

CDP One

Running SQL Queries

Date published: 2022-06-03

Date modified: 2022-08-15

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

About running SQL queries.....	4
Using SSH to access gateway nodes.....	4
Setting up Ranger to run SQL queries.....	5
Running Hive queries.....	6
Running Impala queries.....	8
Setting a workload password.....	10
Registering SSH keys.....	13
Creating an SSH key pair.....	15

About running SQL queries

You can use CDP One to query data using the Hive and Impala SQL engines.

You follow step-by-step instructions to register a Secure Shell (ssh) key pair for logging into the cluster. In a few simple steps, you connect to the cluster gateway. From the command line of the cluster, you start the SQL engine, and run SQL queries.

Using SSH to access gateway nodes

You can use SSH to connect to CDP One gateway nodes. This enables you to access the command line utilities of the analytic components in your CDP cluster and perform client tasks, such as querying Hive or Impala remotely from the command line. You use the Secure Shell (SSH) protocol to connect to a node from a terminal utility. Using SSH, you log into the node using a key pair for authentication instead of a user name and password.

Before you begin

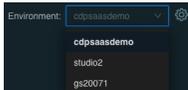
- Set a workload password. See [Setting a workload password](#) on page 10.
- Register your SSH key pair for authentication. See [Registering SSH keys](#) on page 13.



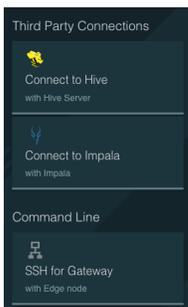
Note: You can only SSH into gateway nodes.

Procedure

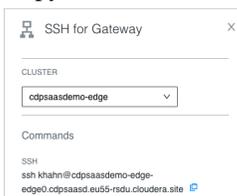
1. Log into CDP One.
2. In the Environment drop-down list, accept the default environment or select another environment.



3. Click All Services.
4. Under Command Line in the UI, click SSH for Gateway.



5. Copy the SSH command.



- Open a terminal, and paste the command.

```
$ ssh myname@cdpsaasdemo-edge-edge0.cdpsaasd.eu55-rsdu.cloudera.site
```

- At the password prompt, enter your workload password.
The connection to the gateway succeeds. The output looks something like this:

```
Last login: Mon Jun 27 21:12:10 2022 from 10.19.9.93
Welcome to Cloudera!
=====
```

Related Information

- [Registering SSH keys](#)
- [Setting a workload password](#)

Setting up Ranger to run SQL queries

As administrator, you set up authorization in Ranger for users to connect to Hive or Impala, and run queries.

Before you begin

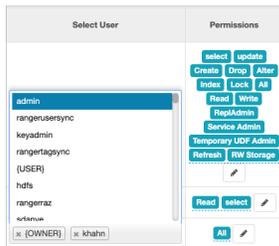
- You obtained the administrator role that allows you to set up Ranger policies.
- You logged into CDP One, and accepted the default or selected another environment.

Procedure

- In Security, click Manage Policies.
The Ranger Service Manager appears:



- Click Hadoop SQL.
- In Hadoop SQL Policies, in Allow Conditions, edit the all - database, table, column.
- In Select User, in All permissions, select your user name.



- Click Save.

Running Hive queries

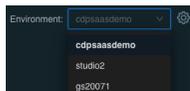
You learn how to query Hive tables from the command line of your cluster. Before attempting to run queries, you must meet a few prerequisites.

Before you begin

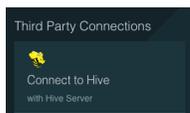
- You have SSH access to the command line of the cluster.
- As administrator, you set up a Ranger policy for running SQL queries.

Procedure

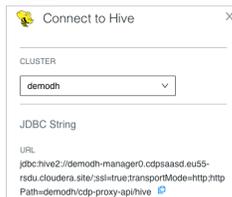
1. In Environment, accept the default environment or select another environment.



