
Before you begin: Log in to your Education sandbox

You have a shared sandbox environment to learn about, use, and experiment with MicroStrategy Web.

- 1 In your browser, navigate to **education.microstrategy.com/MicroStrategy/servlet/mstrWeb**.
- 2 Log in to the sandbox with your **MicroStrategy Account** credentials.
- 3 The MicroStrategy Landing page opens. Click the **MicroStrategy Tutorial** project.
- 4 Click **Go to MicroStrategy Web**.

The Shared Reports folder of the MicroStrategy Tutorial project opens.

In the sandbox environment, you cannot save items in the Shared Reports folder.

Exercise: Join datasets to determine the attribute elements that display

Pollinators such as bees are essential to food production, since three out of every four food crops worldwide depend on pollinators. Bee populations are in a steep decline, likely impacting the production and costs of fruits, vegetables, nuts, seeds, and oils pollinated by bees. HOBOS (HoneyBee Online Studies) teaches schoolchildren about the interconnections of different sciences, using the relationship between bees and agriculture. The project tracks the movements of bees equipped with microchips and monitors environmental measures such as the interior of beehives, climate, vegetation, and soil.

In this exercise, you review a week of data about:

- The number of departures and arrivals from a beehive (referred to as flow). Positive numbers are arrivals to the beehive; negative numbers are departures.
- The humidity of the beehive, as a percentage.
- The temperature of the beehive, in Celsius.

Each of the monitoring systems saves data in a CSV (comma separated values) file, and you must combine them to create a grid that displays the data for effective analysis, such as determining the time range of highest hive flow so decisions can be made about not disrupting hives during this time.

Create a new document and import datasets

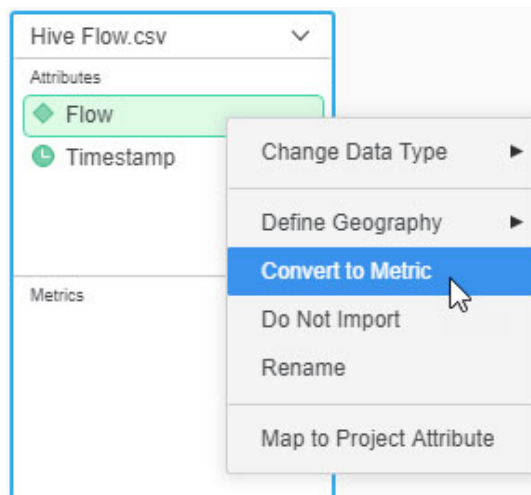
- 1 Access your Education sandbox. For a reminder on how to do this, see *Exercise: Access your Education sandbox*.
- 2 On the My Reports page in MicroStrategy Web, click **Create**, point to **New Document**, and select **01 Blank Dashboard**.

Add Hive Flow.csv as a dataset

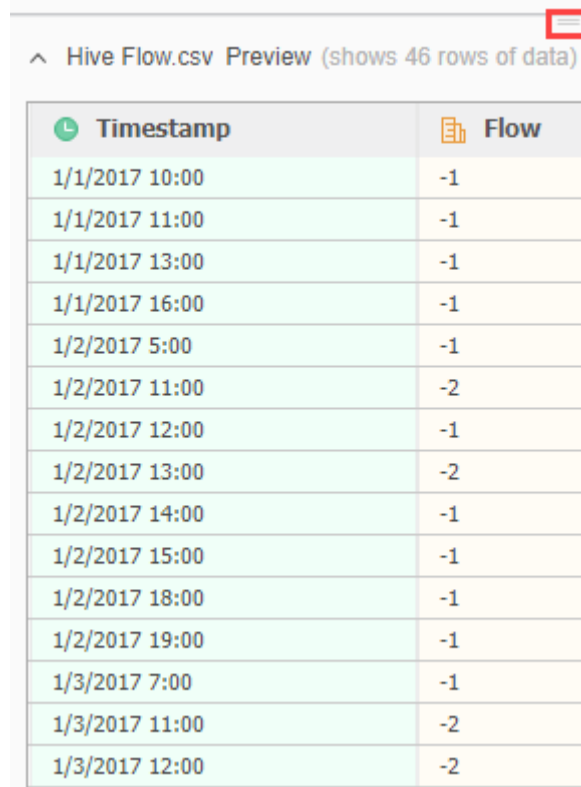
- 3 From the **Data** menu, select **Add Dataset**.
- 4 On the Select Dataset window, click **Add External Data**.
- 5 On the Connect to Your Data page, click **File from Disk**.
- 6 On the Upload your Files page, click **Choose Files**.
- 7 On the Open window, navigate to the folder where you saved the exercise files, select **Hive Flow.csv**, and click **Open**.

