This documentation provides step-by-step instructions for starting the OpenSearch-Metadata-Hub project using Docker. Follow the guidelines below to get started:

Prerequisites

Before starting the OpenSearch-Metadata-Hub, ensure that you have Docker installed on your machine. If Docker is not installed, please refer to the Docker documentation for instructions specific to your operating system: link to docker

Important Linux Settings

For production workloads on Linux, it is important to configure the virtual memory settings correctly. Follow these steps:

1. Check the current value of `vm.max_map_count` by running the following command:

   ```
   cat /proc/sys/vm/max_map_count
   ```

   Ensure that the value is set to at least 262144. If the value is lower, continue with the following steps to increase it.

2. Open the `/etc/sysctl.conf` file using a text editor:

   ```
   sudo <text-editor-of-choice> /etc/sysctl.conf
   ```

3. Add the following line to the end of the file:

   ```
   vm.max_map_count=262144
   ```

4. Save the file and exit the text editor.

5. Run the following command to reload the sysctl settings:

   ```
   sudo sysctl -p
   ```

   This will apply the changes made in the `/etc/sysctl.conf` file.

Getting Started

1. Start “Docker Desktop” on your machine. If you haven’t installed it yet, just install it before.

2. Navigate to our repository and download the latest release of the OpenSearch-Metadata-Hub project: link to the release

3. **Create a `.env` file in the `src/` folder** of the project directory. This file should contain key-value pairs for `URL_CORE_1` and `PW_USER_CORE_1`, where you can store your credentials. In addition, you can enter the credentials for OpenSearch-Dashboards. They are both by default set to `admin`. Here is an example of how the contents should look:

   ```
   URL_CORE_1=https://metadahub.de/example
   PW_USER_CORE_1=testpw
   OS_USER=admin
   OS_PASSWORD=admin
   ```

4. Navigate to the `src/` folder in the project directory.

5. Execute the following command to start the services using Docker Compose:

   ```
   docker-compose up
   ```

   Alternatively, if you want to run the services in detached mode, use the following command:

   ```
   docker-compose up -d
   ```

6. Once the services are up and running, you can access the OpenSearch-Metadata-Hub website on your local machine at `http://localhost:8000`.

Import Pipeline Cronjob

By default, a cronjob is set up to execute the import pipeline every night at 2 am. This cronjob ensures that data is regularly imported into the OpenSearch-Metadata-Hub. The logs of the cronjob are written to /var/log/cronjob_log. If you wish to change the time interval or customize the cronjob, follow these steps:

1. Open the cronjob file under src/application/res/import-script/cronjob
2. Modify the cron expression to set the desired time interval for executing the import pipeline.
3. Save the file.

Stopping the Services

To stop all services, follow these steps:

1. Open a terminal or command prompt.

2. Navigate to the `src/` folder in the project directory.
3. Execute the following command to stop the services:

   ```
   docker-compose down
   ```

Clearing Persistent Data

If you need to clear imported data or any persistent data, you can use the `-v` option in the `docker-compose down` command. However, please note that this action cannot be undone. To clear persistent data, follow these steps:

1. Open a terminal or command prompt.
2. Navigate to the `src/` folder in the project directory.
3. Execute the following command to stop the services and clear persistent data:

   ```
   docker-compose down -v
   ```

Please refer to the OpenSearch-Metadata-Hub documentation for additional configuration options and advanced usage.

Configurable Settings

There are some settings, which can be easily configurated in the Config-File, which can be found here: `src/application/res/config.ini`

Here is a list with all configurable settings and their default settings. Note, that the default values are applied, if you leave one setting out.

Setting	Description	Default Value	Example
`localhost`	Set this to `True` to test and debug the software locally. It is not recommended to use this mode in other purposes.	`False`	`True`
`index_name`	The name of the OpenSearch-Node, where the imported data is stored.	`amoscore`	`amoscore`
`search_size`	Specify, how much search results should be displayed.	`24`	`25`

Setting	Description	Default Value	Example
`limit`	Specify the maximum amount of imported data per period.	unlimited (don't use this setting in the config file)	`100`
`selected_tags`	Specify, which tags should be imported. Note, that this setting will only be applied, if `only_selected_tags` is set to `True`.	`FileName;FileSize;FileType;SourceFile`	`FileName;FileSize;FileType`
`only_selected_tags`	See above.	`False`	`True`
`only_new_data`	Only import data, which wasn't yet imported.	`True`	`False`
`file_types`	Specify, which file types should be imported. For example, if you have an import limit of 100 and you just want to import `.xml`- and `.jpeg`-files, then the import pipeline will import 50 `.xml`- and 50 `.jpeg`-Files at maximum.	all file types, encoded as	`XML;JPEG`

Here is an example configuration file:

```
[General]
localhost = False
index_name = amoscore
search_size = 24
#limit = 100
selected_tags = FileName;FileSize;FileType;SourceFile
only_selected_tags = False
only_new_data = True
file_types =
```