暨南大学本科实验报告专用纸

课程名称 云计算实验 成绩评定

实验项目名称 分布式日志数据采集工具 flume 的部署与使用

指导教师 魏林锋

实验项目编号 0806030805 实验项目类型 综合 实验地点 线上

学生姓名 陈宇 学号 2020101642

学院 信息科学技术学院 系 计算机 专业 软件工程

实验时间 2022 年 10 月 26 日上午～10 月 26 日上午 温度 ℃湿度

5.1 实验目的

1）理解 Flume 工作原理。

2）通过实验掌握分布式日志数据采集工具 Flume 的安装过程。

5.2 实验内容

1）完成分布式日志数据采集工具 Flume 的安装。

5.3 实验环境

已经配置完成的 Hadoop 伪分布式或完全分布式环境。环境配置如下：

Hadoop01：192.168.24.91 Hadoop02：192.168.24.92 Hadoop03：192.168.24.93 管理员用户：root / admin@1 Hadoop 用户：hadoop / hadoop

5.4 实验步骤

1、使用 xftp 工具将 apache-flume-1.9.0-bin.tar.gz 文件上传到服务器上，解压 apache-flume-1.9.0-bin.tar.gz 文件，并重命名 flume 文件夹。命令如下：

```
[root@master ~]# tar -zvxf apache-flume-1.9.0-bin.tar.gz -C /usr/
[root@master ~]# mv /usr/apache-flume-1.9.0-bin/ /usr/flume
```

```
apache-flume-1.9.0-bin/lib/xalan-2.7.2.jar
apache-flume-1.9.0-bin/lib/flume-scribe-source-1.9.0.jar
apache-flume-1.9.0-bin/lib/flume-jms-source-1.9.0.jar
apache-flume-1.9.0-bin/lib/geronimo-jms_1.1_spec-1.1.1.jar
apache-flume-1.9.0-bin/lib/flume-twitter-source-1.9.0.jar
apache-flume-1.9.0-bin/lib/twitter4j-core-3.0.3.jar
apache-flume-1.9.0-bin/lib/twitter4j-media-support-3.0.3.jar
apache-flume-1.9.0-bin/lib/twitter4j-stream-3.0.3.jar
apache-flume-1.9.0-bin/lib/flume-kafka-source-1.9.0.jar
apache-flume-1.9.0-bin/lib/flume-avro-source-1.9.0.jar
apache-flume-1.9.0-bin/lib/flume-thrift-source-1.9.0.jar
apache-flume-1.9.0-bin/lib/flume-taildir-source-1.9.0.jar
apache-flume-1.9.0-bin/lib/flume-ng-environment-variable-config-filter-1.9.0.jar
apache-flume-1.9.0-bin/lib/flume-ng-hadoop-credential-store-config-filter-1.9.0.
jar
apache-flume-1.9.0-bin/lib/flume-ng-external-process-config-filter-1.9.0.jar
apache-flume-1.9.0-bin/lib/flume-ng-log4jappender-1.9.0.jar
apache-flume-1.9.0-bin/lib/flume-tools-1.9.0.jar
apache-flume-1.9.0-bin/lib/slf4j-log4j12-1.7.25.jar
apache-flume-1.9.0-bin/lib/apache-log4j-extras-1.1.jar
apache-flume-1.9.0-bin/tools/flume-ng-log4jappender-1.9.0-jar-with-dependencies.
jar
[root@master 云计算]# mv /usr/apache-flume-1.9.0-bin/ /usr/flume
[root@master 云计算]#
```

2、添加环境变量，并使其生效。命令如下：

```
[root@master ~]# vi /etc/profile
export FLUME_HOME=/usr/flume

export PATH=$FLUME_HOME/bin:$PATH
[root@master ~]# source /etc/profile
```

```
export FLUME_HOME=/usr/flume
export PATH=$FLUME_HOME/bin:$PATH
```

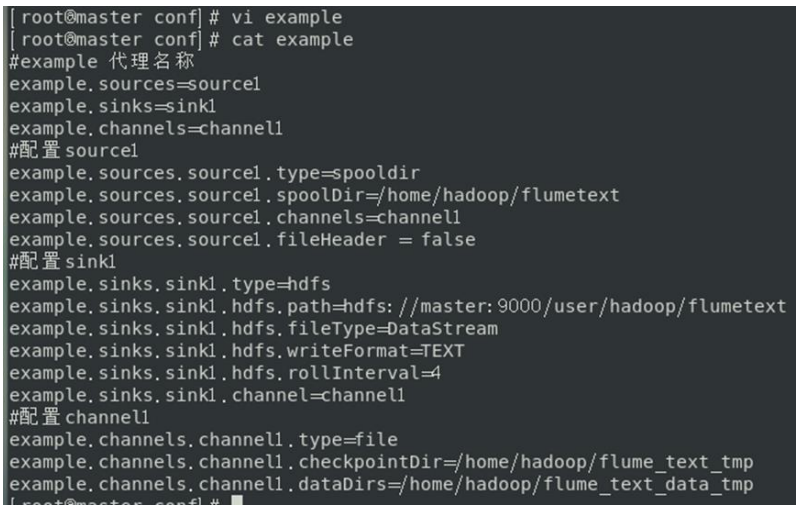
3、进入/usr/flume/conf/目录，创建 flume-env.sh 文件。命令如下：

```
[root@master ~]# cd /usr/flume/conf/
[root@master conf]# cp flume-env.sh.template flume-env.sh
[root@master conf]# vi flume-env.sh
#指定 JAVA 的安装路径
export JAVA_HOME=/usr/java/jdk1.8.0_144/
```

```
[root@master ~]# cd /usr/flume/conf/
[root@master conf]# cp flume-env.sh.template flume-env.sh
[root@master conf]# vi flume-env.sh
[root@master conf]#
```

