the 'Properties' tab for the 'exp\_rename\_fields' node. The 'Expression' tab is selected. It contains a table with six rows, each mapping an incoming field to an output expression:

| Field Name    | Expression      |
|---------------|-----------------|
| orderkey      | o_orderkey      |
| custkey       | o_custkey       |
| orderdate     | o_orderdate     |
| orderpriority | o_orderpriority |
| orderstatus   | o_orderstatus   |
| totalprice    | o_totalprice    |

7. Click **Save** to periodically save work in progress.

### Step 11

Lastly the target table is going to be in Snowflake.

1. Click **Target** to set a target properties.
2. In General tab, enter **tgt\_Snowflake** in the Name field.
3. In Incoming Fields tab, double click on **All Fields** under the **Field Selection Criteria**.

- Click off of the select box (otherwise the **Configure** option will not appear)
- Click **Configure**.

The screenshot shows the 'Properties' tab selected in the top navigation bar. Below it, a table titled 'Incoming Fields' is displayed under the 'Field Rules' section. The table has four columns: 'Operator' (containing 'Include'), 'Field Selection Criteria' (containing 'Named Fields'), 'Detail' (containing a red box around 'Configure...'), and 'Actions'. A red box highlights the 'Incoming Fields' tab in the left sidebar. A note above the table states: 'All incoming fields are included by default. You can configure field rules to exclude incoming fields. You can also rename incoming fields to avoid field name conflicts. The Preview Fields table lists all included and excluded fields.' There are three icons in the top right corner: a document, a grid, and a refresh.

- Select the following:

- itemcount
- o\_orderkey
- o\_custkey
- o\_orderdate
- o\_orderpriority
- o\_orderstatus
- o\_totalprice
- total\_calc

- Select **Rename Fields** at the top. To rename a field click next to it under the **Rename** column. Rename the following:

- |                    |                  |
|--------------------|------------------|
| a. o_orderkey      | -> orderkey      |
| b. o_custkey       | -> custkey       |
| c. o_orderdate     | -> orderdate     |
| d. o_orderpriority | -> orderpriority |
| e. o_orderstatus   | -> orderstatus   |
| f. o_totalprice    | -> totalprice    |

## Configure Field Rules



Rule Details **Rename Fields**

Rename selected fields:  Bulk  Individual

| Field Name      | Rename        | Origin                          |
|-----------------|---------------|---------------------------------|
| itemcount       |               | Σ aggr_item_count_and_order_... |
| o_custkey       | custkey       | dataforingestion/orders.tbl     |
| o_orderdate     | orderdate     | dataforingestion/orders.tbl     |
| o_orderpriority | orderpriority | dataforingestion/orders.tbl     |

8. When completed, the Incoming Fields tab should look like this:

Properties Preview | tgt\_Snowflake

General All incoming fields are included by default. You can configure field rules to exclude incoming fields. You can also rename incoming fields to avoid field name conflicts. The Preview Fields table lists all included and excluded fields.

**Incoming Fields**

▼ Field Rules (+)

| Operator | Field Selection Criteria | Detail                                 | Actions |
|----------|--------------------------|--|---------|
| Include  | Named Fields             | Included: 7 Fields   Renamed: 5 fields |         |

Target

Target Fields

Field Mapping

**Included Fields**    **Excluded Fields**

| Field Name    | Type    | Precision | Scale | Origin                 |
|---------------|---------|-----------|-------|------------------------|
| custkey       | string  | 256       | 0     | dataforingestion/or... |
| itemcount     | integer | 10        | 0     | Σ aggr_item_count_a... |
| orderdate     | string  | 256       | 0     | dataforingestion/or... |
| orderpriority | string  | 256       | 0     | dataforingestion/or... |
| orderstatus   | string  | 256       | 0     | dataforingestion/or... |
| total_col     | decimal | 38        | 2     | Σ aggr_item_count_a... |
| totalprice    | decimal | 38        | 0     | dataforingestion/or... |

9. Click **Target** on the left-hand side, and select the connection prefixed with **Snowflake**.

Properties Preview | tgt\_Snowflake

**General**

**Incoming Fields**

**Target** (highlighted with a red box)

**Target Fields**

**Field Mapping**

**Details**

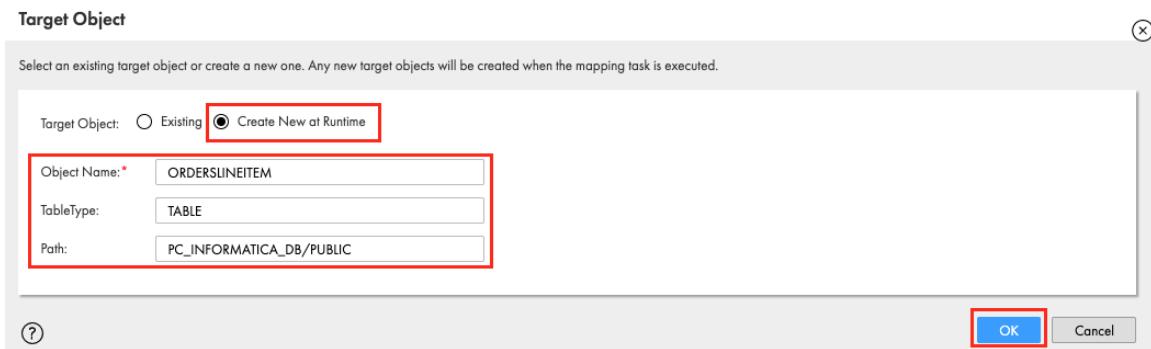
Connection: Snowflake informaticapartner (Snowflake...) ▾ View... New Connection... New Parameter...

Target Type: Single Object ▾

Object: Click Select... **Select...** (highlighted with a red box) Preview Data...

Operation: Insert ▾

10. For the **Object**, click on **Select**.
11. Select **Create New at Runtime** for Target Object.
12. Enter **ORDERSLINEITEM** in Object Name field.
13. Enter **TABLE** in the TableType field.
14. Enter **PC\_INFORMATICA\_DB/PUBLIC** in Path field.



15. The Target Fields tab should look like this:

Properties Preview | tgt\_Snowflake

**General**

**Incoming Fields**

**Target**

**Target Fields** (highlighted with a red box)

**Field Mapping**

The following target fields are included in the data flow.

|   | Name          | Type    | Precision | Scale | Origin        | Is Nullable                         | Default Column Value |
|---|---------------|---------|-----------|-------|---------------|-------------------------------------|----------------------|
| 1 | orderkey      | string  | 256       | 0     | PC_INFORMA... | <input checked="" type="checkbox"/> |                      |
| 2 | custkey       | string  | 256       | 0     | PC_INFORMA... | <input checked="" type="checkbox"/> |                      |
| 3 | orderdate     | string  | 256       | 0     | PC_INFORMA... | <input checked="" type="checkbox"/> |                      |
| 4 | orderpriority | string  | 256       | 0     | PC_INFORMA... | <input checked="" type="checkbox"/> |                      |
| 5 | orderstatus   | string  | 256       | 0     | PC_INFORMA... | <input checked="" type="checkbox"/> |                      |
| 6 | totalprice    | decimal | 38        | 2     | PC_INFORMA... | <input checked="" type="checkbox"/> |                      |
| 7 | itemcount     | integer | 10        | 0     | PC_INFORMA... | <input checked="" type="checkbox"/> |                      |
| 8 | total_calc    | decimal | 38        | 2     | PC_INFORMA... | <input checked="" type="checkbox"/> |                      |

16. The Field Mapping tab should look like this:

Properties Preview tgt\_Snowflake

**General**

Fields will be mapped automatically, as you have chosen to create a new target object when you run the mapping task.

**Incoming Fields**

**Target Fields**

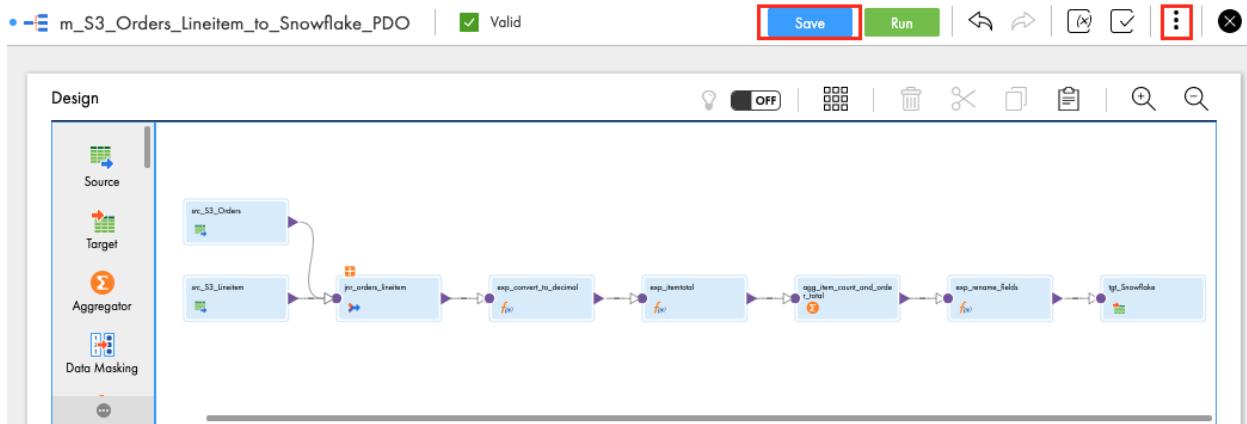
**Field Mapping**

| Field Name    | Mapped Field  |
|---------------|---------------|
| orderkey      | orderkey      |
| custkey       | custkey       |
| orderdate     | orderdate     |
| orderpriority | orderpriority |
| orderstatus   | orderstatus   |
| totalprice    | totalprice    |
| itemcount     | itemcount     |
| total_calc    | total_calc    |

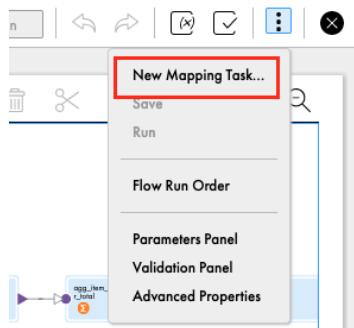
## Step 12

Let's configure Pushdown Optimization (PDO) in the Mapping Task and execute it.

