td-client-python

Release 1.0.1

Arm Treasure Data

Oct 10, 2019

CONTENTS

Treasure Data API library for Python

CHAPTER

ONE

REQUIREMENTS

`td-client` supports the following versions of Python.

- Python 3.5+
- PyPy

CHAPTER

TWO

INSTALL

You can install the releases from PyPI.

```
$ pip install td-client
```

It'd be better to install certifi to enable SSL certificate verification.

```
$ pip install certifi
```

CHAPTER

THREE

EXAMPLES

Please see also the examples at Treasure Data Documentation.

If you want to find API reference, see also API document.

3.1 Listing jobs

Treasure Data API key will be read from environment variable `TD_API_KEY`, if none is given via `apikey=` argument passed to `tdclient.Client`.

Treasure Data API endpoint `https://api.treasuredata.com` is used by default. You can override this with environment variable `TD_API_SERVER`, which in turn can be overridden via `endpoint=` argument passed to `tdclient.Client`. List of available Treasure Data sites and corresponding API endpoints can be found here.

```
import tdclient

with tdclient.Client() as td:
    for job in td.jobs():
        print(job.job_id)
```

3.2 Running jobs

Running jobs on Treasure Data.

```
import tdclient

with tdclient.Client() as td:
    job = td.query("sample_datasets", "SELECT COUNT(1) FROM www_access", type="hive")
    job.wait()
    for row in job.result():
        print(repr(row))
```

3.3 Running jobs via DBAPI2

td-client-python implements PEP 0249 Python Database API v2.0. You can use td-client-python with external libraries which supports Database API such like pandas.

```
import pandas
import tdclient

def on_waiting(cursor):
    print(cursor.job_status())

with tdclient.connect(db="sample_datasets", type="presto", wait_callback=on_waiting)␣
↪as td:
    data = pandas.read_sql("SELECT symbol, COUNT(1) AS c FROM nasdaq GROUP BY symbol",
↪ td)
    print(repr(data))
```

We offer another package for pandas named pytd with some advanced features. You may prefer it if you need to do complicated things, such like exporting result data to Treasure Data, printing job's progress during long execution, etc.

3.4 Importing data

Importing data into Treasure Data in streaming manner, as similar as fluentd is doing.

```
import sys
import tdclient

with tdclient.Client() as td:
    for file_name in sys.argv[:1]:
        td.import_file("mydb", "mytbl", "csv", file_name)
```

3.5 Bulk import

Importing data into Treasure Data in batch manner.

```
import sys
import tdclient
import time
import warnings

if len(sys.argv) <= 1:
    sys.exit(0)

with tdclient.Client() as td:
    session_name = "session-%d" % (int(time.time()),)
    bulk_import = td.create_bulk_import(session_name, "mydb", "mytbl")
    try:
        for file_name in sys.argv[1:]:
            part_name = "part-%s" % (file_name,)
            bulk_import.upload_file(part_name, "json", file_name)
        bulk_import.freeze()
    except:
        bulk_import.delete()
        raise
    bulk_import.perform(wait=True)
    if 0 < bulk_import.error_records:
        warnings.warn("detected %d error records." % (bulk_import.error_records,))
```

(continues on next page)

(continued from previous page)

```
    if 0 < bulk_import.valid_records:
        print("imported %d records." % (bulk_import.valid_records,))
    else:
        raise(RuntimeError("no records have been imported: %s" % (repr(bulk_import.
↪name),)))
    bulk_import.commit(wait=True)
    bulk_import.delete()
```

CHAPTER

FOUR

DEVELOPMENT

4.1 Running tests

Run tests.

```
$ python setup.py test
```

4.2 Running tests (tox)

You can run tests against all supported Python versions. I'd recommend you to install pyenv to manage Pythons.

```
$ pyenv shell system
$ for version in $(cat .python-version); do [ -d "$(pyenv root)/versions/${version}"
↪] || pyenv install "${version}"; done
$ pyenv shell --unset
```

Install tox.

```
$ pip install tox
```

Then, run `tox`.

```
$ tox
```

4.3 Release

Release to PyPI. Ensure you installed twine.

```
$ python setup.py bdist_wheel sdist
$ twine upload dist/*
```

CHAPTER

FIVE

LICENSE

Apache Software License, Version 2.0

5.1 API Reference

5.1.1 Client

tdclient.client.Client class is a public interface for *tdclient*. It provides methods for executions for REST API.

tdclient.client

class tdclient.client.**Client**(**args*, ***kwargs*)

Bases: object

API Client for Treasure Data Service

add_apikey(*name*)

Parameters **name** (*str*) – name of the user

Returns *True* if success

add_user(*name*, *org*, *email*, *password*)

Add a new user

Parameters

- **name** (*str*) – name of the user
- **org** (*str*) – organization
- **email** – (str): e-mail address
- **password** (*str*) – password

Returns *True* if success

bulk_import(*name*)

Get a bulk import session

Parameters **name** (*str*) – name of a bulk import session

Returns *tdclient.models.BulkImport*

bulk_import_delete_part(*name*, *part_name*)

Delete a part from a bulk import session

Parameters

- **name** (*str*) – name of a bulk import session
- **part_name** (*str*) – name of a part of the bulk import session

Returns *True* if success

bulk_import_error_records(*name*)

Parameters **name** (*str*) – name of a bulk import session

Returns an iterator of error records

bulk_import_upload_file(*name*, *part_name*, *format*, *file*)
Upload a part to Bulk Import session, from an existing file on filesystem.

Parameters

- **name** (*str*) – name of a bulk import session
- **part_name** (*str*) – name of a part of the bulk import session
- **format** (*str*) – format of data type (e.g. "msgpack", "json")
- **file** (*str or file-like*) – a name of a file, or a file-like object contains the data

bulk_import_upload_part(*name*, *part_name*, *bytes_or_stream*, *size*)
Upload a part to a bulk import session

Parameters

- **name** (*str*) – name of a bulk import session
- **part_name** (*str*) – name of a part of the bulk import session
- **bytes_or_stream** (*file-like*) – a file-like object contains the part
- **size** (*int*) – the size of the part

bulk_imports()
List bulk import sessions

Returns a list of *tdclient.models.BulkImport*

change_database(*db_name*, *table_name*, *new_db_name*)
Move a target table from it's original database to new destination database.

Parameters

- **db_name** (*str*) – Target database name.
- **table_name** (*str*) – Target table name.
- **new_db_name** (*str*) – Destination database name to be moved.

Returns *True* if succeeded.

Return type bool

close()
Close opened API connections.

commit_bulk_import(*name*)
Commit a bulk import session

Parameters **name** (*str*) – name of a bulk import session

Returns *True* if success

create_bulk_import(*name*, *database*, *table*, *params=None*)
Create new bulk import session

Parameters

- **name** (*str*) – name of new bulk import session
- **database** (*str*) – name of a database
- **table** (*str*) – name of a table

Returns *tdclient.models.BulkImport*

create_database(*db_name*, ***kwargs*)

Parameters **db_name** (*str*) – name of a database to create

Returns *True* if success

create_log_table(*db_name*, *table_name*)

Parameters

- **db_name** (*str*) – name of a database
- **table_name** (*str*) – name of a table to create

Returns *True* if success

create_result(*name*, *url*, *params=None*)
Create a new authentication with the specified name.

Parameters

- **name** (*str*) – Authentication name.
- **url** (*str*) – Url of the authentication to be created. e.g. "ftp://test.com/"
- **params** (*dict, optional*) – Extra parameters.

Returns True if succeeded.

Return type bool

create_schedule(*name*, *params=None*)
Create a new scheduled query with the specified name.

Parameters

- **name** (*str*) – Scheduled query name.
- **params** (*dict, optional*) – Extra parameters.
  - **type (str):** Query type. {"presto", "hive"}. Default: "hive"
  - **database (str):** Target database name.
  - **timezone (str):** Scheduled query's timezone. e.g. "UTC" For details, see also: https://gist.github.com/frsyuki/4533752
  - **cron (str, optional):** Schedule of the query. {`"@daily"`, `"@hourly"`, `"10 * * * *"` (custom cron)} See also: https://support.treasuredata.com/hc/en-us/articles/360001451088-Scheduled-Jobs-Web-Console
  - **delay (int, optional):** A delay ensures all buffered events are imported before running the query. Default: 0

- **query (str):** Is a language used to retrieve, insert, update and modify data. See also: https://support.treasuredata.com/hc/en-us/articles/360012069493-SQL-Examples-of-Scheduled-Queries
- **priority (int, optional):** Priority of the query. Range is from -2 (very low) to 2 (very high). Default: 0
- **retry_limit (int, optional):** Automatic retry count. Default: 0
- **engine_version (str, optional):** Engine version to be used. If none is specified, the account's default engine version would be set. {"stable", "experimental"}
- **pool_name (str, optional):** For Presto only. Pool name to be used, if not specified, default pool would be used.
- **result (str, optional):** Location where to store the result of the query. e.g. 'tableau://user:password@host.com:1234/datasource'

Returns Start date time.

