目录

[1. 实验一：Flume组件安装配置 2](#_Toc11982)

[1.1. 实验目的 2](#_Toc4994)

[1.2. 实验要求 2](#_Toc10035)

[1.3. 实验环境 2](#_Toc18215)

[1.4. 实验过程 2](#_Toc13540)

[1.4.1. 实验任务一：下载和解压Flume 2](#_Toc9940)

[1.4.2. 实验任务二：Flume组件部署 3](#_Toc29630)

[1.4.3. 实验任务三：使用Flume发送和接受信息 4](#_Toc26688)

# 实验一：Flume组件安装配置

## 实验目的

完成本实验，您应该能够：

* 掌握下载和解压Flume
* 掌握Flume组件部署
* 掌握使用Flume发送和接受信息

## 实验要求

* 了解Flume相关知识
* 熟悉Flume功能应用
* 熟悉Flume组件设置

## 实验环境

本实验所需之主要资源环境如表1-1所示。

表1-1 资源环境

|  |  |
| --- | --- |
| **服务器集群** | 3个以上节点，节点间网络互通，各节点最低配置：双核CPU、8GB内存、100G硬盘 |
| **运行环境** | CentOS 7.4 |
| **大数据平台** | H3C DataEngine-E0104-RHEL6-X86\_64 |
| **服务和组件** | Hadoop、JDK等，其他服务根据实验需求安装 |

## 实验过程

### 实验任务一：下载和解压Flume

可以从官网下载Flume组件安装包，下载地址如下URL链接所示[https://archive.apache.org/dist/flume/1.6.0/](https://archive.apache.org/dist/flume/1.7.0/)。

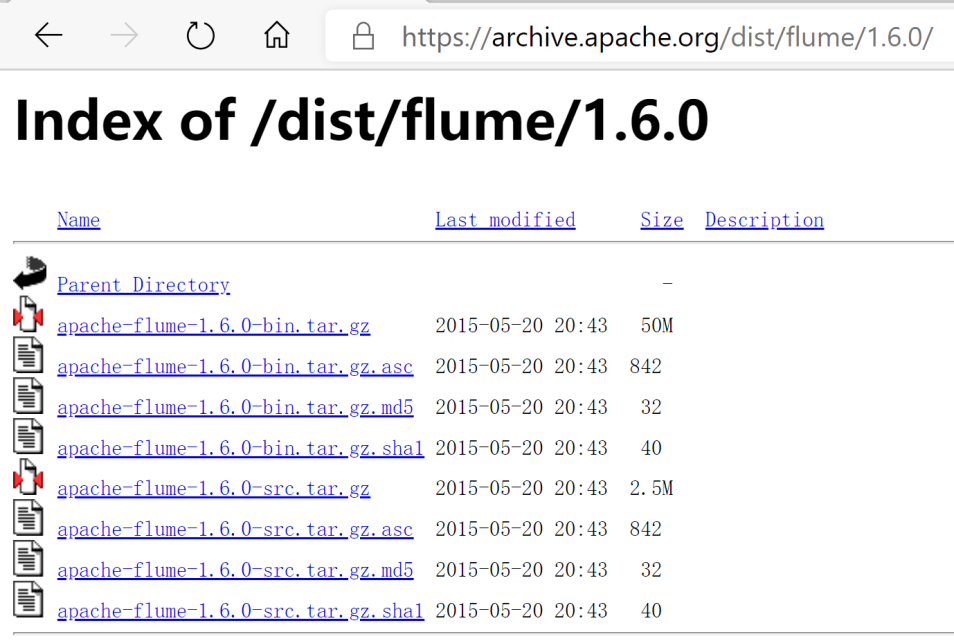


图10-2 Flume组件下载地址

使用root用户解压Flume安装包到“/usr/local/src”路径，并修改解压后文件夹名为flume。

[root@master ~]#tar zxvf /opt/software/apache-flume-1.6.0-bin.tar.gz -C /usr/local/src