1. Click **Save** to save and validate mapping.
2. Click 3 dots to create **Mapping task** from the mapping



3. Select **New Mapping Task...**



4. In the New mapping task window, enter **mct\_S3\_Orders\_Lineitem\_to\_Snowflake\_PDO** in the Name field.

5. Select **Hands-on Lab** for Location.
  6. Select **Informatica Cloud Hosted Agent** for Runtime Environment.
  7. Click **Next**.

**Task Details**

Task Name: \* **mct\_S3\_Orders\_Lineitem\_to\_Snowflake\_PDO**

Location: \* Hands-on Lab

Description:

Runtime Environment: \* **Informatica Cloud Hosted Agent**

Mapping: \* **m\_S3\_Orders\_Lineitem\_to\_Snowflake(PDO)**

**Mapping Image: m\_S3\_Orders\_Lineitem\_to\_Snowflake(PDO)**

```

graph LR
    S3_Orders[aws_S3_Orders] --> Join1{Join}
    S3_Lineitems[aws_S3_Lineitems] --> Join1
    Join1 --> Map1{map_convert_to_desired}
    Map1 --> Map2{map_derived}
    Map2 --> Map3{map_derive_and_map}
    Map3 --> Map4{map_rename_fields}
    Map4 --> Snowflake[Snowflake]
  
```

The mapping diagram illustrates the data flow from two S3 sources to a single Snowflake target. It begins with two source boxes: 'aws\_S3\_Orders' and 'aws\_S3\_Lineitems'. Arrows point from both sources to a 'Join' node ('jr\_orders\_lineitem'). From the 'Join' node, an arrow points to a 'map\_convert\_to\_desired' node. This is followed by a 'map\_derived' node, which then leads to a 'map\_derive\_and\_map' node. Finally, an arrow points from this node to a 'map\_rename\_fields' node, which concludes with an arrow pointing to the 'Snowflake' target box.

8. Click **Add** in Advanced Session Properties.
  9. Select **Pushdown Optimization** from the Session Property Name dropdown list.
  10. Select **Full** from the Session Property Value dropdown list.
  11. Click **Finish**.

**Advanced Session Properties**

| Add                                   | Session Property Name*                             | Session Property Value*           |
|---------------------------------------|--|-----------------------------------|
| <input type="button" value="Remove"/> | <input type="text" value="Pushdown Optimization"/> | <input type="text" value="Full"/> |

Enable cross-schema pushdown optimization [?](#)

Allow the mapping task to be executed simultaneously. [?](#)

12. Click **Run** to execute the mapping task.

**mct\_S3\_Orders\_Lineitem\_to\_Snowflake\_PDO**

**Task Details**

Task Name: mct\_S3\_Orders\_Lineitem\_to\_Snowflake\_PDO  
Location: Hands-on Lab  
Description:  
Runtime Environment: Informatica Cloud Hosted Agent

### Step 13

View job execution progress.

1. Click **My Jobs** to monitor the job execution.

2. Click **Refresh** icon when “Updates available” message appears.  
3. When job is completed, make sure Status is **Success**.

Jobs (3 of 134) **Updates available**

| Instance Name        | Location     | Subtasks | Start Time            | End Time | Rows Processed | Status  |
|----------------------|--------------|----------|-----------------------|----------|----------------|---------|
| mct_S3_Orders_Lin... | Hands-on Lab |          | Apr 5, 2021, 11:42 AM |          |                | Running |

Jobs (3 of 134) **Up to date**

| Instance Name        | Location     | Subtasks | Start Time            | End Time              | Rows Processed | Status  |
|----------------------|--------------|----------|-----------------------|-----------------------|----------------|---------|
| mct_S3_Orders_Lin... | Hands-on Lab |          | Apr 5, 2021, 11:42 AM | Apr 5, 2021, 11:43 AM | 150000         | Success |

4. You can download job session log to view the log. In the log you will see this line indicating PDO is enabled.

```
OPT_63309 [2021-04-02 22:13:33.415] Pushdown Optimization was  
successfully enabled.
```

5. And an INSERT SQL statement.

```
INSERT INTO  
"PC_INFORMATICA_DB"."PUBLIC"."ORDERSLINEITEM"("orderkey","custkey","ord  
erdate","orderpriority","orderstatus","totalprice","itemcount","total_c  
alc") SELECT t5.t5c6, t5.t5c7, t5.t5c10, t5.t5c11, t5.t5c8, t5.t5c9,  
t5.t5c12::NUMBER(18,0), t5.t5c13 FROM (SELECT t3.t3c0, t3.t3c1,  
t3.t3c2, t3.t3c3, t3.t3c4, t3.t3c5, t3.t3c0 c0, t3.t3c1 c1, t3.t3c2 c2,  
t3.t3c3 c3, t3.t3c4 c4, t3.t3c5 c5, COUNT(t1.t1c0)::NUMBER(10,0),  
SUM(((t1.t1c1) * (1 - (t1.t1c2))) * (1 + (t1.t1c3))) FROM (SELECT  
t0."l_orderkey":VARCHAR(256), t0."l_extendedprice":VARCHAR(256),  
t0."l_discount":VARCHAR(256), t0."l_tax":VARCHAR(256) FROM  
"PC_INFORMATICA_DB"."PUBLIC"."ORDERSLINEITEM_1617648173588" AS t0) AS  
t1(t1c0 , t1c1 , t1c2 , t1c3) Join (SELECT  
t2."o_orderkey":VARCHAR(256), t2."o_custkey":VARCHAR(256),  
t2."o_orderstatus":VARCHAR(256),  
(t2."o_totalprice":NUMBER(38,2))::DOUBLE,  
t2."o_orderdate":VARCHAR(256), t2."o_orderpriority":VARCHAR(256) FROM  
"PC_INFORMATICA_DB"."PUBLIC"."ORDERSLINEITEM_1617648173277" AS t2) AS  
t3(t3c0 , t3c1 , t3c2 , t3c3 , t3c4 , t3c5) ON t3.t3c0 = t1.t1c0 GROUP  
BY 1, 2, 3, 4, 5, 6) AS t5(t5c0 , t5c1 , t5c2 , t5c3 , t5c4 , t5c5 ,  
t5c6 , t5c7 , t5c8 , t5c9 , t5c10 , t5c11 , t5c12 , t5c13)
```

#### Step 14

1. In Snowflake UI, you should see 150,000 rows inserted in **ORDERSLINEITEM** table.

The screenshot shows the Snowflake Worksheets interface. At the top, there are navigation icons for Databases, Shares, Data Marketplace, Warehouses, Worksheets (selected), and History. On the right, there are links for Preview App, Partner Connect, Help, and user information (EWIDJAJA, SYSADMIN). Below the header, the current workspace is EddyW, and the active tab is Webinar.

In the center, a query is running:

```

1 USE PC_INFORMATICA_DB;
2

```

The results are shown in a Data Preview table for the ORDERSLINEITEM table:

| Row | orderkey | custkey | orderdate  | orderpriority | orderstatus | totalprice | itemcount | total_calc |
|-----|----------|---------|------------|---------------|-------------|------------|-----------|------------|
| 1   | 1        | 3691    | 1996-01-02 | 5-LOW         | O           | 194029.55  | 6         | 194029.59  |
| 2   | 2        | 7801    | 1996-12-01 | 1-URGENT      | O           | 60951.63   | 1         | 60951.64   |
| 3   | 3        | 12332   | 1993-10-14 | 5-LOW         | F           | 247296.05  | 6         | 247296.09  |
| 4   | 4        | 13678   | 1995-10-11 | 5-LOW         | O           | 53829.87   | 1         | 53829.88   |
| 5   | 5        | 4450    | 1994-07-30 | 5-LOW         | F           | 139660.54  | 3         | 139660.56  |
| 6   | 6        | 5563    | 1992-02-21 | 4-NOT SPEC... | F           | 65843.52   | 1         | 65843.53   |
| 7   | 7        | 3914    | 1996-01-10 | 2-HIGH        | O           | 231037.28  | 7         | 231037.36  |
| 8   | 33       | 6697    | 1993-10-27 | 3-MEDIUM      | F           | 118518.56  | 4         | 118518.59  |
| 9   | 34       | 6101    | 1998-07-21 | 3-MEDIUM      | O           | 75662.77   | 3         | 75662.79   |
| 10  | 36       | 11527   | 1995-11-03 | 1-URGENT      | O           | 72196.43   | 1         | 72196.43   |
| 11  | 37       | 8612    | 1992-06-03 | 3-MEDIUM      | F           | 156440.15  | 3         | 156440.18  |
| 12  | 39       | 8177    | 1996-09-20 | 3-MEDIUM      | O           | 307811.89  | 6         | 307811.98  |
| 13  | 64       | 3212    | 1994-07-16 | 3-MEDIUM      | F           | 30616.90   | 1         | 30616.90   |
| 14  | 66       | 12920   | 1994-01-20 | 5-LOW         | F           | 100991.26  | 2         | 100991.28  |
| 15  | 68       | 2855    | 1998-04-18 | 3-MEDIUM      | O           | 305815.83  | 7         | 305815.91  |

----- Congratulations. You have successfully completed this lab -----

## Module 6: Transform Semi-Structured JSON Data

JSON (JavaScript Object Notation) is a text-based data format that is commonly used between servers and web applications / web-connected devices. Because it is text-based, it is readable by both humans and machines. JSON semi-structured data can be stored in Snowflake variant columns alongside relational data. In IICS, the hierarchy parser transformation parses and transforms hierarchy data to relational data.