Return type `datetime.datetime`

database(*db_name*)

Parameters **db_name** (*str*) – name of a database

Returns *tdclient.models.Database*

databases()

Returns a list of *tdclient.models.Database*

delete_bulk_import(*name*)

Delete a bulk import session

Parameters **name** (*str*) – name of a bulk import session

Returns *True* if success

delete_database(*db_name*)

Parameters **db_name** (*str*) – name of database to delete

Returns *True* if success

delete_result(*name*)

Delete the authentication having the specified name.

Parameters **name** (*str*) – Authentication name.

Returns True if succeeded.

Return type bool

delete_schedule(*name*)

Delete the scheduled query with the specified name.

Parameters **name** (*str*) – Target scheduled query name.

Returns Tuple of cron and query.

Return type (str, str)

delete_table(*db_name*, *table_name*)

Delete a table

Parameters

- **db_name** (*str*) – name of a database
- **table_name** (*str*) – name of a table

Returns a string represents the type of deleted table

export_data(*db_name*, *table_name*, *storage_type*, *params=None*)
Export data from Treasure Data Service

Parameters

- **db_name** (*str*) – name of a database
- **table_name** (*str*) – name of a table
- **storage_type** (*str*) – type of the storage
- **params** (*dict*) – optional parameters. Assuming the following keys:
  - **access_key_id (str):** ID to access the information to be exported.
  - **secret_access_key (str):** Password for the *access_key_id*.
  - **file_prefix (str, optional):** Filename of exported file. Default: "<database_name>/<table_name>"
  - **file_format (str, optional):** File format of the information to be exported. {"jsonl.gz", "tsv.gz", "json.gz"}
  - **from (int, optional):** From Time of the data to be exported in Unix epoch format.
  - **to (int, optional):** End Time of the data to be exported in Unix epoch format.
  - assume_role (str, optional): Assume role.
  - **bucket (str):** Name of bucket to be used.
  - **domain_key (str, optional):** Job domain key.
  - **pool_name (str, optional):** For Presto only. Pool name to be used, if not specified, default pool would be used.

Returns *tdclient.models.Job*

freeze_bulk_import(*name*)
Freeze a bulk import session

Parameters **name** (*str*) – name of a bulk import session

Returns *True* if success

history(*name*, *_from=None*, *to=None*)
Get the history details of the saved query for the past 90days.

Parameters

- **name** (*str*) – Target name of the scheduled query.
- **_from** (*int, optional*) – Indicates from which nth record in the run history would be fetched. Default: 0. Note: Count starts from zero. This means that the first record in the list has a count of zero.
- **to** (*int, optional*) – Indicates up to which nth record in the run history would be fetched. Default: 20

Returns [*tdclient.models.ScheduledJob*]

import_data(*db_name*, *table_name*, *format*, *bytes_or_stream*, *size*, *unique_id=None*)
Import data into Treasure Data Service

Parameters

- **db_name** (*str*) – name of a database
- **table_name** (*str*) – name of a table
- **format** (*str*) – format of data type (e.g. "msgpack.gz")
- **bytes_or_stream** (*str or file-like*) – a byte string or a file-like object contains the data
- **size** (*int*) – the length of the data
- **unique_id** (*str*) – a unique identifier of the data

Returns second in float represents elapsed time to import data

import_file(*db_name*, *table_name*, *format*, *file*, *unique_id=None*)
Import data into Treasure Data Service, from an existing file on filesystem.

This method will decompress/deserialize records from given file, and then convert it into format acceptable from Treasure Data Service ("msgpack.gz").

Parameters

- **db_name** (*str*) – name of a database
- **table_name** (*str*) – name of a table
- **format** (*str*) – format of data type (e.g. "msgpack", "json")
- **file** (*str or file-like*) – a name of a file, or a file-like object contains the data
- **unique_id** (*str*) – a unique identifier of the data

Returns float represents the elapsed time to import data

job(*job_id*)
Get a job from *job_id*

Parameters **job_id** (*str*) – job id

Returns *tdclient.models.Job*

job_result(*job_id*)

Parameters **job_id** (*str*) – job id

Returns a list of each rows in result set

job_result_each(*job_id*)

Parameters **job_id** (*str*) – job id

Returns an iterator of result set

job_result_format(*job_id*, *format*)

Parameters

- **job_id** (*str*) – job id
- **format** (*str*) – output format of result set

Returns a list of each rows in result set

job_result_format_each(*job_id*, *format*)

Parameters

- **job_id** (*str*) – job id
- **format** (*str*) – output format of result set

Returns an iterator of rows in result set

job_status(*job_id*)

Parameters **job_id** (*str*) – job id

Returns a string represents the status of the job ("success", "error", "killed", "queued", "running")

jobs(*_from=None*, *to=None*, *status=None*, *conditions=None*)

List jobs

Parameters

- **_from** (*int, optional*) – Gets the Job from the nth index in the list. Default: 0.
- **to** (*int, optional*) – Gets the Job up to the nth index in the list. By default, the first 20 jobs in the list are displayed
- **status** (*str, optional*) – Filter by given status. {"queued", "running", "success", "error"}
- **conditions** (*str, optional*) – Condition for `TIMESTAMPDIFF()` to search for slow queries. Avoid using this parameter as it can be dangerous.

Returns a list of *tdclient.models.Job*

kill(*job_id*)

Parameters **job_id** (*str*) – job id

Returns a string represents the status of killed job ("queued", "running")

list_apikeys(*name*)

Parameters **name** (*str*) – name of the user

Returns a list of string of API key

list_bulk_import_parts(*name*)

List parts of a bulk import session

Parameters **name** (*str*) – name of a bulk import session

Returns a list of string represents the name of parts

partial_delete(*db_name*, *table_name*, *to*, *_from*, *params=None*)

Create a job to partially delete the contents of the table with the given time range.

Parameters

- **db_name** (*str*) – Target database name.
- **table_name** (*str*) – Target table name.
- **to** (*int*) – Time in Unix Epoch format indicating the End date and time of the data to be deleted. Should be set only by the hour. Minutes and seconds values will not be accepted.

- **_from** (*int*) – Time in Unix Epoch format indicating the Start date and time of the data to be deleted. Should be set only by the hour. Minutes and seconds values will not be accepted.
- **params** (*dict, optional*) – Extra parameters.
  - **pool_name (str, optional):** Indicates the resource pool to execute this job. If not provided, the account's default resource pool would be used.
  - **domain_key (str, optional):** Domain key that will be assigned to the partial delete job to be created

Returns *tdclient.models.Job*

perform_bulk_import(*name*)
Perform a bulk import session

Parameters **name** (*str*) – name of a bulk import session

Returns *tdclient.models.Job*

query(*db_name*, *q*, *result_url=None*, *priority=None*, *retry_limit=None*, *type='hive'*, ***kwargs*)
Run a query on specified database table.

Parameters

- **db_name** (*str*) – name of a database
- **q** (*str*) – a query string
- **result_url** (*str*) – result output URL. e.g., `postgresql://<username>:<password>@<hostname>:<port>/<database>/<table>`
- **priority** (*int or str*) – priority (e.g. "NORMAL", "HIGH", etc.)
- **retry_limit** (*int*) – retry limit
- **type** (*str*) – name of a query engine

Returns *tdclient.models.Job*

Raises **ValueError** – if unknown query type has been specified

remove_apikey(*name*, *apikey*)

Parameters

- **name** (*str*) – name of the user
- **apikey** (*str*) – an API key to remove

Returns *True* if success

remove_user(*name*)
Remove a user

Parameters **name** (*str*) – name of the user

Returns *True* if success

results()
Get the list of all the available authentications.

Returns a list of *tdclient.models.Result*

run_schedule(*name*, *time*, *num*)
Execute the specified query.

Parameters

- **name** (*str*) – Target scheduled query name.
- **time** (*int*) – Time in Unix epoch format that would be set as TD_SCHEDULED_TIME
- **num** (*int*) – Indicates how many times the query will be executed. Value should be 9 or less.

Returns [*tdclient.models.ScheduledJob*]

schedules()

Get the list of all the scheduled queries.

Returns [*tdclient.models.Schedule*]

server_status()

Returns a string represents current server status.

swap_table(*db_name*, *table_name1*, *table_name2*)

Parameters

- **db_name** (*str*) – name of a database
- **table_name1** (*str*) – original table name
- **table_name2** (*str*) – table name you want to rename to

Returns *True* if success

table(*db_name*, *table_name*)

Parameters

- **db_name** (*str*) – name of a database
- **table_name** (*str*) – name of a table

Returns *tdclient.models.Table*

Raises **tdclient.api.NotFoundError** – if the table doesn't exist

tables(*db_name*)

List existing tables

Parameters **db_name** (*str*) – name of a database

Returns a list of *tdclient.models.Table*

tail(*db_name*, *table_name*, *count*, *to=None*, *_from=None*, *block=None*)

Get the contents of the table in reverse order based on the registered time (last data first).

