
Tamr - Python Client Documentation

Release 0.11

Tamr

Apr 05, 2020

CONTENTS

1 Example	3
2 User Guide	5
2.1 FAQ	5
2.2 Installation	5
2.3 Quickstart	7
2.4 Secure Credentials	8
2.5 Workflows	9
2.6 Creating and Modifying Resources	11
2.7 Logging	12
2.8 Geospatial Data	13
2.9 Advanced Usage	15
3 Contributor Guide	19
3.1 Contributor guide	19
4 Reference	25
4.1 Reference	25
5 BETA	71
5.1 BETA	71
Python Module Index	79
Index	81

[View on Github](#)

EXAMPLE

```
from tamr_unify_client import Client
from tamr_unify_client.auth import UsernamePasswordAuth
import os

# grab credentials from environment variables
username = os.environ['TAMR_USERNAME']
password = os.environ['TAMR_PASSWORD']
auth = UsernamePasswordAuth(username, password)

host = 'localhost' # replace with your Tamr host
tamr = Client(auth, host=host)

# programmatically interact with Tamr!
# e.g. refresh your project's Unified Dataset
project = tamr.projects.by_resource_id('3')
ud = project.unified_dataset()
op = ud.refresh()
assert op.succeeded()
```


2.1 FAQ

2.1.1 What version of the Python Client should I use?

The Python Client just cares about features, and will try everything it knows to implement those features correctly, independent of the API version.

If you are starting a new project or your existing project does not yet use the Python Client, we encourage you to use the **latest stable version** of the Python Client.

Otherwise, check the [Change Log](#) to see:

- what new features and bug fixes are available in newer versions
- which breaking changes (if any) will require changes in your code to get those new features and bug fixes

Note: You do not need to reason about the Tamr API version nor the the Tamr app/server version.

2.1.2 How do I call custom endpoints, e.g. endpoints outside the Tamr API?

To call a custom endpoint *within* the Tamr API, use the `client.request()` method, and provide an endpoint described by a path relative to `base_path`.

For example, if `base_path` is `/api/versioned/v1/` (the default), and you want to get `/api/versioned/v1/projects/1`, you only need to provide `projects/1` (the relative ID provided by the project) as the endpoint, and the Client will resolve that into `/api/versioned/v1/projects/1`.

There are various APIs outside the `/api/versioned/v1/` prefix that are often useful or necessary to call - e.g. `/api/service/health`, or other un-versioned / unsupported APIs. To call a custom endpoint *outside* the Tamr API, use the `client.request()` method, and provide an endpoint described by an *absolute* path (a path starting with `/`). For example, to get `/api/service/health` (no matter what `base_path` is), call `client.request()` with `/api/service/health` as the endpoint. The Client will ignore `base_path` and send the request directly against the absolute path provided.

For additional detail, see [Raw HTTP requests and Unversioned API Access]([user-guide/advanced-usage:Raw HTTP requests and Unversioned API Access](#))

2.2 Installation

`tamr-unify-client` is compatible with Python 3.6 or newer.

2.2.1 Stable releases

Installation is as simple as:

```
pip install tamr-unify-client
```

Or:

```
poetry add tamr-unify-client
```

Note: If you don't use `poetry`, we recommend you use a virtual environment for your project and install the Python Client into that virtual environment.

You can create a virtual environment with Python 3 via:

```
python3 -m venv my-venv
```

For more, see [The Hitchhiker's Guide to Python](#).

2.2.2 Latest (unstable)

Note: This project uses the new `pyproject.toml` file, not a `setup.py` file, so make sure you have the latest version of `pip` installed: `pip install -U pip`.

To install the bleeding edge:

```
git clone https://github.com/Datatamer/tamr-client
cd tamr-client
pip install .
```

2.2.3 Offline installs

First, download `tamr-unify-client` and its dependencies on a machine with online access to PyPI:

```
pip download tamr-unify-client -d tamr-unify-client-requirements
zip -r tamr-unify-client-requirements.zip tamr-unify-client-requirements
```

Then, ship the `.zip` file to the target machine where you want `tamr-unify-client` installed. You can do this via email, cloud drives, `scp` or any other mechanism.

Finally, install `tamr-unify-client` from the saved dependencies:

```
unzip tamr-unify-client-requirements.zip
pip install --no-index --find-links=tamr-unify-client-requirements tamr-unify-client
```

If you are not using a virtual environment, you may need to specify the `--user` flag if you get permissions errors:

```
pip install --user --no-index --find-links=tamr-unify-client-requirements tamr-unify-
↪client
```

2.3 Quickstart

2.3.1 Client configuration

Start by importing the Python Client and authentication provider:

```
from tamr_unify_client import Client
from tamr_unify_client.auth import UsernamePasswordAuth
```

Next, create an authentication provider and use that to create an authenticated client:

```
import os

username = os.environ['TAMR_USERNAME']
password = os.environ['TAMR_PASSWORD']

auth = UsernamePasswordAuth(username, password)
tamr = Client(auth)
```

Warning: For security, it's best to read your credentials in from environment variables or secure files instead of hardcoding them directly into your code.

For more, see [User Guide > Secure Credentials](#).

By default, the client tries to find the Tamr instance on `localhost`. To point to a different host, set the `host` argument when instantiating the `Client`.

For example, to connect to `10.20.0.1`:

```
tamr = Client(auth, host='10.20.0.1')
```

2.3.2 Top-level collections

The Python Client exposes 2 top-level collections: `Projects` and `Datasets`.

You can access these collections through the client and loop over their members with simple `for`-loops.

E.g.:

```
for project in tamr.projects:
    print(project.name)

for dataset in tamr.datasets:
    print(dataset.name)
```

2.3.3 Fetch a specific resource

If you know the identifier for a specific resource, you can ask for it directly via the `by_resource_id` methods exposed by collections.

E.g. To fetch the project with ID `'1'`:

```
project = tamr.projects.by_resource_id('1')
```

Similarly, if you know the name of a specific resource, you can ask for it directly via the `by_name` methods exposed by collections.

E.g. To fetch the project with name 'Number 1':

```
project = tamr.projects.by_name('Number 1')
```

Note: If working with projects across Tamr instances for migrations or promotions, use external IDs (via `by_external_id`) instead of name (via `by_name`).

2.3.4 Resource relationships

Related resources (like a project and its unified dataset) can be accessed through specific methods.

E.g. To access the Unified Dataset for a particular project:

```
ud = project.unified_dataset()
```

2.3.5 Kick-off Tamr Operations

Some methods on Model objects can kick-off long-running Tamr operations.

Here, kick-off a “Unified Dataset refresh” operation:

```
operation = project.unified_dataset().refresh()
assert op.succeeded()
```

By default, the API Clients expose a synchronous interface for Tamr operations.

2.4 Secure Credentials

This section discusses ways to pass credentials securely to `UsernamePasswordAuth`. Specifically, you **should not** hardcode your password(s) in your source code. Instead, you should use environment variables or secure files to store your credentials and simple Python code to read your credentials.

2.4.1 Environment variables

You can use `os.environ` to read in your credentials from environment variables:

```
# my_script.py
import os

from tamr_unify_client.auth import UsernamePasswordAuth

username = os.environ['TAMR_USERNAME'] # replace with your username environment_
↳variable name
password = os.environ['TAMR_PASSWORD'] # replace with your password environment_
↳variable name
```

(continues on next page)

(continued from previous page)

```
auth = UsernamePasswordAuth(username, password)
```

You can pass in the environment variables from the terminal by including them before your command:

```
TAMR_USERNAME="my Tamr username" TAMR_PASSWORD="my Tamr password" python my_script.py
```

You can also create an `.sh` file to store your environment variables and simply source that file before running your script.

2.4.2 Config files

You can also store your credentials in a secure credentials file:

```
# credentials.yaml
---
username: "my tamr username"
password: "my tamr password"
```

Then `pip install pyyaml` read the credentials in your Python code:

```
# my_script.py
from tamr_unify_client.auth import UsernamePasswordAuth
import yaml

with open("path/to/credentials.yaml") as f: # replace with your credentials.yaml path
    creds = yaml.safe_load(f)

auth = UsernamePasswordAuth(creds['username'], creds['password'])
```

As in this example, we recommend you use YAML as your format since YAML has support for comments and is more human-readable than JSON.

Important: You **should not** check these credentials files into your version control system (e.g. `git`). Do not share this file with anyone who should not have access to the password stored in it.

2.5 Workflows

2.5.1 Continuous Categorization

```
from tamr_unify_client import Client
from tamr_unify_client.auth import UsernamePasswordAuth
import os

username = os.environ['TAMR_USERNAME']
password = os.environ['TAMR_PASSWORD']
auth = UsernamePasswordAuth(username, password)

host = 'localhost' # replace with your host
tamr = Client(auth)
```

(continues on next page)

(continued from previous page)

```
project_id = "1" # replace with your project ID
project = tamr.projects.by_resource_id(project_id)
project = project.as_categorization()

unified_dataset = project.unified_dataset()
op = unified_dataset.refresh()
assert op.succeeded()

model = project.model()
op = model.train()
assert op.succeeded()

op = model.predict()
assert op.succeeded()
```

2.5.2 Continuous Mastering

```
from tamr_unify_client import Client
from tamr_unify_client.auth import UsernamePasswordAuth
import os

username = os.environ['TAMR_USERNAME']
password = os.environ['TAMR_PASSWORD']
auth = UsernamePasswordAuth(username, password)

host = 'localhost' # replace with your host
tamr = Client(auth)

project_id = "1" # replace with your project ID
project = tamr.projects.by_resource_id(project_id)
project = project.as_mastering()

unified_dataset = project.unified_dataset()
op = unified_dataset.refresh()
assert op.succeeded()

op = project.pairs().refresh()
assert op.succeeded()

model = project.pair_matching_model()
op = model.train()
assert op.succeeded()

op = model.predict()
assert op.succeeded()

op = project.record_clusters().refresh()
assert op.succeeded()

op = project.published_clusters().refresh()
assert op.succeeded()
```

2.6 Creating and Modifying Resources

2.6.1 Creating resources

Resources, such as projects, dataset, and attribute configurations, can be created through their respective collections. Each `create` function takes in a dictionary that conforms to the [Tamr Public Docs](#) for creating that resource type:

```
spec = {
    "name": "project",
    "description": "Mastering Project",
    "type": "DEDUP",
    "unifiedDatasetName": "project_unified_dataset"
}
project = tamr.projects.create(spec)
```

2.6.2 Using specs

These dictionaries can also be created using spec classes.

Each `Resource` has a corresponding `ResourceSpec` which can be used to build an instance of that resource by specifying the value for each property.

The spec can then be converted to a dictionary that can be passed to `create`.

For instance, to create a project:

```
spec = (
    ProjectSpec.new()
    .with_name("Project")
    .with_type("DEDUP")
    .with_description("Mastering Project")
    .with_unified_dataset_name("Project_unified_dataset")
    .with_external_id("tamrProject1")
)
project = tamr.projects.create(spec.to_dict())
```

Calling `with_*` on a spec creates a new spec with the same properties besides the modified one. The original spec is unaltered, so it could be used multiple times:

```
base_spec = (
    ProjectSpec.new()
    .with_type("DEDUP")
    .with_description("Mastering Project")
)

specs = []
for name in project_names:
    spec = (
        base_spec.with_name(name)
        .with_unified_dataset_name(name + "_unified_dataset")
    )
    specs.append(spec)

projects = [tamr.projects.create(spec.to_dict()) for spec in specs]
```

2.6.3 Creating a dataset

Datasets can be created as described above, but the dataset's schema and records must then be handled separately.

To combine all of these steps into one, `DatasetCollection` has a convenience function `create_from_dataframe` that takes a `Pandas DataFrame`. This makes it easy to create a Tamr dataset from a CSV:

```
import pandas as pd

df = pd.read_csv("my_data.csv", dtype=str)      # string is the recommended data type
dataset = tamr.datasets.create_from_dataframe(df, primary_key_name="primary key name",
↪ dataset_name="My Data")
```

This will create a dataset called “My Data” with the specified primary key, an attribute for each column of the `DataFrame`, and the `DataFrame`'s rows as records.

2.6.4 Modifying a resource

Certain resources can also be modified using specs.

After getting a spec corresponding to a resource and modifying some properties, the updated resource can be committed to Tamr with the `put` function:

```
updated_dataset = (
    dataset.spec()
    .with_description("Modified description")
    .put()
)
```

Each spec class has many properties that can be changed, but refer to the [Public Docs](#) for which properties will actually be updated in Tamr. If an immutable property is changed in the update request, the new value will simply be ignored.

2.7 Logging

IMPORTANT Make sure to configure logging BEFORE importing from 3rd party libraries. Logging will use the first configuration it finds, and if a library configures logging before you, your configuration will be ignored.

To configure logging, simply follow the [official Python logging HOWTO](#).

For example:

```
# script.py
import logging

logging.basicConfig(filename="script.log", level=logging.INFO)

# configure logging before other imports

from tamr_unify_client import Client
from tamr_unify_client.auth import UsernamePasswordAuth

auth = UsernamePasswordAuth("my username", "my password")
```

(continues on next page)

(continued from previous page)

```
tamr = Client(auth, host="myhost")

for p in tamr.projects:
    print(p)

for d in tamr.datasets:
    print(d)

# should cause an HTTP error
tamr.get("/invalid/api/path").successful()
```

This will log all API requests made and print the response bodies for any requests with HTTP error codes.

If you want to **only** configure logging for the Tamr Client:

```
import logging
logger = logging.getLogger('tamr_unify_client')
logger.setLevel(logging.INFO)
logger.addHandler(logging.FileHandler('tamr-client.log'))

# configure logging before other imports

from tamr_unify_client import Client
from tamr_unify_client import UsernamePasswordAuth

# rest of script goes here
```

2.8 Geospatial Data

2.8.1 What geospatial data is supported?

In general, the Python Geo Interface is supported; see <https://gist.github.com/sgillies/2217756>.

There are three layers of information, modeled after GeoJSON (see <https://tools.ietf.org/html/rfc7946>):

- The outermost layer is a FeatureCollection
- Within a FeatureCollection are Features, each of which represents one “thing”, like a building or a river. Each feature has:
 - type (string; required)
 - id (object; required)
 - geometry (Geometry, see below; optional)
 - bbox (“bounding box”, 4 doubles; optional)
 - properties (map[string, object]; optional)
- Within a Feature is a Geometry, which represents a shape, like a point or a polygon. Each geometry has:
 - type (one of “Point”, “MultiPoint”, “LineString”, “MultiLineString”, “Polygon”, “MultiPolygon”; required)
 - coordinates (doubles; exactly how these are structured depends on the type of the geometry)

Although the Python Geo Interface is non-prescriptive when it comes to the data types of the id and properties, Tamr has a more restricted set of supported types. See <https://docs.tamr.com/reference#attribute-types>.

The `Dataset` class supports the `__geo_interface__` property. This will produce one `FeatureCollection` for the entire dataset.

There is a companion iterator `itergeofeatures()` that returns a generator that allows you to stream the records in the dataset as Geospatial features.

To produce a GeoJSON representation of a dataset:

```
dataset = client.datasets.by_name("my_dataset")
with open("my_dataset.json", "w") as f:
    json.dump(dataset.__geo_interface__, f)
```

By default, `itergeofeatures()` will use the first dataset attribute with geometry type to fill in the feature geometry. You can override this by specifying the geometry attribute to use in the `geo_attr` parameter to `itergeofeatures`.

`Dataset` can also be updated from a feature collection that supports the Python Geo Interface:

```
import geopandas
geodataframe = geopandas.GeoDataFrame(...)
dataset = client.dataset.by_name("my_dataset")
dataset.from_geo_features(geodataframe)
```

By default the features' geometries will be placed into the first dataset attribute with geometry type. You can override this by specifying the geometry attribute to use in the `geo_attr` parameter to `from_geo_features`.

2.8.2 Rules for converting from Tamr records to Geospatial Features

The record's primary key will be used as the feature's `id`. If the primary key is a single attribute, then the value of that attribute will be the value of `id`. If the primary key is composed of multiple attributes, then the value of the `id` will be an array with the values of the key attributes in order.

Tamr allows any number of geometry attributes per record; the Python Geo Interface is limited to one. When converting Tamr records to Python Geo Features, the first geometry attribute in the schema will be used as the geometry; all other geometry attributes will appear as properties with no type conversion. In the future, additional control over the handling of multiple geometries may be provided; the current set of capabilities is intended primarily to support the use case of working with `FeatureCollections` within Tamr, and `FeatureCollection` has only one geometry per feature.

An attribute is considered to have geometry type if it has type `RECORD` and contains an attribute named `point`, `multiPoint`, `lineString`, `multiLineString`, `polygon`, or `multiPolygon`.

If an attribute named `bbox` is available, it will be used as `bbox`. No conversion is done on the value of `bbox`. In the future, additional control over the handling of `bbox` attributes may be provided.

All other attributes will be placed in `properties`, with no type conversion. This includes all geometry attributes other than the first.

2.8.3 Rules for converting from Geospatial Features to Tamr records

The Feature's `id` will be converted into the primary key for the record. If the record uses a simple key, no value translation will be done. If the record uses a composite key, then the value of the Feature's `id` must be an array of values, one per attribute in the key.

If the Feature contains keys in `properties` that conflict with the record keys, `bbox`, or geometry, those keys are ignored (omitted).

If the Feature contains a `bbox`, it is copied to the record's `bbox`.

All other keys in the Feature's `properties` are propagated to the same-name attribute on the record, with no type conversion.