In this lab, you will use the weather forecast data available in the Snowflake sample database (for most deployments) or that you loaded separately (for deployments where it wasn't available). You will create a hierarchical schema, then use it in a mapping to parse and transform the JSON weather forecast data, join them, and add an expression to convert the temperature, then write to a new table.

\*\*\*

**NOTE: The lab assumes you are accessing the data from the SNOWFLAKE\_SAMPLE\_DATA database. If you needed to load the data manually, just change all references to SNOWFLAKE\_SAMPLE\_DATA to DEMO\_DB (or whichever database you loaded the data to).**

\*\*\*

### Step 1

Firstly we will make sure **PC\_INFORMATICA\_ROLE** has access to **DEMO\_DB** data. In Snowflake as the **ACCOUNTADMIN** role go to **Databases**. Select **DEMO\_DB** and grant usage to **PC\_INFORMATICA\_ROLE**

#### Grant privileges on database DEMO\_DB

Grant privileges on  (−) (+)

Privileges to grant

Grant privileges to

with Grant Option

Cancel Grant

Now we can copy JSON data from the Snowflake table and save it locally in your computer.

1. In Snowflake worksheet, find **SNOWFLAKE\_SAMPLE\_DATA** or **DEMO\_DB** database (whichever is available)
2. Expand **WEATHER** schema.
3. Click on **DAILY\_14\_TOTAL** table.
4. Click **Preview Data**.
5. Then click the first record in column **V** and the details window pops up.

Find database objects

Starting with...

- SNOWFLAKE\_SAMPLE\_DATA**
  - INFORMATION\_SCHEMA
  - TPCDS\_SF100TCL
  - TPCDS\_SF10TCL
  - TPCH\_SF1
  - TPCH\_SF10
  - TPCH\_SF100
  - TPCH\_SF1000
  - WEATHER**
    - Tables
  - DAILY\_14\_TOTAL

89,230,179 rows 36.4 GB

Cluster by

Columns Data Type

V VARIANT

T TIMESTAMP\_NTZ(9)

Preview Data

Results Data Preview

Table: SNOWFLAKE\_SAMPLE\_DATA.WEATHER.DAILY\_14\_TOTAL

Data Details

Filter result...

| Row | V  | T                       |
|-----|--|-------------------------|
| 1   | { "city": { "coord": { "lat": 27.716669, "lon": 80.900002 }, "count... }       | 2016-11-21 10:20:47.000 |
| 2   | { "city": { "coord": { "lat": 26.19833, "lon": 78.945 }, "country": "I... }    | 2016-11-21 10:20:47.000 |
| 3   | { "city": { "coord": { "lat": 29.99444, "lon": 77.044441 }, "countr... }       | 2016-11-21 10:20:47.000 |
| 4   | { "city": { "coord": { "lat": 27.65, "lon": 74.383331 }, "country": "I... }    | 2016-11-21 10:20:47.000 |
| 5   | { "city": { "coord": { "lat": 27.816669, "lon": 75.033333 }, "count... }       | 2016-11-21 10:20:47.000 |
| 6   | { "city": { "coord": { "lat": 8.31667, "lon": 77.183327 }, "country": "I... }  | 2016-11-21 10:20:47.000 |
| 7   | { "city": { "coord": { "lat": 10.7, "lon": 79.533333 }, "country": "I... }     | 2016-11-21 10:20:47.000 |
| 8   | { "city": { "coord": { "lat": 21.623329, "lon": 69.981667 }, "count... }       | 2016-11-21 10:20:47.000 |
| 9   | { "city": { "coord": { "lat": 15.76667, "lon": 76.199997 }, "country": "I... } | 2016-11-21 10:20:47.000 |
| 10  | { "city": { "coord": { "lat": 11.56667, "lon": 79.599998 }, "count... }        | 2016-11-21 10:20:47.000 |
| 11  | { "city": { "coord": { "lat": 18.08333, "lon": 75.433327 }, "countr... }       | 2016-11-21 10:20:47.000 |

6. Click **Copy**.
7. Click **Done**.

Find database objects

Starting with...

- SNOWFLAKE\_SAMPLE\_DATA**
  - INFORMATION\_SCHEMA
  - TPCDS\_SF100TCL
  - TPCDS\_SF10TCL
  - TPCH\_SF1
  - TPCH\_SF10
  - TPCH\_SF100
  - TPCH\_SF1000
  - WEATHER**
    - Tables
  - DAILY\_14\_TOTAL

89,230,179 rows 36.4 GB

Cluster by

Columns Data Type

V VARIANT

T TIMESTAMP\_NTZ(9)

Preview Data

Results Data Preview

Table: SNOWFLAKE\_SAMPLE\_DATA.WEATHER.DAILY\_14\_TOTAL

Data Details

```

1 {
2   "city": {
3     "coord": {
4       "lat": 27.716669,
5       "lon": 80.900002
6     },
7     "country": "IN",
8     "id": 1265310,
9     "name": "Laharpur"
10   },
11   "data": [
12     {
13       "clouds": 0
14     }
15   ]
16 }
17
18
19 { "city": { "coord": { "lat": 15.76667, "lon": 76.199997 }, "country": "I... }
20 { "city": { "coord": { "lat": 11.56667, "lon": 79.599998 }, "country": "I... }
21 { "city": { "coord": { "lat": 18.08333, "lon": 75.433327 }, "country": "I...

```

Copy

Done

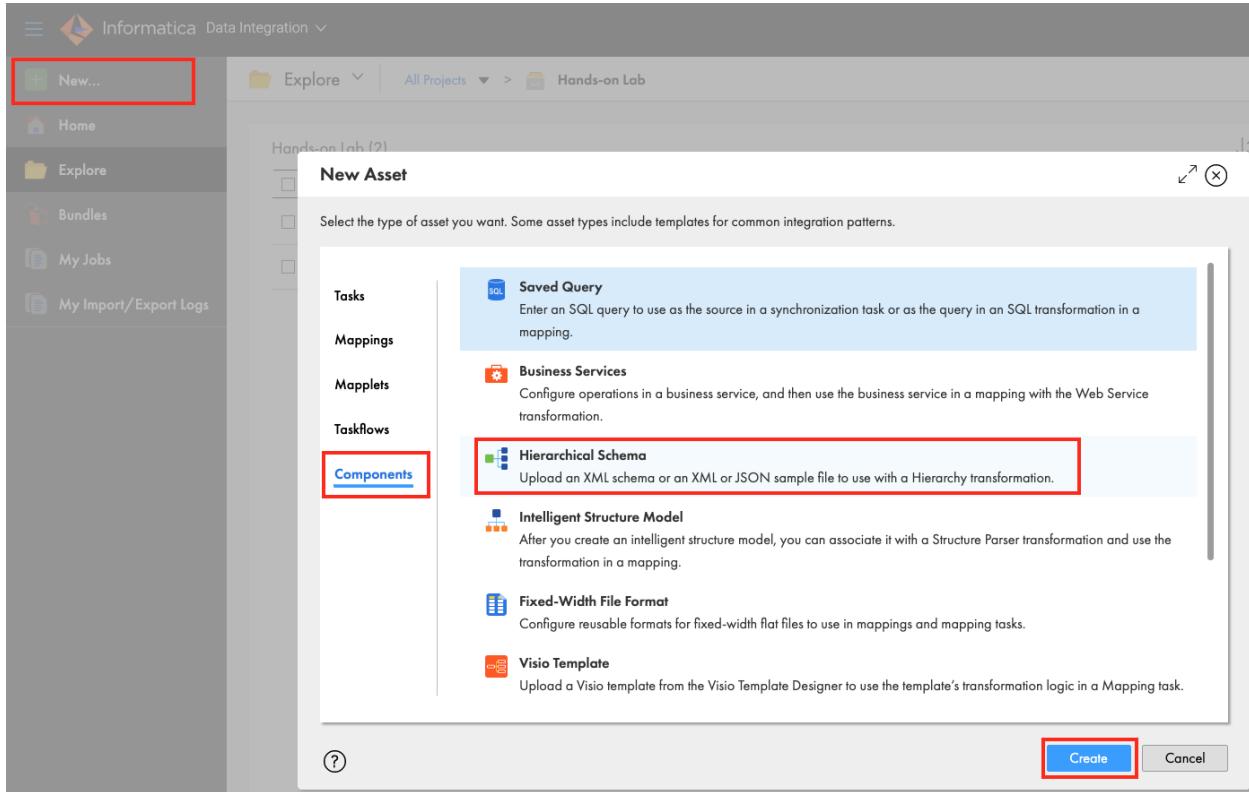
| Row | V  | T                       |
|-----|--|-------------------------|
| 1   | { "city": { "coord": { "lat": 27.716669, "lon": 80.900002 }, "count... }       | 2016-11-21 10:20:47.000 |
| 2   | { "city": { "coord": { "lat": 26.19833, "lon": 78.945 }, "country": "I... }    | 2016-11-21 10:20:47.000 |
| 3   | { "city": { "coord": { "lat": 29.99444, "lon": 77.044441 }, "countr... }       | 2016-11-21 10:20:47.000 |
| 4   | { "city": { "coord": { "lat": 27.65, "lon": 74.383331 }, "country": "I... }    | 2016-11-21 10:20:47.000 |
| 5   | { "city": { "coord": { "lat": 27.816669, "lon": 75.033333 }, "count... }       | 2016-11-21 10:20:47.000 |
| 6   | { "city": { "coord": { "lat": 8.31667, "lon": 77.183327 }, "country": "I... }  | 2016-11-21 10:20:47.000 |
| 7   | { "city": { "coord": { "lat": 10.7, "lon": 79.533333 }, "country": "I... }     | 2016-11-21 10:20:47.000 |
| 8   | { "city": { "coord": { "lat": 21.623329, "lon": 69.981667 }, "count... }       | 2016-11-21 10:20:47.000 |
| 9   | { "city": { "coord": { "lat": 15.76667, "lon": 76.199997 }, "country": "I... } | 2016-11-21 10:20:47.000 |
| 10  | { "city": { "coord": { "lat": 11.56667, "lon": 79.599998 }, "count... }        | 2016-11-21 10:20:47.000 |
| 11  | { "city": { "coord": { "lat": 18.08333, "lon": 75.433327 }, "country": "I... } | 2016-11-21 10:20:47.000 |

8. Save the copied JSON in a text file locally on your laptop. Filename: **daily\_14.json**.

## Step 2

Create a Hierarchical Schema in IICS.

