

SFTP Server Import Integration

[Learn more about SFTP Server Export Integration.](#)

The Data Connector for SFTP enables you to import files stored on your SFTP server to Treasure Data.

For sample workflows of importing files from your SFTP server, view [Treasure Boxes](#).

- [Prerequisites](#)
- [Limitation](#)
- [Use the TD Console to Create your Connection](#)
 - [Create a New Connection](#)
 - [Transfer Data into Treasure Data](#)

Prerequisites

- Basic knowledge of Treasure Data.
- Requires that your private key has an [OpenSSH 7.8](#) format.
- Requires that the OpenSSH format private key was generated using the '-m PEM' option.
- The default format of the private key after OpenSSH.
- Before using this connector, determine valid protocols for your environment.
If you intend to *SFTP*, you can use this integration for SFTP.
If *FTP/FTPS*, try connect with [FTP Import Integration](#).
 - If you are using a firewall, check your accepted IP range/port. Server administrators sometimes change the default port number from TCP /22 for security reasons.
 - "PuTTY" and other formats are not supported.
- After installation and configuration, review the job log. Warning and errors provide information about the success of your import. For example, you can [identify the source file names associated with import errors](#).

Limitation

- Support only the STORED and DEFLATE compression methods.
- Multi-part gzip file may not work

Use the TD Console to Create your Connection

You can use TD Console to create your data connector.

Create a New Connection

When you configure a data connection, you provide authentication to access the integration. In Treasure Data, you configure the authentication and then specify the source information.

1. Open TD Console.
2. Navigate to **Integrations Hub > Catalog**.
3. Search and select SFTP.



4. Select Create
5. The following dialog opens.

New Authentication

SFTP

✕

1 Credentials >
 2 Details

Host:

Port:

User:

Authentication mode:

Secret key file:

Passphrase for secret key file:

Retry limit:

[Learn more](#)
Continue

6. Enter the required credentials for your remote SFTP instance. Set the following parameters.

- **Host:** The host information of the remote SFTP instance, for example an IP address.
- **Port:** The connection port on the remote SFTP instance, the default is 22.
- **User:** The user name used to connect to the remote FTP instance.
- **Authentication mode:** The way you choose to authenticate with your SFTP server.
- **Secret key file:** Required if 'public / private key pair' is selected from 'Authentication Mode'. (ecdsa key type is supported.)
- **Passphrase for secret key file:** (Optional) If required, provide a passphrase for the provided secret file.
- **Retry limit:** Number of times to retry a failed connection (default 10).
- **Timeout:** Connection timeout in seconds (default 600).

7. Select **Continue**. Type a name for your connection.

New Authentication

FTP

✕

✔ Credentials >
 2 Details

Name

Share with others

Back
[New Source](#)
Done

8. If you would like to share this connection with other users in your organization, check the **Share with others** checkbox. If this box is unchecked this connection is visible only to you.

9. Select **Done**.

Transfer Data into Treasure Data

To get the data from your SFTP server into Treasure Data, you can set up an ad hoc one-time transfer or a recurring transfer at a regular interval. After creating the authenticated connection, you are automatically taken to Authentications.

1. Search for the connection you created.
2. Select **New Source**.

Connection

1. Type a name for your **Source** in the Data Transfer field.

2. Click **Next**.

Source Table

1. The Source dialog opens.

2. Edit the following parameters

Parameters	Description
User directory root	
Path prefix	Prefix of target files (string, required)
Path match pattern	Type a regular expression to query file paths. If a file path doesn't match with the specified pattern, the file is skipped. For example, if you specify the pattern <code>.csv\$ #</code> , then a file is skipped if its path doesn't match the pattern.

Incremental	Enables incremental loading (boolean, optional. default: true. If incremental loading is enabled, the config diff for the next execution will include last_path parameter so that the next execution skips files before the path. Otherwise, last_path is not included.
Start after path	Only paths lexicographically greater than this will be imported.

