Cloudera Octopai Data Lineage 1.0.0

Octopai Client

Date published: 2025-10-09
Date modified: 2025-10-20



https://docs.cloudera.com/

Legal Notice

© Cloudera Inc. 2026. 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

General system requirements for Octopai Client.....4

Installation..... 4

Octopai Client - Linux Installation..... 4
Octopai Client - Installation guide.....8
Configuring Proxy settings.....9
How to Install Curl.....10

Automating Cloudera Data Lineage metadata extractions with Microsoft Task Scheduler.....14

General system requirements for Octopai Client

Learn about the general system requirements for Octopai Client.

Note: Contact Cloudera Support for more details and instructions.

Windows server

- 2019 OS and higher on the organization's domain
- .NET Framework 6 and higher (offline installer)
- .NET Core 8 Runtime (installer)
- 8 Cores (CPU); 16 GB – RAM; 250 GB - Hard Disk
- Reserved ports for local use 5656 & 5657
- Curl is installed. For instructions, see How to Install Curl.
- Continue to Octopai Client - Installation guide.

Linux server

Contact Cloudera Support for more details and instructions.

- Linux distribution - Minimum requirement: RHEL 8 GUI / Oracle 8 GUI
- .NET Core 8 Runtime (link)
- 8 Cores (CPU); 16 GB – RAM; 250 GB - Hard Disk
- Reserved ports for local use 5656 & 5657
- Curl is installed. For instructions, see How to Install Curl.

Browser compatibility - Cloudera Data Lineage (formerly Octopai) application

- Chrome or MS Edge

Important remarks

- All metadata extractions must be done from the same environment, such as Sandbox or QA
- Run the extractions in admin mode
- If you need assistance, contact Cloudera Support.

Installation

Install Octopai Client on Windows or Linux, install Curl for metadata extraction, and configure proxy settings when your network routes traffic through a proxy.

Octopai Client - Linux Installation

Learn about installing Octopai Client on supported Linux systems.

About this task

Warning:

The following tools from the supported integrations list are not supported by the Octopai Client Linux version:

- Azure Analysis Services
- DB2
- Looker
- Power BI Report Server
- SSISSQLS
- Tabular
- SSAS-TABULAR
- SSIS - ISC
- Tableau PAT
- TABLEU

Before you begin

- You must verify that your environment meets the general system requirements before starting the installation. For instructions, see General System Requirements for Octopai Client.
- You must have root or sudo access to the Linux system that must comply with the Octopai Client system requirements.

Procedure

1. Install required dependencies.

```
sudo yum update -y
sudo yum install libicu -y
sudo yum install keyutils -y
```

2. Install .NET 8 by running the following script to install .NET 8 in your preferred directory, for example, ~/.dotnet.

```
sudo wget https://dot.net/v1/dotnet-install.sh -O dotnet-install.sh
sudo chmod +x dotnet-install.sh
sudo ./dotnet-install.sh --channel 8.0 --install-dir ~/.dotnet
```

3. Add .NET to the PATH.

```
echo 'export DOTNET_ROOT=$HOME/.dotnet' >> ~/.bashrc
echo 'export PATH=$HOME/.dotnet:$PATH' >> ~/.bashrc
source ~/.bashrc
dotnet --version
```

4. Configure firewall rules.

   Use an existing or newly installed firewall and open the required ports.

```
sudo firewall-cmd --zone=public --permanent --add-port=80/tcp
sudo firewall-cmd --zone=public --permanent --add-port=5656/tcp
```

   Note: For external firewalls, ensure that ports 80 and 5656 are open for both inbound and outbound traffic.

5. Set SELinux to disabled or permissive.

```
sudo nano /etc/selinux/config
```

a) Locate the line that starts with SELINUX= and update it to one of the following values:

```
SELINUX=disabled
```

```
SELINUX=permissive
```

b) Reboot the machine, then verify the status by using the following command:

```
sestatus
```

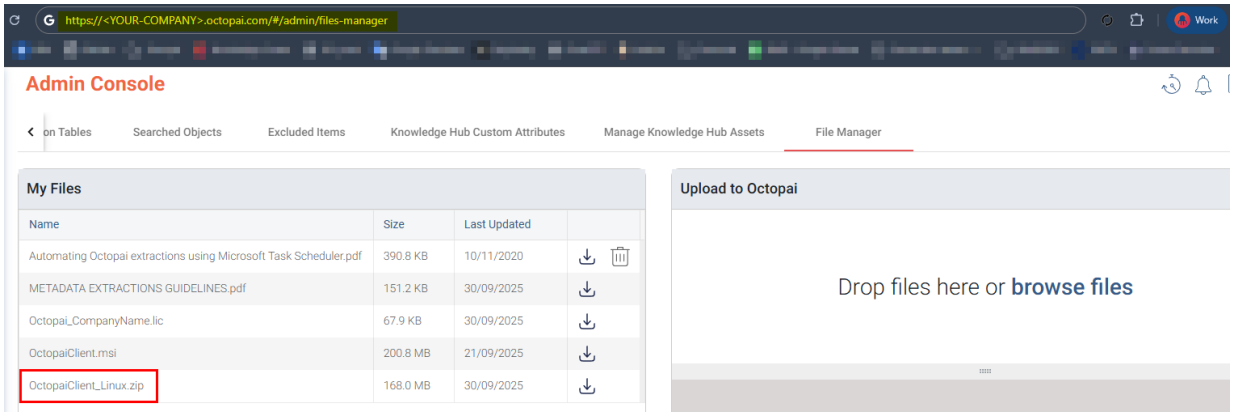
The output should look like this:

```
SELinux status: disabled
```

6. Download TAR or zip archive file from File Manager, and extract the Octopai Client package at /opt/Octopai-Client.

```
sudo tar xvf Octopai-Client.tar.gz
sudo yum install unzip -y
sudo unzip Octopai-Client.zip -d /opt/
```

Figure 1: License file example



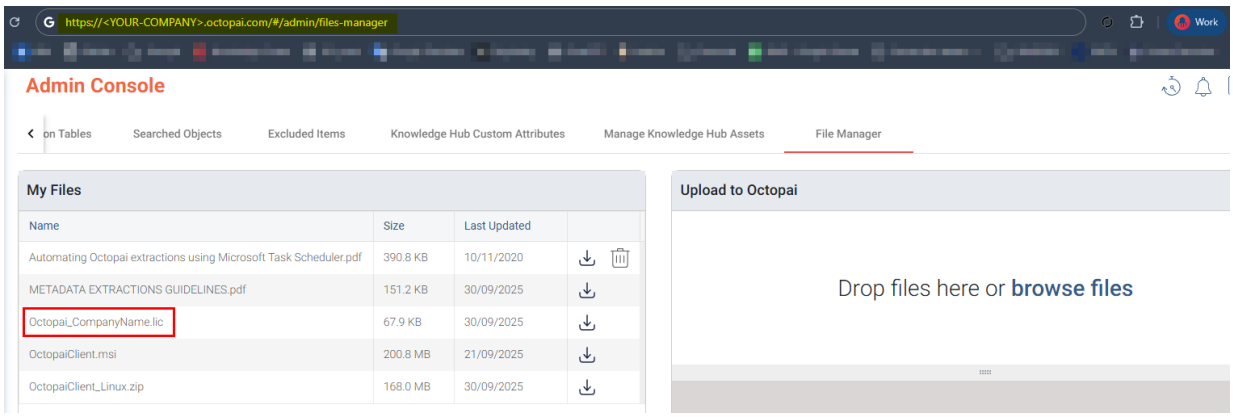
7. Move the client folder and set permissions by using the following command:

```
sudo mv Octopai-Client /opt/
cd /opt/
sudo chmod -R 777 Octopai-Client
```

Attention:

Apply full control using the chmod -R 777 command only to the Octopai Client folder so that the application can read, write, and execute the files it requires.