Parameters

- **db_name** (*str*) – Target database name.
- **table_name** (*str*) – Target table name.
- **count** (*int*) – Number for record to show up from the end.
- **to** – Deprecated parameter.
- **_from** – Deprecated parameter.
- **block** – Deprecated parameter.

Returns Contents of the table.

Return type [dict]

unfreeze_bulk_import (*name*)
Unfreeze a bulk import session

Parameters **name** (*str*) – name of a bulk import session

Returns *True* if success

update_expire (*db_name*, *table_name*, *expire_days*)
Set expiration date to a table

Parameters

- **db_name** (*str*) – name of a database
- **table_name** (*str*) – name of a table
- **epire_days** (*int*) – expiration date in days from today

Returns *True* if success

update_schedule (*name*, *params=None*)
Update the scheduled query.

Parameters

- **name** (*str*) – Target scheduled query name.
- **params** (*dict*) – Extra parameteres.
  - **type (str):** Query type. {"presto", "hive"}. Default: "hive"
  - **database (str):** Target database name.
  - **timezone (str):** Scheduled query's timezone. e.g. "UTC" For details, see also: https://gist.github.com/frsyuki/4533752
  - **cron (str, optional):** Schedule of the query. {`"@daily"`, `"@hourly"`, `"10 * * * *"` (custom cron)} See also: https://support.treasuredata.com/hc/en-us/articles/360001451088-Scheduled-Jobs-Web-Console
  - **delay (int, optional):** A delay ensures all buffered events are imported before running the query. Default: 0
  - **query (str):** Is a language used to retrieve, insert, update and modify data. See also: https://support.treasuredata.com/hc/en-us/articles/360012069493-SQL-Examples-of-Scheduled-Queries
  - **priority (int, optional):** Priority of the query. Range is from -2 (very low) to 2 (very high). Default: 0
  - **retry_limit (int, optional):** Automatic retry count. Default: 0
  - **engine_version (str, optional):** Engine version to be used. If none is specified, the account's default engine version would be set. {"stable", "experimental"}
  - **pool_name (str, optional):** For Presto only. Pool name to be used, if not specified, default pool would be used.
  - **result (str, optional):** Location where to store the result of the query. e.g. 'tableau://user:password@host.com:1234/datasource'

update_schema (*db_name*, *table_name*, *schema*)
Updates the schema of a table

Parameters

- **db_name** (*str*) – name of a database
- **table_name** (*str*) – name of a table
- **schema** (*list*) – a dictionary object represents the schema definition (will be converted to JSON) e.g.

```
[
    ["member_id", # column name
     "string", # data type
     "mem_id", # alias of the column name
    ],
    ["row_index", "long", "row_ind"],
    ...
]
```

Returns *True* if success

users()
List users

Returns a list of *tdclient.models.User*

property api
an instance of *tdclient.api.API*

property apikey
API key string.

5.1.2 DB API

tdclient

tdclient.**Binary**(*string*)

tdclient.**DateFromTicks**(*ticks*)

tdclient.**TimeFromTicks**(*ticks*)

tdclient.**TimestampFromTicks**(*ticks*)

tdclient.**connect**(**args*, ***kwargs*)
Returns a DBAPI compatible connection object

Parameters

- **type** (*str*) – query engine type. "hive" by default.
- **db** (*str*) – the name of database on Treasure Data
- **result_url** (*str*) – result output URL
- **priority** (*str*) – job priority
- **retry_limit** (*int*) – job retry limit
- **wait_interval** (*int*) – job wait interval to check status
- **wait_callback** (*callable*) – a callback to be called on every ticks of job wait

Returns *tdclient.connection.Connection*

tdclient.connection

class tdclient.connection.**Connection**(*type=None*, *db=None*, *result_url=None*, *priority=None*, *retry_limit=None*, *wait_interval=None*, *wait_callback=None*, ***kwargs*)

Bases: object

close()

commit()

cursor()

rollback()

property api

tdclient.cursor

class tdclient.cursor.**Cursor**(*api*, *wait_interval=5*, *wait_callback=None*, ***kwargs*)

Bases: object

callproc(*procname*, **parameters*)

close()

execute(*query*, *args=None*)

executemany(*operation*, *seq_of_parameters*)

fetchall()
Fetch all (remaining) rows of a query result, returning them as a sequence of sequences (e.g. a list of tuples). Note that the cursor's arraysize attribute can affect the performance of this operation.

fetchmany(*size=None*)
Fetch the next set of rows of a query result, returning a sequence of sequences (e.g. a list of tuples). An empty sequence is returned when no more rows are available.

fetchone()
Fetch the next row of a query result set, returning a single sequence, or *None* when no more data is available.

job_result()
Fetch job results

Returns Job result in list

job_status()
Show job status

Returns The status information of the given job id at last execution.

nextset()

setinputsizes(*sizes*)

setoutputsize(*size*, *column=None*)

show_job()
Returns detailed information of a Job

Returns Detailed information of a job

Return type dict

property api

property description

property rowcount

5.1.3 Model

Some methods of *tdclient.client.Client* returns model object which represents results from REST API.

tdclient.model

class tdclient.model.**Model**(*client*)
Bases: object

property client
a *tdclient.client.Client* instance

Type Returns

tdclient.models

tdclient.models.**BulkImport = <class 'tdclient.bulk_import_model.BulkImport'>**
Bulk-import session on Treasure Data Service

tdclient.models.**Database = <class 'tdclient.database_model.Database'>**
Database on Treasure Data Service

tdclient.models.**Schema = <class 'tdclient.job_model.Schema'>**
Schema of a database table on Treasure Data Service

tdclient.models.**Job = <class 'tdclient.job_model.Job'>**
Job on Treasure Data Service

tdclient.models.**Result = <class 'tdclient.result_model.Result'>**
Result on Treasure Data Service

tdclient.models.**ScheduledJob = <class 'tdclient.schedule_model.ScheduledJob'>**
Scheduled job on Treasure Data Service

tdclient.models.**Schedule = <class 'tdclient.schedule_model.Schedule'>**
Schedule on Treasure Data Service

tdclient.models.**Table = <class 'tdclient.table_model.Table'>**
Database table on Treasure Data Service

tdclient.models.**User = <class 'tdclient.user_model.User'>**
User on Treasure Data Service

tdclient.bulk_import_model

class tdclient.bulk_import_model.**BulkImport**(*client*, ***kwargs*)
Bases: *tdclient.model.Model*

Bulk-import session on Treasure Data Service

commit(*wait=False*, *wait_interval=5*, *timeout=None*)
Commit bulk import

delete()
Delete bulk import

delete_part(*part_name*)
Delete a part of a Bulk Import session

Parameters **part_name** (*str*) – name of a part of the bulk import session

Returns True if succeeded.

error_record_items()
Fetch error record rows.

Yields Error record

freeze()
Freeze bulk import

list_parts()
Return the list of available parts uploaded through `bulk_import_upload_part()`.

Returns The list of bulk import part name.

Return type [str]

perform(*wait=False*, *wait_interval=5*, *wait_callback=None*)
Perform bulk import

Parameters

- **wait** (*bool, optional*) – Flag for wait bulk import job. Default *False*
- **wait_interval** (*int, optional*) – wait interval in second. Default *5*.
- **wait_callback** (*callable, optional*) – A callable to be called on every tick of wait interval.

unfreeze()
Unfreeze bulk import

update()

upload_file(*part_name*, *fmt*, *file_like*)
Upload a part to Bulk Import session, from an existing file on filesystem.

Parameters

- **part_name** (*str*) – name of a part of the bulk import session
- **fmt** (*str*) – format of data type (e.g. "msgpack", "json")
- **file_like** (*str or file-like*) – a name of a file, or a file-like object contains the data

upload_part(*part_name*, *bytes_or_stream*, *size*)
Upload a part to bulk import session

Parameters

- **part_name** (*str*) – name of a part of the bulk import session
- **bytes_or_stream** (*file-like*) – a file-like object contains the part
- **size** (*int*) – the size of the part

STATUS_COMMITTED = 'committed'

STATUS_COMMITTING = 'committing'

STATUS_PERFORMING = 'performing'

STATUS_READY = 'ready'

STATUS_UPLOADING = 'uploading'

property database
A database name in a string which the bulk import session is working on

property error_parts
The number of error parts.

property error_records
The number of error records.

property job_id
Job ID

property name
A name of the bulk import session

property status
The status of the bulk import session in a string

property table
A table name in a string which the bulk import session is working on

property upload_frozen
The number of upload frozen.

property valid_parts
The number of valid parts.

property valid_records
The number of valid records.

tdclient.database_model

class tdclient.database_model.**Database**(*client*, *db_name*, ***kwargs*)
Bases: *tdclient.model.Model*

Database on Treasure Data Service

create_log_table(*name*)

Parameters **name** (*str*) – name of new log table

Returns tdclient.model.Table

delete()
Delete the database

Returns *True* if success

query(*q*, ***kwargs*)
Run a query on the database

Parameters **q** (*str*) – a query string

Returns tdclient.model.Job

table(*table_name*)

Parameters **table_name** (*str*) – name of a table

Returns `tdclient.model.Table`

tables()

Returns a list of `tdclient.model.Table`

PERMISSIONS = ['administrator', 'full_access', 'import_only', 'query_only']

PERMISSION_LIST_TABLES = ['administrator', 'full_access']

property count
Total record counts in a database.