[root@master ~]#cd /usr/local/src/

[root@master ~]#mv apache-flume-1.6.0-bin/ flume # 修改Flume安装路径文件夹名称

[root@master ~]#chown -R hadoop:hadoop flume/ # 修改文件夹归属用户和归属组为hadoop用户和hadoop组

### 实验任务二：Flume组件部署

#### 步骤一：使用root用户设置Flume环境变量，并使环境变量对所有用户生效。

[root@master ~]#vi /etc/profile # 编辑系统环境变量配置文件

# set flume environment

export FLUME\_HOME=/usr/local/src/flume # Flume安装路径

export PATH=$PATH:$FLUME\_HOME/bin # 添加系统PATH环境变量

#### 步骤二：修改Flume相应配置文件。

首先，切换到hadoop用户，并切换当前工作目录到Flume的配置文件夹。

[root@master ~]#su - hadoop

[hadoop@master ~]$source /etc/profile

[hadoop@master ~]$cd /usr/local/src/flume/conf

拷贝flume-env.sh.template文件并重命名为flume-env.sh。

[hadoop@master ~]$cp flume-env.sh.template flume-env.sh

#### 步骤三：修改并配置flume-env.sh文件。

删除JAVA\_HOME变量前的注释，修改为JDK的安装路径。

[hadoop@master ~]$ vi /usr/local/src/flume/conf/flume-env.sh

# 修改JAVA\_HOME参数值为jdk安装路径

export JAVA\_HOME=/usr/local/src/java

使用flume-ng version命令验证安装是否成功，若能够正常查询Flume组件版本为1.6.0，则表示安装成功。

[hadoop@master ~]$ flume-ng version

**Flume 1.6.0 # 查询到Flume版本为1.6.0**

**Source code repository: https://git-wip-us . apache . org/ repos/asf/ flume. git**

**Revision: 2561a23240a71ba20bf288c7c2cda88f443c2080**

**Compiled by hshreedharan on Mon May 11 11:15:44 PDT 2015**

**From source with checksum b29e416802ce9ece3269d34233baf43f**

### 实验任务三：使用Flume发送和接受信息

通过Flume将Web服务器中数据传输到HDFS中。

#### 步骤一：在Flume安装目录中创建simple-hdfs-flume.conf文件。

[hadoop@master ~]$ cd /usr/local/src/flume/

[hadoop@master ~]$ vi /usr/local/src/flume/simple-hdfs-flume.conf

#添加以下内容

a1.sources=r1

a1.sinks=k1

a1.channels=c1

a1.sources.r1.type=spooldir

a1.sources.r1.spoolDir=/usr/local/src/hadoop/logs/

a1.sources.r1.fileHeader=true

a1.sinks.k1.type=hdfs

a1.sinks.k1.hdfs.path=hdfs://master:9000/tmp/flume

a1.sinks.k1.hdfs.rollsize=1048760

a1.sinks.k1.hdfs.rollCount=0

a1.sinks.k1.hdfs.rollInterval=900

a1.sinks.k1.hdfs.useLocalTimeStamp=true

a1.channels.c1.type=file

a1.channels.c1.capacity=1000

a1.channels.c1.transactionCapacity=100

a1.sources.r1.channels = c1

a1.sinks.k1.channel = c1

#### 步骤二：使用flume-ng agent命令加载simple-hdfs-flume.conf配置信息，启动flume传输数据。

[hadoop@master flume] $ flume-ng agent --conf-file simple-hdfs-flume.conf --name a1

ctrl+c退出flume传输

#### 步骤三：查看Flume传输到HDFS的文件，若能查看到HDFS上/tmp/flume目录有传输的数据文件，则表示数据传输成功。

[hadoop@master flume] $ hdfs dfs -ls /tmp/f1ume # 查看HDFS文件系统/tmp/f1ume目录下文件

**-rw-r--r-- 2 root super group 1325 2019-06-05 11:14 /tmp/f1ume/F lumeData. 1559747635008**

**-rw-r--r-- 2 root super group 1344 2019-06-05 11:14 /tmp/f1ume/F1umeData. 1559747635009**

**-rw-r--r-- 2 root super qroup 1442 2019-06-05 11:14 /tmp/f1ume/F lumeData.1559747635010**