

Configure Type Inference

Contents:

- *Configure Type Inference for Schematized Sources*
 - *Enable*
 - *Configure Load Limits for Inference*
 - *Use*
 - *Define for individual connections*
 - *Specify on dataset import*
 - *Configure Type Inference in the Data Grid*
 - *Type Inference on Export*
-

By default, the Designer Cloud powered by Trifacta® platform applies its own type inference to datasets when they are imported and again when new steps are applied to the data. This section provides information on how you can configure where type inference is applied in the platform.

Data types are inferred by the Designer Cloud powered by Trifacta platform when:

- Imported datasets are originally loaded.
- A new transformation step is added in a recipe.
- Non-inferred types are imported as String type.

Tip: You can use the Change Column Type transformation to override the data type inferred for a column. However, if a new transformation step is added, the column data type is re-inferred, which may override your specific typing. You should consider applying Change Column Type transformations as late as possible in your recipes.

For more information on how the Designer Cloud powered by Trifacta platform applies data types to specific sources of data on import, see *Type Conversions*.

Configure Type Inference for Schematized Sources

Optionally, you can choose to disable type inference for schematized sources. A **schematized source** includes column data type information as part of the object definition. The following schematized sources are supported for import into the Designer Cloud powered by Trifacta platform :

- All JDBC sources

NOTE: You cannot disable type inference for Oracle sources. This is a known issue.

- Hive
- Redshift

- Avro file format

Enable

Type inference on schematized sources	Setting	Behavior
Enabled	"webapp.connectivity.disableRelationalTypeInference": false,	<p>All imported datasets from schematized sources are automatically inferred by the type system in the Designer Cloud powered by Trifacta platform .</p> <p>The inferred data types may be different from those in the source. When the dataset is loaded, data types can be applied to individual columns through the application.</p> <p>Users can apply overrides for:</p> <ul style="list-style-type: none">• Individual connections• Individual datasets at time of import
Disabled	"webapp.connectivity.disableRelationalTypeInference": true,	<p>For schematized data sources, type inference is not automatically inferred by Designer Cloud powered by Trifacta platform .</p> <p>Data type information is taken from the source schema and applied where applicable to the dataset. If there is no corresponding data type in the Designer Cloud powered by Trifacta platform , the data is imported as String type.</p> <p>Users can apply overrides for:</p> <ul style="list-style-type: none">• Individual connections• Individual datasets at time of import

Please perform the following configuration change to disable type inference of schematized sources at the global level.

Steps:

1. You can apply this change through the *Admin Settings Page* (recommended) or `trifacta-conf.json`. For more information, see *Platform Configuration Methods*.
2. Change the following configuration setting to `true`:

```
"webapp.connectivity.disableRelationalTypeInference": false,
```

3. Save your changes.

Configure Load Limits for Inference

When a dataset is imported into the Designer Cloud application , a volume of data is read from the source, up to the parameterized limits below. These limits define the maximum size of the data read for:

- **Split row inference:** data read for determining where each row ends in the dataset.
- **Type inference:** data read for determining the data types of each column.

Tip: You can raise these limits gradually if you are noticing issues with either data inference or row splits. Raising these values significantly can impact load performance in the Transformer page.

Parameter	Description
-----------	-------------

webapp.loadLimitForSplitInference	Maximum number of bytes to be read from an imported dataset for initial inference for splitting rows. Default value is 20000.
webapp.loadLimitForTypeInference	Maximum number of bytes to be read from an imported dataset for initial inference of column data types. Default value is 524288.

Use

In the application, type inference can be applied to your imported data through the following mechanisms.

Define for individual connections

You can specify individual connections to apply or not apply Trifacta type inference when the connection is created or edited.

NOTE: When Default Column Data Type Inference is disabled for an individual connection, Trifacta type inference can still be applied on import of individual datasets.

For more information, see *Create Connection Window*.

Specify on dataset import

When type inference has been disabled globally for schematized sources, you can choose to enable or disable it for individual source import.

Tip: To compare how data types are imported from the schematized source or when applied by the Designer Cloud powered by Trifacta platform, you can import the same schematized source twice. The first instance of the source can be imported with type inference enabled, and the second can be imported with it disabled.

In the Import Data page, click **Edit Settings** on the data source card.

For more information, see *Import Data Page*.

Configure Type Inference in the Data Grid

Type inference is automatically enabled in the data grid. It cannot be disabled.

Tip: You can override the Trifacta data type by applying a Change Column Type transformation.

When a new transformation step is applied, each column is re-inferred for its Trifacta data type.

Type Inference on Export

When you generate results, the current data types in the data grid are applied to the generated results.

If the publishing destination is a schematized environment, the generated results are written to the target environment based on the environment type. These data type mappings cannot be modified.

For more information on output types, see *Type Conversions*.