Type int

property created_at
`datetime.datetime`

property name
a name of the database

Type str

property org_name
organization name

Type str

property permission
permission for the database (e.g. "administrator", "full_access", etc.)

Type str

property updated_at
`datetime.datetime`

tdclient.job_model

class tdclient.job_model.**Job**(*client*, *job_id*, *type*, *query*, ***kwargs*)
Bases: *tdclient.model.Model*

Job on Treasure Data Service

error()

Returns *True* if the job has been finished in error

finished()

Returns *True* if the job has been finished in success, error or killed

kill()
Kill the job

Returns a string represents the status of killed job ("queued", "running")

killed()

Returns *True* if the job has been finished in killed

queued()

Returns *True* if the job is queued

result()

Yields an iterator of rows in result set

result_format(*fmt*)

Parameters **fmt** (*str*) – output format of result set

Yields an iterator of rows in result set

running()

Returns *True* if the job is running

status()

Returns a string represents the status of the job ("success", "error", "killed", "queued", "running")

Return type str

success()

Returns *True* if the job has been finished in success

update()
Update all fields of the job

wait(*timeout=None*, *wait_interval=5*, *wait_callback=None*)
Sleep until the job has been finished

Parameters

- **timeout** (*int, optional*) – Timeout in seconds. No timeout by default.
- **wait_interval** (*int, optional*) – wait interval in second. Default 5 seconds.
- **wait_callback** (*callable, optional*) – A callable to be called on every tick of wait interval.

FINISHED_STATUS = ['success', 'error', 'killed']

JOB_PRIORITY = {-2: 'VERY LOW', -1: 'LOW', 0: 'NORMAL', 1: 'HIGH', 2: 'VERY HIGH'

STATUS_BOOTING = 'booting'

STATUS_ERROR = 'error'

STATUS_KILLED = 'killed'

STATUS_QUEUED = 'queued'

STATUS_RUNNING = 'running'

STATUS_SUCCESS = 'success'

property database
a string represents the name of a database that job is running on

property debug
a `dict` of debug output (e.g. "cmdout", "stderr")

property id
a string represents the identifier of the job

property job_id
a string represents the identifier of the job

property linked_result_export_job_id
Linked result export job ID from query job

property num_records
the number of records of job result

property org_name
organization name

property priority
a string represents the priority of the job (e.g. “NORMAL”, “HIGH”, etc.)

property query
a string represents the query string of the job

property result_export_target_job_id
Associated query job ID from result export job ID

property result_schema
an array of array represents the type of result columns (Hive specific) (e.g. [[“_c1”, “string”], [“_c2”, “bigint”]])

property result_size
the length of job result

property result_url
a string of URL of the result on Treasure Data Service

property retry_limit
a number for automatic retry count

property type
a string represents the engine type of the job (e.g. “hive”, “presto”, etc.)

property url
a string of URL of the job on Treasure Data Service

property user_name
executing user name

class tdclient.job_model.**Schema**(*fields=None*)
Bases: object

Schema of a database table on Treasure Data Service

class Field(*name*, *type*)
Bases: object

property name
add docstring

Type TODO

property type
add docstring

Type TODO

add_field(*name*, *type*)
TODO: add docstring

property fields
add docstring

Type TODO

tdclient.result_model

class tdclient.result_model.**Result**(*client*, *name*, *url*, *org_name*)
Bases: *tdclient.model.Model*

Result on Treasure Data Service

property name
a name for a authentication

Type str

property org_name
organization name

Type str

property url
a result output URL

Type str

tdclient.schedule_model

class tdclient.schedule_model.**Schedule**(*client*, **args*, ***kwargs*)
Bases: *tdclient.model.Model*

Schedule on Treasure Data Service

run(*time*, *num=None*)
Run a scheduled job

Parameters

- **time** (*int*) – Time in Unix epoch format that would be set as TD_SCHEDULED_TIME
- **num** (*int*) – Indicates how many times the query will be executed. Value should be 9 or less.

Returns [*tdclient.models.ScheduledJob*]

property created_at
Create date

Type datetime.datetime

property cron
The configured schedule of a scheduled job.

Returns a string represents the schedule in cron form, or *None* if the job is not scheduled to run (saved query)

property database
The target database of a scheduled job

property delay
A delay ensures all buffered events are imported before running the query.

property name
The name of a scheduled job

property next_time
Schedule for next run

Type datetime.datetime

property org_name
add docstring

Type TODO

property priority
The priority of a scheduled job

property query
The query string of a scheduled job

property result_url
The result output configuration in URL form of a scheduled job

property retry_limit
Automatic retry count.

property timezone
The time zone of a scheduled job

property type
Query type. {"presto", "hive"}.

property user_name
User name of a scheduled job

class tdclient.schedule_model.**ScheduledJob**(*client*, *scheduled_at*, *job_id*, *type*, *query*, ***kwargs*)
Bases: *tdclient.job_model.Job*

Scheduled job on Treasure Data Service

property scheduled_at
a datetime.datetime represents the schedule of next invocation of the job

tdclient.table_model

class tdclient.table_model.**Table**(**args*, ***kwargs*)
Bases: *tdclient.model.Model*

Database table on Treasure Data Service

delete()
a string represents the type of deleted table

export_data(*storage_type*, ***kwargs*)
Export data from Treasure Data Service

Parameters

- **storage_type** (*str*) – type of the storage
- ****kwargs** (*dict*) – optional parameters. Assuming the following keys:
  - **access_key_id (str):** ID to access the information to be exported.
  - **secret_access_key (str):** Password for the *access_key_id*.
  - **file_prefix (str, optional):** Filename of exported file. Default: "<database_name>/<table_name>"

- **file_format (str, optional):** File format of the information to be exported. {"jsonl.gz", "tsv.gz", "json.gz"}
- **from (int, optional):** From Time of the data to be exported in Unix epoch format.
- **to (int, optional):** End Time of the data to be exported in Unix epoch format.
- **assume_role (str, optional):** Assume role.
- **bucket (str):** Name of bucket to be used.
- **domain_key (str, optional):** Job domain key.
- **pool_name (str, optional):** For Presto only. Pool name to be used, if not specified, default pool would be used.

Returns *tdclient.models.Job*

import_data (*format*, *bytes_or_stream*, *size*, *unique_id=None*)
Import data into Treasure Data Service

Parameters

- **format** (*str*) – format of data type (e.g. "msgpack.gz")
- **bytes_or_stream** (*str or file-like*) – a byte string or a file-like object contains the data
- **size** (*int*) – the length of the data
- **unique_id** (*str*) – a unique identifier of the data

Returns second in float represents elapsed time to import data

import_file (*format*, *file*, *unique_id=None*)
Import data into Treasure Data Service, from an existing file on filesystem.

This method will decompress/deserialize records from given file, and then convert it into format acceptable from Treasure Data Service ("msgpack.gz").

Parameters

- **file** (*str or file-like*) – a name of a file, or a file-like object contains the data
- **unique_id** (*str*) – a unique identifier of the data

Returns float represents the elapsed time to import data

tail (*count*, *to=None*, *_from=None*)

Parameters

- **count** (*int*) – Number for record to show up from the end.
- **to** – Deprecated parameter.
- **_from** – Deprecated parameter.

Returns the contents of the table in reverse order based on the registered time (last data first).

property count
total number of the table

Type int

property created_at
Created datetime

Type datetime.datetime

property database_name
a string represents the name of the database

property db_name
a string represents the name of the database

property estimated_storage_size
estimated storage size

property estimated_storage_size_string
a string represents estimated size of the table in human-readable format

property expire_days
an int represents the days until expiration

property identifier
a string identifier of the table

property last_import
datetime.datetime

property last_log_timestamp
datetime.datetime

property name
a string represents the name of the table

property permission
permission for the database (e.g. "administrator", "full_access", etc.)

