Build Sequence of Datasets

In some situations, you may need to create a sequence of datasets, in which the output of one recipe becomes the input of another recipe.

Potential uses:

1. You may want to handle data cleanup tasks in one set, before that data is made available to other users for customization for their needs.
2. Columns or rows of data may need to be dropped before the dataset is made available to other users.
3. You may want to have different individuals working on each phase of the data transformation process. For example, one individual may be responsible for cleansing the data, while another may be responsible for transforming the data into final format.

Depending on your situation, you can apply one of the following solutions.

Chain Recipes in Same Flow

Within a flow, you can chain together recipes. For example, you may wish to use the first recipe for cleansing and then second recipe for transforming. This method is useful if you are using a single imported dataset for multiple types of transformations within the same flow.

Steps:

1. Click the imported dataset. Click **Add new recipe**.
2. Click the new recipe. Name it, `Cleanse`.
3. With the new recipe selected, click **Add new recipe**.
4. Click the new recipe. Name it, `Transform`.

The output of `Cleanse` recipe becomes the input of `Transform` recipe.

Figure: Chained recipes

Create Reference Objects

If you need to make the output of a recipe available in other flows, you can create a reference object. This reference is available in other flows that you control.

Steps:

1. In Flow View, select the recipe whose output you wish to make available to other flows.

2. Click the Create Reference icon:

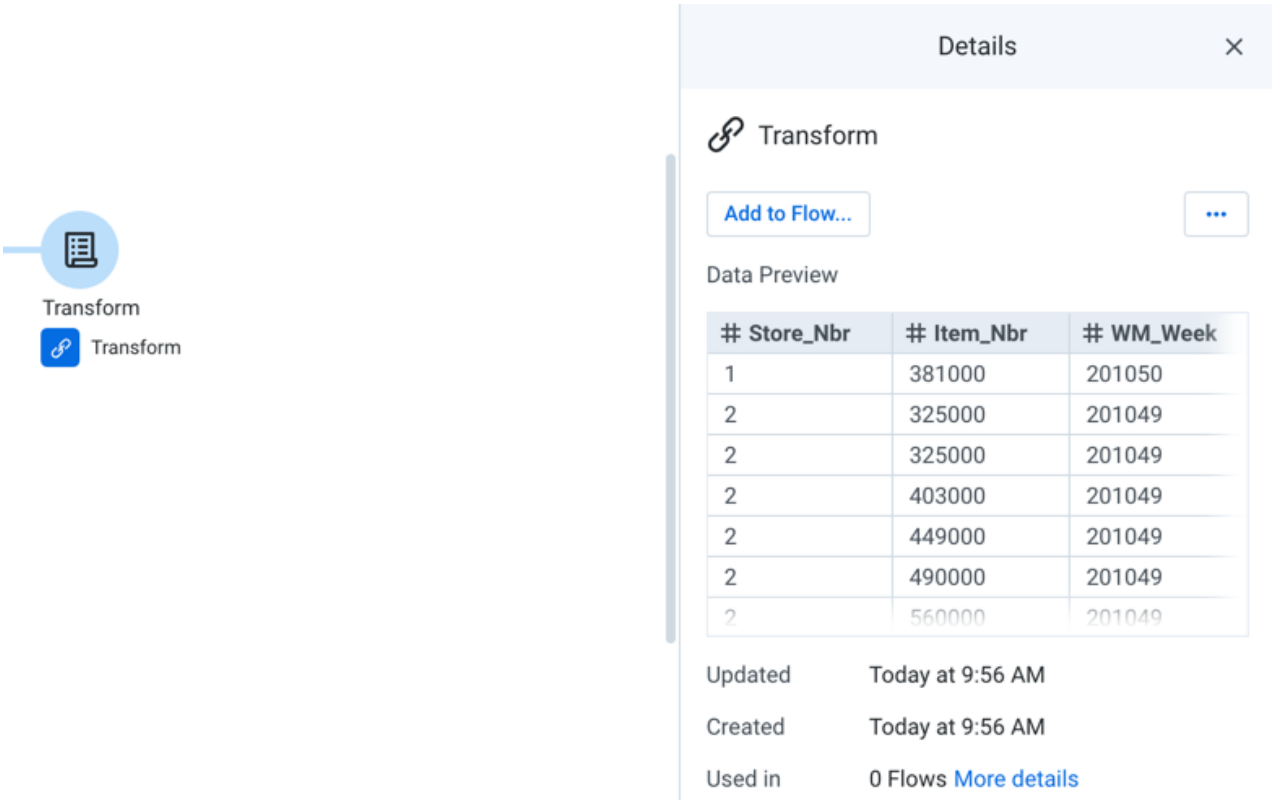

Figure: Create reference object

3. To use it in one of your other flows, click **Add to Flow...**.
4. In the target flow, the reference object appears as a **reference dataset**. It works like an imported dataset with the following considerations.

Key Considerations:

- When you run a job in a flow that contains a reference dataset, all upstream dependencies of that reference dataset are executed. For the source reference object, all imported datasets and recipes are gathered and executed to populate the reference dataset with fresh data.
- The above has the following implications:
  - If the user running the job in flow #2 does not have permissions to access all of the upstream dependencies of the reference dataset, the job may fail. These dependencies include imported datasets and any connections.
  - If the upstream objects are owned by other users, you may not be able to review these items. For example, if the source recipe is changed by another user, your downstream recipe may break without notice. If you cannot review that recipe, then you can see what was changed and how to fix it.

Create Imported Dataset from Output

If any of the above considerations are a concern, you can create an imported dataset from the job results of flow #1.

In the Job Details page, click the Output Destinations tab. For the generated output, select **Create imported dataset** from its context menu.From the results of wrangling your first dataset, you can create a new dataset. This dataset is wrangled in a separate recipe, the output of which can become a third dataset. In this manner, you can create sequences of datasets.

Key Considerations:

- The imported dataset in flow #2 is not refreshed until you run the job that generates it in flow #1.
- If the output of flow #1 uses the same filename each time, you may not know if the data has been refreshed. When the job is executed in flow #2, it collects the source imported dataset and executes, whether the data is new or not. Workarounds:
  -
  - After the job in flow #2 executes, rename or remove the output of flow #1 from its target location. That way, whenever job #2 executes again, any data that it collects from the source location is likely to be newer.

See *Job Details Page*.