2. Click All Services.
3. Click Connect to Hive.



4. Copy the JDBC String URL.



5. Connect to the gateway node of the cluster using SSH.
6. On the command line of the cluster, start Beeline.
For example:

```
$ beeline
```

7. At the Beeline prompt, enter ! connect, a space, and then paste the JDBC String URL.

```
! connect jdbc:hive2://demodh-manager0.cdpsaasd.eu55-rsdu.cloudera.site/
;ssl=true;transportMode=http;httpPath=demodh/cdp-proxy-api/hive
```

8. At the username prompt enter your workload user name, and then enter your workload password.
The Hive prompt appears when the connection succeeds:

```
Connected to: Apache Hive (version 3.1.3000.7.2.12.4-1)
Driver: Hive JDBC (version 3.1.3000.7.2.12.4-1)
Transaction isolation: TRANSACTION_REPEATABLE_READ
0: jdbc:hive2://demodh-manager0.cdpsaasd.eu55>
```

9. Run Hive queries.

For example:

```
show databases;
```

The output looks something like this:

```
+-----+
|          database_name          |
+-----+
| airlines_demo                  |
| default                        |
| information_schema             |
| sys                            |
+-----+
```

10. Query the airlines_demo data:

For example:

```
USE airlines_demo;

SELECT f.month, a.iata, a.airport, a.city, a.state, a.country
FROM flights f,
airports a
WHERE f.origin = a.iata
GROUP BY
f.month,
a.iata,
a.airport,
a.city,
a.state,
a.country
HAVING COUNT(*) > 10000
ORDER BY AVG(f.DepDelay) DESC
LIMIT 10;
```

The output looks something like this:

```
+-----+-----+-----+-----+
| f.month | a.iata | a.airport | a.city | a.state |
| a.country | | | | |
+-----+-----+-----+-----+
| 12      | ORD    | Chicago O'Hare | Chicago | NULL    |
| USA    | | | | |
| 6       | EWR    | Newark Intl    | Newark  | NULL    |
| USA    | | | | |
| 7       | JFK    | John F Kennedy Intl | New York | NULL    |
| USA    | | | | |
| 6       | IAD    | Washington Dulles | Chantilly | NULL    |
| USA    | | | | |
| 7       | EWR    | Newark Intl    | Newark  | NULL    |
| USA    | | | | |
| 6       | PHL    | Philadelphia Intl | Philadelphia | NULL    |
| USA    | | | | |
| 1       | ORD    | Chicago O'Hare | Chicago | NULL    |
| USA    | | | | |
| 6       | ORD    | Chicago O'Hare | Chicago | NULL    |
| USA    | | | | |
| 7       | ATL    | William B Hartsfield-Atlanta | Atlanta | N      |
| NULL   | USA    | | | |
```

```

| 12 | MDW | Chicago Midway | Chicago | NULL
| USA |
+-----+-----+-----+-----+
10 rows selected (103.812 seconds)

```

Related Information

[Using SSH to access gateway nodes](#)

[Setting up Ranger to run SQL queries](#)

Running Impala queries

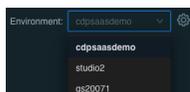
You learn how to query Impala tables from the command line of your cluster. Before attempting to run queries, you must meet a few prerequisites.

Before you begin

- You have SSH access to the command line of the cluster.
- As administrator, you set up a Ranger policy for running SQL queries.

Procedure

1. In Environment, accept the default environment or select another environment.



2. Click All Services.
3. Click Connect to Impala.



4. Copy the host name.



5. Connect to the gateway node of the cluster using SSH.


```

FROM flights f,
airports a
WHERE f.origin = a.iata
GROUP BY
f.month,
a.iata,
a.airport,
a.city,
a.state,
a.country
HAVING COUNT(*) > 10000
ORDER BY AVG(f.DepDelay) DESC
LIMIT 10;

```

Output looks something like this:

Query progress can be monitored at: https://demodh-coordinator0.cdpsaasd.eu55-rsdu.cloudera.site:25000/query_plan?query_id=624836910191fe2e:c195d9dd00000000

```

+-----+-----+-----+-----+-----+
+-----+
| month | iata | airport | city | state |
| country |
+-----+-----+-----+-----+-----+
+-----+
| 12 | ORD | Chicago O'Hare International | Chicago | NULL |
| USA |
| 6 | EWR | Newark Intl | Newark | NULL |
| USA |
| 7 | JFK | John F Kennedy Intl | New York | NULL |
| USA |
| 6 | IAD | Washington Dulles International | Chantilly | NULL |
| USA |
| 7 | EWR | Newark Intl | Newark | NULL |
| USA |
| 6 | PHL | Philadelphia Intl | Philadelphia | NULL |
| USA |
| 1 | ORD | Chicago O'Hare International | Chicago | NULL |
| USA |
| 6 | ORD | Chicago O'Hare International | Chicago | NULL |
| USA |
| 7 | ATL | William B Hartsfield-Atlanta Intl | Atlanta | NULL |
| USA |
| 12 | MDW | Chicago Midway | Chicago | NULL |
| USA |
+-----+-----+-----+-----+-----+
+-----+
Fetched 10 row(s) in 21.06s