8. Copy the license file by downloading the license file from File Manager and placing it in the /opt/Octopai-Client directory on the Linux host.

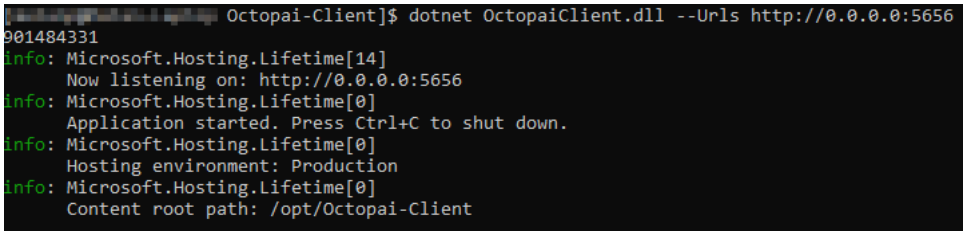


9. Run the Octopai Client by using the following command:

```
cd /opt/Octopai-Client
dotnet OctopaiClient.dll --Urls http://0.0.0.0:5656
```

The following example shows a sample Octopai Client output.

Figure 2: Sample Octopai Client output



10. Open the Octopai Client interface.

Go to http://localhost:5656/index.html to begin connecting systems for metadata analysis.

What to do next

Create a Linux service

Configure a `systemd` service so the Octopai Client starts automatically after reboots.

1. Create the service definition.
   a. Edit the service file.

```
sudo nano /etc/systemd/system/octopai.service
```

   b. Paste and adjust the following content to match your environment.

```
[Unit]
Description=Octopai Client
[Service]
ExecStart=/usr/bin/dotnet /opt/Octopai-Client/OctopaiClient.dll --Urls h
ttp://0.0.0.0:5656
Restart=always
RestartSec=10
KillSignal=SIGINT
SyslogIdentifier=octopai-client
User=octopai
Group=octopai
WorkingDirectory=/opt/Octopai-Client/
Environment=USER=octopai
```

```
[Install]
WantedBy=multi-user.target
```

2. Reload `systemd` configuration by using the following command:

```
sudo systemctl daemon-reload
```

3. Enable the service by using the following command:

```
sudo systemctl enable octopai.service
```

4. Start or restart the service by using the following command:

```
sudo systemctl restart octopai.service
```

5. Check the service log by using the following command:

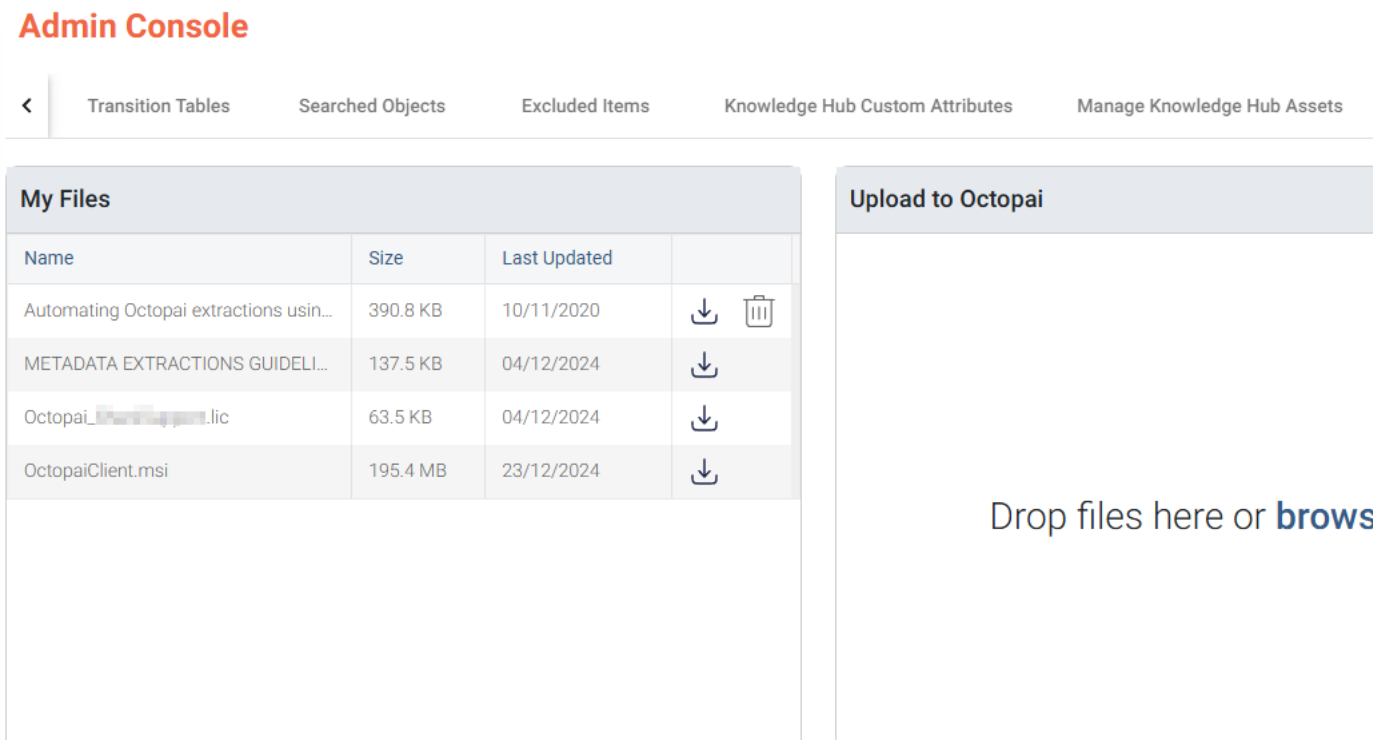
```
sudo journalctl -u octopai.service
```

Octopai Client - Installation guide

Learn about installing your Octopai Client.

Procedure

1. On the server, where the Octopai Client is installed or will be installed, go to your platform at [***CUSTOMER NAME***].octopai.com, then go to the File Manager tab on the Admin Console.
2. From File Manager, download the latest OctopaiClient.msi installer and license files (.lic).



