Object Overview

Contents:

Flow Structure and Objects

Within Trifacta® Wrangler, the basic unit for organizing your work is the flow. The following diagram illustrates the component objects of a flow and how they are related:
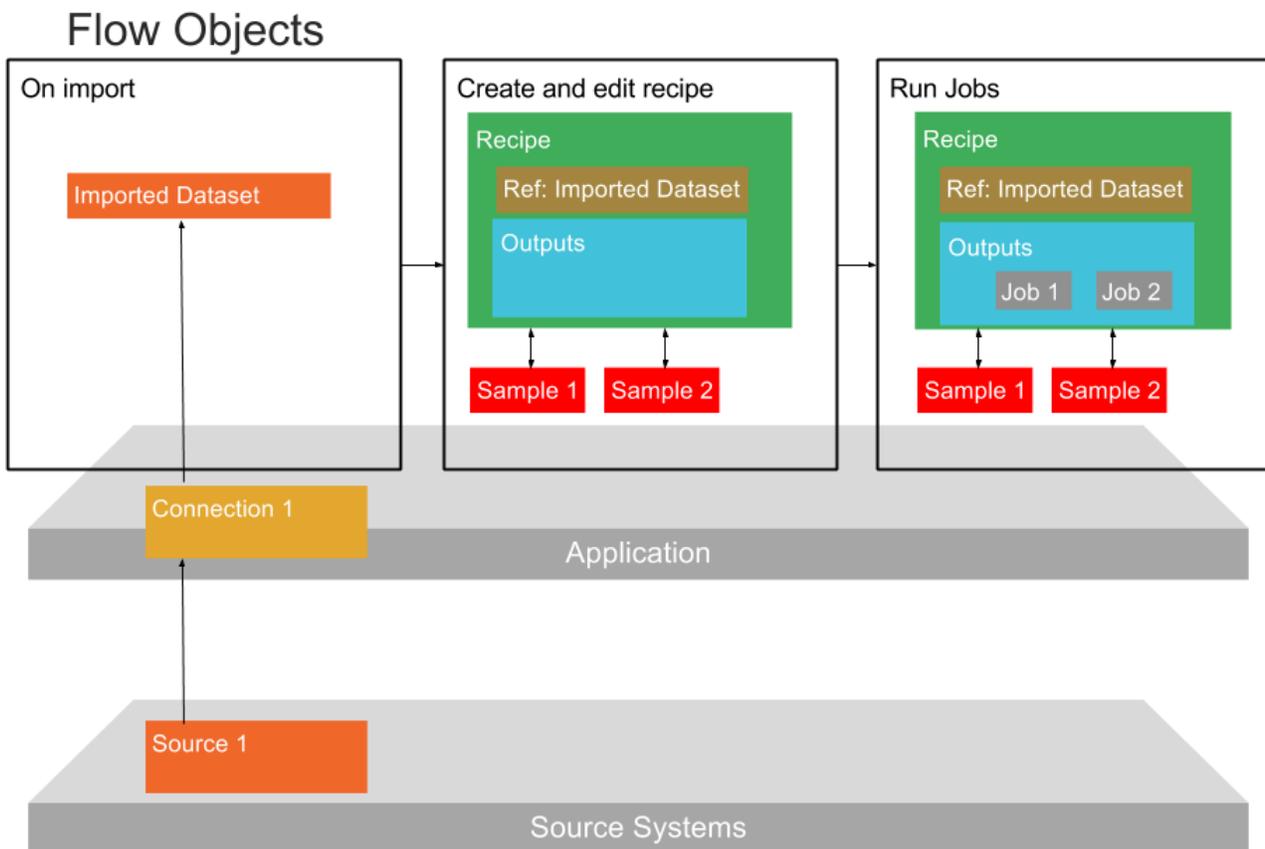

Figure: Objects in a Flow

Flow

A **flow** is a container for holding one or more datasets, associated recipes and other objects. This container is a means for packaging Trifacta objects for the following types of actions:

- Creating relationships between datasets, their recipes, and other datasets.
-
- Copying
- Execution of pre-configured jobs
- Creating references between recipes and external flows
-

Imported Dataset

Data that is imported to the platform is referenced as an imported dataset. An **imported dataset** is simply a reference to the original data; the data does not exist within the platform. An imported dataset can be a reference to a file, multiple files, database table, or other type of data.

> **NOTE:** An imported dataset is a pointer to a source of data. It cannot be modified or stored within Trifacta Wrangler.

> **NOTE:** External sources of data are not supported in Trifacta Wrangler. All sources must be uploaded files.

- An imported dataset can be referenced in recipes.
- Imported datasets are created through the *Import Data Page*.
- For more information on the process, see *Import Basics*.

After you have created an imported dataset, it becomes usable after it has been added to a flow. You can do this as part of the import process or later.

Recipe

A **recipe** is a user-defined sequence of steps that can be applied to transform a dataset.

