

Repro Import Integration

[Learn more about Repro Export Integration.](#)

You can use Repro Import Integration to ingest files from your Amazon S3 buckets with customized parameters for an easy configuration.

- [Prerequisites](#)
- [Limitations](#)
- [Using the TD Console to Create Your Connection](#)
 - [Create a New Connection](#)
 - [Transfer Your Repro Account Data to Treasure Data](#)
 - [Create Your Source](#)
 - [Configure Your Data](#)
 - [Preview Your Data](#)
 - [Data Placement](#)

Prerequisites

- Basic knowledge of Treasure Data, including the TD [Toolbelt](#).
- A Repro application id, Access key ID, and Secret access key.

Limitations

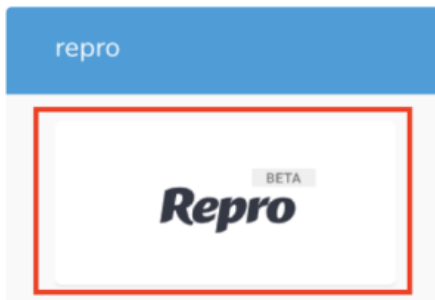
- If you enter File name patterns and select Incremental?, the data does not load. The data does not load because Repro does not put the data into the old folder, it creates a new folder every time.

Using the TD Console to Create Your Connection

Create a New Connection

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

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



The following dialog opens:

New Authentication ×
Repro

1 Credentials > 2 Details

Region:

Authentication Method:

Access key ID:

Secret access key:

[Learn more](#)

4. Enter the required information:
 - Region. The region of your Repro's application (for example, ap-northeast-1, us-east-1 ...)
 - Authentication Method. Select basic.
 - Access key ID. Enter the key you obtained from Repro.
 - Secret access key. Enter the secret access key you obtained from Repro.
5. Select **Continue**.
6. Enter a name for your connection.

New Authentication ×
repro_authentication

✓ Credentials > 2 Details

Name

is a reserved name

Share with others

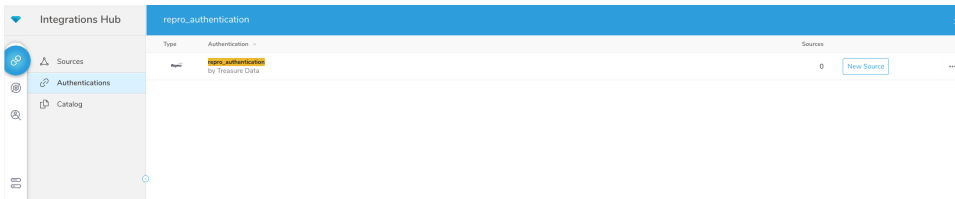
[New Source](#)

7. Select **Done**.

Transfer Your Repro Account Data to Treasure Data

After creating the authenticated connection, you are automatically taken to Authentications.

1. Search for the connection you created.



2. Select **New Source**.

Create Your Source

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

Create Source
Using repro_authentication

1 Connection | Data Transfer Name: my_repro | Authentication: repro_authentication

2 Source Table
3 Data Settings
4 Data Preview
5 Data Placement

Cancel Back Next

2. Select **Next**.

Create Source
Using repro_authentication

1 Connection | Bucket: repro-data-for-outer-production

2 Source Table | App ID: | Folder: custom_event_histories

3 Data Settings | Upload Time (optional): YYYYMMDDHH

4 Data Preview | File name pattern (optional):

5 Data Placement | Filter by Modified Time: | Incremental?:

When run repeatedly, attempt to only import new data since the last import. (Assumes lexicographic order)

Cancel Back Next

1. Edit the following parameters in your source table.

Parameters	Description
Bucket	The bucket where your Repro application is located, for example, repro-data-for-outer-production.
App ID	Your Repro application id.
Upload Time	The specific time you would like to ingest the data (YYYYMMDDHH format).
Filename pattern	Use regexp to match file paths. If a file path doesn't match the specified pattern, the file is skipped. For example, if you specify the pattern .csv\$# , then a file is skipped if its path doesn't match the pattern. Read more about regular expressions .
Filter by Modified Time	Select if you would like to use modified time as the main criteria to load data. Insert these parameters so that the first executions skip files that were modified before that specified timestamp. For example, 2019-06-03T10:30:19.806Z.
Incremental by Modified Time <i>(Available if Filter by Modified Time is selected)</i>	Select to ingest only new data since the previous ingestion.

Incremental? <i>(Available if Filter by Modified Time is selected)</i>	Select to ingest only new data since the previous ingestion.
--	--

Configure Your Data

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

Create Source
Using repro_authentication

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

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

DECODERS

[Add](#)

Gzip

PARSER

Type: CSV

Delimiter: ,
Delimiter character such as , for CSV, "I" 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: 0
Skip this number of lines first. Set to 1 if the file has a header line.

Null string: null
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

Cancel Back Next

Preview Your Data

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-3714c86...	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@otrprivacy.com	2020-08-06 17:51:12...	NULL	0
7	5ef9542c-315d-4b56-ad1c-c63ad0...	NULL	MichaelWhite@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@mail.com	2020-09-09 20:01:46...	NULL	0
10	4fd8983a-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.

The screenshot shows the 'Data Placement' configuration window. On the left, a sidebar lists steps: 1 Connection, 2 Source Table, 3 Data Settings, 4 Data Preview, and 5 Data Placement (highlighted). The main area is split into two sections:

- 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**.