Overview of Scheduling

Contents:

As needed, you can schedule the execution of recipes in your flows on a recurring basis. For example, if the source file of your flow is updated outside of the application on a weekly basis, you can define a schedule to execute the recipe associated with the related imported dataset after the data has been refreshed. When the scheduled job successfully executes, you can collect the wrangled output in the specified output location, where it is available in the published form that you have specified.

To schedule a job, you must create the following configuration objects:

1. **Define a schedule** - For each flow you can define a schedule. A **schedule** specifies one or more recurring times when scheduled jobs for the flow are executed. For example, in a single schedule, you can specify daily execution times for incremental updates and monthly execution times for rollups.

> ✅ **Tip:** The scheduler supports a modified form of cron job syntax. For more information, see *cron Schedule Syntax Reference*.

2. **Define one or more scheduled destinations** - When you specify a **scheduled destination** for a recipe, the recipe is executed whenever one of the schedule's execution times occurs. Scheduled destinations are specified like regular destinations in flow view.

Limitations

- You can create one schedule per flow. A schedule may have one or more execution times.
- You cannot create schedules for individual wrangled datasets within a flow.
- Only a flow owner can create or modify a flow's schedule.

Data Management

> ⓘ **NOTE:** Since scheduled destinations are re-populated with each scheduled execution, you must determine how you wish to manage the data that is published to each location. Data management should be done outside of Trifacta.

- **Import:** Before each scheduled execution, you should refresh the source of the imported dataset with new data outside of Trifacta.
- **Execution:** Please verify that the publishing settings for your scheduled destination are consistent with how you are using the results. For example, if the scheduled destination creates a new file with the same name for each execution (replace), you must move the generated file out of the output location before the next scheduled execution.

- **Output:** You must collect the generated results. While you can export the job's results through the Jobs page, you may find it easier to use an external scheduler to gather the results and forward to the downstream consumer of them.

Flows for scheduling

✓ **Tip:** When a schedule is executed, all outputs in a flow are generated, even if they are unused. For better performance on larger flows, you can create a separate flow that contains only the references back to the objects in the source flow that you wish to have scheduled. As an additional benefit, this separation keeps development and scheduled execution in separate flows.

Schedule a Job

Schedules and scheduled destinations are defined through Flow View.

✓ **Tip:** You can create schedules for datasets with parameters and apply overrides through Flow View at runtime. See *Flow View Page*.

For more information, see *Schedule a Job*.

Track job execution

You can monitor a scheduled job like any other job in the application. See *Jobs Page*.

Configure

See *Configure Scheduling*.

Automation

Automation of scheduled jobs can be managed outside of the user interface using the following methods.

via CLI

You can use an external scheduler to manage the repeated executions of a specified job. See *CLI Example - Parameterize Job Runs*.

via API

Not supported in this release.