2.8.4 Streaming data access

The Dataset method `itergeofeatures()` returns a generator that allows you to stream the records in the dataset as Geospatial features:

```
my_dataset = client.datasets.by_name("my_dataset")
for feature in my_dataset.itergeofeatures():
    do_something(feature)
```

Note that many packages that consume the Python Geo Interface will be able to consume this iterator directly. For example::

```
from geopandas import GeoDataFrame
df = GeoDataFrame.from_features(my_dataset.itergeofeatures())
```

This allows construction of a `GeoDataFrame` directly from the stream of records, without materializing the intermediate dataset.

2.9 Advanced Usage

2.9.1 Asynchronous Operations

You can opt-in to an asynchronous interface via the `asynchronous` keyword argument for methods that kick-off Tamr operations.

E.g.:

```
op = project.unified_dataset().refresh(asynchronous=True)
# do asynchronous stuff here while operation is running
op = op.wait() # hangs until operation finishes
assert op.succeeded()
```

2.9.2 Raw HTTP requests and Unversioned API Access

We encourage you to use the high-level, object-oriented interface offered by the Python Client. If you aren't sure whether you need to send low-level HTTP requests, you probably don't.

But sometimes it's useful to directly send HTTP requests to Tamr; for example, Tamr has many APIs that are not covered by the higher-level interface (most of which are neither versioned nor supported). You can still call these endpoints using the Python Client, but you'll need to work with raw `Response` objects.

Custom endpoint

The client exposes a `request` method with the same interface as `requests.request`:

```
# import Python Client library and configure your client

tamr = Client(auth)
# do stuff with the `tamr` client

# now I NEED to send a request to a specific endpoint
response = tamr.request('GET', 'relative/path/to/resource')
```

This will send a request relative to the `base_path` registered with the client. If you provide an absolute path to the resource, the `base_path` will be ignored when composing the request:

```
# import Python Client library and configure your client

tamr = Client(auth)

# request a resource outside the configured base_path
response = tamr.request('GET', '/absolute/path/to/resource')
```

You can also use the `get`, `post`, `put`, `delete` convenience methods:

```
# e.g. `get` convenience method
response = tamr.get('relative/path/to/resource')
```

Custom Host / Port / Base API path

If you need to repeatedly send requests to another port or base API path (i.e. not `/api/versioned/v1/`), you can simply instantiate a different client.

Then just call `request` as described above:

```
# import Python Client library and configure your client

tamr = api.Client(auth)
# do stuff with the `tamr` client

# now I NEED to send requests to a different host/port/base API path etc..
# NOTE: in this example, we reuse `auth` from the first client, but we could
# have made a new Authentication provider if this client needs it.
custom_client = api.Client(
    auth,
    host="10.10.0.1",
    port=9090,
    base_path="/api/some_service/",
)
response = custom_client.get('relative/path/to/resource')
```

One-off authenticated request

All of the Python Client Authentication providers adhere to the `requests.auth.BaseAuth` interface.

This means that you can pass in an Authentication provider directly to the `requests` library:

```
from tamr_unify_client.auth import UsernamePasswordAuth
import os
import requests
```

(continues on next page)

(continued from previous page)

```
username = os.environ['TAMR_USERNAME']
password = os.environ['TAMR_PASSWORD']
auth = UsernamePasswordAuth(username, password)

response = requests.request('GET', 'some/specific/endpoint', auth=auth)
```


CONTRIBUTOR GUIDE

3.1 Contributor guide

3.1.1 Submitting Bug Reports and Feature Requests

Submit bug reports and feature requests as [Github issues](#) .

Check through existing issues (open and closed) to confirm that the bug hasn't been reported before.

If the bug/feature has been submitted already, leave a like on the Github Issue.

3.1.2 Code Migrations

Some of the codebase is old and outdated.

To know which patterns to follow and which to avoid, you can check out [ongoing code migrations](#)

3.1.3 Configure your Text Editor

- [Atom](#)

3.1.4 Install

This project uses `pyenv` and `poetry`. If you do not have these installed, checkout the [toolchain guide](#).

1. Clone your fork and `cd` into the project:

```
git clone https://github.com/<your-github-username>/tamr-client
cd tamr-client
```

2. Set a Python version for this project. Must be Python 3.6+ (e.g. 3.7.3):

```
pyenv local 3.7.3
```

3. Check that your Python version matches the version specified in `.python-version`:

```
cat .python-version
python --version
```

4. Install dependencies via poetry:

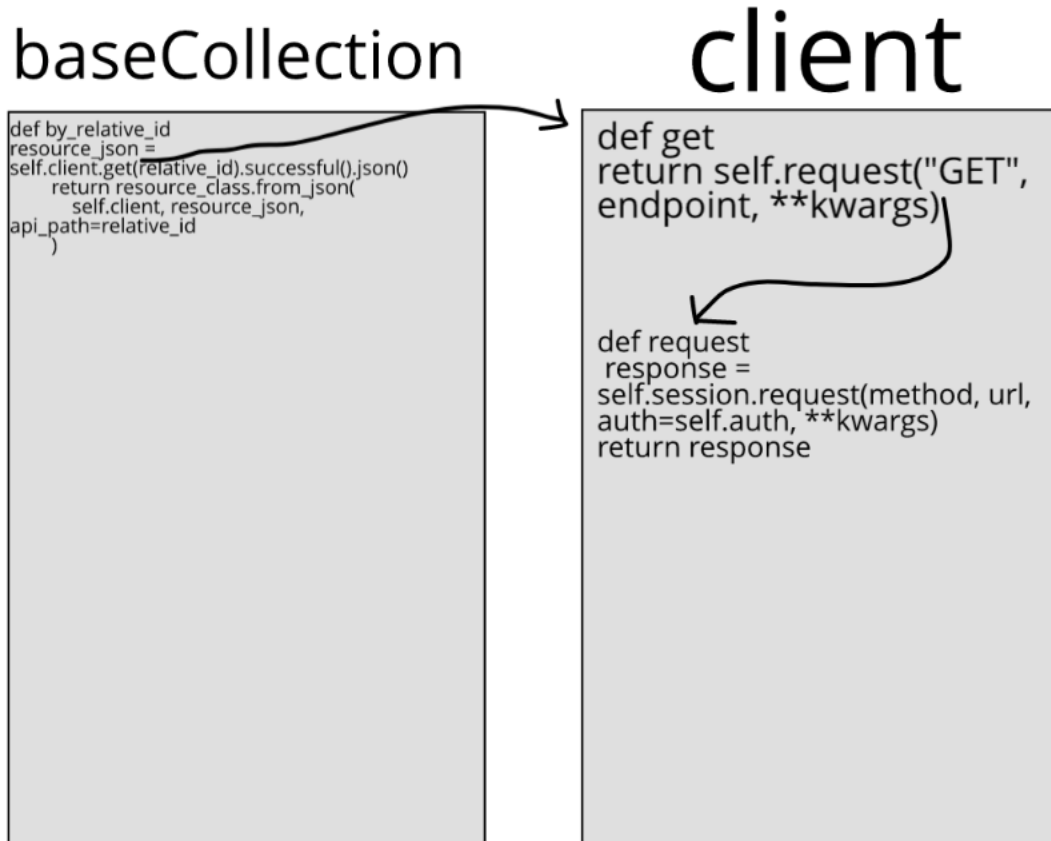
```
poetry install
```

3.1.5 Navigating Inheritance

Older parts of the codebase heavily use inheritance. We are in the process of [migrating to dataclasses](#) to simplify the codebase, but in the meantime you might want to know how the inheritance machinery we have works.

`yourResource` and `yourCollection` are files that inherit from `baseResource` and `baseCollection`. Examples of such files would be `resource.py` and `collection.py` in the `attribute_configuration` folder under project.





Step 1 (red): `yourCollection`'s `by_relative_id` returns `super.by_relative_id`, which comes from `baseCollection`

Step 1a (black): within `by_relative_id`, variable `resource_json` is defined as `self.client.get`. [etc]. `Client`'s `.get` returns `self.request`

Step 1b (black): `client`'s `.request` makes a request to the provided URL (this is the method actually fetching the data)

Step 2 (orange): `baseCollection`'s `by_relative_id` returns `resource_class.from_json`, which is the `from_json` defined in `yourResource`

Step 3 (yellow): `yourResource`'s `from_json` returns `super.from_data`, which comes from `baseResource`

Step 4 (green): `baseResource`'s `from_data` returns `cls`, one of the parameters entered for `from_data`. `cls` is a `yourResource`, because in `from_json` the return type is specified to be a `yourResource`. When `cls` is returned, a `yourResource` that has been filled with the data retrieved in `client`'s `.request` is what comes back.

3.1.6 Pull Requests

For larger, new features:

[Open an RFC issue](#). Discuss the feature with project maintainers to be sure that your change fits with the project vision and that you won't be wasting effort going in the wrong direction.

Once you get the green light from maintainers, you can proceed with the PR.

Contributions / PRs should follow the [Forking Workflow](#):

1. Fork it: [https://github.com/\[your-github-username\]/tamr-client/fork](https://github.com/[your-github-username]/tamr-client/fork)
2. Create your feature branch:

```
git checkout -b my-new-feature
```

3. Commit your changes:

```
git commit -am 'Add some feature'
```

4. Push to the branch:

```
git push origin my-new-feature
```

5. Create a new Pull Request
-

We optimize for PR readability, so please squash commits before and during the PR review process if you think it will help reviewers and onlookers navigate your changes.

Don't be afraid to push `-f` on your PRs when it helps our eyes read your code.

Remember to check for any [ongoing code migrations](#) that may be relevant to your PR.

3.1.7 Run and Build

This project uses `invoke` as its task runner.

Since `invoke` will be running inside of a `poetry` environment, we recommend adding the following alias to your `.bashrc` / `.zshrc` to save you some keystrokes:

```
alias pri='poetry run invoke'
```

Linting & Formatting

To run linter:

```
pri lint # with alias  
poetry run invoke lint # without alias
```

To run formatter:

```
pri format # with alias  
poetry run invoke format # without alias
```

Run the formatter with the `--fix` flag to autofix formatting:

```
pri format --fix # with alias  
poetry run invoke format --fix # without alias
```

Typechecks

To run typechecks:

```
pri typecheck # with alias
poetry run invoke typecheck # without alias
```

Tests

To run all tests:

```
pri test # with alias
poetry run invoke test # without alias
```

To run specific tests, see [these pytest docs](#) and run `pytest` explicitly:

```
poetry run pytest tests/unit/test_attribute.py
```

Docs

To build the docs:

```
pri docs # with alias
poetry run invoke docs # without alias
```

After docs are build, view them by:

```
open -a 'firefox' docs/_build/index.html # open in Firefox
open -a 'Google Chrome' docs/_build/index.html # open in Chrome
```

3.1.8 Toolchain

This project uses `poetry` as its package manager. For details on `poetry`, see the [official documentation](#).

1. Install `pyenv`:

```
curl https://pyenv.run | bash
```

2. Use `pyenv` to install a compatible Python version (3.6 or newer; e.g. 3.7.3):

```
pyenv install 3.7.3
```

3. Set that Python version to be your version for this project(e.g. 3.7.3):

```
pyenv shell 3.7.3
python --version # check that version matches your specified version
```

4. Install `poetry` as described [here](#):

```
curl -sSL https://raw.githubusercontent.com/sdispater/poetry/master/get-poetry.py ↵
↵ | python
```


REFERENCE

4.1 Reference

4.1.1 Attributes

Attribute

class `tamr_unify_client.attribute.resource.Attribute` (*client, data, alias=None*)
A Tamr Attribute.

See <https://docs.tamr.com/reference#attribute-types>

property `relative_id`

Type `str`

property `name`

Type `str`

property `description`

Type `str`

property `type`

Type `AttributeType`

property `is_nullable`

Type `bool`

spec ()

Returns a spec representation of this attribute.

Returns The attribute spec.

Return type `AttributeSpec`

delete ()

Deletes this resource. Some resources do not support deletion, and will raise a 405 error if this is called.

Returns HTTP response from the server

Return type `requests.Response`

property `resource_id`

Type `str`

Attribute Spec

class `tamr_unify_client.attribute.resource.AttributeSpec` (*client, data, api_path*)

A representation of the server view of an attribute

static of (*resource*)

Creates an attribute spec from an attribute.

Parameters `resource` (*Attribute*) – The existing attribute.

Returns The corresponding attribute spec.

Return type *AttributeSpec*

static new ()

Creates a blank spec that could be used to construct a new attribute.

Returns The empty spec.

Return type *AttributeSpec*

from_data (*data*)

Creates a spec with the same client and API path as this one, but new data.

Parameters `data` (*dict*) – The data for the new spec.

Returns The new spec.

Return type *AttributeSpec*

to_dict ()

Returns a version of this spec that conforms to the API representation.

Returns The spec's dict.

Return type *dict*

with_name (*new_name*)

Creates a new spec with the same properties, updating name.

Parameters `new_name` (*str*) – The new name.

Returns The new spec.

Return type *AttributeSpec*

with_description (*new_description*)

Creates a new spec with the same properties, updating description.

Parameters `new_description` (*str*) – The new description.

Returns The new spec.

Return type *AttributeSpec*

with_type (*new_type*)

Creates a new spec with the same properties, updating type.

Parameters `new_type` (*AttributeTypeSpec*) – The spec of the new type.

Returns The new spec.

Return type *AttributeSpec*

with_is_nullable (*new_is_nullable*)

Creates a new spec with the same properties, updating is nullable.

Parameters `new_is_nullable` (*bool*) – The new is nullable.

Returns The new spec.

Return type *AttributeSpec*

put ()

Commits the changes and updates the attribute in Tamr.

Returns The updated attribute.

Return type *Attribute*

Attribute Collection

class `tamr_unify_client.attribute.collection.AttributeCollection` (*client*, *api_path*)

Collection of *Attribute*s.

Parameters

- **client** (*Client*) – Client for API call delegation.
- **api_path** (*str*) – API path used to access this collection. E.g. "datasets/1/attributes".

by_resource_id (*resource_id*)

Retrieve an attribute by resource ID.

Parameters **resource_id** (*str*) – The resource ID. E.g. "AttributeName"

Returns The specified attribute.

Return type *Attribute*

by_relative_id (*relative_id*)

Retrieve an attribute by relative ID.

Parameters **relative_id** (*str*) – The resource ID. E.g. "datasets/1/attributes/AttributeName"

Returns The specified attribute.

Return type *Attribute*

by_external_id (*external_id*)

Retrieve an attribute by external ID.

Since attributes do not have external IDs, this method is not supported and will raise a `NotImplementedError`.

Parameters **external_id** (*str*) – The external ID.

Returns The specified attribute, if found.

Return type *Attribute*

Raises

- **KeyError** – If no attribute with the specified `external_id` is found
- **LookupError** – If multiple attributes with the specified `external_id` are found

stream ()

Stream attributes in this collection. Implicitly called when iterating over this collection.

Returns Stream of attributes.

Return type Python generator yielding *Attribute*

Usage:

```

>>> for attribute in collection.stream(): # explicit
>>>     do_stuff(attribute)
>>> for attribute in collection: # implicit
>>>     do_stuff(attribute)
```

by_name (*attribute_name*)

Lookup a specific attribute in this collection by exact-match on name.

Parameters **attribute_name** (*str*) – Name of the desired attribute.

Returns Attribute with matching name in this collection.

Return type *Attribute*

create (*creation_spec*)

Create an Attribute in this collection

Parameters **creation_spec** (*dict[str, str]*) – Attribute creation specification should be formatted as specified in the [Public Docs for adding an Attribute](#).

Returns The created Attribute

Return type *Attribute*

delete_by_resource_id (*resource_id*)

Deletes a resource from this collection by resource ID.

Parameters **resource_id** (*str*) – The resource ID of the resource that will be deleted.

Returns HTTP response from the server.

Return type *requests.Response*

Attribute Type

class `tamr_unify_client.attribute.type.AttributeType` (*data*)

The type of an *Attribute* or *SubAttribute*.

See <https://docs.tamr.com/reference#attribute-types>

Parameters **data** (*dict*) – JSON data representing this type

property **base_type**

Type *str*

property **inner_type**

Type *AttributeType*

property **attributes**

Type *list[SubAttribute]*

spec ()

Returns a spec representation of this attribute type.

Returns The attribute type spec.

Return type *AttributeTypeSpec*

Attribute Type Spec

class tamr_unify_client.attribute.type.**AttributeTypeSpec** (*data*)

static of (*resource*)

Creates an attribute type spec from an attribute type.

Parameters **resource** (*AttributeType*) – The existing attribute type.

Returns The corresponding attribute type spec.

Return type *AttributeTypeSpec*

static new ()

Creates a blank spec that could be used to construct a new attribute type.

Returns The empty spec.

Return type *AttributeTypeSpec*

to_dict ()

Returns a version of this spec that conforms to the API representation.

Returns The spec's dict.

Return type *dict*

with_base_type (*new_base_type*)

Creates a new spec with the same properties, updating the base type.

Parameters **new_base_type** (*str*) – The new base type.

Returns The new spec.

Return type *AttributeTypeSpec*

with_inner_type (*new_inner_type*)

Creates a new spec with the same properties, updating the inner type.

Parameters **new_inner_type** (*AttributeTypeSpec*) – The spec of the new inner type.

Returns The new spec.

Return type *AttributeTypeSpec*

with_attributes (*new_attributes*)

Creates a new spec with the same properties, updating attributes.

Parameters **new_attributes** (list[*AttributeSpec*]) – The specs of the new attributes.

Returns The new spec.

Return type *AttributeTypeSpec*

SubAttribute

class tamr_unify_client.attribute.subattribute.**SubAttribute** (*name*, *type*,
is_nullable, *_json*,
description=None)

An attribute which is itself a property of another attribute.

