Object Overview

**Contents:**

- *Flow Structure and Objects*
  - *Flow*
  - *Imported Dataset*
  - *Recipe*
  - *Flow Example*
- *Working with recipes*
- *Connections*
- *Flow Schedules*
- *Plans*

Explore the objects that you create and their relationships. Flows, imported datasets, and recipes are created to transform your sampled data. After you build your output objects, you can run a job to transform the entire dataset based on your recipe and deliver the results according to your output definitions.

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:
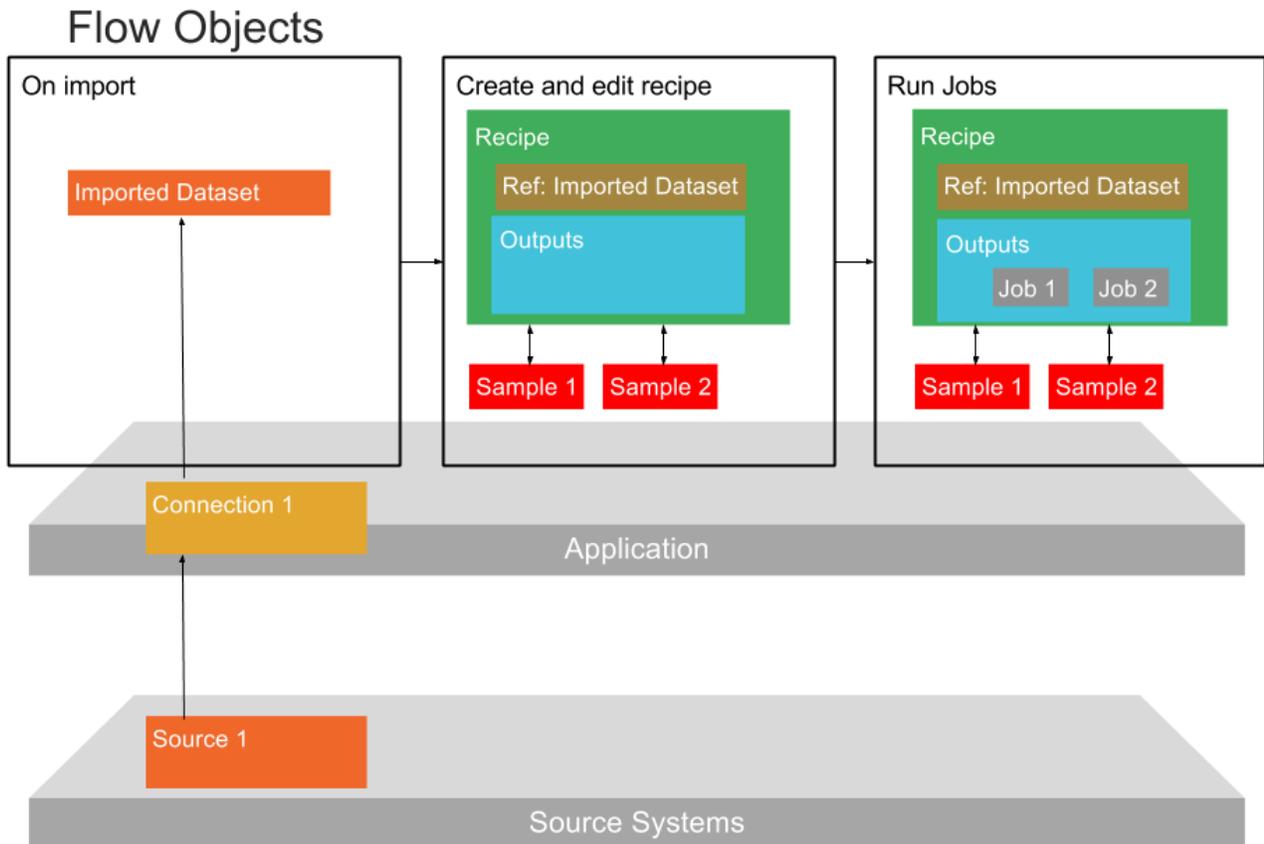
Figure: Objects in a Flow

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.

- 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.

Reference Datasets

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.