```

Related Information

[Using SSH to access gateway nodes](#)

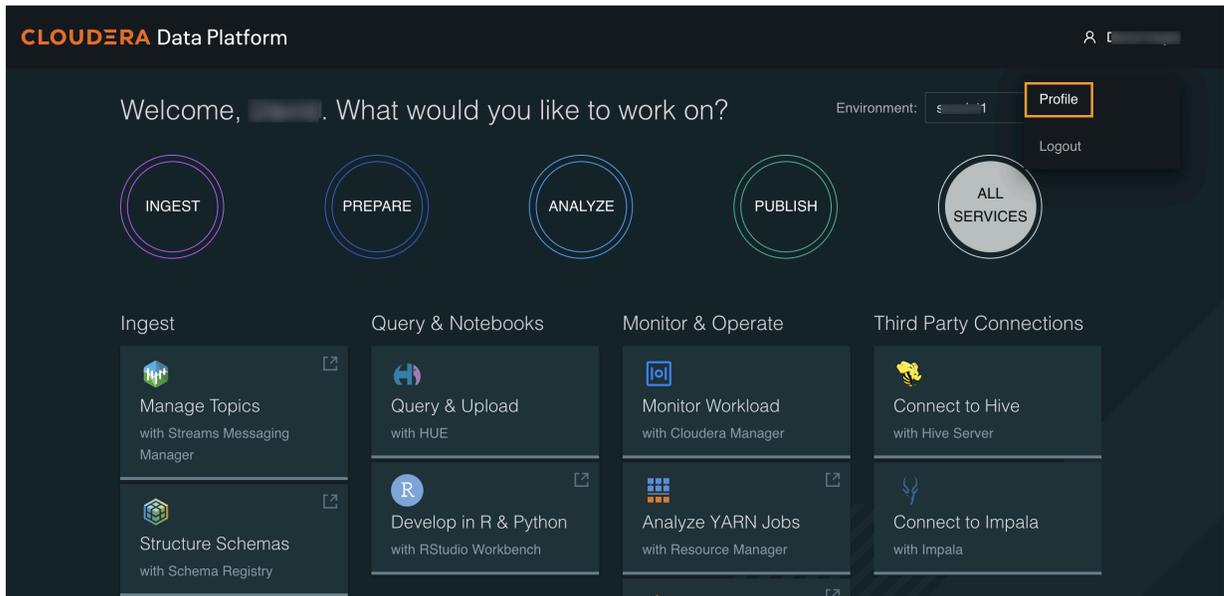
[Setting up Ranger to run SQL queries](#)

Setting a workload password

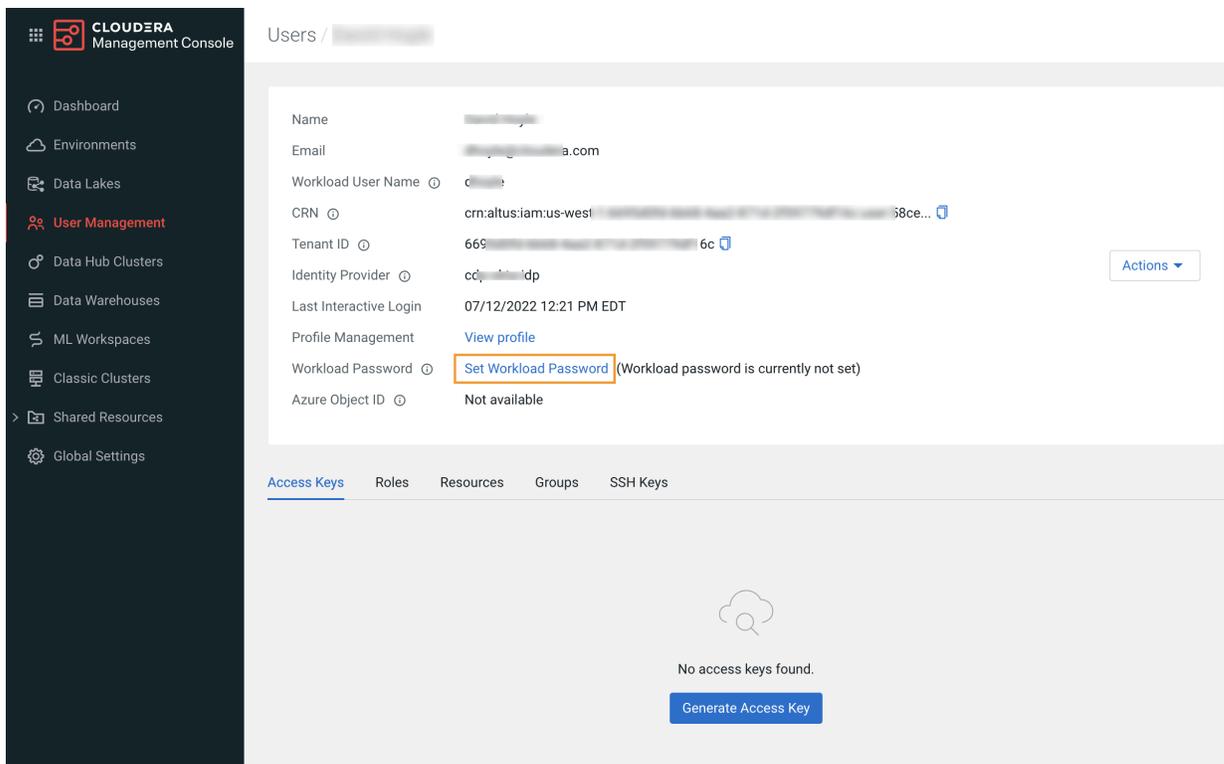
You can use your user profile page to set a workload password. Your workload password is used as the SSH password when accessing the gateway node CLI. You must perform a user sync after setting a workload password.