See <https://docs.tamr.com/reference#attribute-types>

Parameters

- **name** (*str*) – Name of sub-attribute
- **description** (*Optional[str]*) – Description of sub-attribute
- **type** (*AttributeType*) – See <https://docs.tamr.com/reference#attribute-types>
- **is_nullable** (*bool*) – If this sub-attribute can be null

static from_json (*data*)

Create a SubAttribute from JSON data.

Parameters **data** (*Dict[str, Any]*) – JSON data received from Tamr server.

Return type *SubAttribute*

4.1.2 Auth

class `tamr_unify_client.auth.UsernamePasswordAuth` (*username, password*)

Provides username/password authentication for Tamr. Specifically, sets the *Authorization* HTTP header with Tamr's custom *BasicCreds* format.

Parameters

- **username** (*str*) –
- **password** (*str*) –

Usage:

```
>>> from tamr_unify_client.auth import UsernamePasswordAuth
>>> auth = UsernamePasswordAuth('my username', 'my password')
>>> import tamr_unify_client as api
>>> unify = api.Client(auth)
```

4.1.3 Categorization

Categorization Project

class `tamr_unify_client.categorization.project.CategorizationProject` (*client, data, alias=None*)

A Categorization project in Tamr.

model ()

Machine learning model for this Categorization project. Learns from verified labels and predicts categorization labels for unlabeled records.

Returns The machine learning model for categorization.

Return type *MachineLearningModel*

create_taxonomy (*creation_spec*)

Creates a *Taxonomy* for this project.

A taxonomy cannot already be associated with this project.

Parameters **creation_spec** (*dict*) – The creation specification for the taxonomy, which can include name.

Returns The new Taxonomy

Return type *Taxonomy*

taxonomy ()

Retrieves the *Taxonomy* associated with this project. If a taxonomy is not already associated with this project, call *create_taxonomy()* first.

Returns The project's Taxonomy

Return type *Taxonomy*

add_input_dataset (*dataset*)

Associate a dataset with a project in Tamr.

By default, datasets are not associated with any projects. They need to be added as input to a project before they can be used as part of that project

Parameters **dataset** (*Dataset*) – The dataset to associate with the project.

Returns HTTP response from the server

Return type *requests.Response*

as_categorization ()

Convert this project to a *CategorizationProject*

Returns This project.

Return type *CategorizationProject*

Raises **TypeError** – If the *type* of this project is not "CATEGORIZATION"

as_mastering ()

Convert this project to a *MasteringProject*

Returns This project.

Return type *MasteringProject*

Raises **TypeError** – If the *type* of this project is not "DEDUP"

attribute_configurations ()

Project's attribute's configurations.

Returns The configurations of the attributes of a project.

Return type *AttributeConfigurationCollection*

attribute_mappings ()

Project's attribute's mappings.

Returns The attribute mappings of a project.

Return type *AttributeMappingCollection*

property attributes

Attributes of this project.

Returns Attributes of this project.

Return type *AttributeCollection*

delete ()

Deletes this resource. Some resources do not support deletion, and will raise a 405 error if this is called.

Returns HTTP response from the server

Return type *requests.Response*

property description

Type `str`

property external_id

Type `str`

input_datasets ()

Retrieve a collection of this project's input datasets.

Returns The project's input datasets.

Return type `DatasetCollection`

property name

Type `str`

property relative_id

Type `str`

remove_input_dataset (dataset)

Remove a dataset from a project.

Parameters `dataset` (`Dataset`) – The dataset to be removed from this project.

Returns HTTP response from the server

Return type `requests.Response`

property resource_id

Type `str`

spec ()

Returns this project's spec.

Returns The spec for the project.

Return type `ProjectSpec`

property type

A Tamr project type, listed in <https://docs.tamr.com/reference#create-a-project>.

Type `str`

unified_dataset ()

Unified dataset for this project.

Returns Unified dataset for this project.

Return type `Dataset`

Categories

Category

class `tamr_unify_client.categorization.category.resource.Category` (*client, data, alias=None*)

A category of a taxonomy

property name

Type `str`

property description

Type `str`

property path

Type `list[str]`

parent ()

Gets the parent Category of this one, or None if it is a tier 1 category

Returns The parent Category or None

Return type `Category`

spec ()

Returns this category's spec.

Returns The spec for the category.

Return type `CategorySpec`

delete ()

Deletes this resource. Some resources do not support deletion, and will raise a 405 error if this is called.

Returns HTTP response from the server

Return type `requests.Response`

property relative_id

Type `str`

property resource_id

Type `str`

Category Spec

```
class tamr_unify_client.categorization.category.resource.CategorySpec (client,
                                                                    data,
                                                                    api_path)
```

A representation of the server view of a category.

static of (resource)

Creates a category spec from a category.

Parameters **resource** (`Category`) – The existing category.

Returns The corresponding category spec.

Return type `CategorySpec`

static new ()

Creates a blank spec that could be used to construct a new category.

Returns The empty spec.

Return type `CategorySpec`

from data (data)

Creates a spec with the same client and API path as this one, but new data.

Parameters **data** (`dict`) – The data for the new spec.

Returns The new spec.

Return type *CategorySpec*

to_dict ()

Returns a version of this spec that conforms to the API representation.

Returns The spec's dict.

Return type *dict*

with_name (*new_name*)

Creates a new spec with the same properties, updating name.

Parameters **new_name** (*str*) – The new name.

Returns The new spec.

Return type *CategorySpec*

with_description (*new_description*)

Creates a new spec with the same properties, updating description.

Parameters **new_description** (*str*) – The new description.

Returns The new spec.

Return type *CategorySpec*

with_path (*new_path*)

Creates a new spec with the same properties, updating path.

Parameters **new_path** (*list[str]*) – The new path.

Returns The new spec.

Return type *CategorySpec*

Category Collection

class `tamr_unify_client.categorization.category.collection.CategoryCollection` (*client*, *api_path*)

Collection of *Category* s.

Parameters

- **client** (*Client*) – Client for API call delegation.
- **api_path** (*str*) – API path used to access this collection. E.g. "projects/1/taxonomy/categories".

by_resource_id (*resource_id*)

Retrieve a category by resource ID.

Parameters **resource_id** (*str*) – The resource ID. E.g. "1"

Returns The specified category.

Return type *Category*

by_relative_id (*relative_id*)

Retrieve a category by relative ID.

Parameters **relative_id** (*str*) – The relative ID. E.g. "projects/1/categories/1"

Returns The specified category.

Return type *Category*

by_external_id (*external_id*)

Retrieve an attribute by external ID.

Since categories do not have external IDs, this method is not supported and will raise a `NotImplementedError`.

Parameters **external_id** (*str*) – The external ID.

Returns The specified category, if found.

Return type *Category*

Raises

- **KeyError** – If no category with the specified `external_id` is found
- **LookupError** – If multiple categories with the specified `external_id` are found

stream ()

Stream categories in this collection. Implicitly called when iterating over this collection.

Returns Stream of categories.

Return type Python generator yielding *Category*

Usage:

```
>>> for category in collection.stream(): # explicit
>>>     do_stuff(category)
>>> for category in collection: # implicit
>>>     do_stuff(category)
```

create (*creation_spec*)

Creates a new category.

Parameters **creation_spec** (*dict*) – Category creation specification, formatted as specified in the [Public Docs for Creating a Category](#).

Returns The newly created category.

Return type *Category*

bulk_create (*creation_specs*)

Creates new categories in bulk.

Parameters **creation_specs** (*iterable[dict]*) – A collection of creation specifications, as detailed for `create`.

Returns JSON response from the server

Return type *dict*

delete_by_resource_id (*resource_id*)

Deletes a resource from this collection by resource ID.

Parameters **resource_id** (*str*) – The resource ID of the resource that will be deleted.

Returns HTTP response from the server.

Return type `requests.Response`

Taxonomy

class `tamr_unify_client.categorization.taxonomy.Taxonomy` (*client, data, alias=None*)

A project's taxonomy

property name

Type `str`

categories ()

Retrieves the categories of this taxonomy.

Returns A collection of the taxonomy categories.

Return type `CategoryCollection`

delete ()

Deletes this resource. Some resources do not support deletion, and will raise a 405 error if this is called.

Returns HTTP response from the server

Return type `requests.Response`

property relative_id

Type `str`

property resource_id

Type `str`

4.1.4 Client

class `tamr_unify_client.Client` (*auth, host='localhost', protocol='http', port=9100, base_path='/api/versioned/v1/', session=None*)

Python Client for Tamr API.

Each client is specific to a specific origin (protocol, host, port).

Parameters

- **auth** (`AuthBase`) – Tamr-compatible Authentication provider.
Recommended: use one of the classes described in [Authentication](#)
- **host** (`str`) – Host address of remote Tamr instance (e.g. '10.0.10.0')
- **protocol** (`str`) – Either 'http' or 'https'
- **port** (`int`) – Tamr instance main port
- **base_path** (`str`) – Base API path. Requests made by this client will be relative to this path.
- **session** (`Optional[Session]`) – Session to use for API calls. If none is provided, will use a new `requests.Session`.

Example


```
>>> from tamr_unify_client import Client
>>> from tamr_unify_client.auth import UsernamePasswordAuth
>>> auth = UsernamePasswordAuth('my username', 'my password')
>>> tamr_local = Client(auth) # on http://localhost:9100
>>> tamr_remote = Client(auth, protocol='https', host='10.0.10.0') # on https://
↪10.0.10.0:9100
```

property origin

HTTP origin i.e. <protocol>://<host>[:<port>].

For additional information, see [MDN web docs](#) .

Return type `str`

request (method, endpoint, **kwargs)

Sends a request to Tamr.

The URL for the request will be <origin>/<base_path>/<endpoint>. The request is authenticated via `Client.auth`.

Parameters

- **method** (`str`) – The HTTP method to use (e.g. ‘GET’ or ‘POST’)
- **endpoint** (`str`) – API endpoint to call (relative to the Base API path for this client).

Return type `Response`

Returns HTTP response from the Tamr server

get (endpoint, **kwargs)

Calls `request ()` with the "GET" method.

post (endpoint, **kwargs)

Calls `request ()` with the "POST" method.

put (endpoint, **kwargs)

Calls `request ()` with the "PUT" method.

delete (endpoint, **kwargs)

Calls `request ()` with the "DELETE" method.

property projects

Collection of all projects on this Tamr instance.

Return type `ProjectCollection`

Returns Collection of all projects.

property datasets

Collection of all datasets on this Tamr instance.

Return type `DatasetCollection`

Returns Collection of all datasets.

4.1.5 Datasets

Dataset

class `tamr_unify_client.dataset.resource.Dataset` (*client, data, alias=None*)

A Tamr dataset.

property name

Type `str`

property external_id

Type `str`

property description

Type `str`

property version

Type `str`

property tags

Type `list[str]`

property key_attribute_names

Type `list[str]`

property attributes

Attributes of this dataset.

Returns Attributes of this dataset.

Return type `AttributeCollection`

upsert_from_dataframe (*df*, *, *primary_key_name*, *ignore_nan=True*)

Upserts a record for each row of *df* with attributes for each column in *df*.

Parameters

- **df** (`DataFrame`) – The data to upsert records from.
- **primary_key_name** (`str`) – The name of the primary key of the dataset. Must be a column of *df*.
- **ignore_nan** (`bool`) – Whether to convert *NaN* values to *null* before upserting records to Tamr. If *False* and *NaN* is in *df*, this function will fail. Optional, default is *True*.

Return type `dict`

Returns JSON response body from the server.

Raises `KeyError` – If *primary_key_name* is not a column in *df*.

upsert_records (*records*, *primary_key_name*, ***json_args*)

Creates or updates the specified records.

Parameters

- **records** (`iterable[dict]`) – The records to update, as dictionaries.
- **primary_key_name** (`str`) – The name of the primary key for these records, which must be a key in each record dictionary.
- ****json_args** – Arguments to pass to the JSON *dumps* function, as documented [here](#). Some of these, such as *indent*, may not work with Tamr.

Returns JSON response body from the server.

Return type `dict`

delete_records (*records*, *primary_key_name*)

Deletes the specified records.

Parameters

- **records** (*iterable[dict]*) – The records to delete, as dictionaries.
- **primary_key_name** (*str*) – The name of the primary key for these records, which must be a key in each record dictionary.

Returns JSON response body from the server.

Return type *dict*

delete_records_by_id (*record_ids*)

Deletes the specified records.

Parameters **record_ids** (*iterable*) – The IDs of the records to delete.

Returns JSON response body from the server.

Return type *dict*

delete_all_records ()

Removes all records from the dataset.

Returns HTTP response from the server

Return type *requests.Response*

refresh (***options*)

Brings dataset up-to-date if needed, taking whatever actions are required.

Parameters ****options** – Options passed to underlying *Operation* . See *apply_options()* .

Returns The refresh operation.

Return type *Operation*

profile ()

Returns profile information for a dataset.

If profile information has not been generated, call `create_profile()` first. If the returned profile information is out-of-date, you can call `refresh()` on the returned object to bring it up-to-date.

Returns Dataset Profile information.

Return type *DatasetProfile*

create_profile (***options*)

Create a profile for this dataset.

If a profile already exists, the existing profile will be brought up to date.

Parameters ****options** – Options passed to underlying *Operation* . See *apply_options()* .

Returns The operation to create the profile.

Return type *Operation*

records ()

Stream this dataset’s records as Python dictionaries.

Returns Stream of records.

Return type Python generator yielding *dict*

status ()

Retrieve this dataset’s streamability status.

Returns Dataset streamability status.

Return type *DatasetStatus*

usage ()

Retrieve this dataset's usage by recipes and downstream datasets.

Returns The dataset's usage.

Return type *DatasetUsage*

from_geo_features (features, geo_attr=None)

Upsert this dataset from a geospatial FeatureCollection or iterable of Features.

features can be:

- An object that implements `__geo_interface__` as a FeatureCollection (see <https://gist.github.com/sgillies/2217756>)
- An iterable of features, where each element is a feature dictionary or an object that implements the `__geo_interface__` as a Feature
- A map where the "features" key contains an iterable of features

See: `geopandas.GeoDataFrame.from_features()`

If `geo_attr` is provided, then the named Tamr attribute will be used for the geometry. If `geo_attr` is not provided, then the first attribute on the dataset with geometry type will be used for the geometry.

Parameters

- **features** – geospatial features
- **geo_attr (str)** – (optional) name of the Tamr attribute to use for the feature's geometry

upstream_datasets ()

The Dataset's upstream datasets.

API returns the URIs of the upstream datasets, resulting in a list of DatasetURIs, not actual Datasets.

Returns A list of the Dataset's upstream datasets.

Return type `list[DatasetURI]`

spec ()

Returns this dataset's spec.

Returns The spec of this dataset.

Return type *DatasetSpec*

delete (cascade=False)

Deletes this dataset, optionally deleting all derived datasets as well.

Parameters **cascade (bool)** – Whether to delete all datasets derived from this one. Optional, default is *False*. Do not use this option unless you are certain you need it as it can have unintended consequences.

Returns HTTP response from the server

Return type `requests.Response`

itergeofeatures (geo_attr=None)

Returns an iterator that yields feature dictionaries that comply with `__geo_interface__`

See <https://gist.github.com/sgillies/2217756>

Parameters `geo_attr` (*str*) – (optional) name of the Tamr attribute to use for the feature’s geometry

Returns stream of features

Return type Python generator yielding `dict[str, object]`

`property relative_id`

Type `str`

`property resource_id`

Type `str`

Dataset Spec

class `tamr_unify_client.dataset.resource.DatasetSpec` (*client, data, api_path*)

A representation of the server view of a dataset.

static of (*resource*)

Creates a dataset spec from a dataset.

Parameters `resource` (*Dataset*) – The existing dataset.

Returns The corresponding dataset spec.

Return type `DatasetSpec`

static new ()

Creates a blank spec that could be used to construct a new dataset.

Returns The empty spec.

Return type `DatasetSpec`

from_data (*data*)

Creates a spec with the same client and API path as this one, but new data.

Parameters `data` (*dict*) – The data for the new spec.

Returns The new spec.

Return type `DatasetSpec`

to_dict ()

Returns a version of this spec that conforms to the API representation.

Returns The spec’s dict.

Return type `dict`

with_name (*new_name*)

Creates a new spec with the same properties, updating name.

Parameters `new_name` (*str*) – The new name.

Returns A new spec.

Return type `DatasetSpec`

with_external_id (*new_external_id*)

Creates a new spec with the same properties, updating external ID.

Parameters `new_external_id` (*str*) – The new external ID.

Returns A new spec.

Return type *DatasetSpec*

with_description (*new_description*)

Creates a new spec with the same properties, updating description.

Parameters **new_description** (*str*) – The new description.

Returns A new spec.

Return type *DatasetSpec*

with_key_attribute_names (*new_key_attribute_names*)

Creates a new spec with the same properties, updating key attribute names.

Parameters **new_key_attribute_names** (*list[str]*) – The new key attribute names.

Returns A new spec.

Return type *DatasetSpec*

with_tags (*new_tags*)

Creates a new spec with the same properties, updating tags.

Parameters **new_tags** (*list[str]*) – The new tags.

Returns A new spec.

Return type *DatasetSpec*

put ()

Updates the dataset on the server.

Returns The modified dataset.

Return type *Dataset*

Dataset Collection

class `tamr_unify_client.dataset.collection.DatasetCollection` (*client*,
api_path='datasets')

Collection of *Dataset* s.

Parameters

- **client** (*Client*) – Client for API call delegation.
- **api_path** (*str*) – API path used to access this collection. E.g. "projects/1/inputDatasets". Default: "datasets".

by_resource_id (*resource_id*)

Retrieve a dataset by resource ID.