4、创建并配置 example 文件。命令如下：

```
[root@master conf]# vi example
#example 代理名称
example.sources=source1
example.sinks=sink1
example.channels=channel1
#配置 source1
example.sources.source1.type=spooldir
example.sources.source1.spoolDir=/home/hadoop/flumetext
example.sources.source1.channels=channel1
example.sources.source1.fileHeader = false
#配置 sink1
example.sinks.sink1.type=hdfs
example.sinks.sink1.hdfs.path=hdfs://master:9000/user/hadoop/flumetext
example.sinks.sink1.hdfs.fileType=DataStream
example.sinks.sink1.hdfs.writeFormat=TEXT
example.sinks.sink1.hdfs.rollInterval=4
example.sinks.sink1.channel=channel1
#配置 channel1
example.channels.channel1.type=file
example.channels.channel1.checkpointDir=/home/hadoop/flume_text_tmp
example.channels.channel1.dataDirs=/home/hadoop/flume_text_data_tmp
```



5、修改 flume 文件夹权限。命令如下：

```
[root@master conf]# chown -R hadoop:hadoop /usr/flume
```

6、切换到 hadoop 用户，创建 flumetext 文件夹。命令如下：

```
[root@master conf]# su hadoop
[hadoop@master conf]$ mkdir -p /home/hadoop/flumetext
```

```
[root@master conf]# chown -R hadoop:hadoop /usr/flume
[root@master conf]# su hadoop
[hadoop@master conf]$ mkdir -p /home/hadoop/flumetext
[hadoop@master conf]$
```

7、启动 Hadoop 集群。命令如下：

```
[hadoop @master ~]$ start-all.sh
```

8、启动 flume。命令如下：

```
[hadoop @master ~]$ flume-ng agent -n example -c conf -f /usr/flume/conf/example -Dflume.root.logger=INFO,console
```

```
2022-10-26 09:43:51,923 INFO file.LogFile: Opened /home/hadoop/flume_text_data_tmp/log-2
2022-10-26 09:43:51,930 INFO file.Log: Roll end
2022-10-26 09:43:51,930 INFO file.EventQueueBackingStoreFile: Start checkpoint for /home/hadoop/flume_text_tmp/checkpoint, elements to sync = 0
2022-10-26 09:43:51,932 INFO file.EventQueueBackingStoreFile: Updating checkpoint metadata: logWriteOrderID: 1666748631677, queueSize: 0, queueHead: 0
2022-10-26 09:43:51,940 INFO file.Log: Updated checkpoint for file: /home/hadoop/flume_text_data_tmp/log-2 position: 0 logWriteOrderID: 1666748631677
2022-10-26 09:43:51,940 INFO file.FileChannel: Queue Size after replay: 0 [channel=channel1]
2022-10-26 09:43:51,940 INFO node.Application: Starting Sink sink1
2022-10-26 09:43:51,941 INFO node.Application: Starting Source source1
2022-10-26 09:43:51,941 INFO source.SpoolDirectorySource: SpoolDirectorySource source starting with directory: /home/hadoop/flumetext
2022-10-26 09:43:51,942 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SINK, name: sink1: Successfully registered new MBean.
2022-10-26 09:43:51,942 INFO instrumentation.MonitoredCounterGroup: Component type: SINK, name: sink1 started
2022-10-26 09:43:51,949 INFO instrumentation.MonitoredCounterGroup: Monitored counter group for type: SOURCE, name: source1: Successfully registered new MBean.
2022-10-26 09:43:51,949 INFO instrumentation.MonitoredCounterGroup: Component type: SOURCE, name: source1 started
```

9、新建另一个终端，创建/user/hadoop/flumetext 目录。命令如下：

```
[hadoop @master ~]$ hdfs dfs -mkdir -p /user/hadoop/flumetext
```

```
[hadoop@master ~]$ hdfs dfs -mkdir -p /user/hadoop/flumetext
[hadoop@master ~]$
```

10、往 flume 的 source 写入数据。命令如下：

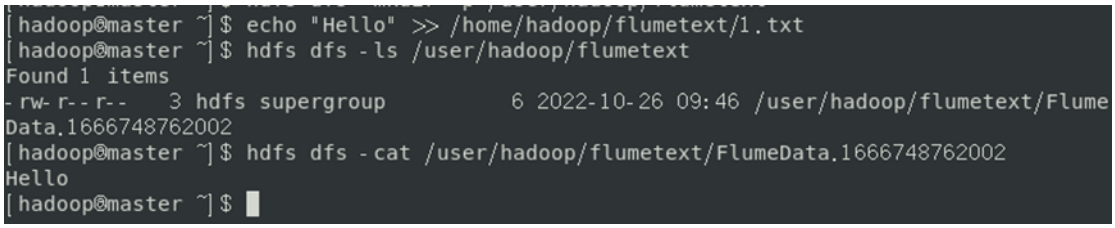
```
[hadoop @master ~]$ echo "Hello" >> /home/hadoop/flumetext/1.txt
```

11、查看 flume 的 sink（即 hdfs 上对应的目录）是否写入成功。命令如下：

```
[hadoop @master ~]$ hdfs dfs -ls /user/hadoop/flumetext

[hadoop @master ~]$ hdfs dfs -cat /user/hadoop/flumetext/FlumeData.1666596349806
```

```
[hadoop@master ~]$ hdfs dfs -cat /user/hadoop/flumetext/FlumeData.1666596349806
hello
```



注意：内容为 hello

实验总结

理解了 Flume 工作原理。

通过实验掌握分布式日志数据采集工具 Flume 的安装过程。

本次实验比较简单，但是还是会出现问题。