

# Flume 的安装部署

案例：使用网络 telnet 命令向一台机器发送一些网络数据，然后通过 flume 采集网络端口数据

## 第一步：下载解压修改配置文件

Flume 的安装非常简单，只需要解压即可，当然，前提是已有 hadoop 环境  
上传安装包到数据源所在节点上

这里我们采用在第三台机器来进行安装

```
tar -zxvf flume-ng-1.6.0-cdh5.14.0.tar.gz -C /export/servers/  
cd /export/servers/apache-flume-1.6.0-cdh5.14.0-bin/conf  
cp flume-env.sh.template flume-env.sh
```

```
[root@node03 conf]# ll  
total 16  
-rw-r--r--. 1 1106 4001 1661 Jan 7 2018 flume-conf.properties.template  
-rw-r--r--. 1 1106 4001 1455 Jan 7 2018 flume-env.ps1.template  
-rw-r--r--. 1 1106 4001 1565 Jan 7 2018 flume-env.sh.template  
-rw-r--r--. 1 1106 4001 3107 Jan 7 2018 log4j.properties  
[root@node03 conf]# cp flume-env.sh.template flume-env.sh  
[root@node03 conf]#
```

```
vim flume-env.sh
```

```
export JAVA_HOME=/export/servers/jdk1.8.0_141
```

```
# "License"); you may not use this file except in compliance  
# with the License. You may obtain a copy of the License at  
#  
# http://www.apache.org/licenses/LICENSE-2.0  
#  
# Unless required by applicable law or agreed to in writing, software  
# distributed under the License is distributed on an "AS IS" BASIS,  
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.  
# See the License for the specific language governing permissions and  
# limitations under the License.  
#  
# If this file is placed at FLUME_CONF_DIR/flume-env.sh, it will be sourced  
# during Flume startup.  
#  
# Environment variables can be set here.  
export JAVA_HOME=/export/servers/jdk1.8.0_141  
# Give Flume more memory and pre-allocate, enable remote monitoring via JMX  
# export JAVA_OPTS="-Xms100m -Xmx2000m -Dcom.sun.management.jmxremote"  
# Let Flume write raw event data and configuration information to its log files for debugging  
# purposes. Enabling these flags is not recommended in production,  
# as it may result in logging sensitive user information or encryption secrets.  
# export JAVA_OPTS="$JAVA_OPTS -Dorg.apache.flume.log.rawdata=true -Dorg.apache.flume.log.printconfig=true "  
# Note that the Flume conf directory is always included in the classpath.  
#FLUME_CLASSPATH=""
```

## 第二步：开发配置文件

根据数据采集的需求配置采集方案，描述在配置文件中(文件名可任意自定义)

配置我们的网络收集的配置文件

在 flume 的 conf 目录下新建一个配置文件（采集方案）

```
vim /export/servers/apache-flume-1.6.0-cdh5.14.0-bin/conf/netcat-logger.conf
```

```
[root@node03 conf]# ll
total 20
-rw-r--r--. 1 1106 4001 1661 Jan  7  2018 flume-conf.properties.template
-rw-r--r--. 1 1106 4001 1455 Jan  7  2018 flume-env.ps1.template
-rw-r--r--. 1 root root 1568 Jul 12 10:26 flume-env.sh
-rw-r--r--. 1 1106 4001 1565 Jan  7  2018 flume-env.sh.template
-rw-r--r--. 1 1106 4001 3107 Jan  7  2018 log4j.properties
[root@node03 conf]# vim /export/servers/apache-flume-1.6.0-cdh5.14.0-bin/conf/netcat-logger.conf
```

```
# agent name
a1.sources = r1
a1.sinks = k1
a1.channels = c1

# desc source:r1
a1.sources.r1.type = netcat
a1.sources.r1.bind = 192.168.52.120
a1.sources.r1.port = 44444

# desc sink:k1
a1.sinks.k1.type = logger

# desc channel
a1.channels.c1.type = memory
a1.channels.c1.capacity = 1000
a1.channels.c1.transactionCapacity = 100

# desc source channel sink
a1.sources.r1.channels = c1
a1.sinks.k1.channel = c1
```

```
[root@node03 conf]# vim /export/servers/apache-flume-1.6.0-cdh5.14.0-bin/conf/netcat-logger.conf
a1.sources = r1
a1.sinks = k1
a1.channels = c1

a1.sources.r1.type = netcat
a1.sources.r1.bind = 192.168.52.120
a1.sources.r1.port = 44444

a1.sinks.k1.type = logger

a1.channels.c1.type = memory
a1.channels.c1.capacity = 1000
a1.channels.c1.transactionCapacity = 100

a1.sources.r1.channels = c1
a1.sinks.k1.channel = c1
```

