Diagnose Failed Jobs

Contents:

Job Types

The following types of jobs can be executed in Trifacta® Wrangler Pro:

- **Transform job:** This type of job executes the steps in your recipe against the dataset to generate results in the specified format. When you configure your job, any set of selected output formats causes a transform job to execute according to the job settings.
- **Profile job:** This type of job builds a visual profile of the generated results. When you configure your job, select **Profile Results** to generate a profile job.

- **Publish job:** This job publishes results generated by the platform to a different location or datastore.
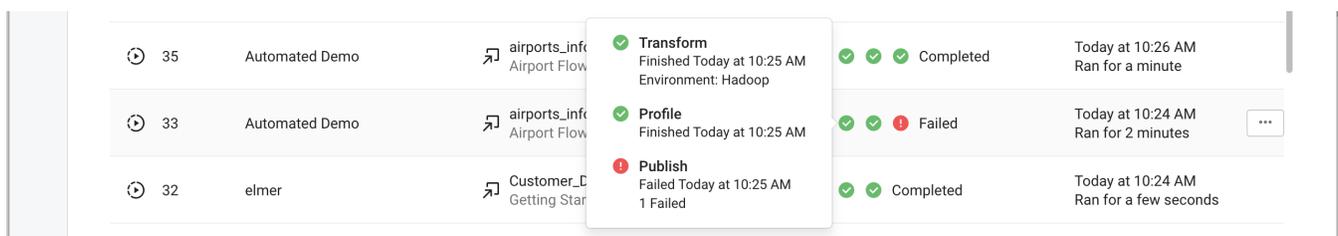
For more information, see *Run Job Page*.

Identify Job Failures

When a job fails to execute, a failure message appears in following locations:

- Jobs tab in Flow View. See *Flow View Page*.
- Listing in the Jobs page. See *Jobs Page*.

The following is an example from the Jobs page:

Figure: Publish job failed

In the above example, the Transform and Profile jobs completed, but the Publish job failed. In this case, the results exist and, if the source of the problem is diagnosed, they can be published separately.

Jobs that Hang

In some cases, a job may stay in a pending state indefinitely. Typically, these errors are related to a failure of the job tracking service. You can try to the following:

- Resubmit the job.

- Submit the job again.

Cloud-based EMR Job Errors

These errors can occur when you are running an EMR job from a cloud-based product edition.

"Runtime AWS" error

There was an error running your job.

"Runtime timeout" error

Your job was terminated for reaching the timeout minutes limit in the product. Job execution time is affected by the size of your input dataset(s) and the complexity of your recipe(s).

Tips:

- Review your recipes to see if you can identify ways to break them up into smaller recipes.
- Operations such as joins and unions can greatly increase the size of your datasets.

Try Other Job Options

You can try to re-execute the job using different options.

Tips:

- **Look to cut data volume.** Some job failures occur due to high data volumes. For jobs that execute across a large dataset, you can re-examine your data to remove unneeded rows and columns of data. Use the Deduplicate transformation to remove duplicate rows. See *Remove Data*.

- **Gather a new sample.** In some cases, jobs can fail when run at scale because the sample displayed in the Transformer page did not include problematic data. If you have modified the number of rows or columns in your dataset, you can generate a new sample, which might illuminate the problematic data. However, gathering a new sample may fail as well, which can indicate a broader problem. See *Samples Panel*.

- **Change the running environment.** If the job failed on Trifacta Photon, try executing it on Spark.

> **Tip:** The Trifacta Photon running environment is not suitable for jobs on large datasets. You should accept the running environment recommended in the Run Job page.

Contact Support

If you are unable to diagnose your job failure, please contact *Trifacta Support*.

> **NOTE:** When you contact support about a job failure, please be sure to download and include the entire zip file, your recipe, and (if possible) your dataset.

Report an Issue

If you believe that your job has failed due to an issue with Trifacta Wrangler Pro, select **Help menu > Report issue** to alert *Trifacta Support*.