Flow Example

The following diagram illustrates the flexibility of object relationships within a flow.
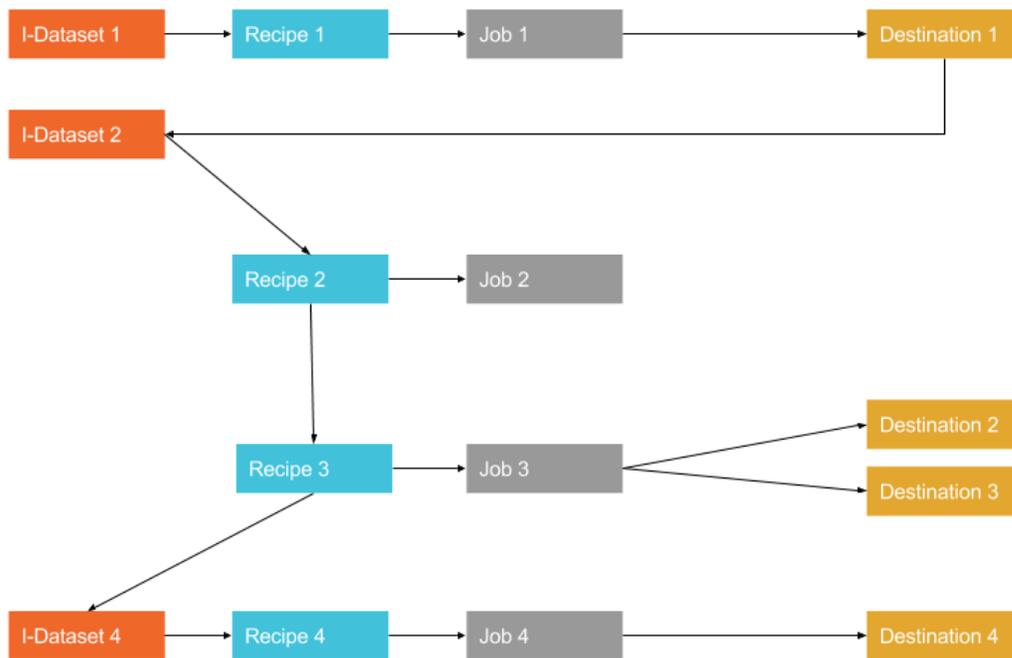
Figure: Flow Example

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) from the 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.

Working with recipes

Recipes are edited in the Transformer page, which provides multiple methods for quickly selecting and building recipe steps.

Samples: Within the Transformer page, you build the steps of your recipe against a **sample** of the dataset.

- A sample is typically a subset of the entire dataset. For smaller datasets, the sample may be the entire dataset.
- As you build or modify your recipe, the results of each modification are immediately reflected in the sampled data. So, you can rapidly iterate on the steps of your recipe within the same interface.
- As needed, you can generate additional samples, which may offer different perspectives on the data.
- See *Sampling Basics*.

Macros: As needed, you can create reusable sequences of steps that can be parameterized for use in other recipes.

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 in the Job Details page.
- For more information on the process, see *Running Job Basics*.

Connections

A **connection** is a configuration object that provides a personal or global integration to an external datastore. Reading data from remote sources and writing results are managed through connections.

- Connections are not associated with individual datasets or flows.
    - Connections are not reflected in the above diagram.
- Most connections can be created by individual users and shared as-needed.
- Depending on the datastore, connections can be read-only, write-only, or both.
- Connections are created in the Connections page.

Flow Schedules

You can associate a schedule with a flow. A **schedule** is a combination of one or more triggers and the outputs that are generated from them.

> **NOTE:** A flow can have only one schedule associated with it.

- A **trigger** is a scheduled time of execution. When a trigger's time occurs, all of the scheduled output destinations are queued for generation.
    - A schedule can have multiple triggers associated with it. Therefore, a flow can be scheduled for execution at multiple intervals.
- A **scheduled destination** is an output associated with a recipe. This output is generated only when the schedule for the flow is triggered.
    - A scheduled destination is not tied to a specific trigger. When a trigger occurs, all scheduled destinations in the flow are generated.
    - A scheduled destination generates one or more publishing actions (outputs) from the recipe when triggered.

- A recipe can have only one scheduled destination.
- Each recipe in a flow can have a scheduled destination.
- If a flow has a trigger but no scheduled destination, nothing is generated at trigger time.

Below, you can see the object hierarchy within a schedule.

```
+ schedule for Flow 1
  + trigger 1
  + trigger 2
  + scheduled destination a
  + scheduled destination b
+ schedule for Flow 2
  + trigger 3
  + scheduled destination c
  + scheduled destination d
```

Schedules are created for a flow through Flow View page.

Plans

A **plan** is a sequence of triggers and tasks that can be executed across multiple flows. A plan is executed on a snapshot of all objects at the time that the plan is triggered.

- A **task** is an executable action that is taken as part of a plan's sequence. For example, task #1 could be to execute a flow that imports all of your source data. Task #2 executes the flow that cleans and combines that data. Example task types:
  - A **flow task** is the execution of the recipes in a specified flow, which result in the generation of one or more selected outputs.
    - A **trigger** for a plan is the schedule for its execution.
    - A **snapshot** is a frozen image of the plan. This snapshot of the plan defines the objects that are executed as part of a plan run.
  - An **HTTP task** is a request submitted by the product to a third-party server as part of the sequence of tasks in a plan. For example, an HTTP task could be the submission of a message to a channel in your enterprise's messaging system.

Plans are created through the Plans page.