Procedure

1. On the CDP One console, move the pointer over the user icon at the top right of the page, then click Profile.

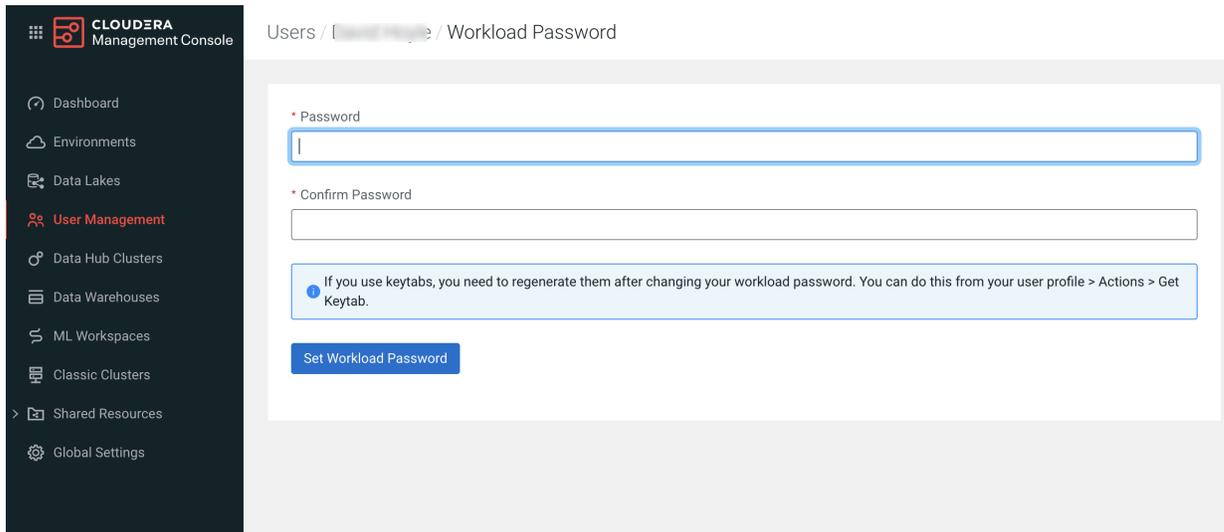


2. On your user profile page, click Set Workload Password.

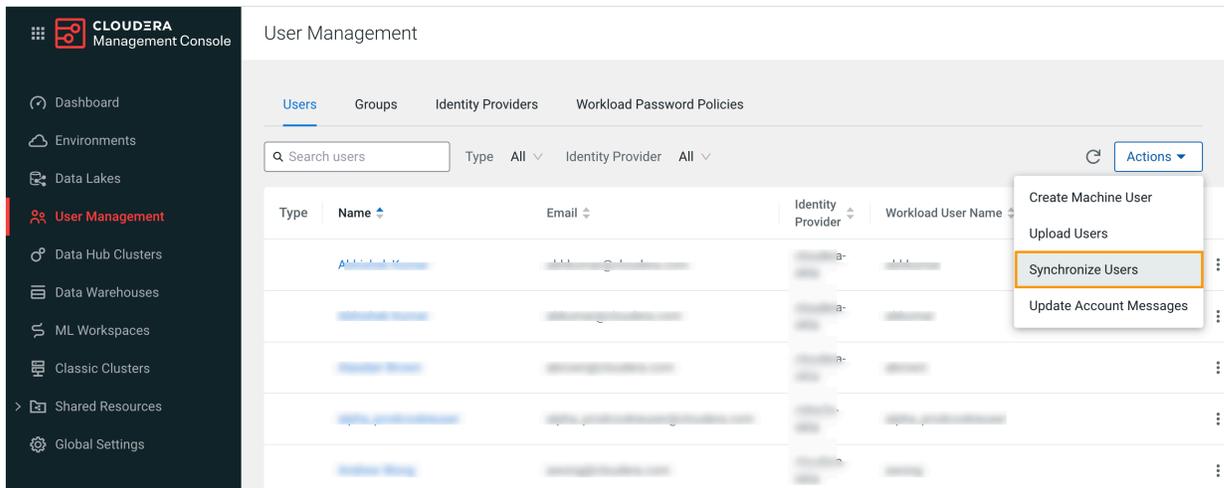


3. On the Workload Password page, type in and confirm a workload password, then click Set Workload Password.

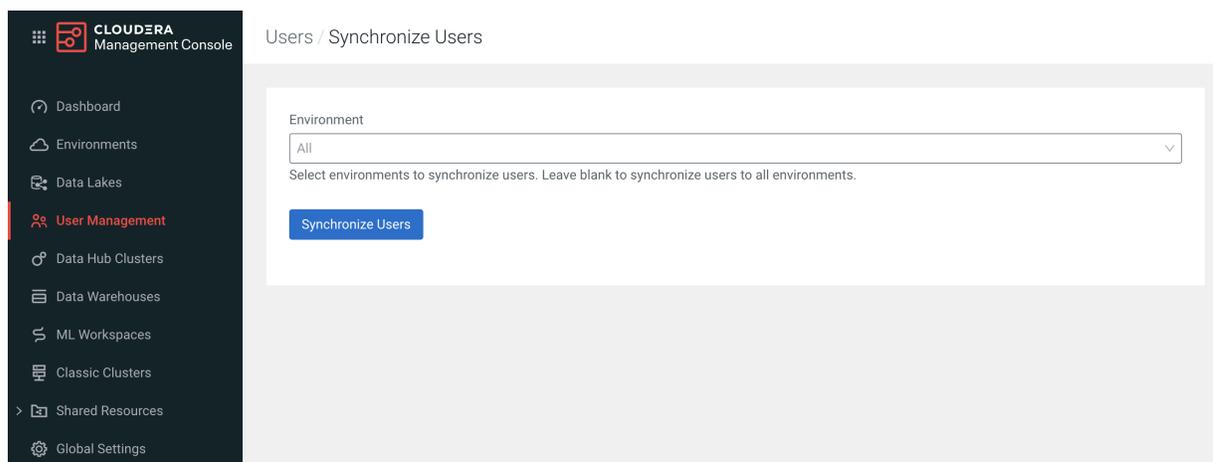
The password must be a minimum of eight characters, and must include at least one upper case character, one lowercase character, one number, and one special character. Supported special characters are "#", "&", "*", "\$", "%", "@", "^", ".", "_", and "!".



4. Click User Management, then select Actions > Synchronize Users.



- On the Synchronize Users page, all environments are selected by default. You can synchronize users in all environments, or select a specific environment. Click Synchronize Users to synchronize users in the specified environments.



Registering SSH keys

You learn how to register an existing Secure Shell (SSH) key pair. Registering the key pair of a user allows the user to access the cluster from the command line. RSA or ED25519 keys are supported.

Before you begin

You must have one of the following roles to complete this task:

- EnvironmentAdmin
- DataSteward
- PowerUser

Procedure

- Go to the root directory on your computer.
For example, on Linux enter the change directory command:

```
$ cd
```

- List hidden directories and files and look for the .ssh directory.
For example, on Linux enter the following command:

```
$ ls -ailg
```

- If you find an .ssh directory, list the files in it.

```
$ cd .ssh
$ ls
```

Output might include a private and public key pair, such as the following pair:

```
id_rsa
id_rsa.pub
```

- If you do not find a .ssh directory, skip the next step, and perform steps in the next topic, “Creating a new key pair”.