Data Settings

1. Select **Next**.
The Data Settings page opens.
2. Optionally, edit the data settings or skip this page of the dialog.

- 1 Connection
- 2 Source Table
- 3 Data Settings
- 4 Filters
- 5 Data Preview
- 6 Data Placement

Optionally, you can modify data settings and then see your changes in Data Preview. [Skip This Step](#)

▼ DECODERS

▼ PARSER

Type:

Delimiter:
Delimiter character such as , for CSV, "\t" for TSV, "|" or any single-byte character

Quote character:
The character surrounding a quoted value. Setting null disables quoting.

Escape character:
Escape character to escape a special character. Set to null to disable escaping.

Skip header lines:
Skip this number of lines first. Set to 1 if the file has a header line.

Null string:
If a value is this string, converts it to NULL. For example, set \N for CSV files created by mysqldump

Trim if not quoted?:
Remove spaces of a value if the value is not surrounded by the quote character

Filters

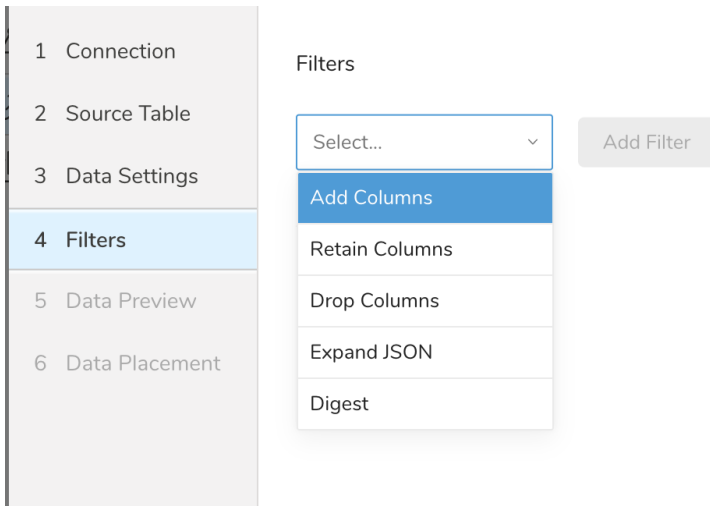
Import Integration Filters enable you to modify your imported data after you have completed [Editing Data Settings](#) for your import.

To apply import integration filters:

Select **Next** in Data Settings.

The Filters dialog opens.

Select the filter option you want to add.



Select **Add Filter**.

The parameter dialog for that filter opens.

Edit the parameters.

For information on each filter type, see one of the following:

- [Retaining Columns Filter](#)
- [Adding Columns Filter](#)
- [Dropping Columns Filter](#)
- [Expanding JSON Filter](#)
- [Digesting Filter](#)

Optionally, to add another filter of the same type, select **Add** within the specific column filter dialog.

Optionally, to add another filter of a different type, select the filter option from the list and repeat the same steps.

After you have added the filters you want, select **Next**.

The Data Preview dialog opens.

Preview

You can see a [preview](#) of your data before running the import by selecting Generate Preview.

Data shown in the data preview is approximated from your source. It is not the actual data that is imported.

1. Select **Next**.
Data preview is optional and you can safely skip to the next page of the dialog if you want.
2. To preview your data, select **Generate Preview**. Optionally, select **Next**.
3. Verify that the data looks approximately like you expect it to.

Create Source
Using onetrust_demo

1 Connection
2 Source Table
3 Data Settings
4 **Data Preview**
5 Data Placement

The preview shows a subset of data from the source based on the data settings. Refer to [help document](#) to learn more about preview data.