### 第三步：启动配置文件

指定采集方案配置文件，在相应的节点上启动 flume agent

先用一个最简单的例子来测试一下程序环境是否正常

启动 agent 去采集数据

```
bin/flume-ng agent -c conf -f conf/netcat-logger.conf -n a1 -
Dflume.root.logger=INFO,console
```

- c conf 指定 flume 自身的配置文件所在目录
- f conf/netcat-logger.conf 指定我们所描述的采集方案
- n a1 指定我们这个 agent 的名字

```
[root@node03 apache-flume-1.6.0-cdh5.14.0-bin]# bin/flume-ng agent -c conf -f conf/netcat-logger.conf -n a1 -Dflume.root.logger=INFO,console
Info: Sourcing environment configuration script /export/servers/apache-flume-1.6.0-cdh5.14.0-bin/conf/flume-env.sh
Info: Including Hadoop libraries found via (/export/servers/hadoop-2.6.0-cdh5.14.0/bin/hadoop) for HDFS access
Info: Including Hive libraries found via () for Hive access
+ exec /export/servers/jdk1.8.0_141/bin/java -Xmx20m -Dflume.root.logger=INFO,console -cp '/export/servers/apache-flume-1.6.0-cdh5.14.0-bin/conf:/export/servers/apache-flume-1.6.0-cdh5.14.0-bin/lib/*:/export/servers/hadoop-2.6.0-cdh5.14.0/etc/hadoop:/export/servers/hadoop-2.6.0-cdh5.14.0/share/hadoop/common/lib/*:/export/servers/hadoop-2.6.0-cdh5.14.0/share/hadoop/common/*:/export/servers/hadoop-2.6.0-cdh5.14.0/share/hadoop/hdfs:/export/servers/hadoop-2.6.0-cdh5.14.0/share/hadoop/hdfs/lib/*:/export/servers/hadoop-2.6.0-cdh5.14.0/share/hadoop/hdfs/*:/export/servers/hadoop-2.6.0-cdh5.14.0/share/hadoop/yarn/lib/*:/export/servers/hadoop-2.6.0-cdh5.14.0/share/hadoop/yarn/*:/export/servers/hadoop-2.6.0-cdh5.14.0/share/hadoop/mapreduce/lib/*:/export/servers/hadoop-2.6.0-cdh5.14.0/share/hadoop/mapreduce/*:/export/servers/hadoop-2.6.0-cdh5.14.0/contrib/capacity-scheduler/*.jar:/lib/*' -Djava.library.path=/export/servers/hadoop-2.6.0-cdh5.14.0/lib/native org.apache.flume.node.Application -f conf/netcat-logger.conf -n a1
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/export/servers/apache-flume-1.6.0-cdh5.14.0-bin/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/export/servers/hadoop-2.6.0-cdh5.14.0/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
2019-07-12 10:30:56,217 (lifecycleSupervisor-1-0) [INFO - org.apache.flume.node.PollingPropertiesFileConfigurationProvider.start(PollingPropertiesFileConfigurationProvider.java:62)] Configuration provider starting
2019-07-12 10:30:56,234 (conf-file-poller-0) [INFO - org.apache.flume.node.PollingPropertiesFileConfigurationProvider.$FileWatcherRunnable.run(PollingPropertiesFileConfigurationProvider.java:134)] Reloading configuration file:conf/netcat-logger.conf
2019-07-12 10:30:56,256 (conf-file-poller-0) [INFO - org.apache.flume.conf.FlumeConfiguration$AgentConfiguration.addProperty(FlumeConfiguration.java:930)] Added sinks: k1 Agent: a1