Type str

property primary_key
add docstring

Type TODO

property primary_key_type
add docstring

Type TODO

property schema
str, alias:str]]: The list of a schema

Type [[column_name

Type str, column_type

property table_name
a string represents the name of the table

property type
a string represents the type of the table

property updated_at
Updated datetime

Type datetime.datetime

tdclient.user_model

class tdclient.user_model.**User**(*client*, *name*, *org_name*, *role_names*, *email*, ***kwargs*)
Bases: *tdclient.model.Model*

User on Treasure Data Service

property email
e-mail address

Type Returns

property name
name of the user

Type Returns

property org_name
organization name

Type Returns

property role_names
add docstring

Type TODO

5.1.4 API

tdclient.api.API class is an internal class represents API.

tdclient.api

class tdclient.api.**API**(*apikey=None*, *user_agent=None*, *endpoint=None*, *headers=None*, *retry_post_requests=False*, *max_cumul_retry_delay=600*, *http_proxy=None*, ***kwargs*)
Bases: *tdclient.bulk_import_api.BulkImportAPI*, *tdclient.connector_api.ConnectorAPI*, *tdclient.database_api.DatabaseAPI*, *tdclient.export_api.ExportAPI*, *tdclient.import_api.ImportAPI*, *tdclient.job_api.JobAPI*, *tdclient.partial_delete_api.PartialDeleteAPI*, *tdclient.result_api.ResultAPI*, *tdclient.schedule_api.ScheduleAPI*, *tdclient.server_status_api.ServerStatusAPI*, *tdclient.table_api.TableAPI*, *tdclient.user_api.UserAPI*

Internal API class

Parameters

- **apikey** (*str*) – the API key of Treasure Data Service. If *None* is given, *TD_API_KEY* will be used if available.
- **user_agent** (*str*) – custom User-Agent.
- **endpoint** (*str*) – custom endpoint URL. If *None* is given, *TD_API_SERVER* will be used if available.
- **headers** (*dict*) – custom HTTP headers.
- **retry_post_requests** (*bool*) – Specify whether allowing API client to retry POST requests. *False* by default.

- **max_cumul_retry_delay** (*int*) – maximum retry limit in seconds. 600 seconds by default.
- **http_proxy** (*str*) – HTTP proxy setting. if *None* is given, *HTTP_PROXY* will be used if available.

build_request(*path=None*, *headers=None*, *endpoint=None*)

checked_json(*body*, *required*)

close()

delete(*path*, *params=None*, *headers=None*, ***kwargs*)

get(*path*, *params=None*, *headers=None*, ***kwargs*)

get_or_else(*hashmap*, *key*, *default_value=None*)

post(*path*, *params=None*, *headers=None*, ***kwargs*)

put(*path*, *bytes_or_stream*, *size*, *headers=None*, ***kwargs*)

raise_error(*msg*, *res*, *body*)

send_request(*method*, *url*, *fields=None*, *body=None*, *headers=None*, ***kwargs*)

DEFAULT_ENDPOINT = 'https://api.treasuredata.com/'

DEFAULT_IMPORT_ENDPOINT = 'https://api-import.treasuredata.com/'

property apikey

property endpoint

tdclient.api.**normalized_msgpack**(*value*)

tdclient.bulk_import_api

class tdclient.bulk_import_api.**BulkImportAPI**

Bases: object

Enable bulk importing of data to the targeted database and table.

This class is inherited by *tdclient.api.API*.

bulk_import_delete_part(*name*, *part_name*, *params=None*)

Delete the imported information with the specified name.

Parameters

- **name** (*str*) – Bulk import name.
- **part_name** (*str*) – Bulk import part name.
- **params** (*dict, optional*) – Extra parameters.

Returns True if succeeded.

bulk_import_error_records(*name*, *params=None*)

List the records that have errors under the specified bulk import name.

Parameters

- **name** (*str*) – Bulk import name.
- **params** (*dict, optional*) – Extra parameters.

Yields Row of the data

bulk_import_upload_file(*name*, *part_name*, *format*, *file*, ***kwargs*)
Upload a file with bulk import having the specified name.

Parameters

- **name** (*str*) – Bulk import name.
- **part_name** (*str*) – Bulk import part name.
- **format** (*str*) – Format name. {msgpack, json, csv, tsv}
- **file** (*file-like*) – Byte string or file-like object contains the data.
- ****kwargs** – Extra argments.

bulk_import_upload_part(*name*, *part_name*, *stream*, *size*)
Upload bulk import having the specified name and part in the path.

Parameters

- **name** (*str*) – Bulk import name.
- **part_name** (*str*) – Bulk import part name.
- **stream** (*str or file-like*) – Byte string or file-like object contains the data
- **size** (*int*) – The length of the data.

commit_bulk_import(*name*, *params=None*)
Commit the bulk import information having the specified name.

Parameters

- **name** (*str*) – Bulk import name.
- **params** (*dict, optional*) – Extra parameters.

Returns True if succeeded.

create_bulk_import(*name*, *db*, *table*, *params=None*)
Enable bulk importing of data to the targeted database and table and stores it in the default resource pool. Default expiration for bulk import is 30days.

Parameters

- **name** (*str*) – Name of the bulk import.
- **db** (*str*) – Name of target database.
- **table** (*str*) – Name of target table.
- **params** (*dict, optional*) – Extra parameters.

Returns True if succeeded

delete_bulk_import(*name*, *params=None*)
Delete the imported information with the specified name

Parameters

- **name** (*str*) – Name of bulk import.
- **params** (*dict, optional*) – Extra parameters.

Returns True if succeeded

freeze_bulk_import(*name*, *params=None*)
Freeze the bulk import with the specified name.

Parameters

- **name** (*str*) – Bulk import name.
- **params** (*dict, optional*) – Extra parameters.

Returns True if succeeded.

list_bulk_import_parts(*name*, *params=None*)
Return the list of available parts uploaded through *bulk_import_upload_part()*.

Parameters

- **name** (*str*) – Name of bulk import.
- **params** (*dict, optional*) – Extra parameteres.

Returns The list of bulk import part name.

Return type [str]

list_bulk_imports(*params=None*)
Return the list of available bulk imports :param params: Extra parameters. :type params: dict, optional

Returns The list of available bulk import details.

Return type [dict]

perform_bulk_import(*name*, *params=None*)
Execute a job to perform bulk import with the indicated priority using the resource pool if indicated, else it will use the account's default.

Parameters

- **name** (*str*) – Bulk import name.
- **params** (*dict, optional*) – Extra parameters.

Returns Job ID

Return type str

show_bulk_import(*name*)
Show the details of the bulk import with the specified name

Parameters **name** (*str*) – Name of bulk import.

Returns Detailed information of the bulk import.

Return type dict

unfreeze_bulk_import(*name*, *params=None*)
Unfreeze bulk_import with the specified name.

Parameters

- **name** (*str*) – Bulk import name.
- **params** (*dict, optional*) – Extra parameters.

Returns True if succeeded.

static validate_part_name(*part_name*)
Make sure the part_name is valid

Parameters **part_name** (*str*) – The part name the user is trying to use

tdclient.connector_api

class tdclient.connector_api.**ConnectorAPI**
Bases: object

Access Data Connector API which handles Data Connector.

This class is inherited by *tdclient.api.API*.

connector_create(*name*, *database*, *table*, *job*, *params=None*)
Create a Data Connector session.

Parameters

- **name** (*str*) – name of the connector job
- **database** (*str*) – name of the database to perform connector job
- **table** (*str*) – name of the table to perform connector job
- **job** (*dict*) – dict representation of *load.yml*
- **params** (*dict, optional*) – Extra parameters
  - **config (str):** Embulk configuration as JSON format. See also https://www.embulk.org/docs/built-in.html#embulk-configuration-file-format
  - **cron (str, optional):** Schedule of the query. {"@daily", "@hourly", "10 * * * *" (custom cron)} See also: https://support.treasuredata.com/hc/en-us/articles/360001451088-Scheduled-Jobs-Web-Console
  - **delay (int, optional):** A delay ensures all buffered events are imported before running the query. Default: 0
  - **database (str):** Target databse for the Data Connector session
  - **name (str):** Name of the Data Connector session
  - **table (str):** Target table for the Data Connector session
  - **time_column (str, optional):** Column in the table for registering config.out.time
  - **timezone (str):** Timezone for scheduled Data Connector session. See here for list of supported timezones https://gist.github.com/frsyuki/4533752

Returns dict

connector_delete(*name*)
Delete a Data Connector session.

Parameters **name** (*str*) – name of the connector job

Returns dict

connector_guess(*job*)
Guess the Data Connector configuration

Parameters **job** (*dict*) – dict representation of *seed.yml*

Returns dict

connector_history(*name*)
Show the list of the executed jobs information for the Data Connector job.

Parameters **name** (*str*) – name of the connector job

Returns list

connector_issue(*db*, *table*, *job*)
Create a Data Connector job.

Parameters

- **db** (*str*) – name of the database to perform connector job
- **table** (*str*) – name of the table to perform connector job
- **job** (*dict*) – dict representation of *load.yml*

Returns job Id

Return type str

connector_list()
Show the list of available Data Connector sessions.

Returns list

connector_preview(*job*)
Show the preview of the Data Connector job.

Parameters **job** (*dict*) – dict representation of *load.yml*

Returns dict

connector_run(*name*, ***kwargs*)
Create a job to execute Data Connector session.

Parameters

- **name** (*str*) – name of the connector job
- ****kwargs** (*optional*) – Extra parameters.
  - **scheduled_time (int):** Time in Unix epoch format that would be set as *TD_SCHEDULED_TIME*.
  - **domain_key (str):** Job domain key which is assigned to a single job.

Returns dict

connector_show(*name*)
Show a specific Data Connector session information.

Parameters **name** (*str*) – name of the connector job

Returns dict

connector_update(*name*, *job*)
Update a specific Data Connector session.

