

Gigya Import Integration

You can ingest data of [Accounts](#), [Profile Management](#), [Data Store](#), and [Audit Log](#) from Gigya (SAP Customer Data Cloud) into Treasure Data.

- [Prerequisites](#)
- [Supported](#)
- [Limitations and Requirements](#)
 - [Query Syntax Limitation](#)
 - [Data Source Limitation](#)
- [Incremental Loading and Numeric and Timestamp Columns](#)
- [Use the TD Console to Create Your Connection](#)
 - [Obtain your API Key, User Key, and Secret Key From Gigya](#)
 - [Create a New Connection](#)
 - [Transfer Your Gigya Accounts Data to Treasure Data](#)
 - [Data Settings](#)
 - [Data Preview](#)
 - [Data Placement](#)

Prerequisites

- Basic knowledge of Treasure Data, including the [TD toolbelt](#)
- A Gigya account
- Basic knowledge about [Query Syntax Specification](#) in order to obtain data from Accounts objects
- Review [About Incremental Loading](#)

Incremental loading uses the maximum value (max value) in the specified incremental column to load all records until the max value for the first execution and subsequent runs imports records from (max value +1) from the previous run to the current time when the job runs, which becomes the new max value.

Supported

- Select clause
- From clause
- Where clause
- Group By clause

Limitations and Requirements

- **START**, **CONTAINS** and **WITH** keyword is not supported
- **COUNTERS** is not supported
- **LIMIT** clause in the query will be removed automatically(**SELECT * FROM ACCOUNTS LIMIT 100**) returns all records ignoring the limit
- Query with aggregate functions (**sum**, **min**, **max**, **avg**, **sum_of_squares**, **variance**, **std**) and **Group By** clause only could ingest the first page
- If you enter an invalid object in **FROM** clause (i.e **SELECT * FROM nonexistent**) the query will be automatically fall back to the accounts object
- Columns names are case-sensitive and Objects name are case-insensitive

Query Syntax Limitation

Treasure Data supports the following SQL query syntax for Gigya:

- Select clause
- From clause
- Where clause
- Group By clause
- **START**, **CONTAINS**, and **WITH** keywords are not supported.
- **COUNTERS** is not supported.
- a **LIMIT** clause in the query is removed automatically.
SELECT * FROM ACCOUNTS LIMIT 100 returns all records and ignores the limit.
- A query with aggregate functions (sum, min, max, avg, sum_of_squares, variance, std) and Group By clause can only ingest the first page.
- Columns names are case-sensitive.
- Object names are case-insensitive.
- The Incremental Column should be numeric or timestamp.

Data Source Limitation

For Account Data Source

- Your data query is limited to two objects: accounts, and email accounts.
- If you enter an invalid object in a FROM clause, the query automatically falls back to the accounts object. For example, SELECT * FROM <does_not_exist> , substitutes accounts for <does_not_exist>.
- Supported incremental columns for accounts objects are:[lastLogin, registered, oldestDataUpdatedTimestamp, lastUpdated, verifiedTimestamp, oldestDataUpdated, lastUpdatedTimestamp, created, createdTimestamp, verified, registeredTimestamp, lastLoginTimestamp, lockedUntil]
- Supported incremental columns for emailAccounts are: [lastUpdated, lastUpdatedTimestamp, created, createdTimestamp]
- Referenced link: [accounts.search REST](#)

For Profile Management Data Source

- Your data query is limited to two objects, accounts, and email accounts.
- If you enter an invalid object in a FROM clause, the query automatically falls back to the accounts object. For example, SELECT * FROM <does_not_exist> , substitutes accounts for <does_not_exist>.
- Supported incremental columns for accounts objects are:[lastLogin, registered, oldestDataUpdatedTimestamp, lastUpdated, verifiedTimestamp, oldestDataUpdated, lastUpdatedTimestamp, created, createdTimestamp, verified, registeredTimestamp, lastLoginTimestamp, lockedUntil]
- Supported incremental columns for emailAccounts are:[lastUpdatedTimestamp, created, createdTimestamp, lastUpdated]
- Referenced link: [ids.search REST](#)

For Data Store Data Source

- If you enter an invalid object in a FROM clause, you will receive error [400006] Invalid parameter value: Invalid argument: accounts type not allowed
- Depending on the schema of the target Data Store then incremental columns will be varied. But you could enter an invalid column name and see the error return from TD Console to know what is an acceptable column name.
- Referenced link: [ds.search REST](#)

For Audit Log Data Source

- Supported incremental columns are: [@timestamp]
- Referenced link: [audit.search](#)

Incremental Loading and Numeric and Timestamp Columns

Incremental loading uses the maximum value (max value) in the specified incremental column to load all records till max value for the first execution and subsequent runs import records from (max value +1) from the previous run to the current time when the job runs (which becomes the new max value).

