

# Flume运维管理

[www.huawei.com](http://www.huawei.com)





# 目录

学完本课程后，您将能够：

- 掌握**Flume**服务健康状态检查方法
- 掌握**Flume**服务基本故障恢复



# 目录

1. 目录介绍
2. 服务检查
3. 日志介绍
4. 常用维护命令
5. 常用性能调优参数
6. 常见问题及定位手段

# Flume安装目录

- **Flume**服务端安装目录：

**{\$BIGDATA\_HOME}/huawei/Bigdata/FusionInsight/ FusionInsight-Flume-1.6.0**

- **FusionInsight-Flume-1.6.0**的目录结构

<b>FusionInsight-Flume-1.6.0</b>	<b>Flume</b> 的安装目录
<b>FusionInsight-Flume-1.6.0 / adapter</b>	<b>Flume</b> 适配 <b>om</b> 目录
<b>FusionInsight-Flume-1.6.0 /flume</b>	<b>Flume</b> 的角色安装目录
<b>FusionInsight-Flume-1.6.0 /flume/bin</b>	<b>Flume</b> 启动、停止、状态检查等脚本目录
<b>FusionInsight-Flume-1.6.0 /flume/conf</b>	<b>Flume</b> 的配置目录
<b>FusionInsight-Flume-1.6.0 /flume/inst</b>	<b>Flume</b> 的安装、卸载等脚本
<b>FusionInsight-Flume-1.6.0 /flume/lib</b>	<b>Flume</b> 运行时 <b>jar</b> 包目录
<b>FusionInsight-Flume-1.6.0 /flume/libexec</b>	<b>Flume</b> 运行时第三 方动态库目录
<b>FusionInsight-Flume-1.6.0 /flume/plugins.d</b>	<b>Flume</b> 插件包目录
<b>FusionInsight-Flume-1.6.0 /flume/plugins.s</b>	<b>Flume</b> 插件脚本目录
<b>FusionInsight-Flume-1.6.0 /flume/tools</b>	<b>Flume java</b> 工具目录

# Flume数据目录

- 数据目录

**Flume**数据目录只有在**channel**是**File**类型时才有数据目录，同时该目录是在**Flume**运行时的配置文件**properties.properties**中如下的参数指定的：

**server.channels.xx.dataDirs =**

**server.channels.xx.checkpointDir =**

# Flume日志目录

- 客户端运行日志目录

默认的日志目录: **`/var/log/Bigdata/flume-client`**

- 服务端运行日志目录

默认的日志目录: **`/var/log/Bigdata/flume`**



# 目录

1. 目录介绍
- 2. 服务检查**
3. 日志介绍
4. 常用维护命令
5. 常用性能调优参数
6. 常见问题及定位手段

# Flume服务检查

- Flume server服务状态检查

登录FusionInsight管理页面，选择 服务管理 > Flume，查看Flume概述中健康状态

。

角色	操作状态	健康状态
Flume	3 已启动	3 良好
MonitorServer	2 已启动	2 良好



# Flume服务检查

- **Flume server**角色实例检查

登录**FusionInsight**管理页面，选择 服务管理 > **Flume** > 实例，查看每个角色实例的健康状态是否为良好。

服务

Flume 实例

服务状态

实例

服务配置

资源贡献排名

+ 添加实例

更多操作

每30秒刷新一次

<input type="checkbox"/>	角色	主机名	管理IP	业务IP	机架	操作状态
<input type="checkbox"/>	Flume	Loader13	192.168.1.10	192.168.1.10	/default/rack0	已启动
<input type="checkbox"/>	Flume	Loader19	192.168.1.11	192.168.1.11	/default/rack0	已启动
<input type="checkbox"/>	MonitorServer	Loader13	192.168.1.13	192.168.1.13	/default/rack0	已启动
<input type="checkbox"/>	MonitorServer	Loader19	192.168.1.15	192.168.1.15	/default/rack0	已启动

当前页: 1

总页数: 1

跳转到

确定

# Flume服务检查

- **Flume** 客户端状态检查

登录客户端所在节点，执行如下的命令查询客户端进程是否存在：

```
ps -ef |grep flume |grep -v grep
```

正常情况下，上述命令应该返回两个进程：一个是**Flume**客户端进程，一个是客户端监控进程。



# 目录

1. 目录介绍
2. 服务检查
3. 日志介绍
4. 常用维护命令
5. 常用性能调优参数
6. 常见问题及定位手段

# 日志介绍——安装日志

- **Flume** 服务端安装日志的存储路径为：  
  
`/var/log/Bigdata/flume/flume/install.log`
- **Flume** 客户端安装日志在用户安装时直接输出在终端上

# 日志介绍——运行日志

- **Flume** 服务端运行日志的存储路径为：

**/var/log/Bigdata/flume/角色**

日志文件名	分析描述
flumeServer-gc.log	Flume运行时GC日志
flumeServer.log	Flume进程运行时产生的大部分日志
prestartDetail.log	Flume进程预启动日志
startDetail.log	Flume启动日志
monitorServer.log	Moitor server运行时日志
startDetail.log	Moitor server启动时日志