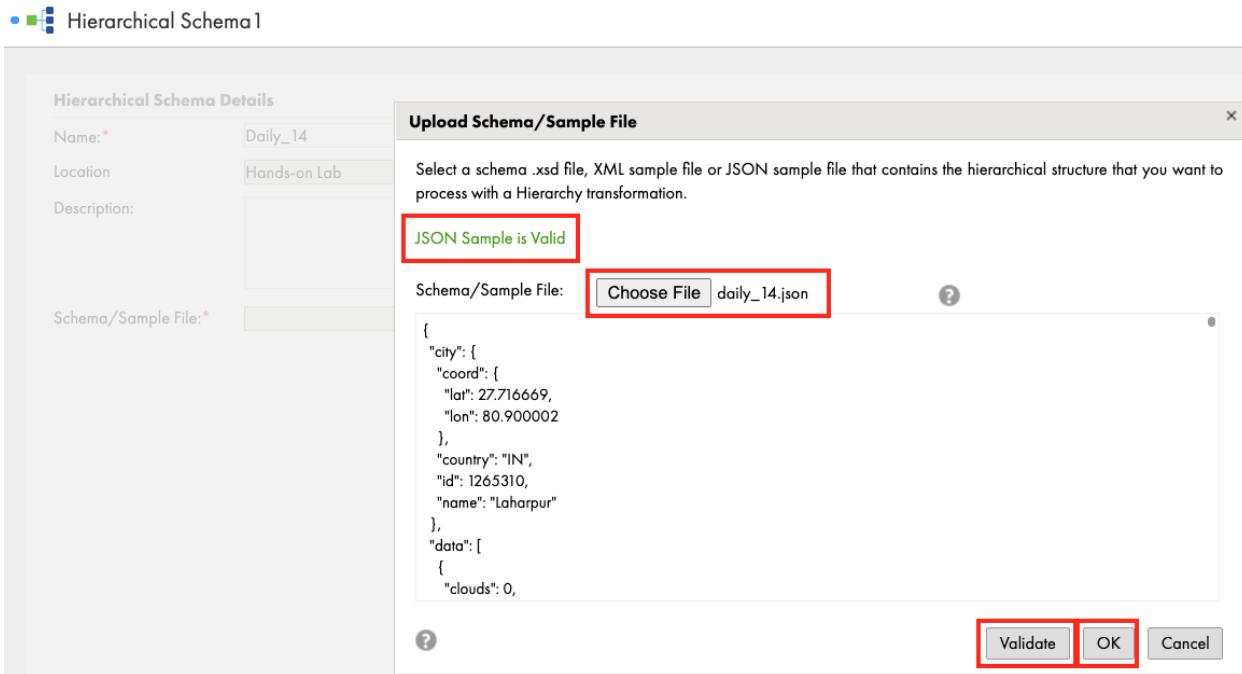
1. In IICS, click **New**.
2. Click **Components**.
3. Select **Hierarchical Schema** and click **Create**.



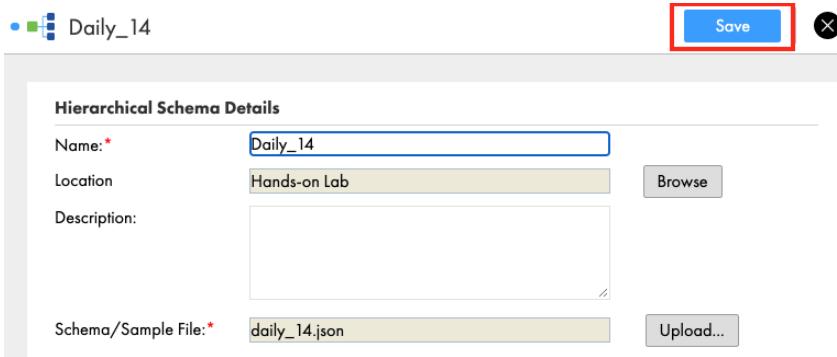
4. Enter **Daily\_14** in the Name field.
5. Select **Hands-on Lab** in the Location field if not already filled in.
6. Click **Upload**.

This screenshot shows the 'Hierarchical Schema Details' dialog. It has fields for 'Name' (containing 'Daily\_14'), 'Location' (containing 'Hands-on Lab'), and 'Description'. Below these is a large text area for 'Schema/Sample File'. At the bottom right is a 'Upload...' button highlighted with a red box.

7. Click **Choose File** and select the JSON file you saved above.
8. Click **Validate** and you should see **JSON Sample is Valid** message.
9. Click **OK**.



10. Click **Save**.



### Step 3

Execute one of the following SQL statements in Snowflake to create a view over one of the shared tables. The one you execute will depend if you have the **WEATHER** data in **SNOWFLAKE\_SAMPLE\_DATA** or **DEMO\_DB**

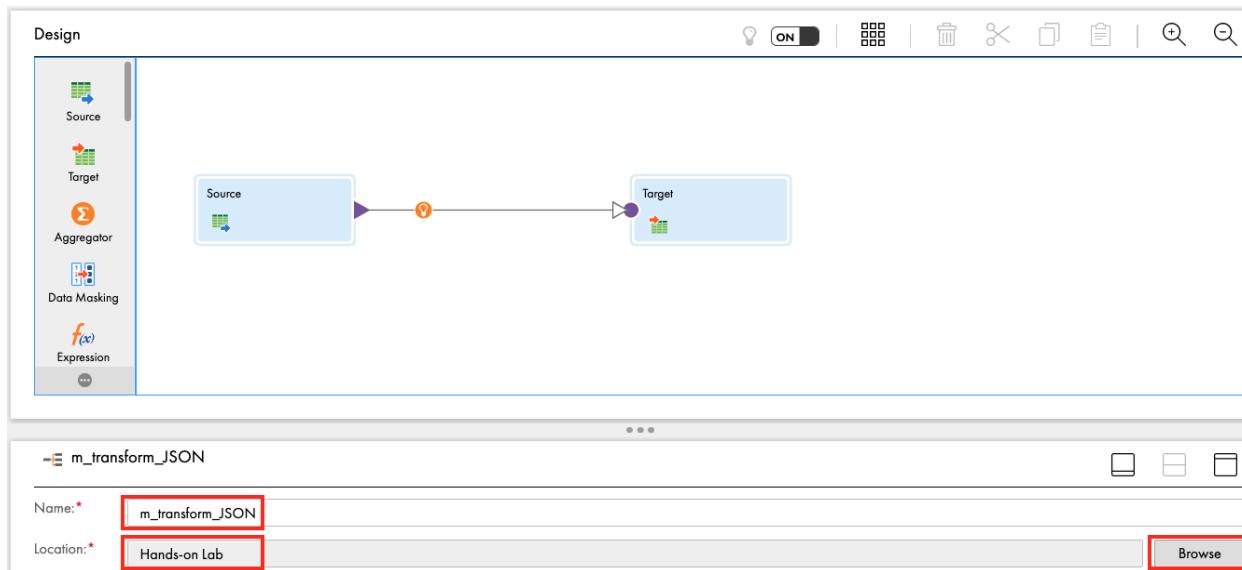
```
create view "PC_INFORMATICA_DB"."PUBLIC"."DAILY_14_TOTAL" as select * from
"SNOWFLAKE_SAMPLE_DATA"."WEATHER"."DAILY_14_TOTAL";
```

or

```
create view "PC_INFORMATICA_DB"."PUBLIC"."DAILY_14_TOTAL" as select * from
"DEMO_DB"."WEATHER"."DAILY_14_TOTAL";
```

Create a mapping to read from the daily\_14\_total table, use a hierarchy parser to parse the JSON data, join the relational data, convert the temperature and write to a new Snowflake table.

1. Click **New...**
2. Click **Mappings**.
3. Select **Mapping**.
4. Click **Create**.
5. Under properties, enter **m\_transform\_JSON** in Name field.
6. Ensure Location is **Hands-on Lab**. If not, click **Browse** and select it.



#### Step 4

Let's configure the data source from Snowflake.

1. Click **Source** transform in the mapping canvas to assign its properties.
2. In General tab, enter **src\_daily\_14** in the Name field.
3. In Source tab, select **Snowflake <account name>** in the Connection dropdown field.
4. Click **Select** to select the source table/object.
5. In Select Source Object window, scroll down to find **PC\_INFORMATICA\_DB** and click it. Then click **PUBLIC** schema.
6. Select **DAILY\_14\_TOTAL** in the tables list on the right pane.
7. Click **OK**.

### Select Source Object

Select a source object, then click OK. You can also search for a source object.