Important:

Only use the OctopaiClient.msi file downloaded from File Manager. Do not use alternate download sources.

Cloudera Support cannot assist with alternate download sources or manually redistributed or split installer files.

Note: Cloudera Data Lineage (formerly Octopai) recommends creating a backup of the Data folder by navigating to C:\Program Files (x86)\Octopai\Service\Data and copying the folder to your desktop.

3. Run the MSI installation file and wait for it to finish.
4. If you aim to run the Octopai Client with a dedicated user or service account, perform the following steps:
   a) Go to Services and locate Octopai Agent.
   b) Right-click Cloudera Agent and select the Properties option.
   c) Go to the Log On tab and set the user.
5. Grant Full Control permission to the Cloudera Data Lineage folder for your dedicated user or group.
   a) Right-click the folder and select the Properties option.
   b) Go to the Security tab, and click Edit.
   c) Choose the user or group and check Full Control.
6. Open the Octopai Client.

   Choose the license by clicking on the green key or pop-up message.
7. Rerun the necessary or desired connections.
   a) Click the metadata source connection on the left-hand bar.
   b) Select Test & Save, click OK, and then click Run.

Note:

If the connections are not visible or some are missing from the Octopai Client after the upgrade, restore the Data folder from your backup. Alternatively, navigate to the C:\ProgramData\Octopai-Backup folder, where Cloudera Data Lineage automatically creates a backup during upgrades, and copy its contents to the C:\Program Files (x86)\Octopai\Service\Data folder.

Configuring Proxy settings

Learn about configuring Proxy settings for the Octopai Client if your server's network is routed through a proxy tunnel.

Before you begin

You must upgrade your Octopai Client to the latest version before configuring the Proxy settings.

Procedure

1. Go to C:\Program Files (x86)\Octopai\Service\Data.
2. Open the **PS** file using your preferred text editor.
3. Update the following parameters for proxy configuration:

   - ProxyEnabled – Set to true to enable or to false to disable the proxy option.
   - ProxyAddress – Specify the address in the "http://YOUR_PROXY_NAME:PORT" format.
   - ProxyUsername – Provide the Username.
   - ProxyPassword – Provide the Password.

Note: You must wrap the above parameters in quotation marks.

Figure 3: Example



If you are upgrading your Octopai Client, these values will be added automatically to the PS file.

Note: The proxy option is disabled by default.

How to Install Curl

Learn about properly installing and configuring Curl for use with the Octopai Client, enabling seamless extraction of metadata.

Option 1: Install Curl with a package manager

1. For Chocolatey, run the `choco install curl` command.

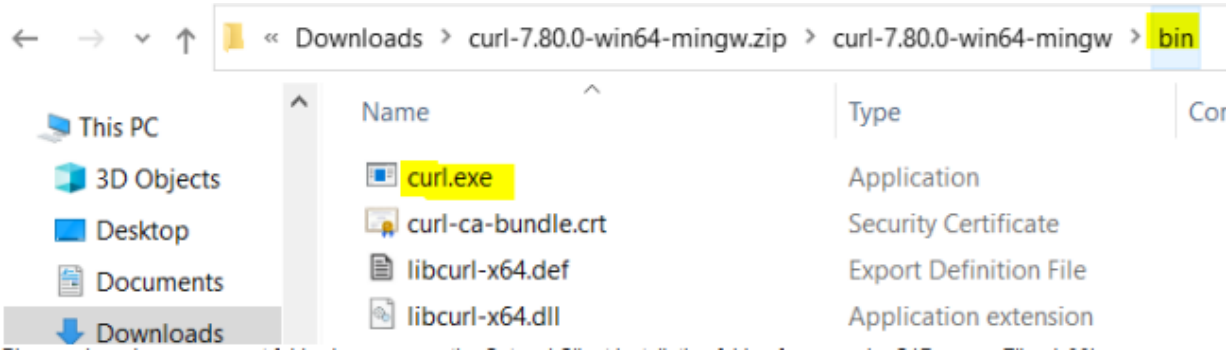
2. For MSYS2, run the `pacman -S curl` command.
3. For Scoop, run the `scoop install curl` command.

Option 2: Install Curl manually

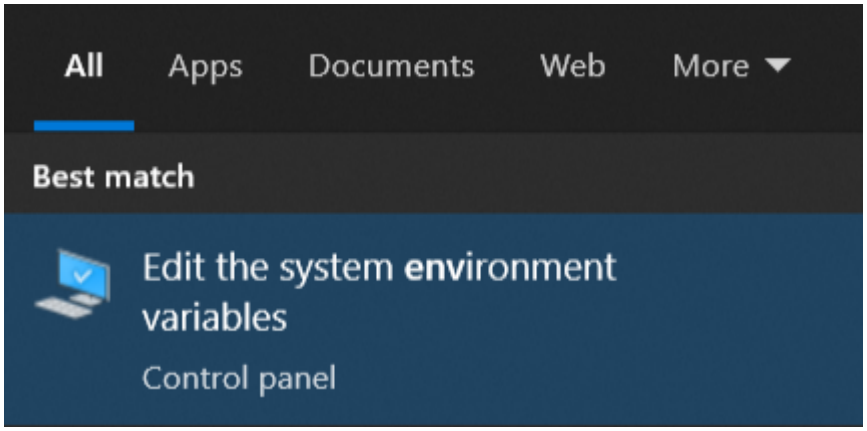
1. Download curl from https://curl.se/windows/ and choose the Windows 64-bit version.