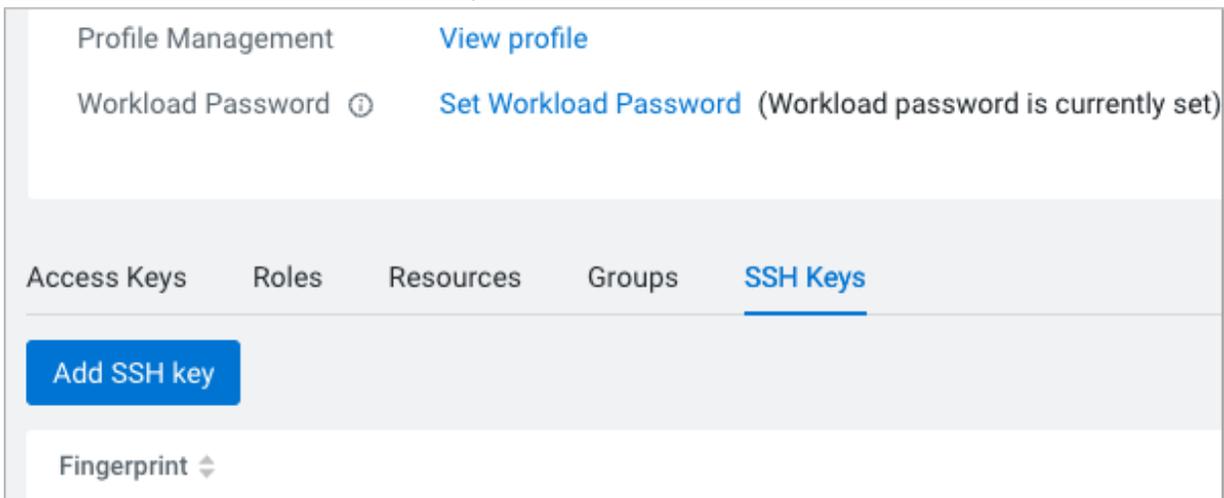
5. Copy your SSH public key to the clipboard.
For example, on Linux, enter the following command:

```
pbcopy < ~/.ssh/id_rsa.pub
```

6. In CDP One, click Profile.



7. In Users, on the SSH tab, click Add SSH key.



8. In Add SSH Public Key, click the SSH public key text box, and paste the contents of your clipboard.

Add SSH Public Key
✕

Description

* SSH Public Key

```
ssh-rsa
AAAAB3NzaC1yc2EAAAADAQABAAQCAOC1SSxUv8OdaU/jRFK/o6R6iqAzmwsXN9LxatPniXFXWzwtG
r1A6xig7vUND/cYiEfexVK1xB5p3Xhm3RGjZzIN7thdXuc5i0Y5gXDqOZsyCUQBnbjKWaRseug4m592P+
D34R0IrKHIWmfNw4t3GYhGiCl...
srUD1ot0JyAhes3av9RkdaHzwE6pAzGzZ0mN94Cy1STc0P59COaYq9Clr88aPoi6JfGFrYC1WNX7p4wl2
HpWKSwwwJP8jEmtgFh0H6wk7GPmnD...
YRfo1hiS+0U3SuXkEb2NGU2vdLkSDMOFaqmphRFTImvem7PZok3uZPRD0ySOgSrZ+AqHIGUHuiUYTM
R2C...
5mvJbWGjzNVnoSWkOGYCB12Y8KFtprV7GrVm118LWtEJEipHR8D5BBTp85l1Y6kion5mkFqexti5fpW
wODUtEGlruO+Y4VNhrcGbZ0+EQjGcmAr8B4ukD.JqkTyfYYIxU3at4hX1lfUTzp1cucb1smdCBYJw==
mulehoofs@gmail.com
```

i Once the SSH key is added, the environment will need to be synced before it can be used for SSH access.

Cancel
Save

9. Click Save.

10. Synchronize users to the environment.

Creating an SSH key pair

Before you begin

OpenSSH is installed on your machine.

You checked for a pre-existing key pair as described in "Registering SSH keys" above, and found none.

Procedure

1. Open a terminal window, and on the command line, type the key generation command: `ssh-keygen`

```
$ ssh-keygen
```

2. Accept the default location for the keys `~/.ssh` (recommended) and file name `id_rsa` or specify another location and name.
3. At the passphrase prompt, create a password for the key pair.
4. Follow steps in "Registering SSH keys" above to register the keys in CDP.