Supported File Formats

Contents:

This section contains information on the fie formats and compression schemes that are supported for input to and output of Trifacta® Wrangler.

> **NOTE:** To work with formats that are proprietary to a desktop application, such as Microsoft Excel, you do not need the supporting application installed on your desktop.

Filenames

> **NOTE:** During import, the Trifacta application identifies file formats based on the extension of the filename. If no extension is provided, the Trifacta application assumes that the submitted file is a text file of some kind. Non-text file formats, such as Avro and Parquet, require filename extensions.

> **NOTE:** Filenames that include special characters can cause problems during import or when publishing to a file-based datastore.

Forbidden characters in import filenames:

The following list of characters present issues in the listed area of the product. If you encounter issues, the following listings may provide some guidance on where the issue occurred.

> **Tip:** You should avoid using any of these characters in your import filenames. This list may not be complete for all available running environments.

- **General:**

```
"/"
```

- **Seb browser:**

```
"\"
```

- **Excel filenames:**

```
"#"
```

- **Spark-based running environment:**

```
"{", "*", "\"
```

Native Input File Formats

Trifacta® Wrangler can read and import directly these file formats:

- CSV
- JSON v1, including nested

> **NOTE:** JSON files can be read natively but often require additional work to properly structure into tabular format. Depending on how the Trifacta application is configured (v1 or v2), JSON files may require conversion before they are available for use in the application. See "Converted file formats" below.

> **NOTE:** Trifacta Wrangler requires that JSON files be submitted with one valid JSON object per line. Consistently malformed JSON objects or objects that overlap linebreaks might cause import to fail.

- Plain Text
- LOG
- TSV
- Parquet

> **NOTE:** When working with datasets sourced from Parquet files, lineage information and the `$sourcerownumber` reference are not supported.

- Avro

> **NOTE:** When working with datasets sourced from Avro files, lineage information and the `$sourcerownumber` reference are not supported.

- Google Sheets

> **NOTE:** Individual users must enable access to their Google Drive. No data other than Google Sheets is read from Google Drive.

Converted file formats

Files of the following type are not read into the product in their native format. Instead, these file types are converted using the Conversion Service into a file format that is natively supported, stored in the base storage layer, and then ingested for use in the product.

> **NOTE:** Compressed files that require conversion of the underlying file format are not supported for use in the product.

Converted file formats:

- Excel (XLS/XLSX)

  > **NOTE:** Other Excel-related formats, such as XLSM format, are not supported.

  > **Tip:** You may import multiple worksheets from a single workbook at one time. See *Import Excel Data* in the User Guide.

- Google Sheets

  > **Tip:** You may import multiple sheets from a single Google Sheet at one time.

- JSON v2

Notes on JSON:

There are two methods of ingesting JSON files for use in the product.

- JSON v2 - This newer version reads the JSON source file through the Conversion Service, which stores a restructured version of the data in tabular format on the base storage layer for quick and simple use within the application.

  > **Tip:** This method is enabled by default and is recommended. For more information, see *Working with JSON v2*.

- JSON v1 - This older version reads JSON files directly into the platform as text files. However, this method often requires additional work to restructure the data into tabular format. For more information, see *Working with JSON v1*.

Native Output File Formats

Trifacta Wrangler can write to these file formats:

- CSV
- JSON

Compression Algorithms

When a file is imported, the Trifacta application attempts to infer the compression algorithm in use based on the filename extension. For example, `.gz` files are assumed to be compressed with GZIP.

> **NOTE:** Import of a compressed file whose underlying format requires conversion through the Conversion Service is not supported.

Read Native File Formats

	GZIP	BZIP	Snappy	Notes
CSV	Supported	Supported	Supported	
JSON v2	Not supported	Not supported	Not supported	A converted file format. See above.
JSON v1	Supported	Supported	Supported	Not a converted file format. See above.

Write Native File Formats

Compression is not supported for outputs in Trifacta.