Import Integrations Filters for Amazon S3, FTP and SFTP

Filters can be used to modify or manipulate the structure of your data during the import process. Although very useful, it is important to know their role in the data-ingestion process, and their limitations. These filter plugins are available for the S3, FTP or SFTP connectors.

All data that you bring into Treasure Data must contain a time column. To optimize performance, all Treasure Data tables require this time column to exist. Internally, we add this time column using a filter. During the preview step, validate that your data contains this column (and that it hasn't been filtered out).

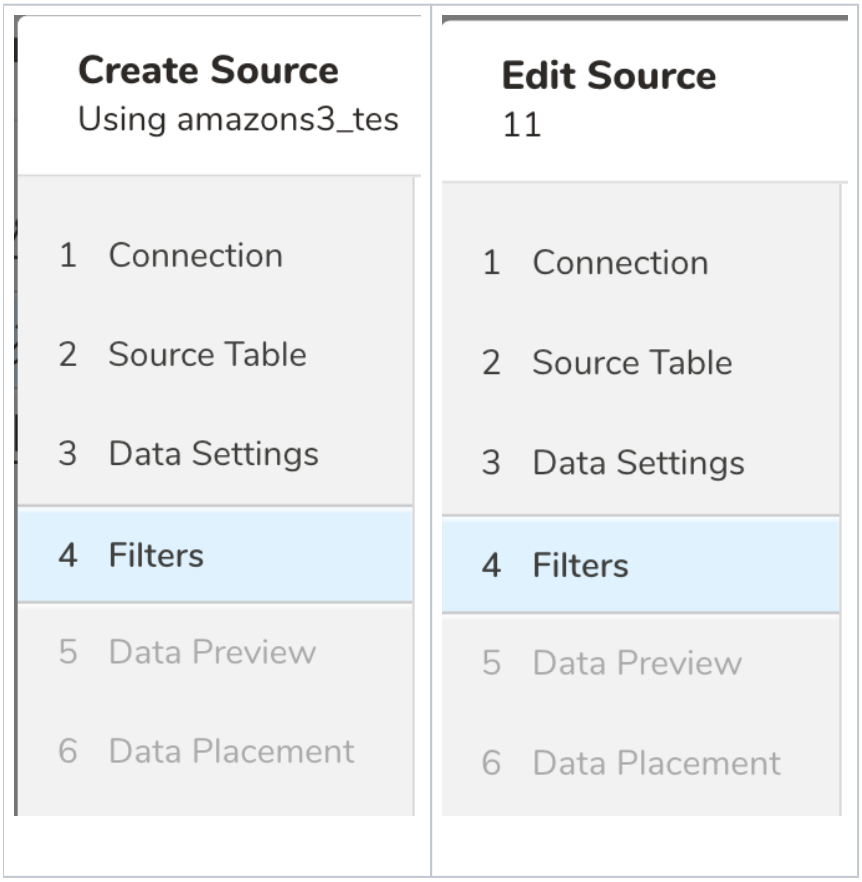
Learn more about Import Integration Filters:

- Import Integration Filters
- Types of Filters
- Order of Operations

See also About Filter Functions .

Import Integration Filters

These filters can be used during import to modify your data. They can be used to retain columns, add columns, drop columns, expand JSON columns, or mask column data. Masking data is typically used to secure sensitive information such as social security numbers. You interact with the filters as part of defining your TD import integrations.



Types of Filters

Filter Type	Description	Use Case Example
Retain Columns	Use to specify the columns that you would like to keep.	Your sales team wants a table with all of your users' names and phone numbers. You can import a data set of user registrations from your website, but that data set has names, phone numbers and 50 other columns that you don't need. The retain filter allows you to specify that only the name and phone number columns are ingested.

Add Columns	Use to duplicate a column or even create a new column with a static value. It can also copy an existing column and rename it to a new value.	You want to add a column named 'source' with a value of "12345" to all rows that come from the job you've setup for your Facebook Ad with the ID of 12345. The Add Columns Filter can do this. It can also copy an existing column and rename it to a new value.
Drop Columns	Use to drop a column from data you are importing.	Your marketing team wants a list of customer information such as name, email, address, phone, and certain preferences and that data exists in a table which also contains some sensitive information or values that are not needed, you can drop those columns.
Expand JSON	Turns JSON objects in your data set into multiple new columns.	If you had a column with a JSON object like: {"first_name": "John", "last_name": "Doe"}, the Expand JSON Filter could extract those values from your JSON and create the following new columns: • first_name • last_name
Digest	Use to protect sensitive information from being exposed. The data you are importing may contain a column that you deem sensitive or private. The digest filter allows you to hash the data in that specific column. You can choose the algorithm that is used to hash each column.	Your marketing team wants a segment list of customer contact information that also includes social security numbers. You can use the digest filter to hide that information.

Learn more about applying import integration filters:

- Applying Import Integration Filters

Order of Operations

Filters run after decoders and parsers. In the TD Console, your data is initially be decoded, parsed and then filtered. Although your source data may have certain schemas or names, the filters only see what comes out of the parser.

In the TD Console, filters receive the data from Data Settings. The schema or settings that come out of Data Settings is what your filters 'see.' For example, if you rename a column in Data Settings to "cats" replacing "dogs", you need to specify the new column name, "cats", in any filters you use.

Additionally, each filter acts on the data it receives from the previous step or filter. For example if one filter renamed a column, the next filter will only see the new, renamed column name from the previous filter.

To illustrate this, let's imagine you:

1. Import a data set with customer_id, name, car_make, and car_vin columns.
2. Apply the Retain Columns Filter to retain only the customer_id, name, and car_make columns.
3. Apply the Digest Filter (after the Retain Columns Filter) to mask (or hash) the car_vin column.
4. This results in a warning message telling you that the column car_vin doesn't exist..

This happens because the Retain Columns Filter already removed the car_vin column and the Digest Filter only sees what came out of the filter before it as follows:

# customer_id	Ab name	Ab car_make	Ab car_vin
1	Hogan Edmeads	Mazda	WBADT33482G881563
2	Salvidor Albasiny	Pontiac	1FTWW3A50AE174799
3	Elwood Rabat	Suzuki	1J4NF1FB3BD317506

RETAIN FILTER: customer_id, name, car_make

# customer_id	Ab name	Ab car_make
1	Hogan Edmeads	Mazda
2	Salvidor Albasiny	Pontiac
3	Elwood Rabat	Suzuki

DIGEST FILTER: car_vin

!

You receive an error message because
car_vin was not retained in previous filter.