DAILY\_14\_TOTAL(PC\_INFORMATICA\_DB/PUBLIC)

| Select                           | Name           | Label          | Description | Type  |
|----------------------------------|----------------|----------------|-------------|-------|
| <input checked="" type="radio"/> | DAILY_14_TOTAL | DAILY_14_TOTAL |             | view  |
| <input type="radio"/>            | ORDERSLINEITEM | ORDERSLINEITEM |             | table |
| <input type="radio"/>            | ORDERSLINEITEM | ORDERSLINEITEM |             | table |

Displaying all 3 objects.

OK Cancel

8. Expand **Query Options**.
9. Click **Configure** for Filter.

Properties Preview Source

General Source Fields Partitions

Details

Connection: Snowflake informaticapartner (Snowflake...) View... New Connection... New Parameter...

Source Type: Single Object

Object: SNOWFLAKE\_SAMPLE\_DATA/WEATHER/DA... Select... Preview Data...

Query Options

Filter: **Configure...**

Sort: Configure...

10. Click Filter Condition dropdown and select **Advanced**.

11. Paste the following in the filter condition:

```
DAILY_14_TOTAL.T      >=      to_date('2021-02-01', 'YYYY-MM-DD')      AND
DAILY_14_TOTAL.T      <=      to_date('2021-02-28', 'YYYY-MM-DD')      AND
DAILY_14_TOTAL.V:city:country='US'    and    DAILY_14_TOTAL.V:city:name    =
'San Francisco'
```

12. Click **OK**.

## Query Options



Filter   Sort

Add conditions to filter records. You can add multiple conditions for each object.

Filter Condition: Advanced

Fields

SNOWFLAKE\_SAMPLE\_DATA/WEATHER/DA

DAILY\_14\_TOTAL.T >= to\_date('2021-02-01','YYYY-MM-DD') AND DAILY\_14\_TOTAL.T <= to\_date('2021-02-28','YYYY-MM-DD') AND DAILY\_14\_TOTAL.V:city:country='US' AND DAILY\_14\_TOTAL.V:city:name = 'San Francisco'

Select a field to see detailed information about it.

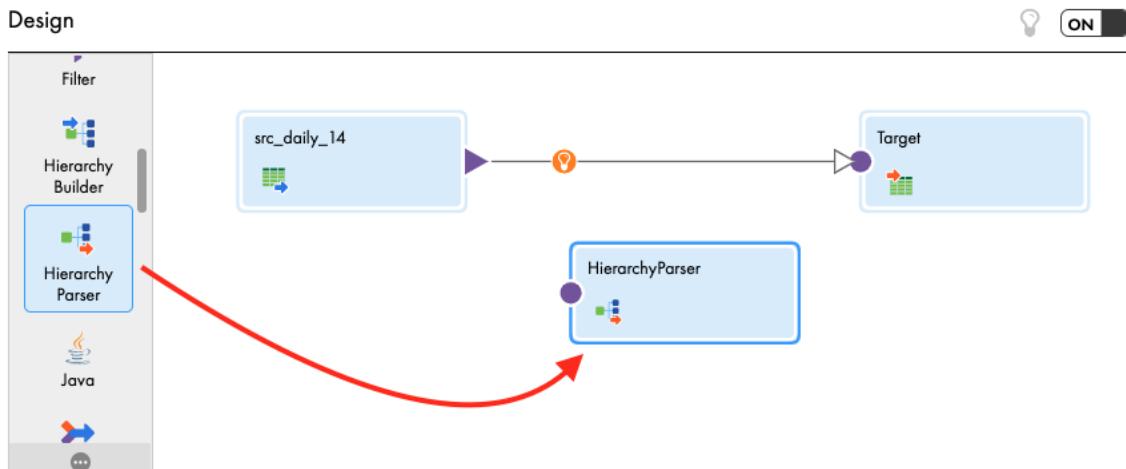
OK   Cancel

13. Click **Save** to periodically save work in progress.

## Step 5

Add HierarchyParser transform and configure it.

1. Drag and drop **Hierarchy Parser** transform on to the canvas.



2. In General tab, enter **hp\_parse\_JSON** in the Name field.
3. In Input Settings tab, click Select and select the **Daily\_14** hierarchical schema. Click **OK**.

**Properties** Preview |  hp\_parse\_JSON

**General**

Input Type:  Buffer  File

Schema: \* **Hands-on Lab\Daily\_14** Select... View...

Description:

**Incoming Fields** Input Settings

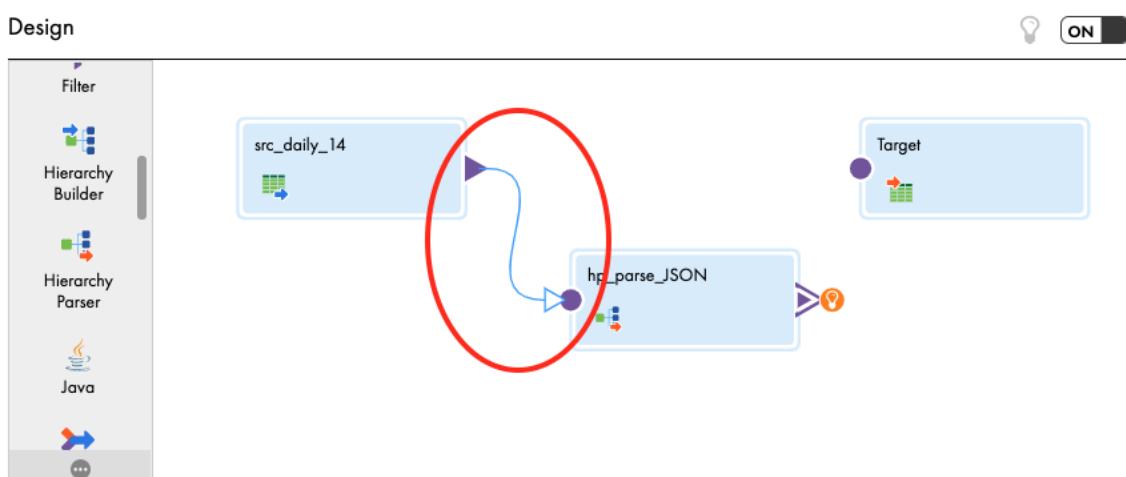
**Input Field Selection**

**Field Mapping**

**Output Fields**

**Advanced**

4. Select the line from **src\_daily\_14** to **Target** and click delete icon .
5. Link **src\_daily\_14** to **hp\_parse\_JSON**.



6. In Input Field Selection tab, drag and drop **V** field from Incoming Fields to Input field in Hierarchical Schema Input Fields.

**Properties** Preview |  hp\_parse\_JSON

**General**

**Incoming Fields**

| Field Name |
|------------|
| V          |
| T          |

**Input Settings**

**Input Field Selection**

**Field Mapping**

**Output Fields**

**Advanced**

**Hierarchical Schema Input Fields: (1 of 1 mapped)**

| Field Name | Mapped Field | Type | Precision | Scale |
|------------|--------------|------|-----------|-------|
| Input      | V            |      | 32768     | 0     |

**Options** Automatch ...

7. In Field Mapping tab, expand root element by clicking the triangle icon or expand/contract icon  .
8. Select (check) root and choose Map All Descendants. Selected fields will automatically show up in the Relational Fields on the right panel. Primary keys and foreign keys are auto created to make the fields relational.

**Schema Structure: (0 of 26 mapped)**

The screenshot shows the 'Schema Structure' interface with a table titled 'Element Name' and 'Cardinal'. A context menu is open over the 'root' node, with the 'Map all descendants' option highlighted by a red box.

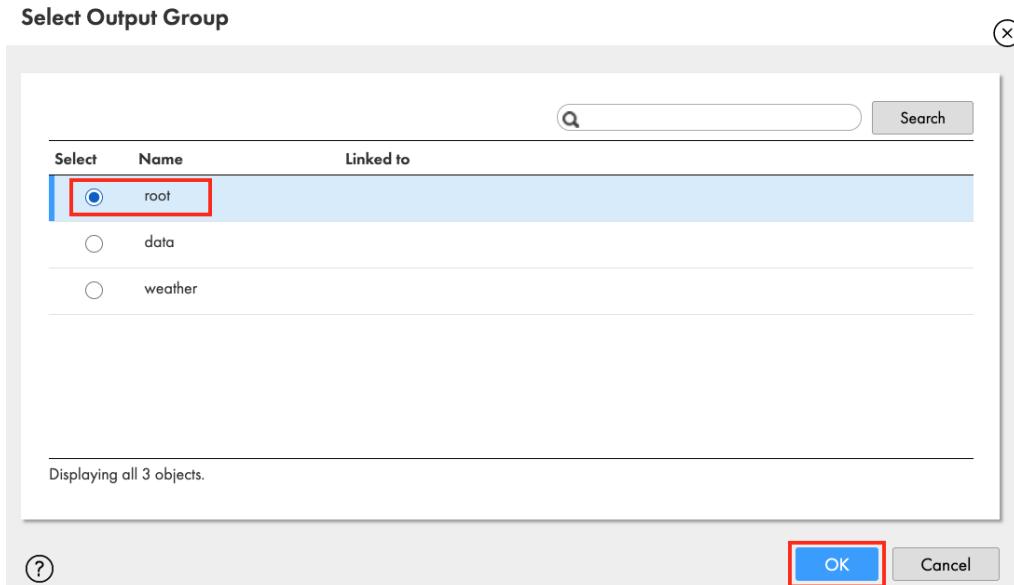
| Element Name             | Cardinal |
|--------------------------|----------|
| root                     | 1-1      |
| Map all descendants      | 0-1      |
| Map immediate children   | 0-1      |
| Unmap all descendants    | 0-1      |
| Unmap immediate children | 0-1      |