Parameters **resource_id** (*str*) – The resource ID. E.g. "1"

Returns The specified dataset.

Return type *Dataset*

by_relative_id (*relative_id*)

Retrieve a dataset by relative ID.

Parameters **relative_id** (*str*) – The resource ID. E.g. "datasets/1"

Returns The specified dataset.

Return type *Dataset*

by_external_id (*external_id*)

Retrieve a dataset by external ID.

Parameters **external_id** (*str*) – The external ID.

Returns The specified dataset, if found.

Return type *Dataset*

Raises

- **KeyError** – If no dataset with the specified external_id is found
- **LookupError** – If multiple datasets with the specified external_id are found

stream ()

Stream datasets in this collection. Implicitly called when iterating over this collection.

Returns Stream of datasets.

Return type Python generator yielding *Dataset*

Usage:

```
>>> for dataset in collection.stream(): # explicit
>>>     do_stuff(dataset)
>>> for dataset in collection: # implicit
>>>     do_stuff(dataset)
```

by_name (*dataset_name*)

Lookup a specific dataset in this collection by exact-match on name.

Parameters **dataset_name** (*str*) – Name of the desired dataset.

Returns Dataset with matching name in this collection.

Return type *Dataset*

Raises **KeyError** – If no dataset with specified name was found.

delete_by_resource_id (*resource_id, cascade=False*)

Deletes a dataset from this collection by resource_id. Optionally deletes all derived datasets as well.

Parameters

- **resource_id** (*str*) – The resource id of the dataset in this collection to delete.
- **cascade** (*bool*) – Whether to delete all datasets derived from the deleted one. Optional, default is *False*. Do not use this option unless you are certain you need it as it can have unintended consequences.

Returns HTTP response from the server.

Return type *requests.Response*

create (*creation_spec*)

Create a Dataset in Tamr

Parameters **creation_spec** (*dict[str, str]*) – Dataset creation specification should be formatted as specified in the [Public Docs for Creating a Dataset](#).

Returns The created Dataset

Return type *Dataset*

create_from_dataframe (*df*, *primary_key_name*, *dataset_name*, *ignore_nan=True*)

Creates a dataset in this collection with the given name, creates an attribute for each column in the *df* (with *primary_key_name* as the key attribute), and upserts a record for each row of *df*.

Each attribute has the default type *ARRAY[STRING]*, besides the key attribute, which will have type *STRING*.

This function attempts to ensure atomicity, but it is not guaranteed. If an error occurs while creating attributes or records, an attempt will be made to delete the dataset that was created. However, if this request errors, it will not try again.

Parameters

- **df** (*pandas.DataFrame*) – The data to create the dataset with.
- **primary_key_name** (*str*) – The name of the primary key of the dataset. Must be a column of *df*.
- **dataset_name** (*str*) – What to name the dataset in Tamr. There cannot already be a dataset with this name.
- **ignore_nan** (*bool*) – Whether to convert *NaN* values to *null* before upserting records to Tamr. If *False* and *NaN* is in *df*, this function will fail. Optional, default is *True*.

Returns The newly created dataset.

Return type *Dataset*

Raises

- **KeyError** – If *primary_key_name* is not a column in *df*.
- **CreationError** – If a step in creating the dataset fails.

```
class tamr_unify_client.dataset.collection.CreationError (error_message)
    An error from create_from_dataframe ()

    with_traceback ()
        Exception.with_traceback(tb) – set self.__traceback__ to tb and return self.
```

Dataset Profile

```
class tamr_unify_client.dataset.profile.DatasetProfile (client, data, alias=None)
    Profile info of a Tamr dataset.
```

```
property dataset_name
    The name of the associated dataset.
```

Type *str*

Return type *str*

```
property relative_dataset_id
    The relative dataset ID of the associated dataset.
```

Type *str*

Return type *str*

```
property is_up_to_date
    Whether the associated dataset is up to date.
```

Type *bool*

Return type *bool*

property profiled_data_version

The profiled data version.

Type `str`

Return type `str`

property profiled_at

Info about when profile info was generated.

Type `dict`

Return type `dict`

property simple_metrics

Simple metrics for profiled dataset.

Type `list`

Return type `list`

property attribute_profiles

Simple metrics for profiled dataset.

Type `list`

Return type `list`

refresh (options)**

Updates the dataset profile if needed.

The dataset profile is updated on the server; you will need to call `profile()` to retrieve the updated profile.

Parameters ****options** – Options passed to underlying `Operation`. See `apply_options()`.

Returns The refresh operation.

Return type `Operation`

delete ()

Deletes this resource. Some resources do not support deletion, and will raise a 405 error if this is called.

Returns HTTP response from the server

Return type `requests.Response`

property relative_id

Type `str`

property resource_id

Type `str`

Dataset Status

class `tamr_unify_client.dataset.status.DatasetStatus (client, data, alias=None)`
 Streamability status of a Tamr dataset.

property dataset_name

The name of the associated dataset.

Type `str`

Return type `str`

property `relative_dataset_id`

The relative dataset ID of the associated dataset.

Type `str`

Return type `str`

property `is_streamable`

Whether the associated dataset is available to be streamed.

Type `bool`

Return type `bool`

delete ()

Deletes this resource. Some resources do not support deletion, and will raise a 405 error if this is called.

Returns HTTP response from the server

Return type `requests.Response`

property `relative_id`

Type `str`

property `resource_id`

Type `str`

Dataset URI

class `tamr_unify_client.dataset.uri.DatasetURI` (`client`, `uri`)

Identifier of a dataset.

Parameters

- **client** (`Client`) – Queried dataset’s client.
- **uri** (`str`) – Queried dataset’s dataset ID.

property `resource_id`

Type `str`

property `relative_id`

Type `str`

property `uri`

Type `str`

dataset ()

Fetch the dataset that this identifier points to.

Returns A Tamr dataset.

Return type

`class ~tamr_unify_client.dataset.resource.Dataset`

Dataset Usage

class `tamr_unify_client.dataset.usage.DatasetUsage` (*client*, *data*, *alias=None*)

The usage of a dataset and its downstream dependencies.

See <https://docs.tamr.com/reference#retrieve-downstream-dataset-usage>

property `relative_id`

Type `str`

property `usage`

Type `DatasetUse`

property `dependencies`

Type `list[DatasetUse]`

delete ()

Deletes this resource. Some resources do not support deletion, and will raise a 405 error if this is called.

Returns HTTP response from the server

Return type `requests.Response`

property `resource_id`

Type `str`

Dataset Use

class `tamr_unify_client.dataset.use.DatasetUse` (*client*, *data*)

The use of a dataset in project steps. This is not a *BaseResource* because it has no API path and cannot be directly retrieved or modified.

See <https://docs.tamr.com/reference#retrieve-downstream-dataset-usage>

Parameters

- **client** (*Client*) – Delegate underlying API calls to this client.
- **data** (*dict*) – The JSON body containing usage information.

property `dataset_id`

Type `str`

property `dataset_name`

Type `str`

property `input_to_project_steps`

Type `list[ProjectStep]`

property `output_from_project_steps`

Type `list[ProjectStep]`

dataset ()

Retrieves the *Dataset* this use represents.

Returns The dataset being used.

Return type `Dataset`

4.1.6 Machine Learning Model

class `tamr_unify_client.base_model.MachineLearningModel` (*client*, *data*, *alias=None*)

A Tamr Machine Learning model.

train (***options*)

Learn from verified labels.

Parameters ****options** – Options passed to underlying *Operation* . See *apply_options()* .

Returns The resultant operation.

Return type *Operation*

predict (***options*)

Suggest labels for unverified records.

Parameters ****options** – Options passed to underlying *Operation* . See *apply_options()* .

Returns The resultant operation.

Return type *Operation*

delete ()

Deletes this resource. Some resources do not support deletion, and will raise a 405 error if this is called.

Returns HTTP response from the server

Return type *requests.Response*

property `relative_id`

Type *str*

property `resource_id`

Type *str*

4.1.7 Mastering

Binning Model

class `tamr_unify_client.mastering.binning_model.BinningModel` (*client*, *data*, *alias=None*)

A binning model object.

records ()

Stream this object's records as Python dictionaries.

Returns Stream of records.

Return type Python generator yielding *dict*

update_records (*records*)

Send a batch of record creations/updates/deletions to this dataset.

Parameters **records** (*iterable[dict]*) – Each record should be formatted as specified in the [Public Docs for Dataset updates](#).

Returns JSON response body from server.

Return type *dict*

delete()

Deletes this resource. Some resources do not support deletion, and will raise a 405 error if this is called.

Returns HTTP response from the server

Return type `requests.Response`

property relative_id

Type `str`

property resource_id

Type `str`

Estimated Pair Counts

```
class tamr_unify_client.mastering.estimated_pair_counts.EstimatedPairCounts (client,
                                                                    data,
                                                                    alias=None)
```

Estimated Pair Counts info for Mastering Project

property is_up_to_date

Whether an estimate pairs job has been run since the last edit to the binning model.

Return type `bool`

property total_estimate

The total number of estimated candidate pairs and generated pairs for the model across all clauses.

Returns

A dictionary containing candidate pairs and estimated pairs mapped to their corresponding estimated counts. For example:

```
{
    "candidatePairCount": "54321",
    "generatedPairCount": "12345"
}
```

Return type `dict[str, str]`

property clause_estimates

The estimated candidate pair count and generated pair count for each clause in the model.

Returns

A dictionary containing each clause name mapped to a dictionary containing the corresponding estimated candidate and generated pair counts. For example:

```
{
    "Clause1": {
        "candidatePairCount": "321",
        "generatedPairCount": "123"
    },
    "Clause2": {
```

```

        "candidatePairCount": "654",
        "generatedPairCount": "456"
    }
}

```

Return type `dict[str, dict[str, str]]`

refresh (**options)

Updates the estimated pair counts if needed.

The pair count estimates are updated on the server; you will need to call `estimate_pairs()` to retrieve the updated estimate.

Parameters ****options** – Options passed to underlying `Operation`. See `apply_options()`.

Returns The refresh operation.

Return type `Operation`

delete ()

Deletes this resource. Some resources do not support deletion, and will raise a 405 error if this is called.

Returns HTTP response from the server

Return type `requests.Response`

property relative_id

Type `str`

property resource_id

Type `str`

Mastering Project

class `tamr_unify_client.mastering.project.MasteringProject` (*client*, *data*, *alias=None*)

A Mastering project in Tamr.

pairs ()

Record pairs generated by Tamr’s binning model. Pairs are displayed on the “Pairs” page in the Tamr UI.

Call `refresh()` from this dataset to regenerate pairs according to the latest binning model.

Returns The record pairs represented as a dataset.

Return type `Dataset`

pair_matching_model ()

Machine learning model for pair-matching for this Mastering project. Learns from verified labels and predicts categorization labels for unlabeled pairs.

Calling `predict()` from this dataset will produce new (unpublished) clusters. These clusters are displayed on the “Clusters” page in the Tamr UI.

Returns The machine learning model for pair-matching.

Return type `MachineLearningModel`

high_impact_pairs ()

High-impact pairs as a dataset. Tamr labels pairs as “high-impact” if labeling these pairs would help it learn most quickly (i.e. “Active learning”).

High-impact pairs are displayed with a lightning bolt icon on the “Pairs” page in the Tamr UI.

Call *refresh ()* from this dataset to produce new high-impact pairs according to the latest pair-matching model.

Returns The high-impact pairs represented as a dataset.

Return type *Dataset*

record_clusters ()

Record Clusters as a dataset. Tamr clusters labeled pairs using pairs model. These clusters populate the cluster review page and get transient cluster ids, rather than published cluster ids (i.e., “Permanent Ids”)

Call *refresh ()* from this dataset to generate clusters based on to the latest pair-matching model.

Returns The record clusters represented as a dataset.

Return type *Dataset*

published_clusters ()

Published record clusters generated by Tamr’s pair-matching model.

Returns The published clusters represented as a dataset.

Return type *Dataset*

published_clusters_configuration ()

Retrieves published clusters configuration for this project.

Returns The published clusters configuration

Return type *PublishedClustersConfiguration*

published_cluster_ids ()

Retrieves published cluster IDs for this project.

Returns The published cluster ID dataset.

Return type *Dataset*

published_cluster_stats ()

Retrieves published cluster stats for this project.

Returns The published cluster stats dataset.

Return type *Dataset*

published_cluster_versions (cluster_ids)

Retrieves version information for the specified published clusters. See <https://docs.tamr.com/reference#retrieve-published-clusters-given-cluster-ids>.

Parameters *cluster_ids* (*iterable[str]*) – The persistent IDs of the clusters to get version information for.

Returns A stream of the published clusters.

Return type Python generator yielding *PublishedCluster*

record_published_cluster_versions (record_ids)

Retrieves version information for the published clusters of the given records. See <https://docs.tamr.com/reference#retrieve-published-clusters-given-record-ids>.

Parameters `record_ids` (*iterable[str]*) – The Tamr IDs of the records to get cluster version information for.

Returns A stream of the relevant published clusters.

Return type Python generator yielding *RecordPublishedCluster*

estimate_pairs ()

Returns pair estimate information for a mastering project

Returns Pairs Estimate information.

Return type *EstimatedPairCounts*

record_clusters_with_data ()

Project's unified dataset with associated clusters.

Returns The record clusters with data represented as a dataset

Return type *Dataset*

published_clusters_with_data ()

Project's unified dataset with associated clusters.

Returns The published clusters with data represented as a dataset

Return type *Dataset*

binning_model ()

Binning model for this project.

Returns Binning model for this project.

Return type *BinningModel*

add_input_dataset (*dataset*)

Associate a dataset with a project in Tamr.

By default, datasets are not associated with any projects. They need to be added as input to a project before they can be used as part of that project

Parameters `dataset` (*Dataset*) – The dataset to associate with the project.

Returns HTTP response from the server

Return type *requests.Response*

as_categorization ()

Convert this project to a *CategorizationProject*

Returns This project.

Return type *CategorizationProject*

Raises **TypeError** – If the *type* of this project is not "CATEGORIZATION"

as_mastering ()

Convert this project to a *MasteringProject*

Returns This project.

Return type *MasteringProject*

Raises **TypeError** – If the *type* of this project is not "DEDUP"

attribute_configurations ()

Project's attribute's configurations.

Returns The configurations of the attributes of a project.

Return type *AttributeConfigurationCollection*

attribute_mappings ()

Project's attribute's mappings.

Returns The attribute mappings of a project.

Return type *AttributeMappingCollection*

property attributes

Attributes of this project.

Returns Attributes of this project.

Return type *AttributeCollection*

delete ()

Deletes this resource. Some resources do not support deletion, and will raise a 405 error if this is called.

Returns HTTP response from the server

Return type *requests.Response*

property description

Type *str*

property external_id

Type *str*

input_datasets ()

Retrieve a collection of this project's input datasets.

Returns The project's input datasets.

Return type *DatasetCollection*

property name

Type *str*

property relative_id

Type *str*

remove_input_dataset (dataset)

Remove a dataset from a project.

Parameters **dataset** (*Dataset*) – The dataset to be removed from this project.

Returns HTTP response from the server

Return type *requests.Response*

property resource_id

Type *str*

spec ()

Returns this project's spec.

Returns The spec for the project.

Return type *ProjectSpec*

property type

A Tamr project type, listed in <https://docs.tamr.com/reference#create-a-project>.

Type `str`

unified_dataset ()

Unified dataset for this project.

Returns Unified dataset for this project.

Return type `Dataset`

Published Clusters

Metric

class `tamr_unify_client.mastering.published_cluster.metric.Metric (data)`

A metric for a published cluster.

This is not a *BaseResource* because it does not have its own API endpoint.

Parameters `data` – The JSON entity representing this cluster.

property name

Type `str`

property value

Type `str`

Published Cluster

class `tamr_unify_client.mastering.published_cluster.resource.PublishedCluster (data)`

A representation of a published cluster in a mastering project with version information. See <https://docs.tamr.com/reference#retrieve-published-clusters-given-cluster-ids>.

This is not a *BaseResource* because it does not have its own API endpoint.

Parameters `data` – The JSON entity representing this *PublishedCluster*.

property id

Type `str`

property versions

Type `list[PublishedClusterVersion]`

Published Cluster Configuration

class `tamr_unify_client.mastering.published_cluster.configuration.PublishedClustersConfiguratio`

The configuration of published clusters in a project.

See <https://docs.tamr.com/reference#the-published-clusters-configuration-object>

property relative_id

Type `str`

property versions_time_to_live

Type `str`

spec()

Returns a spec representation of this published cluster configuration.

Returns The published cluster configuration spec.

Return type :class:`~tamr_unify_client.mastering.published_cluster.configuration.PublishedClustersConfigurationSpec`

delete()

Deletes this resource. Some resources do not support deletion, and will raise a 405 error if this is called.

Returns HTTP response from the server

Return type `requests.Response`

property resource_id

Type `str`

Published Cluster Version

class `tamr_unify_client.mastering.published_cluster.version.PublishedClusterVersion` (*data*)
 A version of a published cluster in a mastering project.

This is not a *BaseResource* because it does not have its own API endpoint.

Parameters *data* – The JSON entity representing this version.

property version

Type `str`

property timestamp

Type `str`

property name

Type `str`

property metrics

Type `list[Metric]`

property record_ids

Type `list[dict[str, str]]`

Record Published Cluster

class `tamr_unify_client.mastering.published_cluster.record.RecordPublishedCluster` (*data*)
 A representation of a published cluster of a record in a mastering project with version information. See <https://docs.tamr.com/reference#retrieve-published-clusters-given-record-ids>.