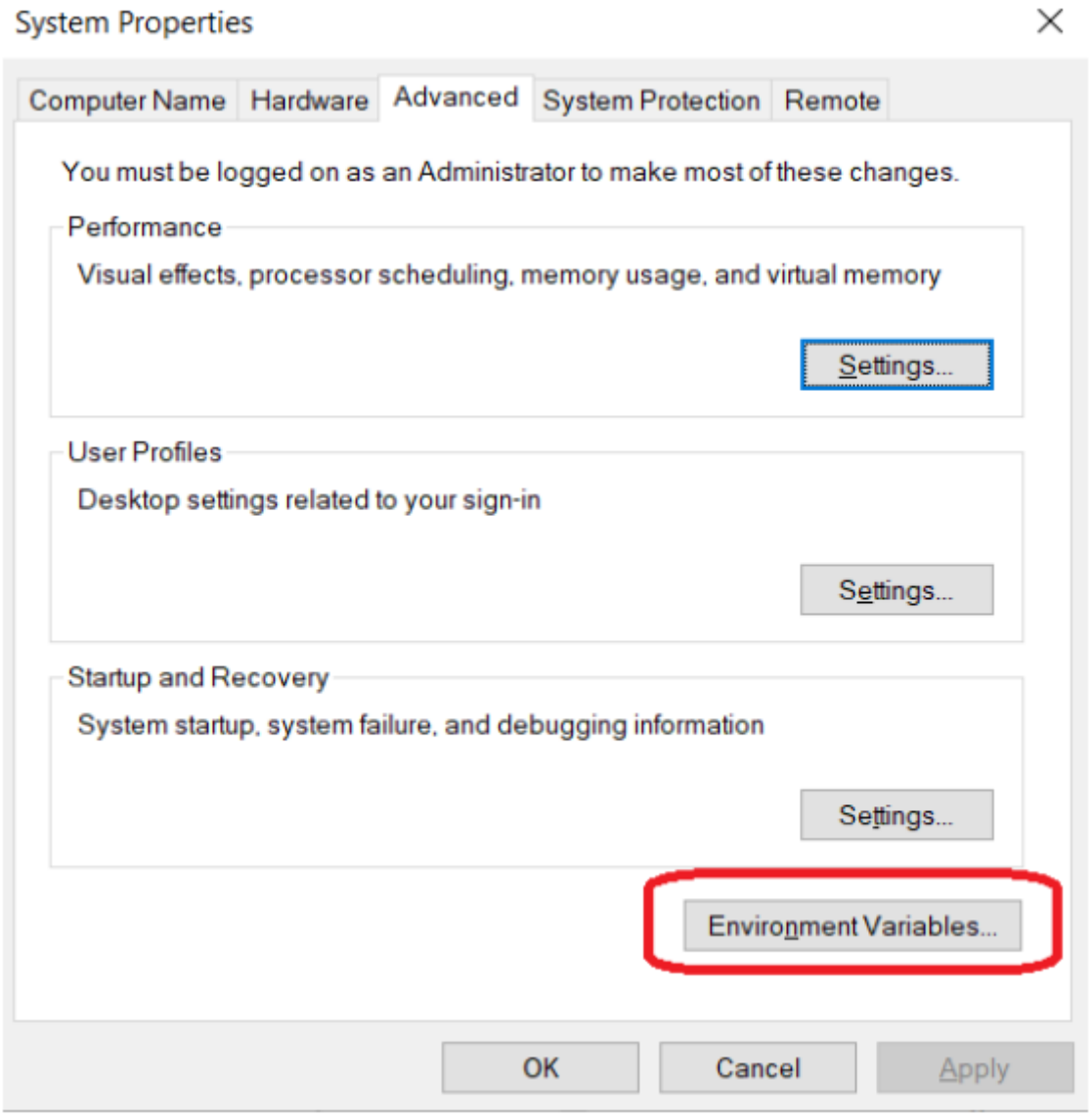
2. Locate the curl.exe file in the \bin folder of the downloaded zip file.



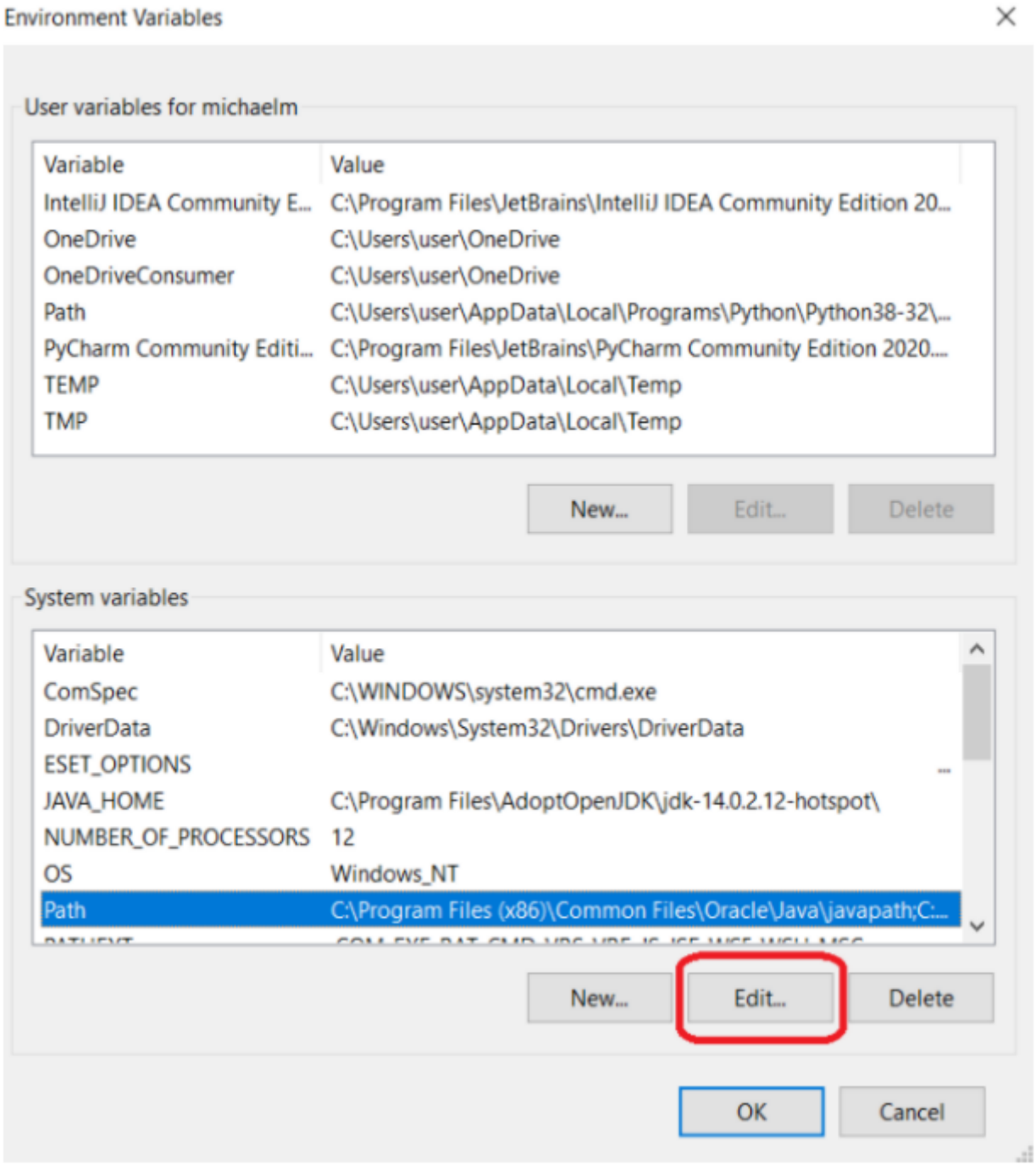
3. Place the curl.exe file in a permanent folder, for example to the Octopai Client installation folder, such as C:\Program Files (x86)\Octopai\Service\Data.
4. Make curl accessible from anywhere on the command line by adding the folder PATH as a new system environment variable.
   a. Open the Start Search, type env, and choose the Edit the system environment variables action.



