

Diagnose Failed Jobs

Contents:

- *Job Types*
- *Identify Job Failures*
 - *Jobs that Hang*
- *Try Other Job Options*
- *Review Logs*
 - *Hadoop logs*

Job Types

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

- **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.
- **Ingest job:** This job manages the import of large volumes of data from a JDBC source into the default datastore for purposes of running a transformation job.

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:

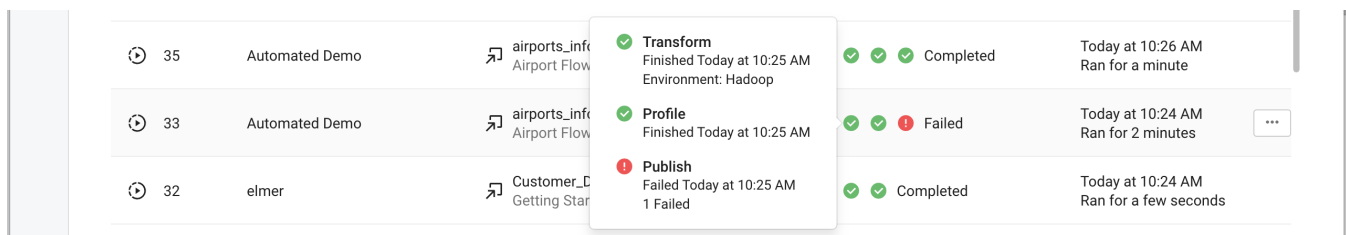


Figure: Publish job failed

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. From the job's context menu, select **Download Logs**. You can download the jobs logs to look for reasons for the failure. See *Review Logs* below.

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.
- Have an administrator restart the platform. See *Start and Stop the Platform*.
- Submit the job again.

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` transform 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 Photon, try executing it on Spark.

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

Review Logs

Job logs

In the listing for the job on the Jobs page, click **Download Logs** to send the job-related logs to your local desktop.

When you unzip the ZIP file, you should see a numbered folder with the internal identifier for your job on it. If you executed a transform job and a profile job, the ZIP contains two numbered folders with the lower number representing the transform job.

`job.log`. Review this log file for information on how the job was handled by the application.

Tip: Search this log file for `error`.

Trifacta node logs

NOTE: You must be an administrator to access these logs.

On the Trifacta node, these logs are located in the following directory:

```
<install_dir>/logs
```

This directory contains the following logs:

- `batch-job-runner.log`. This log contains vital information about the state of any launched jobs.
- `webapp.log`. This log monitors interactions with the web application.

Issues related to jobs running locally on the Trifacta Photon running environment can appear here.

Hadoop logs

In addition to these logs, you can also use the Hadoop job logs to troubleshoot job failures.

- You can find the Hadoop job logs at port 50070 or 50030 on the node where the ResourceManager is installed.
- The Hadoop job logs contain important information about any Hadoop-specific errors that may have occurred at a lower level than the Trifacta application, such as JDK issues or container launch failures.

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.