

Machine Learning

API

Date published: 2020-07-16

Date modified:

CLOUDERA

<https://docs.cloudera.com/>

Legal Notice

© Cloudera Inc. 2024. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

Cloudera Data Science Workbench API v2.....	4
API v2 Usage.....	6

Cloudera Data Science Workbench API v2

Cloudera Data Science Workbench exposes a REST API that you can use to perform operations related to projects, jobs, and runs. You can use API commands to integrate Cloudera Data Science Workbench with third-party workflow tools or to control Cloudera Data Science Workbench from the command line.

API v2 supersedes the existing Jobs API. For more information on the Jobs API, see Jobs API in the Related information section, below.

How to view the API Specification

You can view the comprehensive API specification on the REST API v2 Reference page. See Related information, below, for the link.

You can also obtain the specification of the available API commands directly from Cloudera Data Science Workbench. In a browser, enter the following addresses:

- REST API: <https://<domain name of CML instance>/api/v2/swagger.html>
- Python API: <https://<domain name of CML instance>/api/v2/python.html>

You can also get json formatted output, by specifying `swagger.json`.

Quickstart

API key authentication

To get started, generate an API key. The API key is a randomly generated token that is unique to each user. It must be treated as highly sensitive information because it can be used to start jobs via the API. You need this API key to use in API calls.

1. Sign in to Cloudera Data Science Workbench.
2. In User Settings API Keys , click Create API Key.
3. Copy this API key to the clipboard.

Using `curl` from the command line

To use the `curl` command, it is convenient to store the domain and API key in environmental variables, as shown here:

1. Copy the API key.
2. Open a terminal, and store it to a variable. On unix-based systems:

```
export API_KEY=<paste the API key value here>
```
3. Copy the domain name, which is in the address bar of the browser. On unix-based systems: `export CDSW_DOMAIN=<domain>` (a value like: `ml-xxxx123456.com`).

Using the Python client library

To use the Python API in your own code, first install the Python API client and point it to your cluster.

```
pip3 install https://$CDSW_DOMAIN/api/v2/python.tar.gz
```

Include the following code, and specify the values for `<CDSW_DOMAIN>` and `<API_KEY>` with variables or values for your installation.

```
# In a session:
api_instance = default_client()
# Outside a session:
default_client("https://" + cluster, APIKEY)
```



Note: If you use `default_client()` in a session, no arguments are needed. If you use it outside of a session, you must provide the cluster name and API v2 key.

Then you can use commands in your script, such as a call to list projects:

```
projects = api_instance.list_projects()
```

The API returns objects that store values as attributes. To access the values, use dot notation. Do not use bracket notation as you would with a dictionary. For example:

```
myproj = client.create_project(...)
# This doesn't work:
myproj["id"]

# But this does
myproj.id
```

Check the Python documentation for a full list of the available API commands.

Using the Python client library in the REPL

Here is an example of a stub Python script that contains the environmental variables for your installation. Save it to your computer and run it locally to get a Python prompt for API commands.

demo.py

```
import clap
import argparse

parser = argparse.ArgumentParser(description='Test the generated python package.')
parser.add_argument("-host", type=str, help='The host name of your workspace')
parser.add_argument("-token", type=str, help='Your API key')
args = parser.parse_args()

config = clap.Configuration()
config.host = args.host
client = cmlapi.ApiClient(config)
client.set_default_header("authorization", "Bearer " + args.token)
api = cmlapi.ApiApi(client)
```

Run the script from your command line:

```
python3 -i demo.py -host https://$CDSW_DOMAIN -token $API_KEY
```

This returns a Python prompt with `api` available. You can run `api` calls from the prompt, as follows:

```
>>> api
<cmlapi.api.ApiApi object at 0xlasjoid>
>>> api.api_list_projects()
```

You can specify a search filter, such as the following:

```
api.api_list_projects(searchFilter='demo')

api.api_list_projects(page_size=2)
```

```
api.api_list_projects(page_size=2, page_token='<token value>')
```

Related Information[Starting a Job Run Using the API](#)

API v2 Usage

You can use API v2 to perform actions on Projects, Jobs, Models, and Applications.

See [Getting started with CML API](#) for information on using API v2.

Related Information[Starting a Job Run Using the API](#)[Cloudera Data Science Workbench API v2](#)