Change the Flow attribute to a metric

- 8 You are not sure if the data meets your company's guidelines, so you want to review it before adding it to your project. On the Upload your Files page, click **Prepare Data**.
- 9 On the Preview page, both objects in the file are characterized as attributes. Flow (the number of departures and arrivals from the beehive) should be a number, and therefore a metric. Right-click **Flow** in the table at the top of the page, and select **Convert to Metric**.



-
- 10** Review the preview of the data, at the bottom of the screen. You can expand the preview area by dragging its top border, as shown below.



^ Hive Flow.csv Preview (shows 46 rows of data)

Timestamp	Flow
1/1/2017 10:00	-1
1/1/2017 11:00	-1
1/1/2017 13:00	-1
1/1/2017 16:00	-1
1/2/2017 5:00	-1
1/2/2017 11:00	-2
1/2/2017 12:00	-1
1/2/2017 13:00	-2
1/2/2017 14:00	-1
1/2/2017 15:00	-1
1/2/2017 18:00	-1
1/2/2017 19:00	-1
1/3/2017 7:00	-1
1/3/2017 11:00	-2
1/3/2017 12:00	-2

Q: Is every hour of every day represented in the data?

A: No. Data was saved only when a bee entered or exited the hive.

- 11** Click **Finish**.

- 12** Save the file as **Hive Flow CSV** in the **My Reports** folder.

Add Hive Humidity.csv as a dataset

- 13** Follow the steps in *Add Hive Flow.csv as a dataset*, substituting **Hive Humidity.csv** as the file.

Preview the Hive Humidity data

- 14** On the Upload your Files page, click **Prepare Data**.

- 15** On the Preview page, review the preview of the data. If you wanted to view more than the first 50 rows of data, you could open the data wrangling page,
-

by clicking Data Wrangle at the top of the page. Fifty rows are sufficient in this scenario.

✓ Hive Humidity.csv Preview (shows first 50 rows of data)

Timestamp	Humidity
1/1/2017 5:00	92.40666667
1/1/2017 6:00	92.27
1/1/2017 7:00	92.575
1/1/2017 8:00	92.84
1/1/2017 9:00	93.64
1/1/2017 10:00	93.76833333
1/1/2017 11:00	92.95
1/1/2017 12:00	92.08
1/1/2017 13:00	91.35666667
1/1/2017 14:00	91.07166667
1/1/2017 15:00	90.99666667
1/1/2017 16:00	90.945
1/1/2017 17:00	90.97333333
1/1/2017 18:00	91.01333333
1/1/2017 19:00	91.18
1/1/2017 20:00	91.74166667
1/1/2017 21:00	91.57833333
1/1/2017 22:00	92.55333333

Q: Is every hour of every day represented in the data?

A: Yes, beginning at 5 AM on January 1.

16 Click **Finish**.

17 Save the file as **Hive Humidity CSV** in the **My Reports** folder.

Add Hive Temperature.csv as a dataset



18 Follow the steps in *Add Hive Flow.csv as a dataset*, substituting **Hive Temperature.csv** as the file.

Preview the Hive Temperature data

19 On the Upload your Files page, click **Prepare Data**.

20 On the Preview page, review the preview of the data.

^ Hive Temperature.csv Preview (shows first 50 rows of data)

 Timestamp	 Temperature
1/2/2017 10:00	0.316083333
1/2/2017 11:00	0.413983333
1/2/2017 12:00	0.865566667
1/2/2017 13:00	1.139533333
1/2/2017 14:00	1.44175
1/2/2017 15:00	1.389366667
1/2/2017 16:00	1.308833333
1/2/2017 17:00	1.489983333
1/2/2017 18:00	1.5144
1/2/2017 19:00	1.31995
1/2/2017 20:00	1.392816667
1/2/2017 21:00	1.302
1/2/2017 22:00	1.1772
1/2/2017 23:00	0.843483333
1/3/2017 0:00	0.766383333
1/3/2017 1:00	0.959766667
1/3/2017 2:00	0.706216667

Q: Is every hour of every day represented in the data?