This is not a *BaseResource* because it does not have its own API endpoint.

Parameters *data* – The JSON entity representing this *RecordPublishedCluster*.

property entity_id

Type `str`

```

property source_id
    Type str
property origin_entity_id
    Type str
property origin_source_id
    Type str
property versions
    Type list[RecordPublishedClusterVersion]

```

Record Published Cluster Version

class `tamr_unify_client.mastering.published_cluster.record_version.RecordPublishedClusterVersion`
 A version of a published cluster in a mastering project.

This is not a *BaseResource* because it does not have its own API endpoint.

Parameters `data` – The JSON entity representing this version.

```

property version
    Type str
property timestamp
    Type str
property cluster_id
    Type str

```

4.1.8 Operation

class `tamr_unify_client.operation.Operation` (*client*, *data*, *alias=None*)
 A long-running operation performed by Tamr. Operations appear on the “Jobs” page of the Tamr UI.

By design, client-side operations represent server-side operations *at a particular point in time* (namely, when the operation was fetched from the server). In other words: Operations *will not* pick up on server-side changes automatically. To get an up-to-date representation, refetch the operation e.g. `op = op.poll()`.

classmethod `from_response` (*client*, *response*)

Handle idiosyncrasies in constructing Operations from Tamr responses. When a Tamr API call would start an operation, but all results that would be produced by that operation are already up-to-date, Tamr returns *HTTP 204 No Content*

To make it easy for client code to handle these API responses without checking the response code, this method will either construct an Operation, or a dummy *NoOp* operation representing the 204 Success response.

Parameters

- **client** (*Client*) – Delegate underlying API calls to this client.
- **response** (*requests.Response*) – HTTP Response from the request that started the operation.

Returns Operation

Return type *Operation*

apply_options (*asynchronous=False, **options*)

Applies operation options to this operation.

NOTE: This function **should not** be called directly. Rather, options should be passed in through a higher-level function e.g. *refresh()* .

Synchronous mode: Automatically waits for operation to resolve before returning the operation.

asynchronous mode: Immediately return the 'PENDING' operation. It is up to the user to coordinate this operation with their code via *wait()* and/or *poll()* .

Parameters

- **asynchronous** (*bool*) – Whether or not to run in asynchronous mode. Default: *False*.
- ****options** – When running in synchronous mode, these options are passed to the underlying *wait()* call.

Returns Operation with options applied.

Return type *Operation*

property type

Type *str*

property description

Type *str*

property state

Server-side state of this operation.

Operation state can be unresolved (i.e. *state* is one of: 'PENDING', 'RUNNING'), or resolved (i.e. *state* is one of: 'CANCELED', 'SUCCEEDED', 'FAILED'). Unless opting into asynchronous mode, all exposed operations should be resolved.

Note: you only need to manually pick up server-side changes when opting into asynchronous mode when kicking off this operation.

Usage:

```
>>> op.state # operation is currently 'PENDING'
'PENDING'
>>> op.wait() # continually polls until operation resolves
>>> op.state # incorrect usage; operation object state never changes.
'PENDING'
>>> op = op.poll() # correct usage; use value returned by Operation.poll_
↳ or Operation.wait
>>> op.state
'SUCCEEDED'
```

poll()

Poll this operation for server-side updates.

Does not update the calling *Operation* object. Instead, returns a new *Operation*.

Returns Updated representation of this operation.

Return type *Operation*

wait (*poll_interval_seconds=3, timeout_seconds=None*)
Continuously polls for this operation's server-side state.

Parameters

- **poll_interval_seconds** (*int*) – Time interval (in seconds) between subsequent polls.
- **timeout_seconds** (*int*) – Time (in seconds) to wait for operation to resolve.

Raises **TimeoutError** – If operation takes longer than *timeout_seconds* to resolve.

Returns Resolved operation.

Return type *Operation*

succeeded ()

Convenience method for checking if operation was successful.

Returns True if operation's state is 'SUCCEEDED', False otherwise.

Return type *bool*

delete ()

Deletes this resource. Some resources do not support deletion, and will raise a 405 error if this is called.

Returns HTTP response from the server

Return type *requests.Response*

property *relative_id*

Type *str*

property *resource_id*

Type *str*

4.1.9 Projects

Attribute Configurations

Attribute Configuration

class `tamr_unify_client.project.attribute_configuration.resource.AttributeConfiguration` (*clie*

date

alia

The configurations of Tamr Attributes.

See <https://docs.tamr.com/reference#the-attribute-configuration-object>

property *relative_id*

Type *str*

property *id*

Type *str*

property *relative_attribute_id*

Type *str*

property *attribute_role*

Type `str`

`property similarity_function`

Type `str`

`property enabled_for_ml`

Type `bool`

`property tokenizer`

Type `str`

`property numeric_field_resolution`

Type `list`

`property attribute_name`

Type `str`

`spec()`
Returns this attribute configuration's spec.

Returns The spec of this attribute configuration.

Return type `AttributeConfigurationSpec`

`delete()`
Deletes this resource. Some resources do not support deletion, and will raise a 405 error if this is called.

Returns HTTP response from the server

Return type `requests.Response`

`property resource_id`

Type `str`

Attribute Configuration Spec

`class tamr_unify_client.project.attribute_configuration.resource.AttributeConfigurationSpec`

A representation of the server view of an attribute configuration.

static of (`resource`)

Creates an attribute configuration spec from an attribute configuration.

Parameters `resource` (`AttributeConfiguration`) – The existing attribute configuration.

Returns The corresponding attribute creation spec.

Return type `AttributeConfigurationSpec`

static new ()

Creates a blank spec that could be used to construct a new attribute configuration.

Returns The empty spec.

Return type `AttributeConfigurationSpec`

from_data (`data`)

Creates a spec with the same client and API path as this one, but new data.

Parameters `data` (*dict*) – The data for the new spec.

Returns The new spec.

Return type *AttributeConfigurationSpec*

to_dict ()

Returns a version of this spec that conforms to the API representation.

Returns The spec's dict.

Return type *dict*

with_attribute_role (*new_attribute_role*)

Creates a new spec with the same properties, updating attribute role.

Parameters `new_attribute_role` (*str*) – The new attribute role.

Returns A new spec.

Return type *AttributeConfigurationSpec*

with_similarity_function (*new_similarity_function*)

Creates a new spec with the same properties, updating similarity function.

Parameters `new_similarity_function` (*str*) – The new similarity function.

Returns A new spec.

Return type *AttributeConfigurationSpec*

with_enabled_for_ml (*new_enabled_for_ml*)

Creates a new spec with the same properties, updating enabled for ML.

Parameters `new_enabled_for_ml` (*bool*) – Whether the builder is enabled for ML.

Returns A new spec.

Return type *AttributeConfigurationSpec*

with_tokenizer (*new_tokenizer*)

Creates a new spec with the same properties, updating tokenizer.

Parameters `new_tokenizer` (*str*) – The new tokenizer.

Returns A new spec.

Return type *AttributeConfigurationSpec*

with_numeric_field_resolution (*new_numeric_field_resolution*)

Creates a new spec with the same properties, updating numeric field resolution.

Parameters `new_numeric_field_resolution` (*str*) – The new numeric field resolution.

Returns A new spec.

Return type *AttributeConfigurationSpec*

with_attribute_name (*new_attribute_name*)

Creates a new spec with the same properties, updating new attribute name.

Parameters `new_attribute_name` (*str*) – The new attribute name.

Returns A new spec.

Return type *AttributeConfigurationSpec*

put ()
 Updates the attribute configuration on the server.
Returns The modified attribute configuration.
Return type *AttributeConfiguration*

Attribute Configuration Collection

class `tamr_unify_client.project.attribute_configuration.collection.AttributeConfigurationCollection`

Collection of *AttributeConfiguration*

Parameters

- **client** (*Client*) – Client for API call delegation.
- **api_path** (*str*) – API path used to access this collection. E.g. "projects/1/attributeConfigurations"

by_resource_id (*resource_id*)
 Retrieve an attribute configuration by resource ID.

Parameters **resource_id** (*str*) – The resource ID.

Returns The specified attribute configuration.

Return type *AttributeConfiguration*

by_relative_id (*relative_id*)
 Retrieve an attribute configuration by relative ID.

Parameters **relative_id** (*str*) – The relative ID.

Returns The specified attribute configuration.

Return type *AttributeConfiguration*

by_external_id (*external_id*)
 Retrieve an attribute configuration by external ID.

Since attributes do not have external IDs, this method is not supported and will raise a `NotImplementedError`.

Parameters **external_id** (*str*) – The external ID.

Returns The specified attribute, if found.

Return type *AttributeConfiguration*

Raises

- **KeyError** – If no attribute with the specified `external_id` is found
- **LookupError** – If multiple attributes with the specified `external_id` are found
- **NotImplementedError** – `AttributeConfiguration` does not support `external_id`

stream ()
 Stream attribute configurations in this collection. Implicitly called when iterating over this collection.

Returns Stream of attribute configurations.

Return type Python generator yielding *AttributeConfiguration*

Usage:

```

>>> for attributeConfiguration in collection.stream(): # explicit
>>>     do_stuff(attributeConfiguration)
>>> for attributeConfiguration in collection: # implicit
>>>     do_stuff(attributeConfiguration)

```

create (*creation_spec*)

Create an Attribute configuration in this collection

Parameters **creation_spec** (*dict[str, str]*) – Attribute configuration creation specification should be formatted as specified in the [Public Docs for adding an AttributeConfiguration](#).

Returns The created Attribute configuration

Return type *AttributeConfiguration*

delete_by_resource_id (*resource_id*)

Deletes a resource from this collection by resource ID.

Parameters **resource_id** (*str*) – The resource ID of the resource that will be deleted.

Returns HTTP response from the server.

Return type *requests.Response*

Attribute Mappings

Attribute Mapping

class `tamr_unify_client.project.attribute_mapping.resource.AttributeMapping` (*client*, *data*)

see <https://docs.tamr.com/reference#retrieve-projects-mappings> AttributeMapping and AttributeMappingCollection do not inherit from BaseResource and BaseCollection. BC and BR require a specific URL for each individual attribute mapping (ex: /projects/1/attributeMappings/1), but these types of URLs do not exist for attribute mappings

property `id`

Type *str*

property `relative_id`

Type *str*

property `input_attribute_id`

Type *str*

property `relative_input_attribute_id`

Type *str*

property `input_dataset_name`

Type *str*

property `input_attribute_name`

Type *str*

property `unified_attribute_id`

Type *str*

property `relative_unified_attribute_id`

Type `str`

property `unified_dataset_name`

Type `str`

property `unified_attribute_name`

Type `str`

property `resource_id`

Type `str`

spec ()

Returns a spec representation of this attribute mapping.

Returns The attribute mapping spec.

Return type `AttributeMappingSpec`

delete ()

Delete this attribute mapping.

Returns HTTP response from the server

Return type `requests.Response`

Attribute Mapping Spec

class `tamr_unify_client.project.attribute_mapping.resource.AttributeMappingSpec` (*data*)

A representation of the server view of an attribute mapping

static of (*resource*)

Creates an attribute mapping spec from a attribute mapping.

Parameters `resource` (`AttributeMapping`) – The existing attribute mapping.

Returns The corresponding attribute mapping spec.

Return type `AttributeMappingSpec`

static new ()

Creates a blank spec that could be used to construct a new attribute mapping.

Returns The empty spec.

Return type `AttributeMappingSpec`

to_dict ()

Returns a version of this spec that conforms to the API representation.

Returns The spec's dict.

Return type `dict`

with_input_attribute_id (*new_input_attribute_id*)

Creates a new spec with the same properties, updating the input attribute id.

Parameters `new_input_attribute_id` (*str*) – The new input attribute id.

Returns The new spec.

Return type `AttributeMappingSpec`

- with_relative_input_attribute_id** (*new_relative_input_attribute_id*)
Creates a new spec with the same properties, updating the relative input attribute id.
- Parameters** **new_relative_input_attribute_id** (*str*) – The new relative input attribute Id.
- Returns** The new spec.
- Return type** *AttributeMappingSpec*
- with_input_dataset_name** (*new_input_dataset_name*)
Creates a new spec with the same properties, updating the input dataset name.
- Parameters** **new_input_dataset_name** (*str*) – The new input dataset name.
- Returns** The new spec.
- Return type** *AttributeMappingSpec*
- with_input_attribute_name** (*new_input_attribute_name*)
Creates a new spec with the same properties, updating the input attribute name.
- Parameters** **new_input_attribute_name** (*str*) – The new input attribute name.
- Returns** The new spec.
- Return type** *AttributeMappingSpec*
- with_unified_attribute_id** (*new_unified_attribute_id*)
Creates a new spec with the same properties, updating the unified attribute id.
- Parameters** **new_unified_attribute_id** (*str*) – The new unified attribute id.
- Returns** The new spec.
- Return type** *AttributeMappingSpec*
- with_relative_unified_attribute_id** (*new_relative_unified_attribute_id*)
Creates a new spec with the same properties, updating the relative unified attribute id.
- Parameters** **new_relative_unified_attribute_id** (*str*) – The new relative unified attribute id.
- Returns** The new spec.
- Return type** *AttributeMappingSpec*
- with_unified_dataset_name** (*new_unified_dataset_name*)
Creates a new spec with the same properties, updating the unified dataset name.
- Parameters** **new_unified_dataset_name** (*str*) – The new unified dataset name.
- Returns** The new spec.
- Return type** *AttributeMappingSpec*
- with_unified_attribute_name** (*new_unified_attribute_name*)
Creates a new spec with the same properties, updating the unified attribute name.
- Parameters** **new_unified_attribute_name** (*str*) – The new unified attribute name.
- Returns** The new spec.
- Return type** *AttributeMappingSpec*

Attribute Mapping Collection

class tamr_unify_client.project.attribute_mapping.collection.**AttributeMappingCollection** (*client*, *api*)

Collection of *AttributeMapping*

Parameters

- **client** (*Client*) – Client for API call delegation.
- **api_path** (*str*) – API path used to access this collection.

stream ()

Stream attribute mappings in this collection. Implicitly called when iterating over this collection.

Returns Stream of attribute mappings.

Return type Python generator yielding *AttributeMapping*

by_resource_id (*resource_id*)

Retrieve an item in this collection by resource ID.

Parameters **resource_id** (*str*) – The resource ID.

Returns The specified attribute mapping.

Return type *AttributeMapping*

by_relative_id (*relative_id*)

Retrieve an item in this collection by relative ID.

Parameters **relative_id** (*str*) – The relative ID.

Returns The specified attribute mapping.

Return type *AttributeMapping*

create (*creation_spec*)

Create an Attribute mapping in this collection

Parameters **creation_spec** (*dict[str, str]*) – Attribute mapping creation specification should be formatted as specified in the [Public Docs for adding an AttributeMapping](#).

Returns The created Attribute mapping

Return type *AttributeMapping*

delete_by_resource_id (*resource_id*)

Delete an attribute mapping using its Resource ID.

Parameters **resource_id** (*str*) – the resource ID of the mapping to be deleted.

Returns HTTP response from the server

Return type *requests.Response*

Project

class tamr_unify_client.project.resource.**Project** (*client*, *data*, *alias=None*)

A Tamr project.

property **name**

Type *str*

property **external_id**

Type `str`

property description

Type `str`

property type

A Tamr project type, listed in <https://docs.tamr.com/reference#create-a-project>.

Type `str`

property attributes

Attributes of this project.

Returns Attributes of this project.

Return type `AttributeCollection`

unified_dataset ()

Unified dataset for this project.

Returns Unified dataset for this project.

Return type `Dataset`

as_categorization ()

Convert this project to a `CategorizationProject`

Returns This project.

Return type `CategorizationProject`

Raises `TypeError` – If the `type` of this project is not "CATEGORIZATION"

as_mastering ()

Convert this project to a `MasteringProject`

Returns This project.

Return type `MasteringProject`

Raises `TypeError` – If the `type` of this project is not "DEDUP"

add_input_dataset (dataset)

Associate a dataset with a project in Tamr.

By default, datasets are not associated with any projects. They need to be added as input to a project before they can be used as part of that project

Parameters `dataset` (`Dataset`) – The dataset to associate with the project.

Returns HTTP response from the server

Return type `requests.Response`

remove_input_dataset (dataset)

Remove a dataset from a project.

Parameters `dataset` (`Dataset`) – The dataset to be removed from this project.

Returns HTTP response from the server

Return type `requests.Response`

input_datasets ()

Retrieve a collection of this project's input datasets.

Returns The project's input datasets.

Return type *DatasetCollection*

attribute_configurations ()

Project's attribute's configurations.

Returns The configurations of the attributes of a project.

Return type *AttributeConfigurationCollection*

attribute_mappings ()

Project's attribute's mappings.

Returns The attribute mappings of a project.

Return type *AttributeMappingCollection*

spec ()

Returns this project's spec.

Returns The spec for the project.

Return type *ProjectSpec*

delete ()

Deletes this resource. Some resources do not support deletion, and will raise a 405 error if this is called.

Returns HTTP response from the server

Return type *requests.Response*

property relative_id

Type *str*

property resource_id

Type *str*

Project Spec

class `tamr_unify_client.project.resource.ProjectSpec` (*client, data, api_path*)

A representation of the server view of a project.

static of (*resource*)

Creates a project spec from a project.

Parameters **resource** (*Project*) – The existing project.

Returns The corresponding project spec.

Return type *ProjectSpec*

static new ()

Creates a blank spec that could be used to construct a new project.

Returns The empty spec.

Return type *ProjectSpec*

from_data (*data*)

Creates a spec with the same client and API path as this one, but new data.