Parameters

- **name** (*str*) – name of the connector job
- **job** (*dict*) – dict representation of *load.yml*. For detailed format, see also: https://www.embulk.org/docs/built-in.html#embulk-configuration-file-format

Returns dict

tdclient.database_api

class tdclient.database_api.**DatabaseAPI**
Bases: object

Access to Database of Treasure Data Service.

This class is inherited by *tdclient.api.API*.

create_database(*db*, *params=None*)
Create a new database with the given name.

Parameters

- **db** (*str*) – Target database name.
- **params** (*dict*) – Extra parameters.

Returns *True* if succeeded.

Return type bool

delete_database(*db*)
Delete a database.

Parameters **db** (*str*) – Target database name.

Returns *True* if succeeded.

Return type bool

list_databases()
Get the list of all the databases of the account.

Returns Detailed database information. Each key of the dict is database name.

Return type dict

tdclient.export_api

class tdclient.export_api.**ExportAPI**
Bases: object

Access to Export API.

This class is inherited by *tdclient.api.API*.

export_data(*db*, *table*, *storage_type*, *params=None*)
Creates a job to export the contents from the specified database and table names.

Parameters

- **db** (*str*) – Target database name.
- **table** (*str*) – Target table name.
- **storage_type** (*str*) – Name of storage type. e.g. "s3"
- **params** (*dict*) – Extra parameters. Assuming the following keys:
  - **access_key_id (str):** ID to access the information to be exported.
  - **secret_access_key (str):** Password for the *access_key_id*.
  - **file_prefix (str, optional):** Filename of exported file. Default: "<database_name>/<table_name>"
  - **file_format (str, optional):** File format of the information to be exported. {"jsonl.gz", "tsv.gz", "json.gz"}
  - **from (int, optional):** From Time of the data to be exported in Unix epoch format.

- **to (int, optional):** End Time of the data to be exported in Unix epoch format.
- **assume_role (str, optional):** Assume role.
- **bucket (str):** Name of bucket to be used.
- **domain_key (str, optional):** Job domain key.
- **pool_name (str, optional):** For Presto only. Pool name to be used, if not specified, default pool would be used.

Returns Job ID.

Return type str

tdclient.import_api

class tdclient.import_api.**ImportAPI**
Bases: object

Import data into Treasure Data Service.

This class is inherited by *tdclient.api.API*.

import_data(*db*, *table*, *format*, *bytes_or_stream*, *size*, *unique_id=None*)
Import data into Treasure Data Service

This method expects data from a file-like object formatted with "msgpack.gz".

Parameters

- **db** (*str*) – name of a database
- **table** (*str*) – name of a table
- **format** (*str*) – format of data type (e.g. "msgpack.gz")
- **bytes_or_stream** (*str or file-like*) – a byte string or a file-like object contains the data
- **size** (*int*) – the length of the data
- **unique_id** (*str*) – a unique identifier of the data

Returns float represents the elapsed time to import data

import_file(*db*, *table*, *format*, *file*, *unique_id=None*, ***kwargs*)
Import data into Treasure Data Service, from an existing file on filesystem.

This method will decompress/deserialize records from given file, and then convert it into format acceptable from Treasure Data Service ("msgpack.gz"). This method is a warpper function to *import_data*.

Parameters

- **db** (*str*) – name of a database
- **table** (*str*) – name of a table
- **format** (*str*) – format of data type (e.g. "msgpack", "json")
- **file** (*str or file-like*) – a name of a file, or a file-like object contains the data
- **unique_id** (*str*) – a unique identifier of the data

Returns float represents the elapsed time to import data

tdclient.job_api

class tdclient.job_api.**JobAPI**
Bases: object

Access to Job API

This class is inherited by *tdclient.api.API*.

job_result(*job_id*)
Return the job result.

Parameters **job_id** (*int*) – Job ID

Returns Job result in list

job_result_each(*job_id*)
Yield a row of the job result.

Parameters **job_id** (*int*) – Job ID

Yields Row in a result

job_result_format(*job_id*, *format*)
Return the job result with specified format.

Parameters

- **job_id** (*int*) – Job ID
- **format** (*str*) – Output format of the job result information. “json” or “msgpack”

Returns The query result of the specified job in.

job_result_format_each(*job_id*, *format*)
Yield a row of the job result with specified format.

Parameters

- **job_id** (*int*) – job ID
- **format** (*str*) – Output format of the job result information. “json” or “msgpack”

Yields The query result of the specified job in.

job_status(*job_id*)
“Show job status :param job_id: job ID :type job_id: str

Returns The status information of the given job id at last execution.

kill(*job_id*)
Stop the specific job if it is running.

Parameters **job_id** (*str*) – Job Id to kill

Returns Job status before killing

list_jobs(*_from=0*, *to=None*, *status=None*, *conditions=None*)
Show the list of Jobs.

Parameters

- **_from** (*int*) – Gets the Job from the nth index in the list. Default: 0
- **to** (*int, optional*) – Gets the Job up to the nth index in the list. By default, the first 20 jobs in the list are displayed

- **status** (*str, optional*) – Filter by given status. {"queued", "running", "success", "error"}
- **conditions** (*str, optional*) – Condition for `TIMESTAMPDIFF()` to search for slow queries. Avoid using this parameter as it can be dangerous.

Returns a list of `dict` which represents a job

query(*q*, *type='hive'*, *db=None*, *result_url=None*, *priority=None*, *retry_limit=None*, ***kwargs*)

Create a job for given query.

Parameters

- **q** (*str*) – Query string.
- **type** (*str*) – Query type. *hive*, *presto*, *bulkload*. Default: *hive*
- **db** (*str*) – Database name.
- **result_url** (*str*) – Result output URL. e.g., `postgresql://<username>:<password>@<hostname>:<port>/<database>/<table>`
- **priority** (*int or str*) – Job priority. In str, "Normal", "Very low", "Low", "High", "Very high". In int, the number in the range of -2 to 2.
- **retry_limit** (*int*) – Automatic retry count.
- ****kwargs** – Extra options.

Returns Job ID issued for the query

Return type str

show_job(*job_id*)

Return detailed information of a Job.

Parameters **job_id** (*str*) – job ID

Returns Detailed information of a job

Return type `dict`

JOB_PRIORITY = {'HIGH': 1, 'LOW': -1, 'NORM': 0, 'NORMAL': 0, 'VERY HIGH': 2, 'VERY LO

tdclient.partial_delete_api

class tdclient.partial_delete_api.**PartialDeleteAPI**

Bases: `object`

Create a job to partially delete the contents of the table with the given time range.

This class is inherited by *tdclient.api.API*.

partial_delete(*db*, *table*, *to*, *_from*, *params=None*)

Create a job to partially delete the contents of the table with the given time range.

Parameters

- **db** (*str*) – Target database name.
- **table** (*str*) – Target table name.
- **to** (*int*) – Time in Unix Epoch format indicating the End date and time of the data to be deleted. Should be set only by the hour. Minutes and seconds values will not be accepted.

- **_from** (*int*) – Time in Unix Epoch format indicating the Start date and time of the data to be deleted. Should be set only by the hour. Minutes and seconds values will not be accepted.
- **params** (*dict, optional*) – Extra parameters.
  - **pool_name (str, optional):** Indicates the resource pool to execute this job. If not provided, the account's default resource pool would be used.
  - **domain_key (str, optional):** Domain key that will be assigned to the partial delete job to be created

Returns Job ID.

Return type str

tdclient.result_api

class tdclient.result_api.**ResultAPI**
Bases: object

Access to Result API.

This class is inherited by *tdclient.api.API*.

create_result(*name*, *url*, *params=None*)
Create a new authentication with the specified name.

Parameters

- **name** (*str*) – Authentication name.
- **url** (*str*) – Url of the authentication to be created. e.g. "ftp://test.com/"
- **params** (*dict, optional*) – Extra parameters.

Returns True if succeeded.

Return type bool

delete_result(*name*)
Delete the authentication having the specified name.

Parameters **name** (*str*) – Authentication name.

Returns True if succeeded.

Return type bool

list_result()
Get the list of all the available authentications.

Returns

The list of tuple of name, Result output url, and organization name (always *None* for api compatibility).

Return type [(str, str, None)]

tdclient.schedule_api

class tdclient.schedule_api.**ScheduleAPI**
Bases: object

Access to Schedule API

This class is inherited by *tdclient.api.API*.

create_schedule(*name*, *params=None*)
Create a new scheduled query with the specified name.