Support for:

- Incremental columns of numeric or timestamp type
- Incremental columns for Accounts objects are:[lastUpdated, lastUpdatedTimestamp, created, createdTimestamp]
- Incremental columns for EmailAccounts are:
[lastLogin, registered, oldestDataUpdatedTimestamp, lastUpdated, verifiedTimestamp, oldestDataUpdated, lastUpdatedTimestamp, created, createdTimestamp, verified, registeredTimestamp, lastLoginTimestamp, lockedUntil]
- Supported incremental columns for auditLog are: [@timestamp]

Use the TD Console to Create Your Connection

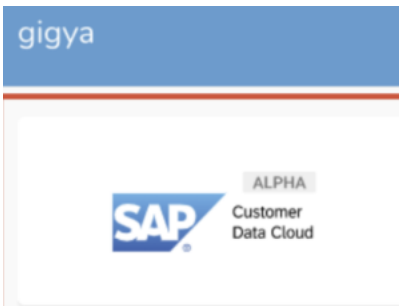
Obtain your API Key, User Key, and Secret Key From Gigya

1. Follow the instruction in [Creating and Managing Applications](#) to create your own application and obtain the **App User Key** and **Secret Key**.
2. Follow the instructions in [API Key and Site Setup](#) to obtain your **API Key**.
3. Follow the instructions to determine your [Data Center](#).

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 Gigya (SAP Customer Data Cloud).



The following dialog opens.

New Authentication ×

Gigya (SAP Customer Data Cloud)

1 Credentials > 2 Details

Data Center

United States Data Center ▼

Authentication Mode

User Key/Secret ▼

API Key

User Key

User Secret

[Learn more](#) [Continue](#)

4. Choose your account Data Center.

5. Type values for the following:

- API Key
- User Key
- User Secret

New Authentication

Gigya (SAP Customer Data Cloud)



1 Credentials > 2 Details

Data Center

United States Data Center

Authentication Mode

User Key/Secret

API Key

••••

User Key

sampleUserKey

User Secret

••••

[Learn more](#)

[Continue](#)

6. Select **Continue**.

7. Enter a name for your connection and select **Done**.

New Authentication

gigya_authentication



✔ Credentials > 2 Details

Name

gigya_authentication

Share with others

Back

New Source

Done

Transfer Your Gigya Accounts Data to Treasure Data

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

1. Search for the connection you created and select **New Source**.

Authentications

Type	Authentication	Sources
	gigya_authentication by Treasure Data	0 New Source ...

2. Name the **Source**.
3. Select **Next**.
4. In the Source Table, edit the parameters.

1 Connection	Data Source:	Accounts
2 Source Table	Query:	select * from accounts;
3 Data Settings	Fields To Exclude:	Comma separated excluding fields. I.e field1,field2 Comma separated excluding fields to be excluded from the output table
4 Data Preview	Batch Size:	1000 Maximum number of records per batch
5 Data Placement	Incremental Loading:	<input type="checkbox"/> When run repeatedly, attempt to only import new data since the last import

Parameters	Description
Data Source	Target data source. Current support: Accounts, Profile Management, Data Store and Audit Log
Query	Gigya's query to ingest data. Depend on your target object, the query would be variant. For Account and Profile Management data source, only support on accounts and emailAccounts object (sample query: Select * From accounts) For Data Store it would be whatever your object is (sample query: Select * From my_data) For Audit Log, only support for auditLog object (sample query: Select * From auditLog)
Fields To Exclude	Due to Gigya's API specification, it is not possible to specify the columns to be included in the SELECT statement. This parameter can be used to remove unnecessary columns.
Batch Size	The maximum number of records to fetch in a single API call. The maximum value is 10000 and the minimum value is 10. When you customize the batch size, consider the following: a smaller value will let the API return faster but will cause more API calls.
Incremental	When running on a schedule, the next import ingests only the data that was updated after the last run based on the value of the Incremental Column.
Incremental Column	Which data object's column on which to perform the incremental transfer. For Account and Profile Management data source, suggested values are: created, createdTimestamp, updated, and updatedTimestamp. For Data Store data source, suggested values are: numeric or datetime column For Audit Log data source, suggested values are: @timestamp

Data Settings

5. In this dialog, you can edit data settings or opt to skip this step.

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

2 Source Table

3 Data Settings

4 Data Preview

5 Data Placement

Max retry count per API call:

Max retry count per API call

Initial retry interval per API call:

Initial retry interval per API call in seconds

Max retry interval per API call:

Max retry interval per API call in seconds

Cancel

Data 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. Click **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, click **Next**.
3. Verify that the data looks approximately like you expect it to.

Create Source
Using onetrust_demo

1 Connection 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.

2 Source Table

3 Data Settings

4 Data Preview

5 Data Placement

8 columns

	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	db558f89-c264-4d82-a246-5939e5...	NULL	example@otrprivacy.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@mail.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	f54bd07c-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

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 a configuration window for data placement. 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 titled 'STORAGE' and contains the following settings:

- 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)

The 'SCHEDULE' section below contains:

- 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**.