A: No. Data collection did not begin until 10 AM on January 2. (The thermometer malfunctioned on New Year's Day, and the human monitoring it was recovering from New Year's Eve and did not notice until the next morning.)

21 Click **Finish**.

22 Save the file as **Hive Temp CSV** in the **My Reports** folder.


Save the document

23 Save the document as **Hive Monitoring** in the **My Reports** folder.

Create the grid to display all the hive data

1 Add a blank grid.

From the **Insert** menu, select **Grid**. Drag and drop in the layout area to position and size the grid.

- 2 Switch to Editable Mode, so that you can view the results as you create the grid. Click the **Editable Mode** icon  in the toolbar.
- 3 Add **Timestamp** from the Hive Flow CSV dataset to the grid.
- 4 Add the following metrics to the grid:
 - **Flow**
 - **Humidity**
 - **Temperature**

The grid looks like the following:

Timestamp	Metrics	Flow	Humidity	Temperature
1/1/2017 10:00:00 AM		-1	14668.91750001	-73.224549999
1/1/2017 11:00:00 AM		-1	14668.91750001	-73.224549999
1/1/2017 1:00:00 PM		-1	14668.91750001	-73.224549999
1/1/2017 4:00:00 PM		-1	14668.91750001	-73.224549999
1/2/2017 5:00:00 AM		-1	14668.91750001	-73.224549999
1/2/2017 11:00:00 AM		-2	14668.91750001	-73.224549999
1/2/2017 12:00:00 PM		-1	14668.91750001	-73.224549999
1/2/2017 1:00:00 PM		-2	14668.91750001	-73.224549999
1/2/2017 2:00:00 PM		-1	14668.91750001	-73.224549999
1/2/2017 3:00:00 PM		-1	14668.91750001	-73.224549999
1/2/2017 6:00:00 PM		-1	14668.91750001	-73.224549999

Q: Is the grid displaying correctly?

A: No, the Humidity and Temperature values calculate incorrectly. Because the grid uses the Timestamp attribute from the Hive Flow dataset, the Flow metric values are correct, while the Humidity and Temperature values are totaled. Notice that only the times recorded in the Hive Flow dataset display. Recall that humidity values started at 5 AM on 1/1, but that Timestamp value is not included.

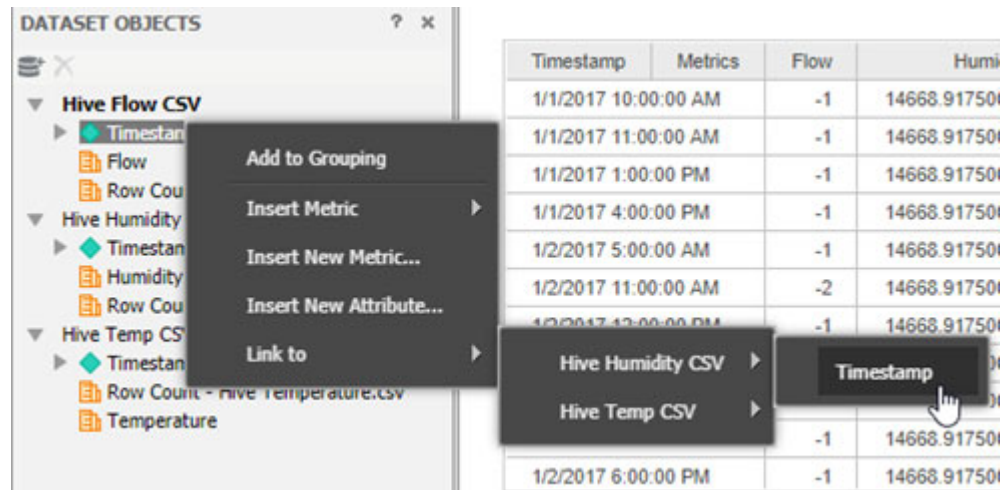
Q: What can you do to ensure that all the grid's values are included, and all the metrics calculate correctly?

A: Link the Timestamp attributes to create a relation between the datasets, allowing the datasets to be joined correctly.