Parameters **data** (*dict*) – The data for the new spec.

Returns The new spec.

Return type *ProjectSpec*

to_dict ()

Returns a version of this spec that conforms to the API representation.

Returns The spec's dict.

Return type dict

with_name (*new_name*)

Creates a new spec with the same properties, updating name.

Parameters **new_name** (*str*) – The new name.

Returns The new spec.

Return type *ProjectSpec*

with_description (*new_description*)

Creates a new spec with the same properties, updating description.

Parameters **new_description** (*str*) – The new description.

Returns The new spec.

Return type *ProjectSpec*

with_type (*new_type*)

Creates a new spec with the same properties, updating type.

Parameters **new_type** (*str*) – The new type.

Returns The new spec.

Return type *ProjectSpec*

with_external_id (*new_external_id*)

Creates a new spec with the same properties, updating external ID.

Parameters **new_external_id** (*str*) – The new external ID.

Returns The new spec.

Return type *ProjectSpec*

with_unified_dataset_name (*new_unified_dataset_name*)

Creates a new spec with the same properties, updating unified dataset name.

Parameters **new_unified_dataset_name** (*str*) – The new unified dataset name.

Returns The new spec.

Return type *ProjectSpec*

put ()

Commits these changes by updating the project in Tamr.

Returns The updated project.

Return type *Project*

Project Collection

```
class tamr_unify_client.project.collection.ProjectCollection (client,  
                                                         api_path='projects')
```

Collection of *Project* s.

Parameters

- **client** (*Client*) – Client for API call delegation.
- **api_path** (*str*) – API path used to access this collection. Default: "projects".

by_resource_id (*resource_id*)

Retrieve a project by resource ID.

Parameters **resource_id** (*str*) – The resource ID. E.g. "1"

Returns The specified project.

Return type *Project*

by_relative_id (*relative_id*)

Retrieve a project by relative ID.

Parameters **relative_id** (*str*) – The resource ID. E.g. "projects/1"

Returns The specified project.

Return type *Project*

by_external_id (*external_id*)

Retrieve a project by external ID.

Parameters **external_id** (*str*) – The external ID.

Returns The specified project, if found.

Return type *Project*

Raises

- **KeyError** – If no project with the specified `external_id` is found
- **LookupError** – If multiple projects with the specified `external_id` are found

stream ()

Stream projects in this collection. Implicitly called when iterating over this collection.

Returns Stream of projects.

Return type Python generator yielding *Project*

Usage:

```
>>> for project in collection.stream(): # explicit
>>>     do_stuff(project)
>>> for project in collection: # implicit
>>>     do_stuff(project)
```

by_name (*project_name*)

Get project by name

Fetches a specific project in this collection by exact-match on name.

Parameters **project_name** (*str*) – Name of the desired project.

Raises **KeyError** – If no project with specified name was found.

Return type *Project*

create (*creation_spec*)

Create a Project in Tamr

Parameters `creation_spec` (*dict[str, str]*) – Project creation specification should be formatted as specified in the [Public Docs for Creating a Project](#).

Returns The created Project

Return type *Project*

delete_by_resource_id (*resource_id*)

Deletes a resource from this collection by resource ID.

Parameters `resource_id` (*str*) – The resource ID of the resource that will be deleted.

Returns HTTP response from the server.

Return type *requests.Response*

Project Step

class `tamr_unify_client.project.step.ProjectStep` (*client, data*)

A step of a Tamr project. This is not a *BaseResource* because it has no API path and cannot be directly retrieved or modified.

See <https://docs.tamr.com/reference#retrieve-downstream-dataset-usage>

Parameters

- **client** (*Client*) – Delegate underlying API calls to this client.
- **data** (*dict*) – The JSON body containing project step information.

property `project_step_id`

Type *str*

property `project_step_name`

Type *str*

property `project_name`

Type *str*

property `type`

A Tamr project type, listed in <https://docs.tamr.com/reference#create-a-project>.

Type *str*

property `project` ()

Retrieves the *Project* this step is associated with.

Returns This step's project.

Return type *Project*

Raises

- **KeyError** – If no project with the specified name is found.
- **LookupError** – If multiple projects with the specified name are found.

5.1 BETA

WARNING: Do not rely on BETA features in production workflows. Support from Tamr may be limited.

5.1.1 Reference

Attributes

Attribute

class `tamr_client.Attribute` (*url, name, type, is_nullable, _json, description=None*)

A Tamr Attribute.

See <https://docs.tamr.com/reference#attribute-types>

Parameters

- **url** (URL) –
- **name** (str) –
- **type** (Union[Boolean, Double, Int, Long, String, Array, Map, Record]) –
- **description** (Optional[str]) –

`tamr_client.attribute.from_resource_id` (*session, dataset, id*)

Get attribute by resource ID

Fetches attribute from Tamr server

Parameters

- **dataset** (Dataset) – Dataset containing this attribute
- **id** (str) – Attribute ID

Raises

- **AttributeNotFound** – If no attribute could be found at the specified URL. Corresponds to a 404 HTTP error.
- **requests.HTTPError** – If any other HTTP error is encountered.

Return type Attribute

`tamr_client.attribute.from_dataset_all` (*session, dataset*)

Get all attributes from a dataset

Parameters `dataset` (Dataset) – Dataset containing the desired attributes

Return type `Tuple[Attribute, ...]`

Returns The attributes for the specified dataset

Raises `requests.HTTPError` – If an HTTP error is encountered.

`tamr_client.attribute.to_json(attr)`
Serialize attribute into JSON

Parameters `attr` (Attribute) – Attribute to serialize

Return type `Dict[str, Any]`

Returns JSON data representing the attribute

`tamr_client.attribute.create(session, dataset, *, name, is_nullable, type=Array(inner_type=String()), description=None)`

Create an attribute

Posts a creation request to the Tamr server

Parameters

- **dataset** (Dataset) – Dataset that should contain the new attribute
- **name** (str) – Name for the new attribute
- **type** (Union[Boolean, Double, Int, Long, String, Array, Map, Record]) – Attribute type for the new attribute
- **is_nullable** (bool) – Determines if the new attribute can contain NULL values
- **description** (Optional[str]) – Description of the new attribute
- **force** – If *True*, skips reserved attribute name check

Return type `Attribute`

Returns The newly created attribute

Raises

- **ReservedAttributeName** – If attribute name is reserved.
- **AttributeExists** – If an attribute already exists at the specified URL. Corresponds to a 409 HTTP error.
- **requests.HTTPError** – If any other HTTP error is encountered.

`tamr_client.attribute.update(session, attribute, *, description=None)`
Update an existing attribute

PUTS an update request to the Tamr server

Parameters

- **attribute** (Attribute) – Existing attribute to update
- **description** (Optional[str]) – Updated description for the existing attribute

Return type `Attribute`

Returns The newly updated attribute

Raises

- **AttributeNotFound** – If no attribute could be found at the specified URL. Corresponds to a 404 HTTP error.

- `requests.HTTPError` – If any other HTTP error is encountered.

`tamr_client.attribute.delete` (*session*, *attribute*)

Deletes an existing attribute

Sends a deletion request to the Tamr server

Parameters `attribute` (`Attribute`) – Existing attribute to delete

Raises

- `AttributeNotFound` – If no attribute could be found at the specified URL. Corresponds to a 404 HTTP error.
- `requests.HTTPError` – If any other HTTP error is encountered.

Exceptions

class `tamr_client.ReservedAttributeName`

Raised when attempting to create an attribute with a reserved name

class `tamr_client.AttributeExists`

Raised when trying to create an attribute that already exists on the server

class `tamr_client.AttributeNotFound`

Raised when referencing (e.g. updating or deleting) an attribute that does not exist on the server.

AttributeType

See <https://docs.tamr.com/reference#attribute-types>

`tamr_client.attribute_type.BOOLEAN` = `Boolean()`

`tamr_client.attribute_type.DOUBLE` = `Double()`

`tamr_client.attribute_type.INT` = `Int()`

`tamr_client.attribute_type.LONG` = `Long()`

`tamr_client.attribute_type.STRING` = `String()`

class `tamr_client.attribute_type.Array` (*inner_type*)

Parameters `inner_type` (`AttributeType`) –

class `tamr_client.attribute_type.Map` (*inner_type*)

Parameters `inner_type` (`AttributeType`) –

class `tamr_client.attribute_type.Record` (*attributes*)

Parameters `attributes` (`Tuple [SubAttribute]`) –

`tamr_client.attribute_type.from_json` (*data*)

Make an attribute type from JSON data (deserialize)

Parameters `data` (`Dict[str, Any]`) – JSON data from Tamr server

Return type `Union[Boolean, Double, Int, Long, String, Array, Map, Record]`

`tamr_client.attribute_type.to_json` (*attr_type*)

Serialize attribute type to JSON

Parameters `attr_type` (`Union[Boolean, Double, Int, Long, String, Array, Map, Record]`) – Attribute type to serialize

Return type `Dict[str, Any]`

Type aliases

```
tamr_client.attributes.type_alias.DEFAULT = Array(inner_type=String())
```

```
tamr_client.attributes.type_alias.GEOSPATIAL = Record(attributes=(SubAttribute(name='point
```

SubAttribute

```
class tamr_client.SubAttribute(name, type, is_nullable, description=None)
```

Parameters

- **name** (`str`) –
- **type** (`AttributeType`) –
- **is_nullable** (`bool`) –
- **description** (`Optional[str]`) –

```
tamr_client.subattribute.from_json(data)
```

Make a SubAttribute from JSON data (deserialize)

Parameters `data` (`Dict[str, Any]`) – JSON data received from Tamr server.

Return type `SubAttribute`

```
tamr_client.subattribute.to_json(subattr)
```

Serialize subattribute into JSON

Parameters `subattr` (`SubAttribute`) – SubAttribute to serialize

Return type `Dict[str, Any]`

Auth

```
class tamr_client.UsernamePasswordAuth(username, password)
```

Provides username/password authentication for Tamr.

Sets the *Authorization* HTTP header with Tamr’s custom *BasicCreds* format.

Parameters

- **username** (`str`) –
- **password** (`str`) –

Example

```
>>> import tamr_client as tc
>>> auth = tc.UsernamePasswordAuth('my username', 'my password')
>>> s = tc.Session(auth)
```

Datasets

Dataset

class `tamr_client.Dataset` (*url, name, key_attribute_names, description=None*)

A Tamr dataset

See <https://docs.tamr.com/reference/dataset-models>

Parameters

- **url** (URL) –
- **key_attribute_names** (Tuple[str,...]) –

`tamr_client.dataset.from_resource_id` (*session, instance, id*)

Get dataset by resource ID

Fetches dataset from Tamr server

Parameters

- **instance** (Instance) – Tamr instance containing this dataset
- **id** (str) – Dataset ID

Raises

- **DatasetNotFound** – If no dataset could be found at the specified URL. Corresponds to a 404 HTTP error.
- **requests.HTTPError** – If any other HTTP error is encountered.

Return type Dataset

Exceptions

class `tamr_client.DatasetNotFound`

Raised when referencing (e.g. updating or deleting) a dataset that does not exist on the server.

Record

See <https://docs.tamr.com/reference/record> “The recommended approach for interacting with records is to use the `upsert()` and `delete()` functions for all use cases they can handle. For more advanced use cases, the underlying `_update` function can be used directly.”

`tamr_client.record.upsert` (*session, dataset, records, *, primary_key_name=None*)

Create or update the specified records.

Parameters

- **dataset** (Dataset) – Dataset to receive record updates
- **records** (Iterable[Dict[~KT, ~VT]]) – The records to update, as dictionaries
- **primary_key_name** (Optional[str]) – The primary key for these records, which must be a key in each record dictionary. By default the `key_attribute_name` of dataset

Return type Dict[str, Any]

Returns JSON response body from server

Raises

- `requests.HTTPError` – If an HTTP error is encountered
- `PrimaryKeyNotFound` – If `primary_key_name` does not match dataset primary key
- `PrimaryKeyNotFound` – If `primary_key_name` not in a record dictionary

`tamr_client.record.delete` (*session, dataset, records, *, primary_key_name=None*)

Deletes the specified records, based on primary key values. Does not check that other attribute values match.

Parameters

- `dataset` (`Dataset`) – Dataset from which to delete records
- `records` (`Iterable[Dict[~KT, ~VT]]`) – The records to update, as dictionaries
- `primary_key_name` (`Optional[str]`) – The primary key for these records, which must be a key in each record dictionary. By default the `key_attribute_name` of dataset

Return type `Dict[str, Any]`

Returns JSON response body from server

Raises

- `requests.HTTPError` – If an HTTP error is encountered
- `PrimaryKeyNotFound` – If `primary_key_name` does not match dataset primary key
- `PrimaryKeyNotFound` – If `primary_key_name` not in a record dictionary

`tamr_client.record._update` (*session, dataset, updates*)

Send a batch of record creations/updates/deletions to this dataset. You probably want to use `upsert()` or `delete()` instead.

Parameters

- `dataset` (`Dataset`) – Dataset containing records to be updated
- `updates` (`Iterable[Dict[~KT, ~VT]]`) – Each update should be formatted as specified in the [Public Docs for Dataset updates](#).

Return type `Dict[str, Any]`

Returns JSON response body from server

Raises `requests.HTTPError` – If an HTTP error is encountered

Exceptions

`class tamr_client.PrimaryKeyNotFound`

Raised when referencing a primary key by name that does not exist.

Dataframe

`tamr_client.dataframe.upsert` (*session, dataset, df, *, primary_key_name=None*)

Upserts a record for each row of `df` with attributes for each column in `df`.

Parameters

- `dataset` (`Dataset`) – Dataset to receive record updates
- `df` (`DataFrame`) – The `DataFrame` containing records to be upserted

- **primary_key_name** (`Optional[str]`) – The primary key of the dataset. Must be a column of *df*. By default the `key_attribute_name` of dataset

Return type `Dict[str, Any]`

Returns JSON response body from the server

Raises

- `requests.HTTPError` – If an HTTP error is encountered
- `PrimaryKeyNotFound` – If *primary_key_name* is not a column in *df* or the index of *df*
- `ValueError` – If *primary_key_name* matches both a column in *df* and the index of *df*

Instance

`class tamr_client.Instance (protocol='http', host='localhost', port=9100)`

`tamr_client.instance.origin (instance)`

HTTP origin i.e. `<protocol>://<host>[:<port>]`.

For additional information, see [MDN web docs](#) .

Return type `str`

Response

Utilities for working with `requests.Response` .

`tamr_client.response.successful (response)`

Ensure response does not contain an HTTP error.

Delegates to `requests.Response.raise_for_status()`

Return type `Response`

Returns The response being checked.

Raises `requests.exceptions.HTTPError` – If an HTTP error is encountered.

`tamr_client.response.ndjson (response, **kwargs)`

Stream newline-delimited JSON from the response body

Analog to `requests.Response.json()` but for `.ndjson`-formatted body.

Recommended: For memory efficiency, use `stream=True` when sending the request corresponding to this response.

Parameters

- **response** (`Response`) – Response whose body should be streamed as newline-delimited JSON.
- ****kwargs** – Keyword arguments passed to underlying `requests.Response.iter_lines()` call.

Returns Each line of the response body, parsed as JSON

Example

```
>>> import tamr_client as tc
>>> s = tc.session.from_auth(...)
>>> r = s.get(..., stream=True)
>>> for data in tc.response.ndjson(r):
...     assert data['my key'] == 'my_value'
```

Return type `Iterator[Dict[str, Any]]`

Session

The `Session` type is an alias for `requests.Session`.

For more information, see the official `requests.Session` docs.

