

Parquet Data Type Conversions

NOTE: The Trifacta® data types listed in this page reflect the raw data type of the converted column. Depending on the contents of the column, the Transformer Page may re-infer a different data type, when a dataset using this type of source is loaded.

Import

NOTE: Trifacta does not support ingest of Parquet files with nested values, which can occur for Map or Object data types.

Parquet Data Type	Trifacta Data Type	Notes
STRING	String	
INT	Integer	
DECIMAL	Decimal	
DATE	Datetime	
TIME	Datetime	
TIMESTAMP	Datetime	
LIST	Array	
MAP	Object	

Limitations on import:

The Parquet data format supports the use of row groups for organizing chunks of data. This row grouping is helpful for processing across distributed systems.

Trifacta places limitations on the volume of data that can be displayed in the browser. By default, these limits are set to 10 MB.

If Parquet row groups are greater than 10 MB:

- You cannot preview data from the file before import.
- When a Parquet-based dataset is loaded in the Transformer page, the screen may be blank.

Tip: You can create a new sample from inside the Transformer page. The sample is displayed normally.

Other product functions work as expected with Parquet format.

Export

On export, Trifacta data types are exported to their corresponding Parquet types, with the following specific mappings:

Trifacta Data Type	Parquet Data Type	Notes
--------------------	-------------------	-------

Boolean	BOOLEAN	
Integer	INT64	
Decimal	DOUBLE	
String	BYTE_ARRAY (STRING)	

The fallback data type on export is BYTE_ARRAY (STRING).