Parameters

- **name** (*str*) – Scheduled query name.
- **params** (*dict, optional*) – Extra parameters.
  - **type (str):** Query type. {"presto", "hive"}. Default: "hive"
  - **database (str):** Target database name.
  - **timezone (str):** Scheduled query's timezone. e.g. "UTC" For details, see also: https://gist.github.com/frsyuki/4533752
  - **cron (str, optional):** Schedule of the query. {"@daily", "@hourly", "10 * * * *" (custom cron)} See also: https://support.treasuredata.com/hc/en-us/articles/360001451088-Scheduled-Jobs-Web-Console
  - **delay (int, optional):** A delay ensures all buffered events are imported before running the query. Default: 0
  - **query (str):** Is a language used to retrieve, insert, update and modify data. See also: https://support.treasuredata.com/hc/en-us/articles/360012069493-SQL-Examples-of-Scheduled-Queries
  - **priority (int, optional):** Priority of the query. Range is from -2 (very low) to 2 (very high). Default: 0
  - **retry_limit (int, optional):** Automatic retry count. Default: 0
  - **engine_version (str, optional):** Engine version to be used. If none is specified, the account's default engine version would be set. {"stable", "experimental"}
  - **pool_name (str, optional):** For Presto only. Pool name to be used, if not specified, default pool would be used.
  - **result (str, optional):** Location where to store the result of the query. e.g. 'tableau://user:password@host.com:1234/datasource'

Returns Start date time.

Return type datetime.datetime

delete_schedule(*name*)
Delete the scheduled query with the specified name.

Parameters **name** (*str*) – Target scheduled query name.

Returns Tuple of cron and query.

Return type (str, str)

history(*name*, *_from=0*, *to=None*)
Get the history details of the saved query for the past 90days.

Parameters

- **name** (*str*) – Target name of the scheduled query.

- **_from** (*int, optional*) – Indicates from which nth record in the run history would be fetched. Default: 0. Note: Count starts from zero. This means that the first record in the list has a count of zero.
- **to** (*int, optional*) – Indicates up to which nth record in the run history would be fetched. Default: 20

Returns History of the scheduled query.

Return type dict

list_schedules()
Get the list of all the scheduled queries.

Returns str, cron:str, query:str, database:str, result_url:str)]

Return type [(name

run_schedule(*name*, *time*, *num=None*)
Execute the specified query.

Parameters

- **name** (*str*) – Target scheduled query name.
- **time** (*int*) – Time in Unix epoch format that would be set as TD_SCHEDULED_TIME
- **num** (*int, optional*) – Indicates how many times the query will be executed. Value should be 9 or less. Default: 1

Returns [(job_id:int, type:str, scheduled_at:str)]

Return type list of tuple

update_schedule(*name*, *params=None*)
Update the scheduled query.

Parameters

- **name** (*str*) – Target scheduled query name.
- **params** (*dict*) – Extra parameteres.
  - **type (str):** Query type. {"presto", "hive"}. Default: "hive"
  - **database (str):** Target database name.
  - **timezone (str):** Scheduled query's timezone. e.g. "UTC" For details, see also: https://gist.github.com/frsyuki/4533752
  - **cron (str, optional):** Schedule of the query. {`"@daily"`, `"@hourly"`, `"10 * * * *"` (custom cron)} See also: https://support.treasuredata.com/hc/en-us/articles/360001451088-Scheduled-Jobs-Web-Console
  - **delay (int, optional):** A delay ensures all buffered events are imported before running the query. Default: 0
  - **query (str):** Is a language used to retrieve, insert, update and modify data. See also: https://support.treasuredata.com/hc/en-us/articles/360012069493-SQL-Examples-of-Scheduled-Queries
  - **priority (int, optional):** Priority of the query. Range is from -2 (very low) to 2 (very high). Default: 0
  - **retry_limit (int, optional):** Automatic retry count. Default: 0

- **engine_version (str, optional):** Engine version to be used. If none is specified, the account's default engine version would be set. {"stable", "experimental"}
- **pool_name (str, optional):** For Presto only. Pool name to be used, if not specified, default pool would be used.
- **result (str, optional):** Location where to store the result of the query. e.g. 'tableau://user:password@host.com:1234/datasource'

tdclient.server_status_api

class tdclient.server_status_api.**ServerStatusAPI**
Bases: object

Access to Server Status API

This class is inherited by *tdclient.api.API*.

server_status()
Show the status of Treasure Data

Returns status

Return type str

tdclient.table_api

class tdclient.table_api.**TableAPI**
Bases: object

Access to Table API

This class is inherited by *tdclient.api.API*.

change_database(*db*, *table*, *dest_db*)
Move a target table from it's original database to new destination database.

Parameters

- **db** (*str*) – Target database name.
- **table** (*str*) – Target table name.
- **dest_db** (*str*) – Destination database name.

Returns *True* if succeeded

Return type bool

create_log_table(*db*, *table*)
Create a new table in the database and registers it in PlazmaDB.

Parameters

- **db** (*str*) – Target database name.
- **table** (*str*) – Target table name.

Returns *True* if succeeded.

Return type bool

delete_table(*db*, *table*)
Delete the specified table.

Parameters

- **db** (*str*) – Target database name.
- **table** (*str*) – Target table name.

Returns Type information of the table (e.g. "log").

Return type str

list_tables(*db*)

Gets the list of table in the database.

Parameters **db** (*str*) – Target database name.

Returns Detailed table information.

Return type dict

Examples

```
>>> td.api.list_tables("my_db")
{ 'iris': {'id': 21039862,
  'name': 'iris',
  'estimated_storage_size': 1236,
  'counter_updated_at': '2019-09-18T07:14:28Z',
  'last_log_timestamp': datetime.datetime(2019, 1, 30, 5, 34, 42,␣
↪tzinfo=tzutc()),
  'delete_protected': False,
  'created_at': datetime.datetime(2019, 1, 30, 5, 34, 42, tzinfo=tzutc()),
  'updated_at': datetime.datetime(2019, 1, 30, 5, 34, 46, tzinfo=tzutc()),
  'type': 'log',
  'include_v': True,
  'count': 150,
  'schema': [['sepal_length', 'double', 'sepal_length'],
   ['sepal_width', 'double', 'sepal_width'],
   ['petal_length', 'double', 'petal_length'],
   ['petal_width', 'double', 'petal_width'],
   ['species', 'string', 'species']],
  'expire_days': None,
  'last_import': datetime.datetime(2019, 9, 18, 7, 14, 28, tzinfo=tzutc())},
}
```

swap_table(*db*, *table1*, *table2*)

Swap the two specified tables with each other belonging to the same database and basically exchanges their names.

Parameters

- **db** (*str*) – Target database name
- **table1** (*str*) – First target table for the swap.
- **table2** (*str*) – Second target table for the swap.

Returns *True* if succeeded.

Return type bool

tail(*db*, *table*, *count*, *to=None*, *_from=None*, *block=None*)

Get the contents of the table in reverse order based on the registered time (last data first).

Parameters

- **db** (*str*) – Target database name.
- **table** (*str*) – Target table name.
- **count** (*int*) – Number for record to show up from the end.
- **to** – Deprecated parameter.
- **_from** – Deprecated parameter.
- **block** – Deprecated parameter.

Returns Contents of the table.

Return type [dict]

update_expire(*db*, *table*, *expire_days*)
Update the expire days for the specified table

Parameters

- **db** (*str*) – Target database name.
- **table** (*str*) – Target table name.
- **expire_days** (*int*) – Number of days where the contents of the specified table would expire.

Returns True if succeeded.

Return type bool

update_schema(*db*, *table*, *schema_json*)
Update the table schema.

Parameters

- **db** (*str*) – Target database name.
- **table** (*str*) – Target table name.
- **schema_json** (*str*) – Schema format JSON string. See also: ~‘Client.update_schema‘ e.g. ‘[[“sep_len”, “long”, “sep_len”], [“sep_wid”, “long”, “sep_wid”]]’

Returns *True* if succeeded.

Return type bool

tdclient.user_api

class tdclient.user_api.**UserAPI**
Bases: object

Access to User API.

This class is inherited by *tdclient.api.API*.

add_apikey(*name*)
Create a new apikey for the specified email address.

Parameters **name** (*str*) – User’s email address

Returns *True* if succeeded.

Return type bool

add_user(*name*, *org*, *email*, *password*)
Add a new user to the current account and sends invitation.

Parameters

- **name** (*str*) – User's name
- **org** (*str*) – Not used
- **email** (*str*) – User's email address
- **password** (*str*) – User's temporary password for logging-in

Returns *True* if succeeded.

Return type bool

authenticate(*user*, *password*)
Authenticate the indicated email address which is not authenticated via SSO.

Parameters

- **user** (*str*) – Email of the user to be authenticated.
- **password** (*str*) – Must contain at least 1 letter, 1 number, and 1 special character such as the following: `r'[!#$%-_=+<>0-9a-zA-Z]'`

Returns API key

Return type str

list_apikeys(*name*)
Get the apikeys of the current user.

Parameters **name** (*str*) – User's email address

Returns List of API keys

Return type [str]

list_users()
Get the list of users for the account.

Returns str,organization:str,[user:str]]

Return type [[name

remove_apikey(*name*, *apikey*)
Delete the apikey for the specified email address.

Parameters

- **name** (*str*) – User's email address
- **apikey** (*str*) – User's apikey to be deleted

Returns *True* if succeeded.