```

```
annelFactory.java:42)] Creating instance of channel c1 type memory
2019-07-12 10:30:56,310 (conf-file-poller-0) [INFO - org.apache.flume.node.AbstractConfigurationProvider.loadChannels(AbstractConfigurationProvider.java:201)] Created channel c1
2019-07-12 10:30:56,310 (conf-file-poller-0) [INFO - org.apache.flume.source.DefaultSourceFactory.create(DefaultSourceFactory.java:41)] Creating instance of source r1, type netcat
2019-07-12 10:30:56,325 (conf-file-poller-0) [INFO - org.apache.flume.sink.DefaultSinkFactory.create(DefaultSinkFactory.java:42)] Creating instance of sink: k1, type: logger
2019-07-12 10:30:56,328 (conf-file-poller-0) [INFO - org.apache.flume.node.AbstractConfigurationProvider.getConfigurations(AbstractConfigurationProvider.java:116)] channel c1 connected to [r1, k1]
2019-07-12 10:30:56,336 (conf-file-poller-0) [INFO - org.apache.flume.node.Application.startAllComponents(Application.java:137)] Starting new configuration: { sourceRunners: {r1=EventDrivenSourceRunner: { source:org.apache.flume.source.NetcatSource{name:r1,state:IDLE} } } sinkRunners: {k1=sinkRunner: { policy:org.apache.flume.sink.DefaultSinkProcessorle32806d counterGroup: { name:null counters: { } } } channels: {c1=org.apache.flume.channel.MemoryChannel{name: c1}} }
2019-07-12 10:30:56,345 (conf-file-poller-0) [INFO - org.apache.flume.node.Application.startAllComponents(Application.java:144)] Starting channel c1
2019-07-12 10:30:56,454 (lifecycleSupervisor-1-0) [INFO - org.apache.flume.instrumentation.MonitoredCounterGroup.register(MonitoredCounterGroup.java:119)] Monitored counter group for type: CHANNEL, name: c1: Successfully registered new MBean.
2019-07-12 10:30:56,454 (lifecycleSupervisor-1-0) [INFO - org.apache.flume.instrumentation.MonitoredCounterGroup.start(MonitoredCounterGroup.java:95)] Component type: CHANNEL, name: c1 started
2019-07-12 10:30:56,464 (conf-file-poller-0) [INFO - org.apache.flume.node.Application.startAllComponents(Application.java:171)] Starting sink k1
2019-07-12 10:30:56,467 (conf-file-poller-0) [INFO - org.apache.flume.node.Application.startAllComponents(Application.java:182)] Starting source r1
2019-07-12 10:30:56,475 (lifecycleSupervisor-1-3) [INFO - org.apache.flume.source.NetcatSource.start(NetcatSource.java:155)] Source starting
2019-07-12 10:30:56,489 (lifecycleSupervisor-1-3) [INFO - org.apache.flume.source.NetcatSource.start(NetcatSource.java:166)] Created serverSocket:sun.nio.ch.ServerSocketChannelImpl[/192.168.52.120:44444]
```