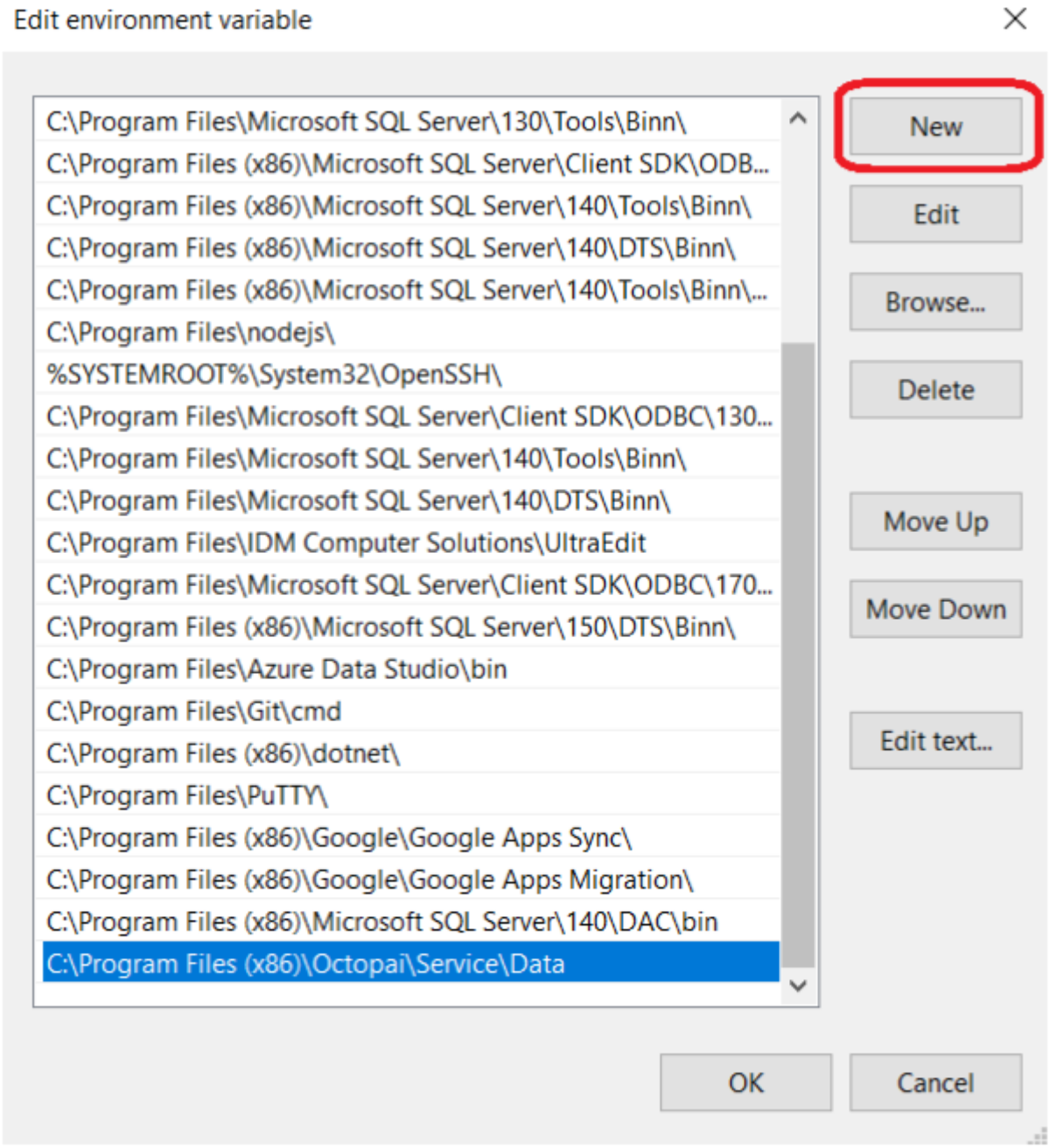
5. Click Environment Variables.



6. In the **System Variables** section, select the Path variable and click Edit.



7. Click New and add the location of the folder that contains the curl.exe file.



8. Click Ok.
9. If applicable, reopen any relevant applications to register the changes.

Automating Cloudera Data Lineage metadata extractions with Microsoft Task Scheduler

Learn about automating Cloudera Data Lineage (formerly Octopai) metadata extractions using Microsoft Task Scheduler on Windows, including task creation, configuration, and scheduling to streamline metadata updates efficiently.

Before you begin

Automating metadata extractions in Cloudera Data Lineage can greatly streamline data management processes. By utilizing Microsoft Task Scheduler, you can schedule the execution of the Cloudera Data Lineage BAT files, which are located by default in the C:\Program Files (x86)\Octopai\Service\BAT directory. Scheduling the execution of BAT files simplifies and streamlines the process, ensuring timely and accurate metadata updates. Leverage the power of Task Scheduler to enhance your Cloudera Data Lineage workflow and improve overall data management efficiency.

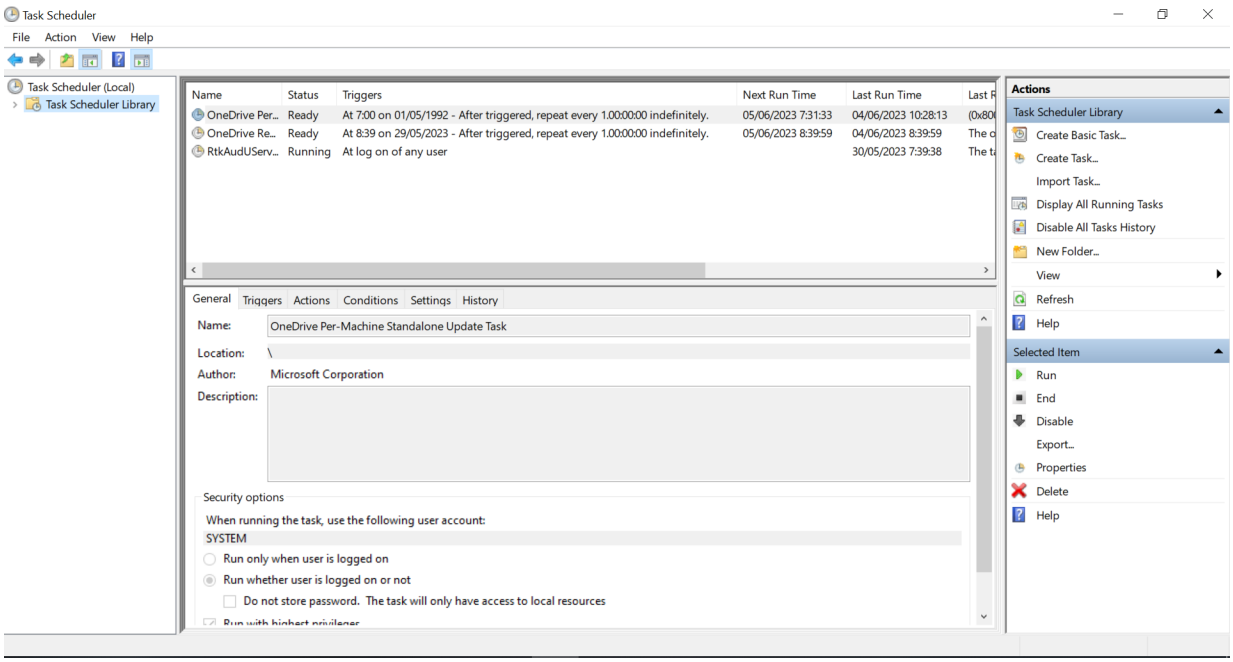
To successfully run the extractions you must have rights to log on as a batch job. For more information, see Log on as a batch job.