- 5 Save the changes to the document, by clicking the **Save** icon .
-


Link the Timestamp attributes

- 1 In the Dataset Objects panel, right-click **Timestamp** in the Hive Flow CSV dataset, point to **Link to**, point to **Hive Humidity CSV**, and select **Timestamp**.



- 2 Click **OK** at the message to disable Flash Mode.

Flash Mode for document execution and MHT export has been deprecated.

The two Timestamp attributes display as linked attributes  in the Dataset Objects panel.

The grid now displays all the dates and times available in both datasets. You can see that the first value is from the Hive Humidity dataset, which contains hourly measures, beginning at 5 AM on 1/1. The blank cells in the Flow column occur because no bee entered or exited the hive in those hours so no data was

collected. The Humidity values calculate correctly, but Temperature still remains a total of all temperatures in the dataset.

DATASET OBJECTS				
-	Hive Flow CSV			
▶	Timestamp			
	Flow			
	Row Count - Hive Flow.csv			
-	Hive Humidity CSV			
▶	Timestamp			
	Humidity			
	Row Count - Hive Humidity.csv			
-	Hive Temp CSV			
▶	Timestamp			
	Temperature			
	Row Count - Hive Temperature.csv			

Timestamp	Metrics	Flow	Humidity	Temperature
1/1/2017 5:00:00 AM			92.40666667	-73.224549999
1/1/2017 6:00:00 AM			92.27	-73.224549999
1/1/2017 7:00:00 AM			92.575	-73.224549999
1/1/2017 8:00:00 AM			92.84	-73.224549999
1/1/2017 9:00:00 AM			93.64	-73.224549999
1/1/2017 10:00:00 AM		-1	93.76833333	-73.224549999
1/1/2017 11:00:00 AM		-1	92.95	-73.224549999
1/1/2017 12:00:00 PM			92.08	-73.224549999
1/1/2017 1:00:00 PM		-1	91.35666667	-73.224549999
1/1/2017 2:00:00 PM			91.07166667	-73.224549999
1/1/2017 3:00:00 PM			90.99666667	-73.224549999

◀ ◁ 1 2 3 4 of 4 pages ▶ ▷

- 3 Link the remaining dataset. In the Dataset Objects panel, right-click **Timestamp** in the Hive Flow CSV dataset, point to **Link to**, point to **Hive Temp CSV**, and select **Timestamp**.

Scroll down to 1/2/2017 so that you can see temperature values. The grid now displays the correct Temperature values, and all Timestamp values for all datasets still display.

DATASET OBJECTS			
▼ Hive Flow CSV			
▶		Timestamp	
		Flow	
		Row Count - Hive Flow.csv	
▼ Hive Humidity CSV			
▶		Timestamp	
		Humidity	
		Row Count - Hive Humidity.csv	
▼ Hive Temp CSV			
▶		Timestamp	
		Row Count - Hive Temperature.csv	
		Temperature	

1/2/2017 6:00:00 AM		92.58166667	
1/2/2017 7:00:00 AM		92.57833333	
1/2/2017 8:00:00 AM		92.38	
1/2/2017 9:00:00 AM		92.905	
1/2/2017 10:00:00 AM		93.25833333	0.316083333
1/2/2017 11:00:00 AM	-2	93.28166667	0.413983333
1/2/2017 12:00:00 PM	-1	93.48666667	0.865566667
1/2/2017 1:00:00 PM	-2	93.29833333	1.139533333
1/2/2017 2:00:00 PM	-1	93.18333333	1.44175
1/2/2017 3:00:00 PM	-1	92.82166667	1.389366667
1/2/2017 4:00:00 PM		92.46166667	1.308833333
1/2/2017 5:00:00 PM		92.605	1.489983333

◀ ◁ 1 2 3 4 ▶ ▷ of 4 pages

- 4 Save** the document.

Q: Why do all the Timestamp values display, no matter which dataset they come from?