`tamr_client.session.from_auth(auth)`

Create a new authenticated session

Parameters `auth` (`HTTPBasicAuth`) – Authentication

Return type `Session`

PYTHON MODULE INDEX

t

`tamr_client.record`, 75

INDEX

Symbols

`_update()` (in module `tamr_client.record`), 76

A

`add_input_dataset()`

(`tamr_unify_client.categorization.project.CategorizationProject` method), 31

`add_input_dataset()`

(`tamr_unify_client.mastering.project.MasteringProject` method), 52

`add_input_dataset()`

(`tamr_unify_client.project.resource.Project` method), 66

`apply_options()` (`tamr_unify_client.operation.Operation`

method), 57

`as_categorization()`

(`tamr_unify_client.categorization.project.CategorizationProject` method), 31

`as_categorization()`

(`tamr_unify_client.mastering.project.MasteringProject` method), 52

`as_categorization()`

(`tamr_unify_client.project.resource.Project` method), 66

`as_mastering()` (`tamr_unify_client.categorization.project.CategorizationProject`

method), 31

`as_mastering()` (`tamr_unify_client.mastering.project.MasteringProject`

method), 52

`as_mastering()` (`tamr_unify_client.project.resource.Project`

method), 66

`Attribute` (class in `tamr_client`), 71

`Attribute` (class in `tamr_unify_client.attribute.resource`), 25

`attribute_configurations()`

(`tamr_unify_client.categorization.project.CategorizationProject` method), 31

`attribute_configurations()`

(`tamr_unify_client.mastering.project.MasteringProject` method), 52

`attribute_configurations()`

(`tamr_unify_client.project.resource.Project` method), 67

`attribute_mappings()`

(`tamr_unify_client.categorization.project.CategorizationProject` method), 31

`attribute_mappings()`

(`tamr_unify_client.mastering.project.MasteringProject` method), 53

`attribute_mappings()`

(`tamr_unify_client.project.resource.Project` method), 67

`attribute_name()` (`tamr_unify_client.project.attribute_configuration`

property), 59

`attribute_profiles()`

(`tamr_unify_client.dataset.profile.DatasetProfile` property), 45

`attribute_role()` (`tamr_unify_client.project.attribute_configuration`

property), 58

`AttributeCollection` (class in

`tamr_unify_client.attribute.collection`), 27

`AttributeConfiguration` (class in

`tamr_unify_client.project.attribute_configuration.resource`),

58

`AttributeConfigurationCollection` (class in

`tamr_unify_client.project.attribute_configuration.collection`),

61

`AttributeConfigurationSpec` (class in

`tamr_unify_client.project.attribute_configuration.resource`),

59

`AttributeExists` (class in `tamr_client`), 73

`AttributeMapping` (class in

`tamr_unify_client.project.attribute_mapping.resource`),

62

`AttributeMappingCollection` (class in

`tamr_unify_client.project.attribute_mapping.collection`),

65

`AttributeMappingSpec` (class in

`tamr_unify_client.project.attribute_mapping.resource`),

63

`AttributeNotFound` (class in `tamr_client`), 73

`attributes()` (`tamr_unify_client.attribute.type.AttributeType`

property), 28

`attributes()` (`tamr_unify_client.categorization.project.Categorization`

property), 31

[attributes\(\) \(tamr_unify_client.dataset.resource.DatasetResource property\), 38](#)
[attributes\(\) \(tamr_unify_client.mastering.project.MasteringProject property\), 53](#)
[attributes\(\) \(tamr_unify_client.project.resource.ProjectResource property\), 66](#)
[AttributeSpec \(class in tamr_unify_client.attribute.resource\), 26](#)
[AttributeType \(class in tamr_unify_client.attribute.type\), 28](#)
[AttributeTypeSpec \(class in tamr_unify_client.attribute.type\), 29](#)
B
[base_type\(\) \(tamr_unify_client.attribute.type.AttributeType property\), 28](#)
[binning_model\(\) \(tamr_unify_client.mastering.project.MasteringProject method\), 52](#)
[BinningModel \(class in tamr_unify_client.mastering.binning_model\), 48](#)
[BOOLEAN \(in module tamr_client.attribute_type\), 73](#)
[bulk_create\(\) \(tamr_unify_client.categorization.category.collection.CategoryCollection method\), 35](#)
[by_external_id\(\) \(tamr_unify_client.attribute.collection.AttributeCollection method\), 27](#)
[by_external_id\(\) \(tamr_unify_client.categorization.category.collection.CategoryCollection method\), 35](#)
[by_external_id\(\) \(tamr_unify_client.dataset.collection.DatasetCollection in tamr_unify_client\), 36](#)
[by_external_id\(\) \(tamr_unify_client.project.attribute_configuration.collection.AttributeConfigurationCollection method\), 61](#)
[by_external_id\(\) \(tamr_unify_client.project.collection.ProjectCollection in tamr_unify_client.attribute.collection.AttributeCollection\), 69](#)
[by_name\(\) \(tamr_unify_client.attribute.collection.AttributeCollection method\), 28](#)
[by_name\(\) \(tamr_unify_client.dataset.collection.DatasetCollection method\), 43](#)
[by_name\(\) \(tamr_unify_client.project.collection.ProjectCollection method\), 69](#)
[by_relative_id\(\) \(tamr_unify_client.attribute.collection.AttributeCollection method\), 27](#)
[by_relative_id\(\) \(tamr_unify_client.categorization.category.collection.CategoryCollection in tamr_unify_client.project.collection.ProjectCollection\), 34](#)
[by_relative_id\(\) \(tamr_unify_client.dataset.collection.DatasetCollection method\), 42](#)
[by_relative_id\(\) \(tamr_unify_client.project.attribute_configuration.collection.AttributeConfigurationCollection method\), 61](#)
[by_relative_id\(\) \(tamr_unify_client.project.attribute_mapping.collection.AttributeMappingCollection method\), 65](#)
[by_relative_id\(\) \(tamr_unify_client.project.collection.ProjectCollection in tamr_unify_client.categorization.project.CategorizationProject\), 69](#)
[by_resource_id\(\) \(tamr_unify_client.attribute.collection.AttributeCollection method\), 27](#)

C
[categories\(\) \(tamr_unify_client.categorization.taxonomy.Taxonomy method\), 36](#)
[CategorizationProject \(class in tamr_unify_client.categorization.project\), 32](#)
[Category \(class in tamr_unify_client.categorization.category.resource\), 32](#)
[CategoryCollection \(class in tamr_unify_client.categorization.category.collection\), 34](#)
[CategoryCollection \(class in tamr_unify_client.categorization.category.resource\), 32](#)
[Collection \(class in tamr_unify_client.attribute.collection\), 27](#)
[Collection \(class in tamr_unify_client.categorization.category.collection\), 34](#)
[Collection \(class in tamr_unify_client.dataset.collection\), 36](#)
[Collection \(class in tamr_unify_client.project.attribute_configuration.collection\), 61](#)
[Collection \(class in tamr_unify_client.project.collection\), 69](#)
[Collection \(class in tamr_unify_client.attribute.collection\), 27](#)
[Collection \(class in tamr_unify_client.categorization.category.collection\), 34](#)
[Collection \(class in tamr_unify_client.dataset.collection\), 42](#)
[Collection \(class in tamr_unify_client.project.attribute_configuration.collection\), 61](#)
[Collection \(class in tamr_unify_client.project.attribute_mapping.collection\), 65](#)
[Collection \(class in tamr_unify_client.project.collection\), 69](#)
[Collection \(class in tamr_unify_client.attribute.collection\), 27](#)

D

- Dataset (class in *tamr_client*), 75
- Dataset (class in *tamr_unify_client.dataset.resource*), 37
- dataset () (*tamr_unify_client.dataset.uri.DatasetURI* method), 46
- dataset () (*tamr_unify_client.dataset.use.DatasetUse* method), 47
- dataset_id () (*tamr_unify_client.dataset.use.DatasetUse* property), 47
- dataset_name () (*tamr_unify_client.dataset.profile.DatasetProfile* property), 44
- dataset_name () (*tamr_unify_client.dataset.status.DatasetStatus* property), 45
- dataset_name () (*tamr_unify_client.dataset.use.DatasetUse* property), 47
- DatasetCollection (class in *tamr_unify_client.dataset.collection*), 42
- DatasetNotFound (class in *tamr_client*), 75
- DatasetProfile (class in *tamr_unify_client.dataset.profile*), 44
- datasets () (*tamr_unify_client.Client* property), 37
- DatasetSpec (class in *tamr_unify_client.dataset.resource*), 41
- DatasetStatus (class in *tamr_unify_client.dataset.status*), 45
- DatasetURI (class in *tamr_unify_client.dataset.uri*), 46
- DatasetUsage (class in *tamr_unify_client.dataset.usage*), 47
- DatasetUse (class in *tamr_unify_client.dataset.use*), 47
- DEFAULT (in module *tamr_client.attributes.type_alias*), 74
- delete () (in module *tamr_client.attribute*), 73
- delete () (in module *tamr_client.record*), 76
- delete () (*tamr_unify_client.attribute.resource.Attribute* method), 25
- delete () (*tamr_unify_client.base_model.MachineLearningModel* method), 48
- delete () (*tamr_unify_client.categorization.category.resource.Category* method), 33
- delete () (*tamr_unify_client.categorization.project.CategorizationProject* method), 31
- delete () (*tamr_unify_client.categorization.taxonomy.Taxonomy* method), 36
- delete () (*tamr_unify_client.Client* method), 37
- delete () (*tamr_unify_client.dataset.profile.DatasetProfile* method), 45
- delete () (*tamr_unify_client.dataset.resource.Dataset* method), 40
- delete () (*tamr_unify_client.dataset.status.DatasetStatus* method), 46
- delete () (*tamr_unify_client.dataset.usage.DatasetUsage* method), 47
- delete () (*tamr_unify_client.mastering.binning_model.BinningModel* method), 48
- delete () (*tamr_unify_client.mastering.estimated_pair_counts.EstimatedPairCounts* method), 50
- delete () (*tamr_unify_client.mastering.project.MasteringProject* method), 53
- delete () (*tamr_unify_client.mastering.published_cluster.configuration.PublishedClusterConfiguration* method), 55
- delete () (*tamr_unify_client.operation.Operation* method), 58
- delete () (*tamr_unify_client.project.attribute_configuration.resource.AttributeConfiguration* method), 59
- delete () (*tamr_unify_client.project.attribute_mapping.resource.AttributeMapping* method), 63
- delete () (*tamr_unify_client.project.resource.Project* method), 67
- delete_all_records () (*tamr_unify_client.dataset.resource.Dataset* method), 39
- delete_by_resource_id () (*tamr_unify_client.attribute.collection.AttributeCollection* method), 28
- delete_by_resource_id () (*tamr_unify_client.categorization.category.collection.CategoryCollection* method), 35
- delete_by_resource_id () (*tamr_unify_client.dataset.collection.DatasetCollection* method), 43
- delete_by_resource_id () (*tamr_unify_client.project.attribute_configuration.collection.AttributeConfigurationCollection* method), 62
- delete_by_resource_id () (*tamr_unify_client.project.attribute_mapping.collection.AttributeMappingCollection* method), 65
- delete_by_resource_id () (*tamr_unify_client.project.collection.ProjectCollection* method), 70
- delete_records () (*tamr_unify_client.dataset.resource.Dataset* method), 38
- delete_records_by_id () (*tamr_unify_client.dataset.resource.Dataset* method), 39
- dependencies () (*tamr_unify_client.dataset.usage.DatasetUsage* property), 47
- description () (*tamr_unify_client.attribute.resource.Attribute* property), 25
- description () (*tamr_unify_client.categorization.category.resource.Category* property), 32
- description () (*tamr_unify_client.categorization.project.CategorizationProject* property), 31
- description () (*tamr_unify_client.dataset.resource.Dataset* property), 38
- description () (*tamr_unify_client.mastering.project.MasteringProject* property), 53

property), 53

description() (*tamr_unify_client.operation.Operation* **GEOSPATIAL** (in module *tamr_client.attributes.type_alias*), 74
property), 57

description() (*tamr_unify_client.project.resource.Project* **get** () (*tamr_unify_client.Client* method), 37
property), 66

DOUBLE (in module *tamr_client.attribute_type*), 73

E

enabled_for_ml() (*tamr_unify_client.project.attribute_configuration.resource.AttributeConfiguration* **method**), 59
property), 59

entity_id() (*tamr_unify_client.mastering.published_cluster.record.RecordPublishedCluster* **id** () (*tamr_unify_client.mastering.published_cluster.resource.PublishedCluster* **property**), 55
property), 55

estimate_pairs() (*tamr_unify_client.mastering.project.MasteringProject* **method**), 52
property), 54

EstimatedPairCounts (class in *tamr_unify_client.mastering.estimated_pair_counts*), 49
id () (*tamr_unify_client.project.attribute_configuration.resource.AttributeConfiguration* **property**), 58
id () (*tamr_unify_client.project.attribute_mapping.resource.AttributeMapping* **property**), 62

external_id() (*tamr_unify_client.categorization.project.CategorizationProject* **inner_type** () (*tamr_unify_client.attribute.type.AttributeType* **property**), 28
property), 32

external_id() (*tamr_unify_client.dataset.resource.Dataset* **input_attribute_id** () (*tamr_unify_client.project.attribute_mapping.resource.AttributeMapping* **property**), 62
property), 38

external_id() (*tamr_unify_client.mastering.project.MasteringProject* **input_attribute_name** () (*tamr_unify_client.project.attribute_mapping.resource.AttributeMapping* **property**), 62
property), 53

external_id() (*tamr_unify_client.project.resource.Project* **input_dataset_name** () (*tamr_unify_client.project.attribute_mapping.resource.AttributeMapping* **property**), 62
property), 65

F

from_auth() (in module *tamr_client.session*), 78
property), 62

from_data() (*tamr_unify_client.attribute.resource.AttributeSpec* **input_datasets** () (*tamr_unify_client.categorization.project.CategorizationProject* **method**), 32
method), 26

from_data() (*tamr_unify_client.categorization.category.resource.CategorySpec* **input_datasets** () (*tamr_unify_client.mastering.project.MasteringProject* **method**), 53
method), 33

from_data() (*tamr_unify_client.dataset.resource.DatasetSpec* **input_datasets** () (*tamr_unify_client.project.resource.Project* **method**), 66
method), 41

from_data() (*tamr_unify_client.project.attribute_configuration.resource.AttributeConfigurationSpec* **input_to_project_steps** () (*tamr_unify_client.dataset.use.DatasetUse* **property**), 47
method), 59

from_data() (*tamr_unify_client.project.resource.ProjectSpec* **Instance** (class in *tamr_client*), 77
method), 67

from_dataset_all() (in module *tamr_client.attribute*), 71
INT (in module *tamr_client.attribute_type*), 73

from_geo_features() (*tamr_unify_client.dataset.resource.Dataset* **is_nullable** () (*tamr_unify_client.attribute.resource.Attribute* **property**), 25
method), 40

from_json() (in module *tamr_client.attribute_type*), 73
is_streamable () (*tamr_unify_client.dataset.status.DatasetStatus* **property**), 46

from_json() (in module *tamr_client.subattribute*), 74
is_up_to_date () (*tamr_unify_client.dataset.profile.DatasetProfile* **property**), 44

from_json() (*tamr_unify_client.attribute.subattribute.SubAttribute* **is_up_to_date** () (*tamr_unify_client.mastering.estimated_pair_counts* **property**), 49
static method), 30

from_resource_id() (in module *tamr_client.attribute*), 71
itergeofeatures () (*tamr_unify_client.dataset.resource.Dataset* **method**), 40

from_resource_id() (in module *tamr_client.dataset*), 75

K

from_response() (*tamr_unify_client.operation.Operation* **key_attribute_names** ()
class method), 56

(*tamr_unify_client.dataset.resource.Dataset* property), 38

L

LONG (in module *tamr_client.attribute_type*), 73

M

MachineLearningModel (class in *tamr_unify_client.base_model*), 48

MasteringProject (class in *tamr_unify_client.mastering.project*), 50

Metric (class in *tamr_unify_client.mastering.published_cluster.metric*), 54

metrics() (*tamr_unify_client.mastering.published_cluster.version.PublishedClusterVersion* property), 55

model() (*tamr_unify_client.categorization.project.CategorizationProject* method), 30

N

name() (*tamr_unify_client.attribute.resource.Attribute* property), 25

name() (*tamr_unify_client.categorization.category.resource.Category* property), 32

name() (*tamr_unify_client.categorization.project.CategorizationProject* property), 32

name() (*tamr_unify_client.categorization.taxonomy.Taxonomy* property), 36

name() (*tamr_unify_client.dataset.resource.Dataset* property), 37

name() (*tamr_unify_client.mastering.project.MasteringProject* property), 53

name() (*tamr_unify_client.mastering.published_cluster.metric.Metric* property), 54

name() (*tamr_unify_client.mastering.published_cluster.version.PublishedClusterVersion* property), 55

name() (*tamr_unify_client.project.resource.Project* property), 65

ndjson() (in module *tamr_client.response*), 77

new() (*tamr_unify_client.attribute.resource.AttributeSpec* static method), 26

new() (*tamr_unify_client.attribute.type.AttributeTypeSpec* static method), 29

new() (*tamr_unify_client.categorization.category.resource.CategorySpec* static method), 33

new() (*tamr_unify_client.dataset.resource.DatasetSpec* static method), 41

new() (*tamr_unify_client.project.attribute_configuration.resource.AttributeConfigurationSpec* static method), 59

new() (*tamr_unify_client.project.attribute_mapping.resource.AttributeMappingSpec* static method), 63

new() (*tamr_unify_client.project.resource.ProjectSpec* static method), 67

numeric_field_resolution() (*tamr_unify_client.project.attribute_configuration.resource.AttributeConfigurationSpec* property), 59

O

of() (*tamr_unify_client.attribute.resource.AttributeSpec* static method), 26

of() (*tamr_unify_client.attribute.type.AttributeTypeSpec* static method), 29

of() (*tamr_unify_client.categorization.category.resource.CategorySpec* static method), 33

of() (*tamr_unify_client.dataset.resource.DatasetSpec* static method), 41

of() (*tamr_unify_client.project.attribute_configuration.resource.AttributeConfigurationSpec* static method), 59

of() (*tamr_unify_client.project.attribute_mapping.resource.AttributeMappingSpec* static method), 63

of() (*tamr_unify_client.project.resource.ProjectSpec* static method), 67

Operation (class in *tamr_unify_client.operation*), 56

origin() (in module *tamr_client.instance*), 77

origin() (*tamr_unify_client.Client* property), 37

origin_entity_id()

(*tamr_unify_client.mastering.published_cluster.record.RecordPublishedClusterRecord* property), 56

origin_source_id()

(*tamr_unify_client.mastering.published_cluster.record.RecordPublishedClusterRecord* property), 56

output_from_project_steps()

(*tamr_unify_client.dataset.use.DatasetUse* property), 47

P

parent() (*tamr_unify_client.categorization.category.resource.Category* property), 33

parent() (*tamr_unify_client.mastering.project.MasteringProject* method), 50

parent() (*tamr_unify_client.mastering.project.MasteringProject* method), 50

parent() (*tamr_unify_client.categorization.category.resource.Category* method), 33

path() (*tamr_unify_client.categorization.category.resource.Category* property), 33

path() (*tamr_unify_client.operation.Operation* method), 57

post() (*tamr_unify_client.Client* method), 37

predict() (*tamr_unify_client.base_model.MachineLearningModel* method), 48

PrimaryKeyNotFound (class in *tamr_client*), 76

profiled_at() (*tamr_unify_client.dataset.resource.Dataset* method), 39

profiled_at() (*tamr_unify_client.dataset.profile.DatasetProfile* property), 45

profiled_data_version()