Procedure

1. Launch the Microsoft Task Scheduler application on your Windows computer.

   It can typically be found by searching for Task Scheduler in the **Start** menu.

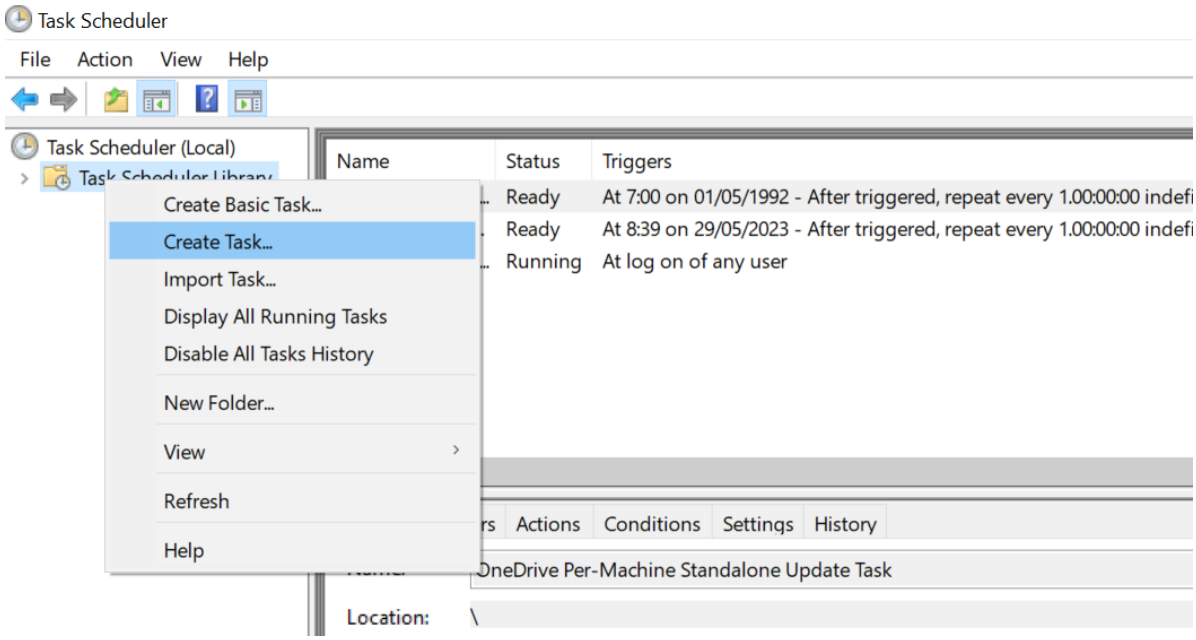
   Figure 4: Microsoft Task Scheduler application



2. Create a new task.

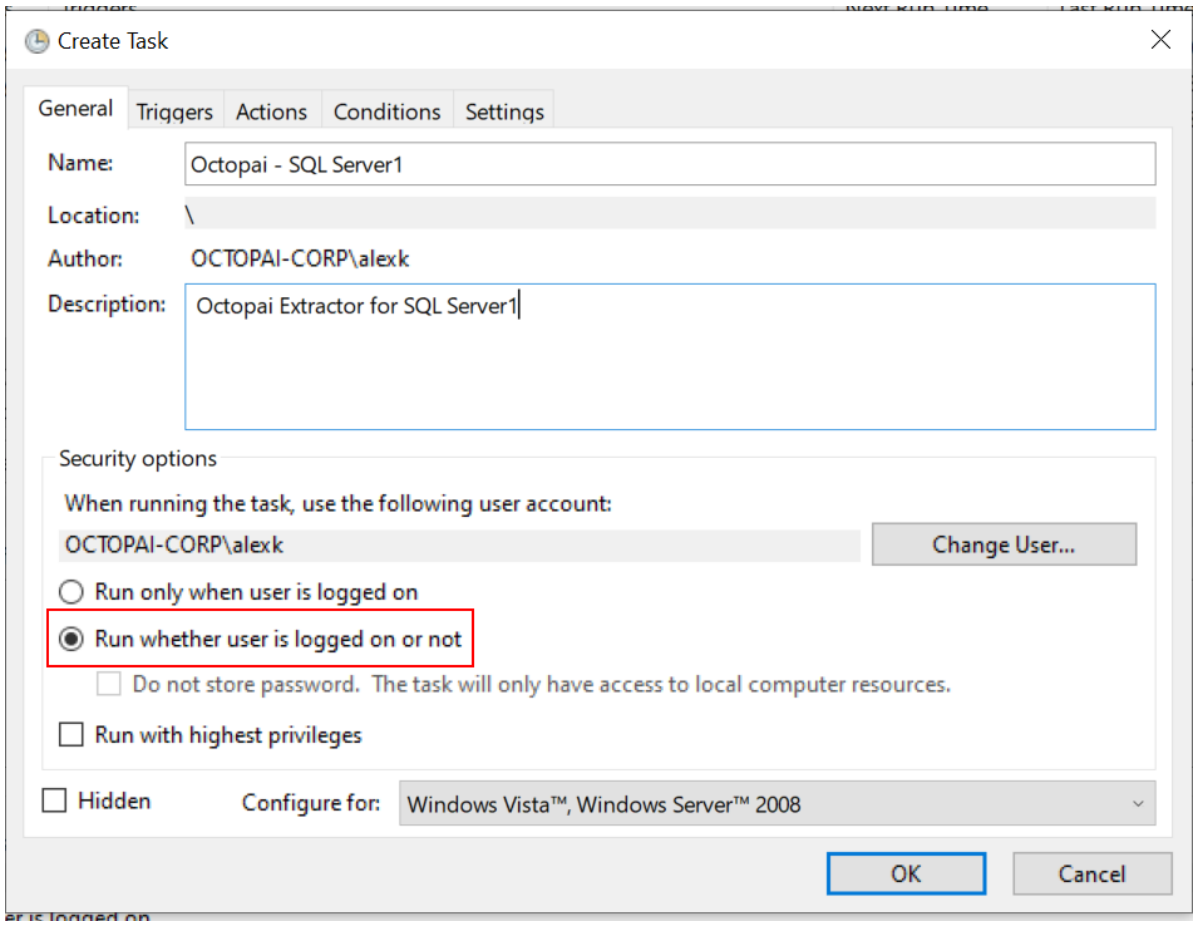
   a) Right-clicking on the **Task Scheduler Library** folder or the desired folder where you want to save the Cloudera Data Lineage schedule task and selecting the Create Task... action.