9. Click **Save** to periodically save work in progress.

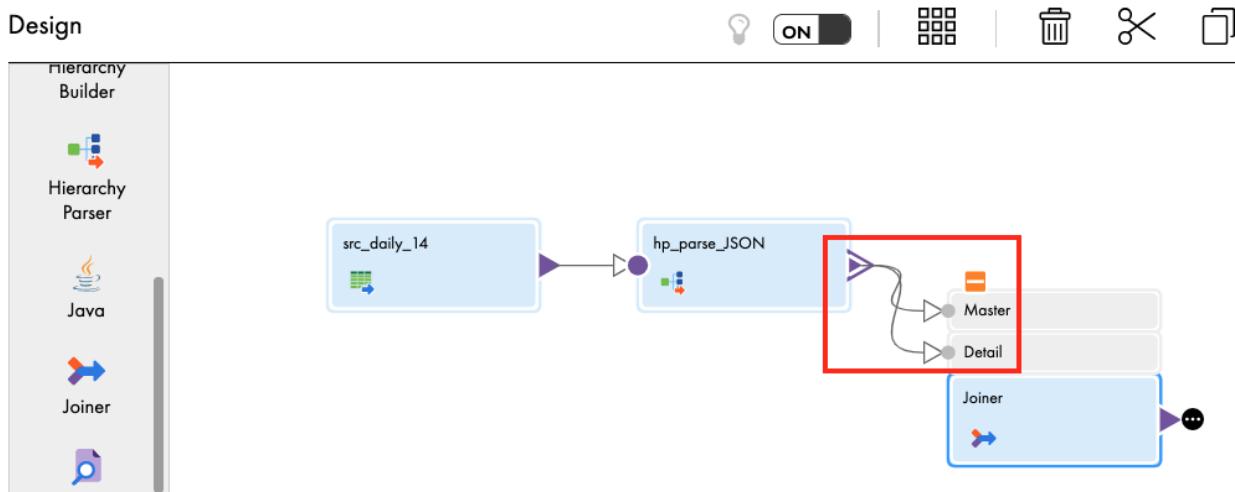
**Step 6**

Add a Joiner transform to link root and data relational field groups and configure it.

1. Drag and drop Joiner transform on the canvas.
2. Link **hp\_parse\_JSON** to the **Master** in Joiner transform.
3. Select Output Group window appears. Select **root** and click **OK**.



4. Link **hp\_parse\_JSON** again but this time to the **Detail** in Joiner transform.
5. Select **data** and click **OK**.



6. In General tab, enter **jnr\_temperature** in the Name field.
7. In Join Condition tab, click add icon
8. Select PK\_root (bigint) in Master column and FK\_root (bigint) in the Detail.

**Properties** Preview jnr\_temperature

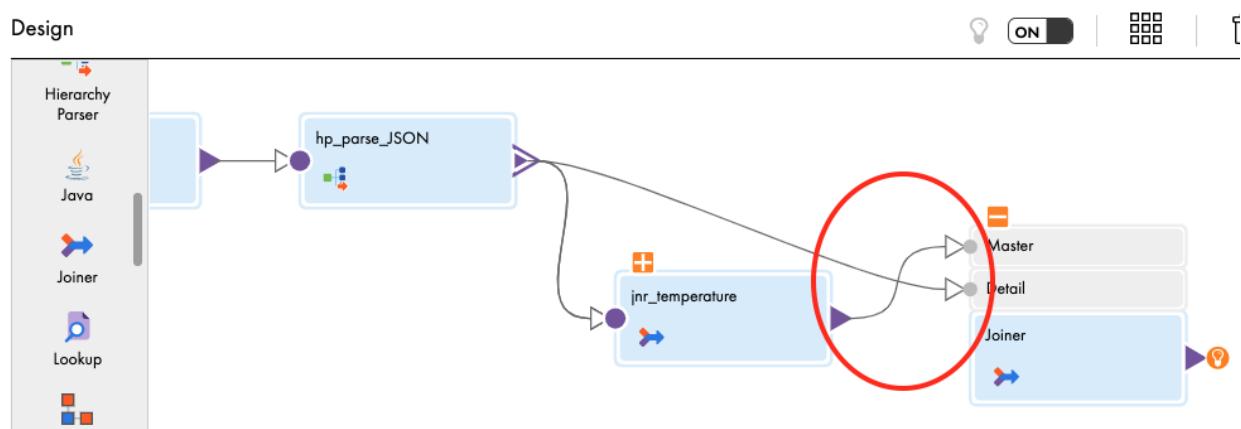
|                        |                          |                       |                          |
|------------------------|--------------------------|-----------------------|--------------------------|
| <b>General</b>         | Join Type: Normal        | Master: hp_parse_JSON | Detail: hp_parse_JSON    |
| <b>Incoming Fields</b> | Join Condition: Simple   |                       |                          |
| <b>Join Condition</b>  | Join Conditions          |                       |                          |
| <b>Advanced</b>        | Master: PK_root (bigint) | Operator: =           | Detail: FK_root (bigint) |

9. In Advanced tab, select **Sorted Input**.
10. Click **Save** to periodically save work in progress.

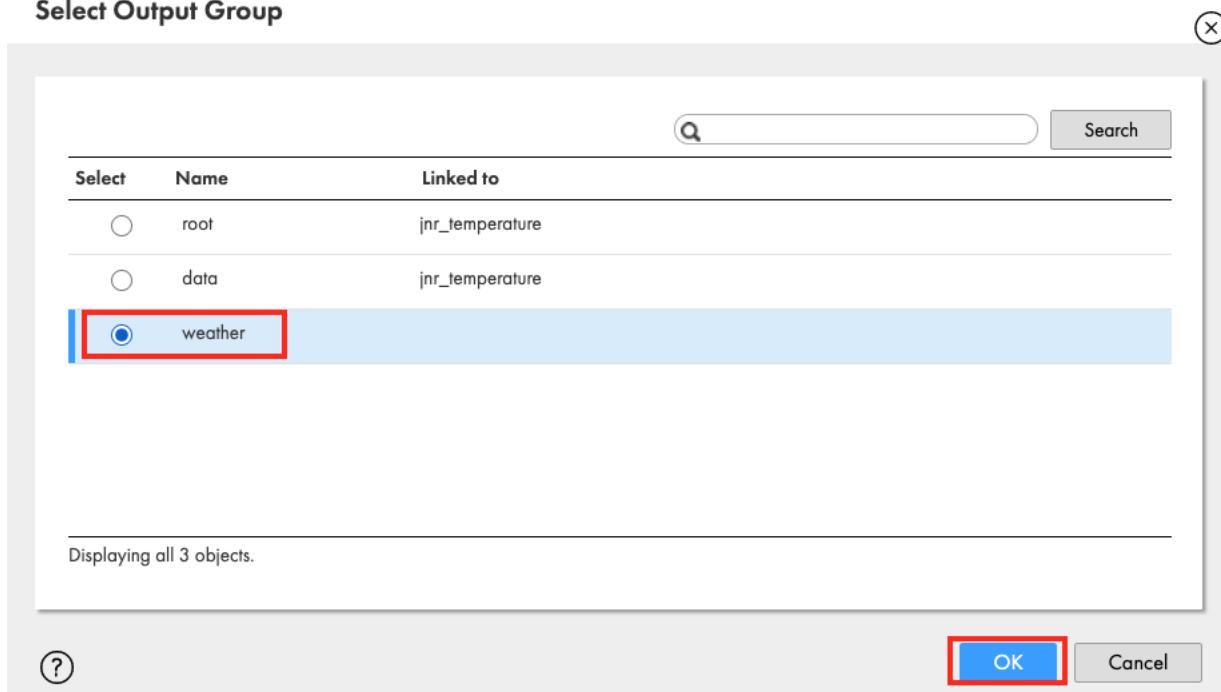
### Step 7 - Optional

Add another Joiner transform to join and configure it.

1. Drag and drop Joiner transform on the canvas.
2. Link **jnr\_temperature** to the **Master** in Joiner transform.
3. Link **hp\_parse\_JSON** to the **Detail** in Joiner transform.



4. Select Output Group window appears. Select **weather** and click **OK**.



5. In Join Condition tab, select **PK\_data (bigint)** in Master and **FK\_data (bigint)** in Detail.

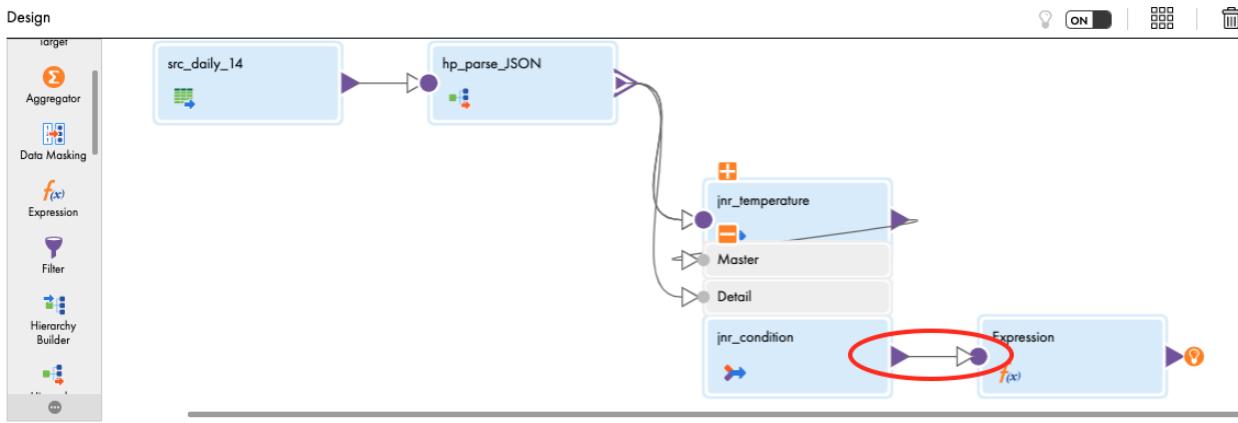
| Master           | Operator | Detail           |
|------------------|----------|------------------|
| PK_data (bigint) | =        | FK_data (bigint) |

6. In Advanced tab, select **Sorted Input**.  
 7. Click **Save** to periodically save work in progress.

#### **Step 8 - Optional**

Add Expression transform to create an ordered fields in the target and convert temperature from Kelvin to Fahrenheit.