# 日志介绍——运行日志

- **Flume** 客户端运行日志的存储路径为：

**/var/log/Bigdata/flume-client/角色**

日志文件名	分析描述
<b>FlumeClient.log</b>	<b>Flume</b> 客户端运行时产生的大部分日志
<b>flumeMonitor.log</b>	<b>Flume</b> 客户端监控进程日志
<b>flumeMonitorChecker.log</b>	<b>Flume</b> 客户端 <b>crontab</b> 监控运行日志



# 目录

1. 目录介绍
2. 服务检查
3. 日志介绍
4. 常用维护命令
5. 常用性能调优参数
6. 常见问题及定位手段

# 常用维护命令

- 客户端安装、卸载命令

在客户端安装包解压目录下，执行如下的命令完成客户端安装：

```
./install.sh -d /opt/FlumeClient -f ip -c  
flume/conf/client.properties.properties -l /var/log/Bigdata
```

在客户端安装目录下，执行如下的命令卸载客户端：

**客户端安装录/fusioninsight-flume-1.6.0/inst/uninstall.sh**

- **Flume**客户端启动、停止、强制启动、强制停止操作命令：

**客户端安装目录/fusioninsight-flume-1.6.0/bin/flume-manage.sh start**

**客户端安装目录/fusioninsight-flume-1.6.0/bin/flume-manage.sh stop**

**客户端安装目录/fusioninsight-flume-1.6.0/bin/flume-manage.sh start force**

**客户端安装目录/fusioninsight-flume-1.6.0/bin/flume-manage.sh start force**





# 目录

1. 目录介绍
2. 服务检查
3. 日志介绍
4. 常用维护命令
- 5. 常用性能调优参数**
6. 常见问题及定位手段

# 常用的性能调优命令或参数

- **Flume**自身配置调优

- 在**Manager**界面调整**Flume**实例的“**GC\_OPTS**”参数值，加入如下参数设定直接内存大小，优化**GC** 算法

**-XX:+UseParallelGC -XX:ParallelGCThreads=20 -XX:+UseParallelOldGC**

- 可通过修改“**properties.properties**”文件实现调优效果，主要修改如下：

**1：针对spooldir source增加event合并，增加参数deserializer.maxBatchLine = 20**

上述表示将**20**行合并成一行

**2：增加source/sink的batchSize（参数名为batchSize）值，将一次提交的event数提高**

**3：增加channel的capacity(参数名为capacity)，提升channel的缓存量，防止sink因为取不到数据而空等待，提升性能**



# 目录

1. 目录介绍
2. 服务检查
3. 日志介绍
4. 常用维护命令或参数
5. 常用性能调优命令或参数
6. 常见问题以及定位手段

# 常见问题1

- 故障：**Flume**服务不可用
- 现象描述

在**FusionInsight Manager** 界面有**Flume**服务不可用告警，**Flume**服务的健康状态为故障

- 可能原因
  - **Flume**服务所依赖的服务故障
  - 网络连接中断导致**Flume**与依赖的服务无法通讯
- 定位思路
  - 查看**Kerberos**与**HDFS**服务是否正常
  - 查看两个**Flume**的进程是否运行

# 常见问题1

## 处理步骤

- 1.登录**FusionInsight Manager**系统。
- 2.在服务列表中查看**Flume**服务所依赖的**Kerberos**和**HDFS**服务健康状态是否为“良好”。  
是，执行步骤4。否，执行步骤3。
- 3.选择健康状态不是良好的服务，查看故障是否恢复。  
是，操作结束。否，执行步骤4。
- 4.使用**PuTTY**工具，以**root**用户登录任意一个**Flume**实例的服务器，  
执行**ps -ef |grep flume |grep -v grep**。  
是，执行步骤6。否，执行步骤5。
- 5.联系系统管理员处理网络故障。若故障恢复，则操作结束，否则执行步骤6。
- 6.选择“系统设置 > 日志下载”，获取相应日志。  
在“服务”下拉框中勾选“**Flume**”，单击“确定”  
设置日志的起始收集时间“开始时间”和“结束时间”，单击“下载”。
- 7.联系技术支持，并发送已收集的故障日志信息

# 常见问题2

- 问题现象

在**HDFS**管理页面，发现**Flume**没有传输数据

- 问题分析

配置文件错误，导致**Flume**不能按照指定的内容采集日志

- 解决办法

- 1) 登录**Flume**集群中其他节点，查看**Flume**运行日志中是否有异常信息打印
- 2) 查看指定的采集源下是否有可采集的数据



## 思考题

1. **Flume**采集数据到**HDFS**上时，**HDFS**上看不到相应的数据，可能存在的原因有哪些。

# 习题

1. **Flume**数据目录只有在**channel**是**File**类型时才有数据目录是（）

- A.是                      B.否

2.下列不是提升**Flume**性能方法是（）

- A.优化**Flume GC**参数
- B. **Envet**在采集时尽量合并
- C.修改**batchSize**和**capacity**
- D. 增加**Flume**级联节点





## 本章总结

- 本章主要介绍了**Flume**的部署安装常用命令、健康检查以及常见故障定位方法，学习完本章后能够清楚**Flume**的安装部署流程，并能够自行安装**Flume**，同时能够进行简单问题的处理。

# Thank you

[www.huawei.com](http://www.huawei.com)