

Create Branching Outputs

From a single collection of datasets, you may need to generate multiple outputs for downstream purposes. Examples:

- You want to preserve the ability to review and profile your source data. For more information, see *Profile Your Source Data*.
- You need different pivot tables produced from the wrangled data.
- You need to filter down the set of rows or columns to deliver to one user community while delivering a different set of columns to another.

Reshaping Transformations

If your next step is to add any of the following transformations and you wish to preserve the existing data for other uses, you should consider adding these steps in a separate dedicated recipe.

| Transformation Name | Description |
|---------------------|--|
| Union | A union appends one or more datasets to your current one. To preserve the original, you may need to create a branching output. See <i>Union Page</i> . |
| Join | A join combines two datasets based on common values in specified columns in both datasets. These types of transformations can greatly change the shape of your data. See <i>Join Window</i> . Similarly, a lookup uses values from a column in your source data to pull in corresponding rows of data from a reference dataset. These transformations add columns to your dataset. See <i>Add Lookup Data</i> . |
| Deduplicate | This transformation removes identical rows from your dataset. However, there may be a set of steps required to standardize values in various columns before applying the de-duplication. You may choose to manage this process in a branching recipe. |
| Delete columns | When a column is removed, it is no longer available for use in any downstream output. See <i>Remove Data</i> . |
| Filter | Rows can be filtered from your dataset to render different perspectives. These changes may be best moved to a secondary, branching recipe. See <i>Filter Data</i> . |
| Pivot data | When you create a pivot table, all source data that is not explicitly specified in the pivot is dropped from the dataset. For more information, see <i>Pivot Data</i> . |
| Group by | You can perform aggregation calculations within a table, which may force column data to be dropped. See <i>Create Aggregations</i> . |

Basic Technique

Whenever you are applying a transformation that destroys data or otherwise reshapes your dataset and you wish to preserve the current state of the dataset, you should do the following:

1. In Flow View, select your current recipe. Click **Add new recipe**.
2. This recipe becomes the source for a branched output. Give the new recipe an appropriate name. For example, `Pivot-SalesPerProductPerStore`.
3. For this recipe, click the Output icon. Specify the appropriate output format and location that you'd like to generate for this branched output.
4. Select your current recipe again. Click **Add new recipe**.
5. This recipe becomes the extension of your current recipe. Give the new recipe an appropriate name. For example, `MyRecipe-Part2`.
6. Select the `Pivot-SalesPerProductPerStore` recipe. Click **Edit recipe**.
7. Build your pivot transformation in this recipe.
8. When ready, run the job. The output should be generated in the appropriate format and location.

The screenshot displays a data flow interface for a project named "2013 POS". The main workspace shows a primary recipe node "POS-r01-Part1" which is connected to several other nodes: "POS-r01-source.txt", "ValuesToCols", "Pivot-SalesPerProductByStore", "POS-r01-Part2", and another "ValuesToCols" node. The "Pivot-SalesPerProductByStore" node is highlighted with a blue circle and a blue arrow, indicating it is the selected node. On the right side, a "Details" panel is open for the "Pivot-SalesPerProductByStore" node. This panel includes a "Run Job" button, a "Destinations" tab, and sections for "Manual Destinations" and "Scheduled Destinations". The "Manual Destinations" section shows a "Create-CSV" destination with a path to an S3 bucket and a "Trifacta Server" environment. The "Scheduled Destinations" section has an "Add" button and a description: "Add a scheduled destination to automatically run the Output when the flow is executed by a schedule."

Figure: Multiple pivot tables sourced from output of a primary recipe for the flow. POS-r01-Part2 can be used for continued wrangling of primary recipe.