   Figure 5: Create a task



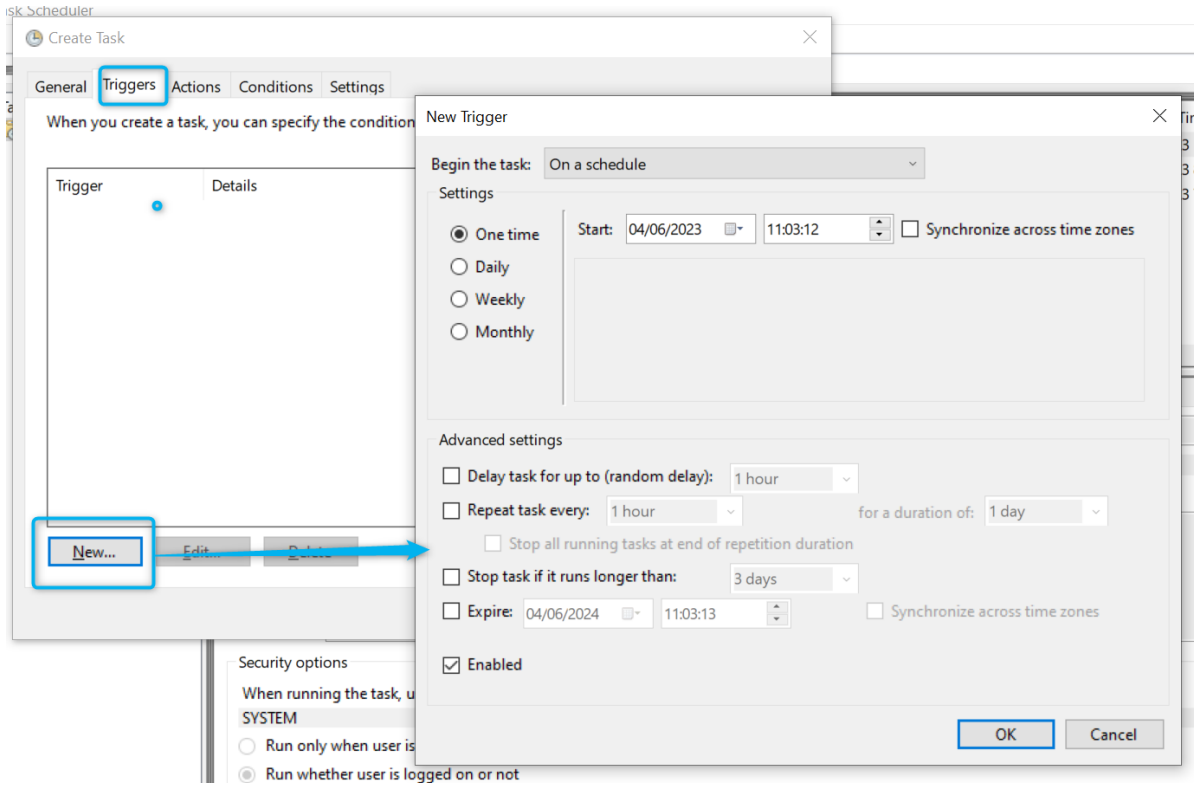
b) In the **General** tab of the **Create Task** window, provide the necessary details.

Figure 6: Create a task General tab



c) In the **Triggers** tab of the **Create Task** window, click New to open the **New Trigger** window and define the schedule recurrence. Cloudera Data Lineage recommends setting it to once a week.

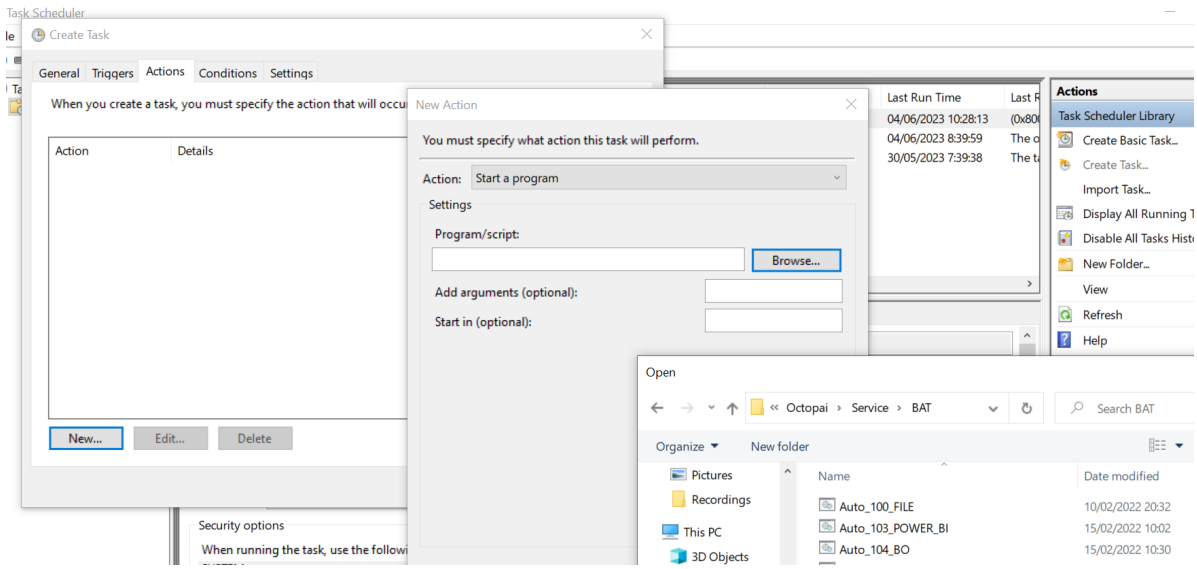
Figure 7: Create a task Triggers tab



d) Configure actions.