Return type bool

remove_user(*name*)
Remove the specified email in the account and revokes the user's access. :param name: User's email address :type name: str

Returns *True* if succeded

Return type bool

5.1.5 Misc

tdclient.errors

exception tdclient.errors.**APIError**
Bases: Exception

exception tdclient.errors.**AlreadyExistsError**
Bases: *tdclient.errors.APIError*

exception tdclient.errors.**AuthError**
Bases: *tdclient.errors.APIError*

exception tdclient.errors.**DataError**
Bases: *tdclient.errors.DatabaseError*

exception tdclient.errors.**DatabaseError**
Bases: *tdclient.errors.Error*

exception tdclient.errors.**Error**
Bases: Exception

exception tdclient.errors.**ForbiddenError**
Bases: *tdclient.errors.APIError*

exception tdclient.errors.**IntegrityError**
Bases: *tdclient.errors.DatabaseError*

exception tdclient.errors.**InterfaceError**
Bases: *tdclient.errors.Error*

exception tdclient.errors.**InternalError**
Bases: *tdclient.errors.DatabaseError*

exception tdclient.errors.**NotFoundError**
Bases: *tdclient.errors.APIError*

exception tdclient.errors.**NotSupportedError**
Bases: *tdclient.errors.DatabaseError*

exception tdclient.errors.**OperationalError**
Bases: *tdclient.errors.DatabaseError*

exception tdclient.errors.**ParameterValidationError**
Bases: Exception

exception tdclient.errors.**ProgrammingError**
Bases: *tdclient.errors.DatabaseError*

tdclient.pseudo_certifi

tdclient.pseudo_certifi.**where**()

5.2 Version History

5.2.1 Unreleased

5.2.2 v1.0.1 (2019-10-10)

- Fix `wait_interval` handling for `BulkImport.perform` appropriately (#74)
- Use `io.TextIOWrapper` to prevent `"x85"` issue creating None (#77)

5.2.3 v1.0.0 (2019-09-27)

- Drop Python 2 support (#60)
- Remove deprecated functions as follows (#76):
  - `TableAPI.create_item_table`
  - `UserAPI.change_email`, `UserAPI.change_password`, and `UserAPI.change_my_password`
  - `JobAPI.hive_query`, and `JobAPI.pig_query`
- Support `TableAPI.tail` and `TableAPI.change_database` (#64, #71)
- Introduce documentation site (#65, #66, #70, #72)

5.2.4 v0.14.0 (2019-07-11)

- Remove ACL and account APIs (#56, #58)
- Fix PyOpenSSL issue which causes pandas-td error (#59)

5.2.5 v0.13.0 (2019-03-29)

- Change msgpack-python to msgpack (#50)
- Dropped 3.3 support as it has already been EOL'd (#52)
- Set urllib3 minimum version as v1.24.1 (#51)

5.2.6 v0.12.0 (2018-05-31)

- Avoided to declare library dependencies too tightly within this project since this is a library project (#42)
- Got rid of all configurations for Python 2.6 completely (#42)

5.2.7 v0.11.1 (2018-05-21)

- Added 3.6 as test target. No functional changes have applied since 0.11.0 (#41)

5.2.8 v0.11.0 (2018-05-21)

- Support missing parameters in JOB API (#39, #40)

5.2.9 v0.10.0 (2017-11-01)

- Ignore empty string in job's `start_at` and `end_at` (#35, #36)

5.2.10 v0.9.0 (2017-02-27)

- Add validation to part names for bulk upload

5.2.11 v0.8.0 (2016-12-22)

- Fix unicode encoding issues on Python 2.x (#27, #28, #29)

5.2.12 v0.7.0 (2016-12-06)

- Fix for tdclient tables data not populating
- `TableAPI.list_tables` now returns a dictionary instead of a tuple

5.2.13 v0.6.0 (2016-09-27)

- Generate universal wheel by default since there's no binary in this package
- Add missing support for `created_time` and `user_name` from `/v3/schedule/list` API (#20, #21)
- Use keyword arguments for initializing model attributes (#22)

5.2.14 v0.5.0 (2016-06-10)

- Prevent retry after PUT request failures. This is the same behavior as https://github.com/treasure-data/td-client-ruby (#16)
- Support HTTP proxy authentication (#17)

5.2.15 v0.4.2 (2016-03-15)

- Catch exceptions on parsing date time string

5.2.16 v0.4.1 (2016-01-19)

- Fix Data Connector APIs based on latest td-client-ruby's implementation (#14)

5.2.17 v0.4.0 (2015-12-14)

- Avoid an exception raised when a `start` is not set for a schedule (#12)
- Fix getting database names of job objects (#13)
- Add Data Connector APIs
- Add deprecation warnings on the usage of "item tables"
- Show `cumul_retry_delay` in retry messages

5.2.18 v0.3.2 (2015-08-01)

- Fix bugs in `ScheduledJob` and `Schedule` models

5.2.19 v0.3.1 (2015-07-10)

- Fix `OverflowError` on importing integer value longer than 64 bit length which is not supported by msgpack specification. Those values will be converted into string.

5.2.20 v0.3.0 (2015-07-03)

- Add Python Database API (PEP 0249) compatible connection and cursor.
- Add varidation to the part name of a bulk import. It should not contain '/'.
- Changed default wait interval of job models from 1 second to 5 seconds.
- Fix many potential problems/warnings found by landscape.io.

5.2.21 v0.2.1 (2015-06-20)

- Set default timeout of API client as 60 seconds.
- Change the timeout of API client from `sum(connect_timeout, read_timeout, send_timeout)` to `max(connect_timeout, read_timeout, send_timeout)`
- Change default user-agent of client from `TD-Client-Python:{version}` to `TD-Client-Python/{version}` to comply RFC2616

5.2.22 v0.2.0 (2015-05-28)

- Improve the job model. Now it retrieves the job values automatically after the invocation of `wait`, `result` and `kill`.
- Add a property `result_schema` to `Job` model to provide the schema of job result
- Improve the bulk import model. Add a convenient method named `upload_file` to upload a part from file-like object.
- Support CSV/TSV format on both streaming import and bulk import
- Change module name; `tdclient.model` -> `tdclient.models`

5.2.23 v0.1.11 (2015-05-17)

- Fix API client to retry POST requests properly if `retry_post_requests` is set to `True` (#5)
- Show warnings if imported data don't have `time` column

5.2.24 v0.1.10 (2015-03-30)

- Fixed a JSON parse error in `job.result_format("json")` with multipe result rows (#4)
- Refactored model classes and tests

5.2.25 v0.1.9 (2015-02-26)

- Stopped using syntax added in recent Python releases

5.2.26 v0.1.8 (2015-02-26)

- Fix SSL verification errors on Python 2.7 on Windows environment. Now it uses `certifi` to verify SSL certificates if it is available.

5.2.27 v0.1.7 (2015-02-26)

- Fix support for Windows environments
- Fix byte encoding problem in `tdclient.api.API#import_file` on Python 3.x

5.2.28 v0.1.6 (2015-02-12)

- Support specifying job priority in its name (e.g. "NORMAL", "HIGH", etc.)
- Convert job priority number to its name (e.g. 0 => "NORMAL", 1 => "HIGH", etc.)
- Fix a broken behavior in `tdclient.model.Job#wait` when specifying timeout
- Fix broken `tdclient.client.Client#database()` which is used from `tdclient.model.Table#permission()`
- Fix broken `tdclient.Client.Client#results()`

5.2.29 v0.1.5 (2015-02-10)

- Fix local variable scope problem in `tdclient.api.show_job` (#2)
- Fix broken multiple assignment in `tdclient.model.Job#_update_status` (#3)

5.2.30 v0.1.4 (2015-02-06)

- Add new data import function of `tdclient.api.import_file` to allow importing data from file-like object or an existing file on filesystem.
- Fix an encoding error in `tdclient.api.import_data` on Python 2.x
- Add missing import to fix broken `tdclient.model.Job#wait`
- Use `td.api.DEFAULT_ENDPOINT` for all requests

5.2.31 v0.1.3 (2015-01-24)

- Support PEP 343 in `tdclient.Client` and remove `contextlib` from example
- Add deprecation warnings to `hive_query` and `pig_query` of `tdclient.api.API`
- Add `tdclient.model.Job#id` as an alias of `tdclient.model.Job#job_id`
- Parse datatime properly returned from `tdclient.Client#create_schedule`
- Changed `tdclient.model.Job#query` as a property since it won't be modified during the execution
- Allow specifying query options from `tdclient.model.Database#query`

5.2.32 v0.1.2 (2015-01-21)

- Fix broken PyPI identifiers
- Update documentation

5.2.33 v0.1.1 (2015-01-21)

- Improve the verification of SSL certificates on RedHat and variants
- Implement `wait` and `kill` in `tdclient.model.Job`
- Change the "Development Status" from Alpha to Beta

5.2.34 v0.1.0 (2015-01-15)

- Initial public release

CHAPTER

SIX

INDICES AND TABLES

- genindex
- modindex
- search

PYTHON MODULE INDEX

t

INDEX

D

E

F

G

H

I

J

T

U

V

W