- A recipe object is created from an imported dataset or another recipe. You can create a recipe from a recipe to chain together recipes.
- Recipes are interpreted by Trifacta Wrangler and turned into commands that can be executed against data.
- When initially created, a recipe contains no steps. Recipes are augmented and modified using the various visual tools in the *Transformer Page*.
- For more information on the process, see *Transform Basics*.

In a flow, the following objects are associated with each recipe, which are described below:

- Outputs
- References

Outputs and Publishing Destinations

Outputs contain one or more publishing destinations, which define the output format, location, and other publishing options that are applied to the results generated from a job run on the recipe.

When you select a recipe's output object in a flow, you can:

- Define the publishing destinations for outputs that are generated when the recipe is executed. **Publishing destinations** specify output format, location, and other publishing actions. A single recipe can have multiple publishing destinations.
- Run an on-demand job using the specified destinations. The job is immediately queued for execution.

References and Reference Datasets

References allow you to create a reference to the output of the recipe's steps in another dataset. References are not depicted in the above diagram.

When you select a recipe's reference object, you can add it to another flow. This object is then added as a reference dataset in the target flow. A **reference dataset** is a read-only version of the output data generated from the execution of a recipe's steps.

Working with recipes

Recipes are edited in the Transformer page, which provides multiple methods for quickly selecting and building recipe steps.**Macros:** As needed, you can create reusable sequences of steps that can be parameterized for use in other recipes. For more information, see *Overview of Macros*.

Run Jobs: When you are satisfied with the recipe that you have created in the Transformer page, you can execute a **job**. A job may be composed of one or more of the following job types:

- **Transform job:** Executes the set of recipe steps that you have defined against your sample(s), generating the transformed set of results across the entire dataset.
- **Profile job:** Optionally, you can choose to generate a visual profile of the results of your transform job. This visual profile can provide important feedback on data quality and can be a key for further refinement of your recipe.
- When a job completes, you can review the resulting data and identify data that still needs fixing. See *Job Details Page*.
- For more information on the process, see *Running Job Basics*.

Flow Example

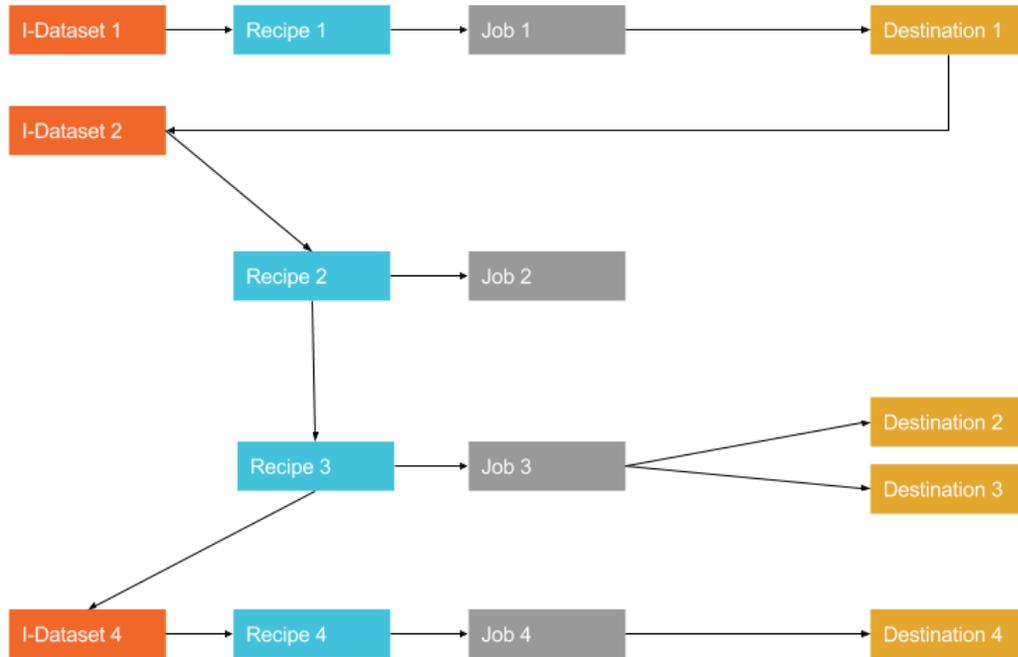The following diagram illustrates the flexibility of object relationships within a flow.

Flow Example

Figure: Flow Example

Type	Datasets	Description
Standard job execution	Recipe 1/Job 1	Results of the job are used to create a new imported dataset (I-Dataset 2). See *Job Details Page*.
Create dataset from generated results	Recipe 2/Job 2	Recipe 2 is created off of I-Dataset 2 and then modified. A job has been specified for it, but the results of the job are unused.
Chaining datasets	Recipe 3/Job 3	Recipe 3 is chained off of Recipe 2. The results of running jobs off of Recipe 2 include all of the upstream changes as specified in I-Dataset 1/Recipe1 and I-Dataset 2/Recipe 2.
Reference dataset	Recipe 4/Job 4	I-Dataset 4 is created as a reference off of Recipe 3. It can have its own recipe, job, destinations, and results.

Flows are created in the Flows page. See *Flows Page*.