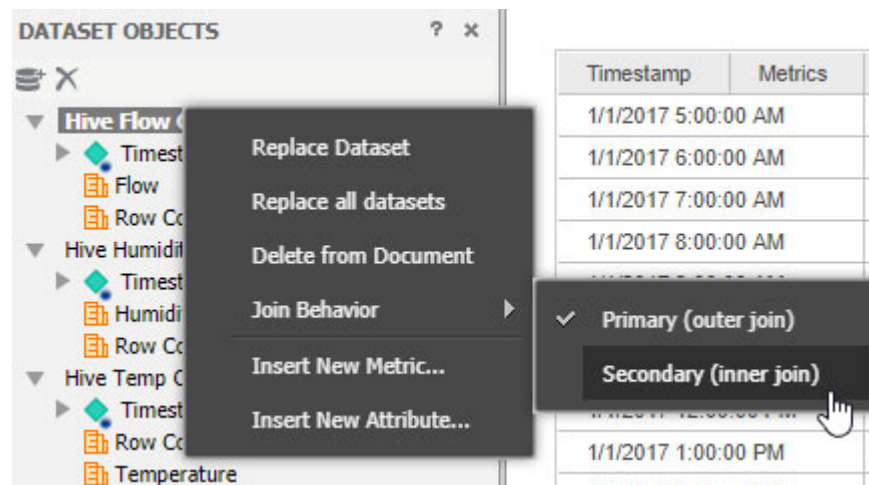
A: All the datasets are, by default, primary datasets. Primary datasets are joined with compound joins, which results in a grid with a row for each Timestamp value in the datasets.

Q: You need to display only the date/times that have flow, humidity, and temperature values. How can you accomplish this?

A: Change all the datasets to be secondary datasets, which use inner joins. When a document contains all secondary datasets, only the rows that are in all the datasets display.

Change the dataset joins

- 1 In the Dataset Objects panel, right-click **Hive Flow CSV**, point to **Join Behavior**, and select **Secondary (inner join)**.



Q: The grid does not change. Why?

A: Hive Humidity contains all the attribute data points, unlike the other two datasets. Since it is still a primary dataset, all the rows/dates and times still display in the grid.

Q: When you set Hive Humidity CSV to be a secondary dataset, will the grid change? And why?

- 2 In the Dataset Objects panel, right-click **Hive Humidity CSV**, point to **Join Behavior**, and select **Secondary (inner join)**.

A: The grid changes and no longer displays all the Timestamp values. Only those rows with Temperature values show, since the Hive Temperature CSV dataset is still a primary dataset.

Timestamp	Metrics	Flow	Humidity	Temperature
1/2/2017 10:00:00 AM			93.25833333	0.316083333
1/2/2017 11:00:00 AM		-2	93.28166667	0.413983333
1/2/2017 12:00:00 PM		-1	93.48666667	0.865566667
1/2/2017 1:00:00 PM		-2	93.29833333	1.139533333
1/2/2017 2:00:00 PM		-1	93.18333333	1.44175
1/2/2017 3:00:00 PM		-1	92.82166667	1.389366667
1/2/2017 4:00:00 PM			92.46166667	1.308833333
1/2/2017 5:00:00 PM			92.605	1.489983333
1/2/2017 6:00:00 PM		-1	93.15	1.5144
1/2/2017 7:00:00 PM		-1	92.54333333	1.31995
1/2/2017 8:00:00 PM			92.30333333	1.392816667

◀ ◁ 1 2 3 of 3 pages ▶ ▷

- 3 In the Dataset Objects panel, right-click **Hive Temp CSV**, point to **Join Behavior**, and select **Secondary (inner join)**.

Now that all the datasets are secondary, only the dates/times that have flow, humidity, and temperature values display.

Timestamp	Metrics	Flow	Humidity	Temperature
1/2/2017 11:00:00 AM		-2	93.28166667	0.413983333
1/2/2017 12:00:00 PM		-1	93.48666667	0.865566667
1/2/2017 1:00:00 PM		-2	93.29833333	1.139533333
1/2/2017 2:00:00 PM		-1	93.18333333	1.44175
1/2/2017 3:00:00 PM		-1	92.82166667	1.389366667
1/2/2017 6:00:00 PM		-1	93.15	1.5144
1/2/2017 7:00:00 PM		-1	92.54333333	1.31995
1/3/2017 7:00:00 AM		-1	92.27333333	1.239883333
1/3/2017 11:00:00 AM		-2	91.80166667	1.583283333
1/3/2017 12:00:00 PM		-2	91.57666667	2.039666667
1/3/2017 2:00:00 PM		-1	91.55833333	2.009666667
1/3/2017 4:00:00 PM		-1	91.425	1.953383333

- 4 **Save** the document.
-