## 第四步：安装 telnet 准备测试

在 node02 机器上面安装 telnet 客户端，用于模拟数据的发送

```
yum -y install telnet
```

```
telnet node03 44444 # 使用 telnet 模拟数据发送
```

```
[root@node02 ~]# yum -y install telnet
Loaded plugins: fastestmirror, security
Setting up Install Process
Loading mirror speeds from cached hostfile
 * base: mirrors.aliyun.com
 * extras: mirrors.aliyun.com
 * updates: mirrors.huaweicloud.com
base                                     3.7 kB    00:00
extras                                 3.4 kB    00:00
updates                               3.4 kB    00:00
updates/primary_db                    5.1 MB    00:04
Resolving Dependencies
--> Running transaction check
--> Package telnet.x86_64 1:0.17-48.el6 will be installed
--> Finished Dependency Resolution

Dependencies Resolved

=====
Package                               Arch      Version              Repository           Size
=====
Installing:
telnet                                x86_64    1:0.17-48.el6        base                 58 k
Transaction Summary
-----
Install      1 Package(s)
Total download size: 58 k
Installed size: 109 k
```

```
[root@node02 ~]# telnet node03 4444
Trying 192.168.52.120...
Connected to node03.
Escape character is '^['.
minglan
OK
```

```
192.168.52.120 | 192.168.52.120 [0]
s(AbstractConfigurationProvider.java:201)] Created channel c1
2019-07-12 10:30:56,310 (conf-file-poller-0) [INFO - org.apache.flume.source.DefaultSourceFactory.create(DefaultSourceFactory.java:41)] Creating instance of source r1, type netcat
2019-07-12 10:30:56,325 (conf-file-poller-0) [INFO - org.apache.flume.sink.DefaultSinkFactory.create(DefaultSinkFactory.java:42)] Creating instance of sink: k1, type: logger
2019-07-12 10:30:56,328 (conf-file-poller-0) [INFO - org.apache.flume.node.AbstractConfigurationProvider.getConfigurations(AbstractConfigurationProvider.java:116)] channel c1 connected to [r1, k1]
2019-07-12 10:30:56,336 (conf-file-poller-0) [INFO - org.apache.flume.node.Application.startAllComponents(Application.java:137)] Starting new configuration:{ sourceRunners:{r1=EventDrivenSourceRunner: { source:org.apache.flume.source.NetcatSource{name:r1,state:IDLE} }} sinkRunners:{k1=SinkRunner: { policy:org.apache.flume.sink.DefaultSinkProcessor@1e32806d counterGroup:{ name:null counters:{} } }} channels:{c1=org.apache.flume.channel.MemoryChannel{name: c1}} }
2019-07-12 10:30:56,345 (conf-file-poller-0) [INFO - org.apache.flume.node.Application.startAllComponents(Application.java:144)] Starting channel c1
2019-07-12 10:30:56,454 (lifecycleSupervisor-1-0) [INFO - org.apache.flume.instrumentation.MonitoredCounterGroup.register(MonitoredCounterGroup.java:119)] Monitored counter group for type: CHANNEL, name: c1: Successfully registered new MBean.
2019-07-12 10:30:56,454 (lifecycleSupervisor-1-0) [INFO - org.apache.flume.instrumentation.MonitoredCounterGroup.start(MonitoredCounterGroup.java:95)] Component type: CHANNEL, name: c1 started
2019-07-12 10:30:56,464 (conf-file-poller-0) [INFO - org.apache.flume.node.Application.startAllComponents(Application.java:171)] Starting Sink k1
2019-07-12 10:30:56,467 (conf-file-poller-0) [INFO - org.apache.flume.node.Application.startAllComponents(Application.java:182)] Starting Source r1
2019-07-12 10:30:56,475 (lifecycleSupervisor-1-3) [INFO - org.apache.flume.source.NetcatSource.start(NetcatSource.java:155)] Source starting
2019-07-12 10:30:56,489 (lifecycleSupervisor-1-3) [INFO - org.apache.flume.source.NetcatSource.start(NetcatSource.java:166)] Created serverSocket:sun.nio.ch.ServerSocketChannelImpl[/192.168.52.120:4444]
2019-07-12 10:34:06,535 (SinkRunner-PollingRunner-DefaultSinkProcessor) [INFO - org.apache.flume.sink.LoggersSinkProcessor(LoggersSink.java:95)] Event: { headers:{} body: 6D 69 6E 67 6C 61 6E 20 0D minglan }
```