

Installing Apache Flume on EMR

This document will guide you through the steps to set up Apache Flume on your EMR cluster.

- First, ensure that you are the root user. After that, run the following command to download the Flume binaries.

```
sudo -i  
wget https://de-emr-flume.s3.amazonaws.com/apache-flume-1.9.0-bin.tar.gz
```

```
[root@ip-172-31-45-198 ~]# wget https://de-emr-flume.s3.amazonaws.com/apache-flume-1.9.0-bin.tar.gz  
--2021-08-12 18:10:19-- https://de-emr-flume.s3.amazonaws.com/apache-flume-1.9.0-bin.tar.gz  
Resolving de-emr-flume.s3.amazonaws.com (de-emr-flume.s3.amazonaws.com)... 52.217.18.92  
Connecting to de-emr-flume.s3.amazonaws.com (de-emr-flume.s3.amazonaws.com)|52.217.18.92|:443... connected.  
HTTP request sent, awaiting response... 200 OK  
Length: 67938106 (65M) [application/x-gzip]  
Saving to: 'apache-flume-1.9.0-bin.tar.gz'  
  
100%[=====]  
938,106 58.4MB/s in 1.1s  
  
2021-08-12 18:10:20 (58.4 MB/s) - 'apache-flume-1.9.0-bin.tar.gz' saved [67938106/67938106]  
  
[root@ip-172-31-45-198 ~]# |
```

- Now you need to untar the file by running the following command.

```
tar -zxvf apache-flume-1.9.0-bin.tar.gz
```

```
[root@ip-172-31-45-198 ~]# tar -zxvf apache-flume-1.9.0-bin.tar.gz  
apache-flume-1.9.0-bin/conf/  
apache-flume-1.9.0-bin/bin/  
apache-flume-1.9.0-bin/doap_Flume.rdf  
apache-flume-1.9.0-bin/README.md  
apache-flume-1.9.0-bin/DEVNOTES  
apache-flume-1.9.0-bin/conf/log4j.properties  
apache-flume-1.9.0-bin/conf/flume-env.sh.template  
apache-flume-1.9.0-bin/conf/flume-conf.properties.template  
apache-flume-1.9.0-bin/conf/flume-env.ps1.template  
apache-flume-1.9.0-bin/NOTICE  
apache-flume-1.9.0-bin/CHANGELOG  
apache-flume-1.9.0-bin/bin/flume-ng  
apache-flume-1.9.0-bin/bin/flume-ng-ps1
```

- You will now need to run the following commands to go to the Flume Binaries directory and copy the flume configuration file.

```
cd apache-flume-1.9.0-bin
cp conf/flume-conf.properties.template conf/flume.conf
```

```
[root@ip-172-31-45-198 ~]# cd apache-flume-1.9.0-bin
[root@ip-172-31-45-198 apache-flume-1.9.0-bin]# ls
CHANGELOG  DEVNOTES  LICENSE  NOTICE  README.md  RELEASE-NOTES  bin  conf  doap_Flume.rdf  docs  lib  tools
[root@ip-172-31-45-198 apache-flume-1.9.0-bin]# cp conf/flume-conf.properties.template conf/flume.conf
[root@ip-172-31-45-198 apache-flume-1.9.0-bin]# |
```

- Now, run the following commands to go to the parent /root directory and open the .bash_profile file.

```
cd
vi .bash_profile
```

```
# .bash_profile

# Get the aliases and functions
if [ -f ~/.bashrc ]; then
    . ~/.bashrc
fi

# User specific environment and startup programs

PATH=$PATH:$HOME/bin

export PATH
~
~
~
```

- Copy the following text to the end of this file. Make sure that you press I first to enter the **INSERT** mode.

```
export PATH=$PATH:/root/apache-flume-1.9.0-bin/bin
```

```
export PATH
export PATH=$PATH:/root/apache-flume-1.9.0-bin/bin|
~
```

After this, press the **Esc** button and then type **:wq!** to exit the vi editor.

- Run the following command to update the bash profile.

```
source .bash_profile
```

```
[root@ip-172-31-45-198 ~]# source .bash_profile  
[root@ip-172-31-45-198 ~]# |
```

- You have now installed Flume on your EMR cluster. To verify whether Flume has been successfully installed, run the following command.

```
flume-ng version
```

You can ignore the error that pops up.

```
[root@ip-172-31-45-198 ~]# flume-ng version  
Error: Could not find or load main class org.apache.flume.tools.GetJavaProperty  
Flume 1.9.0  
Source code repository: https://git-wip-us.apache.org/repos/asf/flume.git  
Revision: d4fcab4f501d41597bc616921329a4339f73585e  
Compiled by fszabo on Mon Dec 17 20:45:25 CET 2018  
From source with checksum 35db629a3bda49d23e9b3690c80737f9
```