1. Drag and drop **Expression** transform on the canvas.
2. Link **jnr\_condition** to the **Expression**.



3. In General tab, enter **exp\_convert\_temperature** in the Name field.
4. In Expression tab, add the following fields and expressions.

| Field Name     | Type; Precision; Scale | Expression   |
|----------------|------------------------|--|
| Date           | Date/time; 29; 9       | Add_To_Date(To_Date('1970-01-01', 'YYYY-MM-DD'), 'SS', dt) |
| City           | String; 255; 0         | name   |
| Country_Name   | String; 255; 0         | country  |
| Min_Temp       | Decimal; 10; 1         | (min - 273.15) * 9/5 + 32                                  |
| Max_Temp       | Decimal; 10; 1         | (max - 273.15) * 9/5 + 32                                  |
| Condition      | String; 100; 0         | description  |
| Humidity_Level | Double; 15; 0          | humidity   |

Properties Preview | **f(x)** exp\_convert\_temperature

---

**General** Create simple expressions. You can also use expression macros to create complex expressions.  
 Allow additional fields and expressions during task creation

**Incoming Fields**

**Expression**

| Field Name     | Expression   | Field Description |
|----------------|--|-------------------|
| Date           | Add_To_Date(To_Date('1970-01-01', 'YYYY-MM-DD'), 'SS', dt) |                   |
| City           | name   |                   |
| Country_Name   | country  |                   |
| Min_Temp       | (min - 273.15) * 9/5 + 32                                  |                   |
| Max_Temp       | (max - 273.15) * 9/5 + 32                                  |                   |
| Condition      | description  |                   |
| Humidity_Level | humidity   |                   |

**Window**

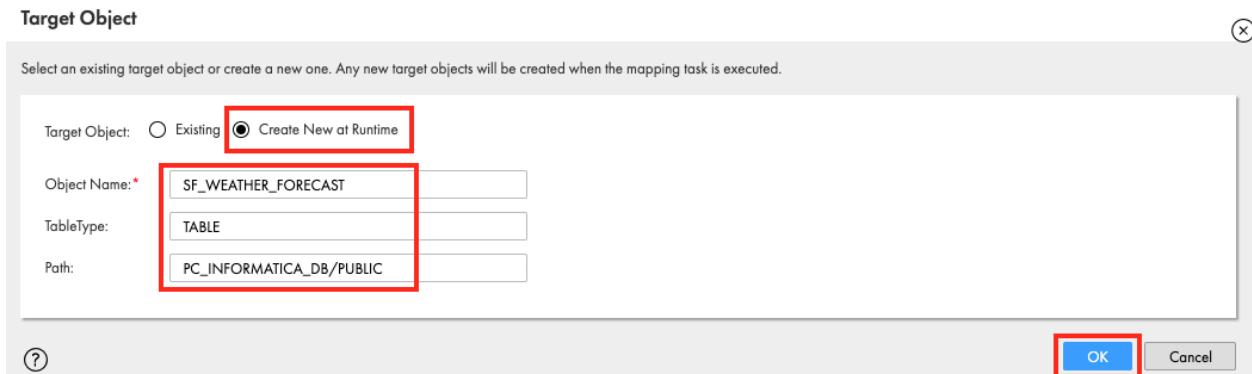
**Advanced**

### Step 9

Finally, let's configure the Target.

1. Link **jnr\_temperature** to Target.
2. In General tab, enter **tgt\_sf\_weather\_forecast** in the Name field.
3. In Target tab, select **Snowflake** connection.
4. Click **Select** to select a table.

- In Target Object window, check **Create New at Runtime**.
- Enter **SF\_WEATHER\_FORECAST** in Object Name field.
- Enter **TABLE** in TableType.
- Enter **PC\_INFORMATICA\_DB/PUBLIC** in Path.
- Click **OK**.



- In Field Mapping tab, the target fields are automatically mapped from the incoming fields.

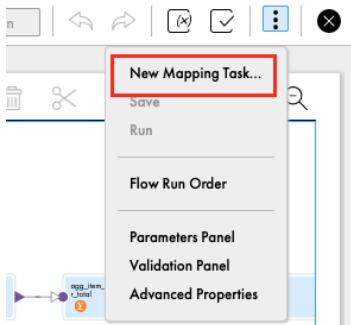
The screenshot shows the 'tgt\_sf\_weather\_forecast' mapping task configuration. The left sidebar has tabs for 'Properties', 'Preview', and 'tgt\_sf\_weather\_forecast'. The main area has tabs for 'General', 'Incoming Fields', 'Target', 'Target Fields', and 'Field Mapping' (which is selected and highlighted with a blue box). On the left, under 'Incoming Fields', there is a table with columns 'Field Name' and 'Mapped Field'. It lists fields: Date (mapped to Date), City (mapped to City), Country\_Name (mapped to Country\_Name), Min\_Temp (mapped to Min\_Temp), Max\_Temp (mapped to Max\_Temp), Condition (mapped to Condition), and Humidity\_Level (mapped to Humidity\_Level). A note above the table says: 'Fields will be mapped automatically, as you have chosen to create a new target object when you run the mapping task.'

- Click **Save**.

### Step 10

Let's configure a Mapping Task and execute it.

- Click 3 dots to create **Mapping task** from the mapping
- Select **New Mapping Task...**



3. In the New mapping task window, enter **mct\_transform\_JSON** in the Name field.
4. Select **Hands-on Lab** for Location.
5. Select **Informatica Cloud Hosted Agent** for Runtime Environment.
6. Click **Finish**.
7. Click **Run** to execute the mapping task.

### Step 11

Validate job execution result.

1. Click **My Jobs** to monitor the job execution.
2. Click **Refresh** icon when “Updates available” message appears.
3. When job is completed, make sure Status is **Success**.
4. **864** rows were created in Snowflake.

| Instance Name        | Location     | Subtasks | Start Time            | End Time         | Rows Processed | Status                                       |
|----------------------|--------------|----------|-----------------------|------------------|----------------|--|
| mct_transform_JSON-1 | Hands-on Lab |          | Apr 8, 2021, 3:32 ... | Apr 8, 2021, ... | 864            | <span style="color: green;">✓ Success</span> |

5. In Snowflake, there are 864 rows as well. Notice the columns label are in the order as configured in the Expression transform.

| Row | PK_root | lat     | lon       | country | id      | name          |
|-----|---------|---------|-----------|---------|---------|---------------|
| 1   | 1       | 37.7749 | -122.4194 | US      | 5391959 | San Francisco |
| 2   | 1       | 37.7749 | -122.4194 | US      | 5391959 | San Francisco |
| 3   | 1       | 37.7749 | -122.4194 | US      | 5391959 | San Francisco |

----- Congratulations. You have successfully completed this lab -----

# Module 7: Cloning, Time Travel in Snowflake

- 7.1.1 Reminder on running SQL commands. We will run these one at a time.

## Many Options to Run Commands.



SQL commands can be executed through the UI (limited), via the Worksheets tab, using our SnowSQL command line tool, a SQL editor of your choice via ODBC/JDBC, or through our Python or Spark connectors.

As mentioned earlier, in this lab we will run some operations via pre-written SQL in the worksheet (as opposed to using the UI) to save time.

- 7.1.2 Run a query by placing your cursor anywhere in the command and clicking the blue “Run” button at the top of the page or by hitting Ctrl/Cmd+Enter on your keyboard.



## Warning

In this lab, never check the “All Queries” box at the top of the worksheet. We want to run SQL queries one at a time in a specific order; not all at once.

```
Run Query (Cmd/Ctr + Return) Worksheet 2 + ->
Run All Queries | Saved 10 minutes ago Context: SYSADMIN COMPUTE_WH (XL) CITIB
1
2
3 -- This SQL file is for the Hands On Lab Guide for the 30-day free Snowflake trial account
4 -- The numbers below correspond to the sections of the Lab Guide in which SQL is to be run in a Snowflake worksheet
5 -- Modules 1 and 2 of the Lab Guide have no SQL to be run
6
7
8 /* ****
9 /* *** MODULE 3 ****
10 /* ****
11
12 -- 3.1.4
13
14 create or replace table trips
15 (tripduration integer,
16 starttime timestamp,
17 stoptime timestamp,
18 start_station_id integer,
19 start_station_name string,
20 start_station_latitude float,
21 start_station_longitude float,
22 end_station_id integer,
23 end_station_name string,
24 end_station_latitude float,
25 end_station_longitude float,
```

- 7.1.3 \*If\* you highlighted the entire SQL text of the command (did not just place your cursor in the command) and ran it, a confirmation box should appear asking “Do you want to run the following queries?”. Click the blue “Run” button in the box. In the future you can keep clicking this “Run” button on this confirmation box or check the “Don’t ask me again (All

Worksheets)" option in this box.