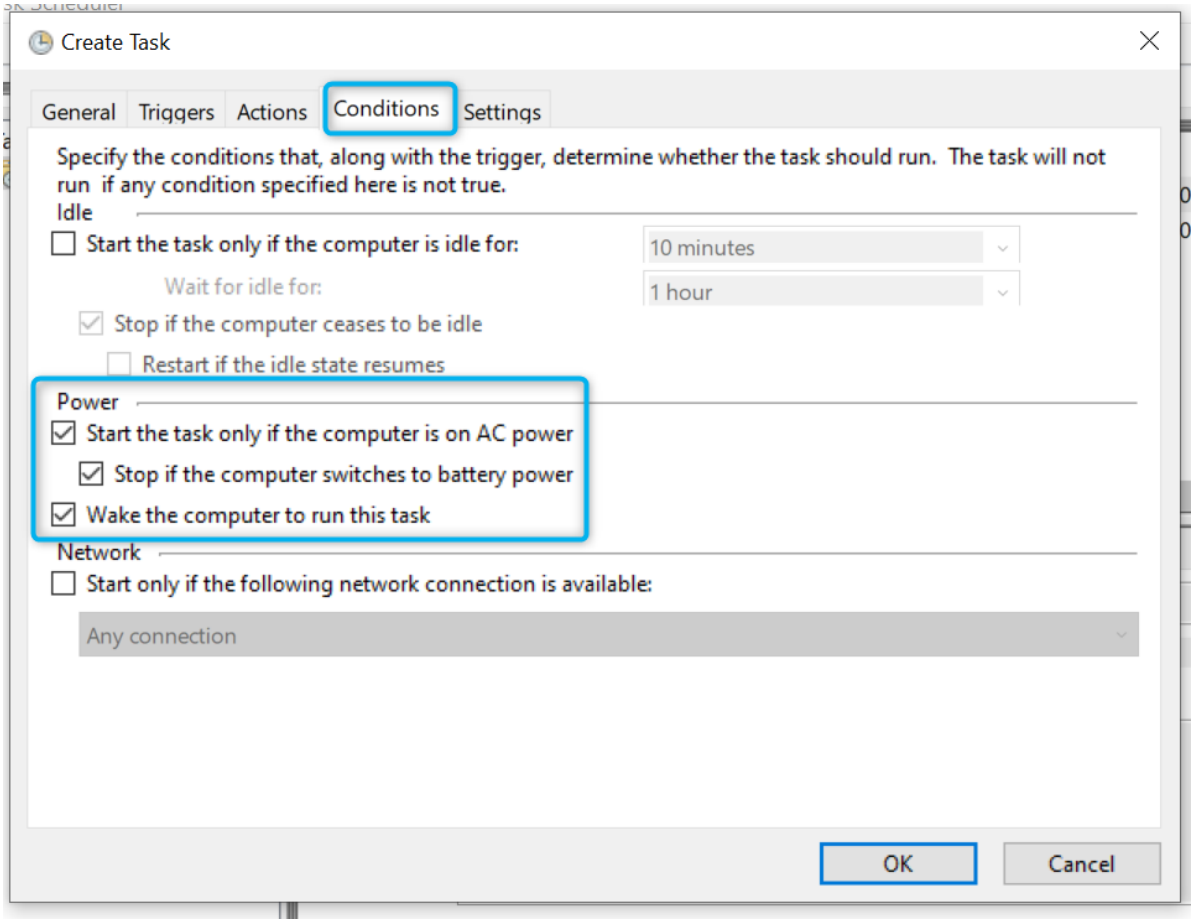
1. In the **Actions** tab of the **Create Task** window, click New to open the **New Action** window.
2. Browse for the relevant BAT file by clicking Browse or manually entering the file path. The default location for BAT files is the folder where Octopai Client is installed that is C:\Program Files (x86)\Octopai \Service\BAT by default.
3. Select the file and click OK to save.

Figure 8: Create a task Actions tab



e) In the **Conditions** tab of the **Create Task** window, select all the checkboxes in the **Power** section and go to the **Settings** tab without clicking OK.

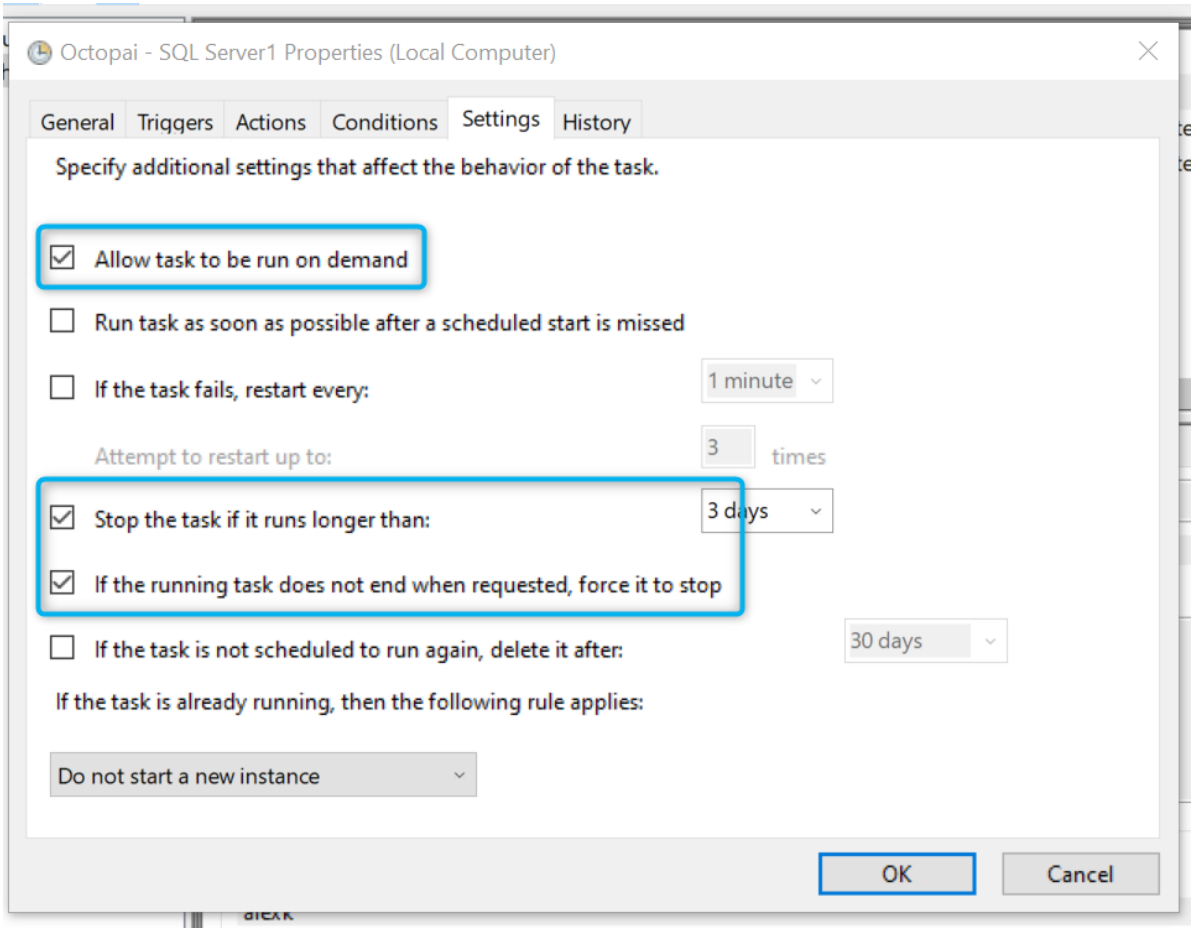
Figure 9: Create a task Conditions tab



f) In the **Settings** tab of the **Create Task** window, select the following checkboxes:

- Allow task to be run on demand
- Stop the task if runs longer than:
- If the running task does not end when requested, force it to stop

Figure 10: Create a task Settings tab



g) Click OK to save the task and the configurations.