

File Settings

When you generate file-based results, you can configure the filename, storage format, compression, number of files, and the updating actions in the right-hand panel.

NOTE: By default, when scheduled or API jobs are executed, no validations are performed of any write settings objects for file-based outputs. Issues with these objects may cause failures during transformation or publishing stages of job execution. Jobs of these types should be tested through the Trifacta application first. A workspace administrator can disable the skipping of these validations.

Create a new file [Parameterize destination](#)

POS-2021

Output Directory

/2227

Data Storage Format

CSV

Create new file every run
Create a new file with an incremental number appended to the name (e.g. POS-2021_2.csv)

Append to this file every run
Append action is not available when publishing as single file to tfs

Replace this file every run
Create it if it doesn't exist.

[Less options](#) ^

Output options

- Include headers as first row on creation
- Include quotes
- Include mismatched values

Delimiter

,

Multi-part options

Single File
Result is a single file in the output location.

Multiple Files
Result is split across multiple files placed in a new folder in the output location.

Compression

None

Cancel

Update



Figure: Output File Settings

Configure the following settings.

1. **Create a new file:** Enter the filename to create. A filename extension is automatically added for you, so you should omit the extension from the filename.
 - a. File output paths can have a maximum length of 2048 characters.
2. **Output directory:** Read-only value for the current directory. To change it, navigate to the proper directory.

NOTE: During job execution, a canary file is written for each set of results to validate the path. For datasets with parameters, if the path includes folder-level parameterization, a separate folder is created for each parameterized path. During cleanup, only the the canary files and the original folder path are removed. The parameterized folders are not removed. This is a known issue.

3. **Data Storage Format:** Select the output format you want to generate for the job.
 - a. **Avro:** This open source format is used widely for data serialization and data exchange between systems.
 - b. **CSV and JSON:** These formats are supported for all types of imported datasets and all running environments.

NOTE: JSON-formatted files that are generated by Trifacta are rendered in JSON Lines format, which is a single line per-record variant of JSON. For more information, see <http://jsonlines.org>.

- c. **Parquet:** This format is a columnar storage format.
- d. **HYPER:** Choose HYPER to generate results that can be imported into Tableau.

Feature Availability: This feature is available in the following editions:

- Trifacta® Enterprise Edition
- Trifacta Professional Edition
- Trifacta Starter Edition
- Trifacta Premium

- e. For more information, see *Supported File Formats*.
4. **Publishing action:** Select one of the following:

NOTE: If multiple jobs are attempting to publish to the same filename, a numeric suffix (`_N`) is added to the end of subsequent filenames (e.g. `filename_1.csv`).

NOTE: If a single user executes two jobs with the same output settings except for different methods (e.g. create vs. replace) on the same output destination, the generated results and potential error conditions are unpredictable. Please wait for the first job to complete execution before changing the configuration for the second job.

- a. **Create new file every run:** For each job run with the selected publishing destination, a new file is created with the same base name with the job number appended to it (e.g. `myOutput_2.csv`, `myOutput_3.csv`, and so on).
- b. **Append to this file every run:** For each job run with the selected publishing destination, the same file is appended, which means that the file grows until it is purged or trimmed.

NOTE: The `append` action is not supported when publishing to S3.

NOTE: Compression of published files is not supported for an `append` action.

- c. **Replace this file every run:** For each job run with the selected publishing destination, the existing file is overwritten by the contents of the new results.

5. **More Options:**

- a. **Include headers as first row on creation:** For CSV outputs, you can choose to include the column headers as the first row in the output. For other formats, these headers are included automatically.

NOTE: Headers cannot be applied to compressed outputs.

- b. **Include quotes:** For CSV outputs, you can choose to include double quote marks around all values, including headers.
- c. **Include mismatched values:** For CSV outputs, you can choose to include any value that is mismatched for its column data type. When disabled, mismatched values are written as null values.
- d. **Delimiter:** For CSV outputs, you can enter the delimiter that is used to separate fields in the output. The default value is the global delimiter, which you can override on a per-job basis in this field.

Tip: If needed for your job, you can enter Unicode characters in the following format: `\uXXXX`

- e. **Single File:** Output is written to a single file. Default setting for smaller, file-based jobs.
 - f. **Multiple Files:** Output is written to multiple files. Default setting for larger file-based jobs.
6. **Compression:** For text-based outputs, compression can be applied to significantly reduce the size of the output. Select a preferred compression format for each format you want to compress.

NOTE: If you encounter errors generating results using Snappy, additional configuration may be required. See *Supported File Formats*.

7. To save the publishing action, click **Add**.