(*tamr_unify_client.dataset.profile.DatasetProfile* property), 45

property), 44
 Project (class in *tamr_unify_client.project.resource*), 65
 project () (*tamr_unify_client.project.step.ProjectStep* method), 70
 project_name () (*tamr_unify_client.project.step.ProjectStep* property), 70
 project_step_id () (*tamr_unify_client.project.step.ProjectStep* property), 70
 project_step_name () (*tamr_unify_client.project.step.ProjectStep* property), 70
 ProjectCollection (class in *tamr_unify_client.project.collection*), 68
 projects () (*tamr_unify_client.Client* property), 37
 ProjectSpec (class in *tamr_unify_client.project.resource*), 67
 ProjectStep (class in *tamr_unify_client.project.step*), 70
 published_cluster_ids () (*tamr_unify_client.mastering.project.MasteringProject* method), 51
 published_cluster_stats () (*tamr_unify_client.mastering.project.MasteringProject* method), 51
 published_cluster_versions () (*tamr_unify_client.mastering.project.MasteringProject* method), 51
 published_clusters () (*tamr_unify_client.mastering.project.MasteringProject* method), 51
 published_clusters_configuration () (*tamr_unify_client.mastering.project.MasteringProject* method), 51
 published_clusters_with_data () (*tamr_unify_client.mastering.project.MasteringProject* method), 52
 PublishedCluster (class in *tamr_unify_client.mastering.published_cluster.resource*), 54
 PublishedClustersConfiguration (class in *tamr_unify_client.mastering.published_cluster.configuration*), 54
 PublishedClusterVersion (class in *tamr_unify_client.mastering.published_cluster.version*), 55
 put () (*tamr_unify_client.attribute.resource.AttributeSpec* method), 27
 put () (*tamr_unify_client.Client* method), 37
 put () (*tamr_unify_client.dataset.resource.DatasetSpec* method), 42
 put () (*tamr_unify_client.project.attribute_configuration.resource.AttributeConfigurationSpec* method), 60
 put () (*tamr_unify_client.project.resource.ProjectSpec* method), 68
R
 record_clusters () (*tamr_unify_client.mastering.project.MasteringProject* method), 51
 record_clusters_with_data () (*tamr_unify_client.mastering.project.MasteringProject* method), 52
 record_ids () (*tamr_unify_client.mastering.published_cluster.version.F* property), 55
 record_published_cluster_versions () (*tamr_unify_client.mastering.project.MasteringProject* method), 51
 RecordPublishedCluster (class in *tamr_unify_client.mastering.published_cluster.record*), 55
 RecordPublishedClusterVersion (class in *tamr_unify_client.mastering.published_cluster.record_version*), 56
 records () (*tamr_unify_client.dataset.resource.Dataset* method), 39
 records () (*tamr_unify_client.mastering.binning_model.BinningModel* method), 48
 refresh () (*tamr_unify_client.dataset.profile.DatasetProfile* method), 45
 refresh () (*tamr_unify_client.dataset.resource.Dataset* method), 39
 refresh () (*tamr_unify_client.mastering.estimated_pair_counts.Estimate* method), 50
 relative_attribute_id () (*tamr_unify_client.project.attribute_configuration.resource.Attribute* property), 58
 relative_dataset_id () (*tamr_unify_client.dataset.profile.DatasetProfile* property), 44
 relative_dataset_id () (*tamr_unify_client.dataset.status.DatasetStatus* property), 46
 relative_id () (*tamr_unify_client.attribute.resource.Attribute* property), 25
 relative_id () (*tamr_unify_client.base_model.MachineLearningMode* property), 48
 relative_id () (*tamr_unify_client.categorization.category.resource.Ca* property), 33
 relative_id () (*tamr_unify_client.categorization.project.Categorization* property), 32
 relative_id () (*tamr_unify_client.categorization.taxonomy.Taxonomy* property), 36
 relative_id () (*tamr_unify_client.dataset.profile.DatasetProfile* property), 45
 relative_id () (*tamr_unify_client.dataset.resource.Dataset* property), 41

relative_id() (*tamr_unify_client.dataset.status.DatasetStatus* property), 46
 property), 46 resource_id() (*tamr_unify_client.dataset.uri.DatasetURI*
 relative_id() (*tamr_unify_client.dataset.uri.DatasetURI* property), 46
 property), 46 resource_id() (*tamr_unify_client.dataset.usage.DatasetUsage*
 relative_id() (*tamr_unify_client.dataset.usage.DatasetUsage* property), 47
 property), 47 resource_id() (*tamr_unify_client.mastering.binning_model.BinningModel*
 relative_id() (*tamr_unify_client.mastering.binning_model.BinningModel*), 49
 property), 49 resource_id() (*tamr_unify_client.mastering.estimated_pair_counts.EstimatedPairCounts*
 relative_id() (*tamr_unify_client.mastering.estimated_pair_counts.EstimatedPairCounts*
 property), 50 resource_id() (*tamr_unify_client.mastering.project.MasteringProject*
 relative_id() (*tamr_unify_client.mastering.project.MasteringProject* property), 53
 property), 53 resource_id() (*tamr_unify_client.mastering.published_cluster.configuration.PublishedClustersConfiguration*
 relative_id() (*tamr_unify_client.mastering.published_cluster.configuration.PublishedClustersConfiguration*
 property), 54 resource_id() (*tamr_unify_client.operation.Operation*
 relative_id() (*tamr_unify_client.operation.Operation* property), 58
 property), 58 resource_id() (*tamr_unify_client.project.attribute_configuration.ResourceAttributeConfiguration*
 relative_id() (*tamr_unify_client.project.attribute_configuration.ResourceAttributeConfiguration*
 property), 58 resource_id() (*tamr_unify_client.project.attribute_mapping.ResourceAttributeMapping*
 relative_id() (*tamr_unify_client.project.attribute_mapping.ResourceAttributeMapping*
 property), 62 resource_id() (*tamr_unify_client.project.resource.Project*
 relative_id() (*tamr_unify_client.project.resource.Project* property), 67
 property), 67

S

relative_input_attribute_id() (*tamr_unify_client.project.attribute_mapping.ResourceAttributeMapping*
 property), 62 unifiable_attribute_name() (*tamr_unify_client.project.attribute_configuration.ResourceAttributeConfiguration*
 relative_unified_attribute_id() (*tamr_unify_client.project.attribute_mapping.ResourceAttributeMapping*
 property), 62 property), 59
 property), 62 simple_usage() (*tamr_unify_client.dataset.profile.DatasetProfile*
 remove_input_dataset() (*tamr_unify_client.categorization.project.CategorizationProject*
 method), 32 source_id() (*tamr_unify_client.mastering.published_cluster.record.Record*
 remove_input_dataset() (*tamr_unify_client.mastering.project.MasteringProject*
 method), 53 spec() (*tamr_unify_client.attribute.resource.Attribute*
 remove_input_dataset() (*tamr_unify_client.project.resource.Project*
 method), 53 spec() (*tamr_unify_client.attribute.type.AttributeType*
 method), 28
 spec() (*tamr_unify_client.categorization.category.resource.Category*
 method), 33
 spec() (*tamr_unify_client.categorization.project.CategorizationProject*
 method), 32
 spec() (*tamr_unify_client.dataset.resource.Dataset*
 resource_id() (*tamr_unify_client.attribute.resource.Attribute* method), 40
 property), 25 spec() (*tamr_unify_client.mastering.project.MasteringProject*
 resource_id() (*tamr_unify_client.base_model.MachineLearningModel*), 53
 property), 48 spec() (*tamr_unify_client.mastering.published_cluster.configuration.PublishedClustersConfiguration*
 resource_id() (*tamr_unify_client.categorization.category.resource.Category*
 method), 53
 property), 33 spec() (*tamr_unify_client.project.attribute_configuration.ResourceAttributeConfiguration*
 resource_id() (*tamr_unify_client.categorization.project.CategorizationProject*
 method), 32
 property), 32 spec() (*tamr_unify_client.project.attribute_mapping.ResourceAttributeMapping*
 resource_id() (*tamr_unify_client.categorization.taxonomy.Taxonomy*
 method), 63
 property), 36 spec() (*tamr_unify_client.project.resource.Project*
 resource_id() (*tamr_unify_client.dataset.profile.DatasetProfile* method), 67
 property), 45 state() (*tamr_unify_client.operation.Operation* prop-
 resource_id() (*tamr_unify_client.dataset.resource.Dataset* erty), 57
 property), 41 status() (*tamr_unify_client.dataset.resource.Dataset*
 resource_id() (*tamr_unify_client.dataset.status.DatasetStatus* method), 39
 property), 39

stream() (*tamr_unify_client.attribute.collection.AttributeCollection* (in module *tamr_client.attribute_type*), 73
 method), 27 to_json() (in module *tamr_client.subattribute*), 74
 stream() (*tamr_unify_client.categorization.category.collection.CategoryCollection* (in module *tamr_unify_client.project.attribute_configuration.resource*
 method), 35 property), 59
 stream() (*tamr_unify_client.dataset.collection.DatasetCollection* (in module *tamr_unify_client.mastering.estimated_pair_count*
 method), 43 property), 49
 stream() (*tamr_unify_client.project.attribute_configuration.collection.AttributeConfigurationCollection* (in module *tamr_unify_client.machine_learning_model*
 method), 61 method), 48
 stream() (*tamr_unify_client.project.attribute_mapping.collection.AttributeMappingCollection* (in module *tamr_unify_client.attribute.resource.Attribute*
 method), 65 property), 25
 stream() (*tamr_unify_client.project.collection.ProjectCollection* (in module *tamr_unify_client.categorization.project.CategorizationProject*
 method), 69 property), 32
 STRING (in module *tamr_client.attribute_type*), 73 type() (*tamr_unify_client.mastering.project.MasteringProject*
 SubAttribute (class in *tamr_unify_client.attribute.subattribute*), type() (*tamr_unify_client.operation.Operation* prop-
 29 erty), 57
 succeeded() (*tamr_unify_client.operation.Operation* type() (*tamr_unify_client.project.resource.Project*
 method), 58 property), 66
 successful() (in module *tamr_client.response*), 77 type() (*tamr_unify_client.project.step.ProjectStep*
 property), 70

T

tags() (*tamr_unify_client.dataset.resource.Dataset*
 property), 38
 tamr_client.attribute_type.Array (built-in
 class), 73
 tamr_client.attribute_type.Map (built-in
 class), 73
 tamr_client.attribute_type.Record (built-
 in class), 73
 tamr_client.record (module), 75
 tamr_client.SubAttribute (built-in class), 74
 Taxonomy (class in *tamr_unify_client.categorization.taxonomy*),
 36
 taxonomy() (*tamr_unify_client.categorization.project.CategorizationProject*
 method), 31
 timestamp() (*tamr_unify_client.mastering.published_cluster.record.ClusterRecordPublishedClusterVersion*
 property), 56
 timestamp() (*tamr_unify_client.mastering.published_cluster.version.ClusterVersion*
 property), 55
 to_dict() (*tamr_unify_client.attribute.resource.AttributeSpec*
 method), 26
 to_dict() (*tamr_unify_client.attribute.type.AttributeTypeSpec*
 method), 29
 to_dict() (*tamr_unify_client.categorization.category.resource.CategorySpec* (in module *tamr_client.dataframe*), 76
 method), 34
 to_dict() (*tamr_unify_client.dataset.resource.DatasetSpec*
 method), 41
 to_dict() (*tamr_unify_client.project.attribute_configuration.resource.AttributeConfigurationSpec*
 method), 60
 to_dict() (*tamr_unify_client.project.attribute_mapping.resource.AttributeMappingSpec*
 method), 63
 to_dict() (*tamr_unify_client.project.resource.ProjectSpec*
 method), 67
 to_json() (in module *tamr_client.attribute*), 72

U

unified_attribute_id() (*tamr_unify_client.project.attribute_mapping.resource.AttributeMapping*
 property), 62
 unified_attribute_name() (*tamr_unify_client.project.attribute_mapping.resource.AttributeMapping*
 property), 63
 unified_dataset() (*tamr_unify_client.categorization.project.CategorizationProject*
 method), 32
 unified_dataset() (*tamr_unify_client.mastering.project.MasteringProject*
 method), 54
 unified_dataset() (*tamr_unify_client.mastering.published_cluster.version.ClusterVersion*
 method), 66
 unified_dataset() (*tamr_unify_client.project.attribute_mapping.resource.AttributeMapping*
 property), 63
 update() (in module *tamr_client.attribute*), 72
 update_records() (*tamr_unify_client.mastering.binning_model.BinningModel*
 method), 48
 upsert() (in module *tamr_client.record*), 75
 upsert_from_dataframe() (*tamr_unify_client.dataset.resource.Dataset*
 method), 40
 upsert_records() (*tamr_unify_client.dataset.resource.Dataset*
 method), 40
 upstream_datasets() (*tamr_unify_client.dataset.resource.Dataset*
 method), 40

uri() (*tamr_unify_client.dataset.uri.DatasetURI property*), 46

usage() (*tamr_unify_client.dataset.resource.Dataset method*), 40

usage() (*tamr_unify_client.dataset.usage.DatasetUsage property*), 47

UsernamePasswordAuth (*class in tamr_client*), 74

UsernamePasswordAuth (*class in tamr_unify_client.auth*), 30

V

value() (*tamr_unify_client.mastering.published_cluster.metric.Metric property*), 54

version() (*tamr_unify_client.dataset.resource.Dataset property*), 38

version() (*tamr_unify_client.mastering.published_cluster.record_version.RecordPublishedClusterVersion property*), 56

version() (*tamr_unify_client.mastering.published_cluster.version.PublishedClusterVersion property*), 55

versions() (*tamr_unify_client.mastering.published_cluster.record_version.RecordPublishedClusterVersion property*), 56

versions() (*tamr_unify_client.mastering.published_cluster.version.PublishedClusterVersion property*), 54

versions_time_to_live() (*tamr_unify_client.mastering.published_cluster.configuration.PublishedClusterConfiguration property*), 55

W

wait() (*tamr_unify_client.operation.Operation method*), 57

with_attribute_name() (*tamr_unify_client.project.attribute_configuration.resource.AttributeConfigurationSpec method*), 60

with_attribute_role() (*tamr_unify_client.project.attribute_configuration.resource.AttributeConfigurationSpec method*), 60

with_attributes() (*tamr_unify_client.attribute.type.AttributeTypeSpec method*), 29

with_base_type() (*tamr_unify_client.attribute.type.AttributeTypeSpec method*), 29

with_description() (*tamr_unify_client.attribute.resource.AttributeSpec method*), 26

with_description() (*tamr_unify_client.categorization.category.resource.CategorySpec method*), 34

with_description() (*tamr_unify_client.dataset.resource.DatasetSpec method*), 42

with_description() (*tamr_unify_client.project.resource.ProjectSpec method*), 68

with_enabled_for_ml() (*tamr_unify_client.project.attribute_configuration.resource.AttributeConfigurationSpec method*), 60

with_external_id() (*tamr_unify_client.dataset.resource.DatasetSpec method*), 41

with_external_id() (*tamr_unify_client.project.resource.ProjectSpec method*), 68

with_inner_type() (*tamr_unify_client.attribute.type.AttributeTypeSpec method*), 29

with_input_attribute_id() (*tamr_unify_client.project.attribute_mapping.resource.AttributeMappingSpec method*), 63

with_input_dataset_name() (*tamr_unify_client.project.attribute_mapping.resource.AttributeMappingSpec method*), 64

with_is_public() (*tamr_unify_client.attribute.resource.AttributeSpec method*), 26

with_key() (*tamr_unify_client.dataset.resource.DatasetSpec method*), 42

with_name() (*tamr_unify_client.attribute.resource.AttributeSpec method*), 26

with_name() (*tamr_unify_client.categorization.category.resource.CategorySpec method*), 34

with_name() (*tamr_unify_client.dataset.resource.DatasetSpec method*), 41

with_name() (*tamr_unify_client.project.resource.ProjectSpec method*), 68

with_numeric_field_resolution() (*tamr_unify_client.project.attribute_configuration.resource.AttributeConfigurationSpec method*), 60

with_path() (*tamr_unify_client.categorization.category.resource.CategorySpec method*), 34

with_relative_input_attribute_id() (*tamr_unify_client.project.attribute_mapping.resource.AttributeMappingSpec method*), 63

with_relative_unified_attribute_id() (*tamr_unify_client.project.attribute_mapping.resource.AttributeMappingSpec method*), 64

with_similarity_function() (*tamr_unify_client.project.attribute_configuration.resource.AttributeConfigurationSpec method*), 60

with_tags() (*tamr_unify_client.dataset.resource.DatasetSpec method*), 42

with_tokenizer() (*tamr_unify_client.project.attribute_configuration.resource.AttributeConfigurationSpec method*), 60

with_traceback() (*tamr_unify_client.dataset.collection.CreationError method*), 60

method), 44
with_type() (*tamr_unify_client.attribute.resource.AttributeSpec*
method), 26
with_type() (*tamr_unify_client.project.resource.ProjectSpec*
method), 68
with_unified_attribute_id()
(*tamr_unify_client.project.attribute_mapping.resource.AttributeMappingSpec*
method), 64
with_unified_attribute_name()
(*tamr_unify_client.project.attribute_mapping.resource.AttributeMappingSpec*
method), 64
with_unified_dataset_name()
(*tamr_unify_client.project.attribute_mapping.resource.AttributeMappingSpec*
method), 64
with_unified_dataset_name()
(*tamr_unify_client.project.resource.ProjectSpec*
method), 68