

# Product Limitations

## Contents:

- *General Limitations*
    - *Sampling*
    - *Internationalization*
    - *Size Limits*
  - *Limitations by Integration*
    - *General*
    - *LDAP*
    - *Hadoop*
    - *Amazon AMI*
    - *Amazon EMR*
    - *Microsoft Azure*
    - *Redshift*
    - *S3*
    - *Hive*
    - *Spark*
    - *JDBC*
  - *Other Limitations*
- 

This section covers key known limitations of Trifacta® Self-Managed Enterprise Edition.

**NOTE:** This list of limitations should not be considered complete.

## General Limitations

### Sampling

- Sample sizes are defined by parameter for each available running environment. See *Sample Size Limits* be low.
- All values displayed or generated in the application are based on the currently displayed sample.
  - Transforms that generate new data may not factor values that are not present in the current sample.
  - When the job is executed, transforms are applied across all rows and values in the source data.
  - Transforms that make changes based on data values, such as `header` and `valuestocols`, will still be configured according to sample data at the time of that the step was added, instead at execution time. For example, all of the values detected in the sample are used to determine the columns of a `valuestocols` transform step based on the selected sample when the step was added.
- Random samples are derived from up to the first 1 GB of the source file.
  - Data from later parts of a multi-part file may not be included in the sample.
  - For more information, see *Samples Panel*.

### Internationalization

- The product supports a variety of global file encoding types for import. For more information, see *Configure Global File Encoding Type*.
- Within the application, UTF-8 encodings are displayed.
  - Limited set of characters allowed in column names.
  - Header does not support all UTF-8 characters.
  - Emoji are not supported in data wrangling operations.

- Umlauts and other international characters are not supported when filtering datasets in browsers of external datastores.
- UTF-8 is generated in output.
  - Some international characters are presented incorrectly in the output.
- States and Zip Code Column Types and the corresponding maps in visual profiling apply only to the United States.
- UTF-32 encoding is not supported

**NOTE:** Some functions do not correctly account for multi-byte characters. Multi-byte metadata values may not be consistently managed.

## Size Limits

### Sample Size Limits

Defaults for each running environment:

- For the Trifacta Photon running environment, samples are limited to 10 MB.
- For more information on changing sample sizes, see *Running Environment Options* .

### Job Size Limits

Execution on a Spark running environment is recommended for any files over 5GB in net data size, including join keys.

## Limitations by Integration

### General

The product requires definition of a base storage layer, which can be HDFS or S3 for this version. This base storage layer must be defined during install and cannot be changed after installation. See *Set Base Storage Layer*.

### LDAP

- If LDAP integration is enabled, the LDAP user [`ldap.user` (default=`trifacta`)] should be created in the same realm.
- See *Configure SSO for AD-LDAP* .

### Hadoop

- Trifacta Self-Managed Enterprise Edition requires an integration with a working Hadoop cluster.
- See *Running Environment Options*.

### Amazon AMI

- For more information, see *Install from AWS Marketplace*.

### Amazon EMR

- For more information, see *Install for AWS*
- For more information, see *Configure for EMR*.

### Microsoft Azure

- For more information, see *Configure for Azure*.

## Redshift

None.

## S3

- S3 integration is supported only over AWS-hosted instances of S3.
- Oracle Java Runtime 1.8 must be installed on the node hosting the product.
- Writing to S3 requires use of S3 as the base storage layer. For more information, see *Set Base Storage Layer*.
- When publishing single files to S3, you cannot apply an `append` publishing action.

## Hive

- Only HiveServer2 is supported.
- You can create only one connection of this type.
- When reading from a partitioned table, the product reads from all partitions, which impacts performance.
- For more information, see *Configure for Hive*.

## Spark

- None.

## JDBC

- The product supports explicit versions of each JDBC source.
  - Some additional installation may be required.
  - See *Enable Relational Connections*.
- Jobs for JDBC sources must be executed on the Trifacta Photon running environment.
- Writing to JDBC sources is not supported in this release.

## Other Limitations

- **File Formats:** Limitations may apply to individual file formats. See *Supported File Formats*.
- **Data Type Conversions:** There are some limitations on how data types are converted during import or export/publication. See *Type Conversions*.