The screenshot shows the Snowflake web interface. At the top, there's a navigation bar with icons for Databases, Shares, Warehouses, Worksheets (which is highlighted), History, and Partner Connect. Below the navigation bar, there are two tabs: "Worksheet 1" and "Worksheet 2". A "Run" button is visible next to the tabs. The main area contains a block of SQL code. A modal dialog box is overlaid on the page, titled "Do you want to run the following queries?". Inside the dialog, there is a code block showing the definition of a table named "trips". At the bottom of the dialog, there is a checkbox labeled "Don't ask me again (All Worksheets)" and two buttons: "Cancel" and "Run". The "Run" button is highlighted with a red box and a cursor arrow pointing to it.

```
4 -- The numbers below correspond to the sections of the Lab Guide in which SQL is to be run in a Snowflake
5 -- Modules 1 and 2 of the Lab Guide have no SQL to be run
6
7
8 /* **** MODULE 3 ****
9 */
10 /*
11
12 -- 3.1.4
13
14 create or replace table trips
15 (tripduration integer,
16 starttime timestamp,
17 stoptime timestamp,
18 start_station_id integer,
19 start_station_name string,
20 start_station_latitude float,
21 start_station_longitude float,
22 end_station_id integer,
23 end_station_name string,
24 end_station_latitude float,
25 end_station_longitude float,
26 bikeid integer,
27 membership_type string,
28 usertype string,
29 birth_year integer,
30 gender integer);
```

## 7.2 Clone a Table

Snowflake allows you to create clones, also known as “zero-copy clones” of tables, schemas, and databases in seconds. A snapshot of data present in the source object is taken when the clone is created, and is made available to the cloned object. The cloned object is writable, and is independent of the clone source. That is, changes made to either the source object or the clone object are not part of the other.

A popular use case for zero-copy cloning is to clone a production environment for use by Development & Testing to do testing and experimentation on without (1) adversely impacting the production environment and (2) eliminating the need to set up and manage two separate environments for production and Development & Testing.

### Zero-Copy Cloning FTW!

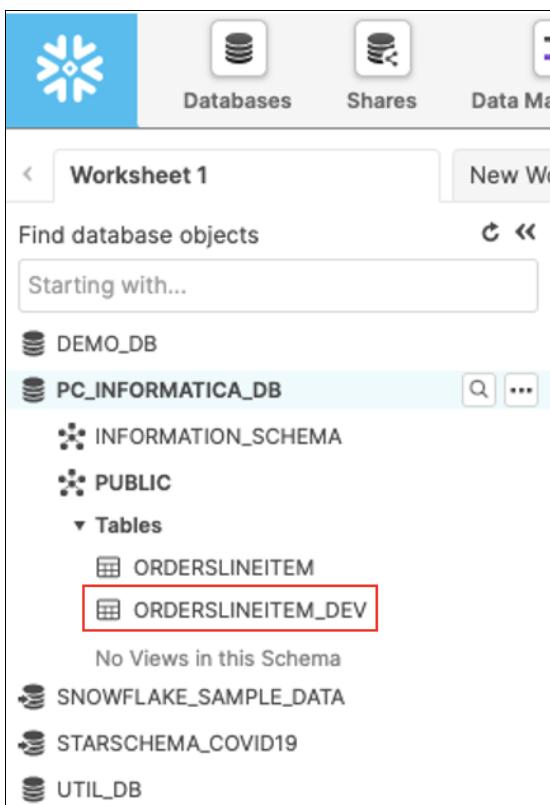


A massive benefit is that the underlying data is not copied; just the metadata/pointers to the underlying data change. Hence “zero-copy” and storage requirements are not doubled when data is cloned. Most data warehouses cannot do this; for Snowflake it is easy!

7.2.1 Run the following command in the worksheet to create a development (dev) table

```
create table orderslineitem_dev clone orderslineitem;
```

7.2.2 If closed, expand the database objects browser on the left of the worksheet. Click the small Refresh button in the left-hand panel and expand the object tree under the PC\_INFORMATICA\_DB database. Check that you can see a new table under the PC\_INFORMATICA\_DB database named ORDERSLINEITEM\_DEV. The development team now can do whatever they want with this table, including even deleting it, without having any impact on the ORDERSLINEITEM table or any other object.



## 7.3 Using Time Travel

Snowflake's Time Travel capability enables accessing historical data at any point within a pre-configurable period of time. The default period of time is 24 hours, and with Snowflake Enterprise Edition it can be up to 90 days. Most data warehouses cannot offer this functionality; with Snowflake it is easy!

Some useful applications of this include:

- Restoring data-related objects (tables, schemas, and databases) that may have been accidentally or intentionally deleted
- Duplicating and backing up data from key points in the past

- Analyzing data usage/manipulation over specified periods of time

## 7.4 Drop and Undrop a Table

First let's see how we can restore data objects that have been accidentally or intentionally deleted.

- 7.4.1 From the worksheet, run the following command which will drop (remove) the ORDERSLINEITEM table:

```
drop table orderslineitem;
```

- 7.4.2 Now run a SELECT statement on the ORDERSLINEITEM table. In the “Results” pane you should see an error because the underlying table has been dropped.

```
select * from orderslineitem limit 10;
```

The screenshot shows the 'Results' tab of a Snowflake interface. At the top, there are tabs for 'Results' (which is selected) and 'Data Preview'. Below the tabs, there are three columns: a red 'X' icon, 'Query ID', and 'SQL'. The 'SQL' column contains the command 'drop table orderslineitem;'. To the right of the SQL command is a progress bar labeled '18ms' and a status message 'SQL compilation error: Object 'ORDERSLINEITEM' does not exist or not authorized.'

- 7.4.3 Now restore the table:

```
undrop table orderslineitem;
```

- 7.4.4 The ORDERSLINEITEM table should be restored.

The screenshot shows the 'Results' tab of a Snowflake interface. At the top, there are tabs for 'Results' (selected) and 'Data Preview'. Below the tabs, there are three columns: a green checkmark '✓', 'Query ID', and 'SQL'. The 'SQL' column contains the command 'undrop table orderslineitem;'. To the right of the SQL command is a progress bar labeled '68ms' and a status message '1 rows'. Below the SQL area are buttons for 'Filter result...', 'Download', and 'Copy'. A table header row follows, with columns 'Row' and 'status'. The data row below shows '1' and 'Table ORDERSLINEITEM successfully restored.'

## 7.5 Roll Back a Table

Now let's look at rolling back a table to a previous state to fix an unintentional DML error that replaces all of the order priorities in the ORDERSLINEITEM table with the word “oops.”

- 7.5.1 Run the following command that replaces all of the order priorities in the table with the word “oops”.

```
update orderslineitem set "orderpriority" = 'oops';
```

- 7.5.2 Now run a query that returns the top order priorities by # of orders - notice how we've screwed up the order priorities so we only get one row:

```
select  
"orderpriority",  
count(*) as orders  
from orderslineitem  
group by 1  
order by 2 desc;
```

The screenshot shows a query results interface with the following details:

- Results tab selected.
- SQL: `select "orderpriority", count(*) as orders from orderslineitem group by 1 order by 2 desc;`
- Time: 219ms
- Rows: 1 rows
- Filter result... button
- Copy button
- Columns dropdown
- Table Data:

| Row | orderpriority | ORDERS |
|-----|---------------|--------|
| 1   | oops          | 150000 |

- 7.5.3 Normally, we would need to scramble and hope we have a backup lying around. But in Snowflake, we can simply run commands to find the query ID of the last UPDATE command & store it in a variable called \$QUERY\_ID...

```
set query_id =  
(select query_id from  
table(information_schema.query_history_by_session (result_limit=>5))  
where query_text like 'update%' order by start_time limit 1);
```

- 7.5.4 Then re-create the table as of before the update:

```
create or replace table orderslineitem as  
(select * from orderslineitem before (statement => $query_id));
```

- 7.5.5 Run the SELECT statement again to check that the station names have been restored:

```
select  
"orderpriority",  
count(*) as orders  
from orderslineitem  
group by 1  
order by 2 desc;
```

Results Data Preview [Open History](#)

✓ Query ID SQL 297ms 5 rows

Filter result... [Download](#) [Copy](#) Columns ▾

| Row | orderpriority   | ORDERS |
|-----|-----------------|--------|
| 1   | 5-LOW           | 30244  |
| 2   | 2-HIGH          | 30172  |
| 3   | 1-URGENT        | 30111  |
| 4   | 4-NOT SPECIFIED | 29910  |
| 5   | 3-MEDIUM        | 29563  |

Congratulations, you are now done with this lab! Now let's wrap things up.

## Summary & Next Steps

This tutorial was designed as a hands-on introduction to the Snowflake Cloud Data Platform, as well as to Informatica Intelligent Cloud Services (IICS) Data Integration functionalities that will help your journey to the cloud. It was designed to teach you how to use both services together, while showcasing some of their key capabilities and differentiators.

We covered how to navigate the Snowflake UI to work with virtual warehouses, set up IICS through Snowflake Partner Connect, load data from AWS S3 and perform a simple mapping with filter and aggregator transformations in IICS, transform semi-structured data, perform zero-copy cloning, and undo user errors in Snowflake.

We encourage you to continue with your free Snowflake trial and IICS by loading in your own sample or production data and by using some of the more advanced capabilities of Snowflake and IICS not covered in this lab. Some suggested next steps include:

- Continue to use your free IICS organization created through Partner Connect to load up to 1 billion rows per month into Snowflake for free, and to try out advanced features such as Push Down Optimization (PDO).
- Visit Informatica's Snowflake solution page at:  
<https://www.informatica.com/solutions/explore-ecosystems/snowflake.html>
- Read the “Definitive Guide to Maximizing Your Free Trial” document at:  
<https://www.snowflake.com/test-driving-snowflake-the-definitive-guide-to-maximizing-your-free-trial/>
- Attend a Snowflake virtual or in-person event to learn more about our capabilities and how customers use us <https://www.snowflake.com/about/events/>
- Contact Sales to learn more <https://www.snowflake.com/free-trial-contact-sales/>