	Ab_id	Ab_language	Ab_identifier	last_updated_date	Ab_link_token	
1	f7abf910-b5da-47c2-bbee-37f4c86...	NULL	Quan3	2020-09-25 22:42:59...	NULL	0
2	9022117f-cf3c-418c-b527-a8bd9a9...	NULL	Quan2	2020-08-05 03:48:19...	NULL	0
3	a432b52f-3d93-483b-b65f-3c7530...	NULL	Quan4	2020-08-05 03:48:19...	NULL	0
4	233ec0c2-70ab-4de4-ac48-a4a048f...	NULL	Quan5	2020-08-05 03:48:19...	NULL	0
5	f78be70b-8b5d-404e-b663-b606a2...	NULL	Quan1	2020-08-05 03:48:19...	NULL	0
6	db5d8f89-c264-4d82-a246-5939e5...	NULL	example@otprivacy.com	2020-08-06 17:51:12...	NULL	0
7	5ef9542c-315d-4b56-ad1c-c63ad0...	NULL	Michael.White@gmail.com	2020-09-09 20:01:45...	NULL	0
8	3f1dfcb9-1904-4517-9087-0cc45f0...	NULL	Robert.Brown@gmail.com	2020-09-09 20:01:45...	NULL	0
9	4a3a88dd-11a3-4c8b-a1d9-d7043f...	NULL	Mary.Anderson@gmail.com	2020-09-09 20:01:46...	NULL	0
10	4f68983a-9e49-46dc-9519-1cf9dea...	NULL	Elizabeth.Scott@gmail.com	2020-09-09 20:01:47...	NULL	0
11	33342e5d-4c95-4cfe-a622-4e91dc5...	NULL	David.Miller@aol.com	2020-09-09 20:01:47...	NULL	0
12	f54b0d7c-df75-4bf3-934a-dc19a96...	NULL	Robert.Anderson@att.com	2020-09-10 04:57:16...	NULL	0
13	43bfe156-dfba-43b8-964d-1b2a4ae...	NULL	Elizabeth.Miller@google.com	2020-09-10 04:57:16...	NULL	0

Cancel Back Next

4. Select **Next**.

Data Placement

For data placement, select the target database and table where you want your data placed and indicate how often the import should run.

1. Select **Next**. Under **Storage** you will create a new or select an existing database and create a new or select an existing table for where you want to place the imported data.

1 Connection
2 Source Table
3 Data Settings
4 Data Preview
5 **Data Placement**

STORAGE

Database: chung_default_db

Table: sftp_v2_devproxy

Method:

- Append: Add records into existing table.
- Always Replace: Always clear the destination table before adding records.
- Replace on new data: When there is new data, delete existing data, and insert new data.

Timestamp-based Partition Key: time

Data Storage Timezone: UTC (default)

SCHEDULE

Repeat: Off On

Scheduling Timezone: Asia/Saigon

2. Select a **Database** > **Select an existing** or **Create New Database**.
3. Optionally, type a database name.
4. Select a **Table** > **Select an existing** or **Create New Table**.
5. Optionally, type a table name.
6. Choose the method for importing the data.
 - **Append** (default)-Data import results are appended to the table. If the table does not exist, it will be created.
 - **Always Replace**-Replaces the entire content of an existing table with the result output of the query. If the table does not exist, a new table is created.
 - **Replace on New Data**-Only replace the entire content of an existing table with the result output when there is new data.
7. Select the **Timestamp-based Partition Key** column. If you want to set a different partition key seed than the default key, you can specify the long or timestamp column as the partitioning time. As a default time column, it uses upload_time with the add_time filter.
8. Select the **Timezone** for your data storage.
9. Under **Schedule**, you can choose when and how often you want to run this query.

- Run once:
 - a. Select **Off**.
 - b. Select **Scheduling Timezone**.
 - c. Select **Create & Run Now**.
- Repeat the query:
 - a. Select **On**.
 - b. Select the **Schedule**. The UI provides these four options: *@hourly*, *@daily* and *@monthly* or custom *cron*.
 - c. You can also select **Delay Transfer** and add a delay of execution time.
 - d. Select **Scheduling Timezone**.
 - e. Select **Create & Run Now**.

After your transfer has run, you can see the results of your transfer in **Data Workbench > Databases**.