

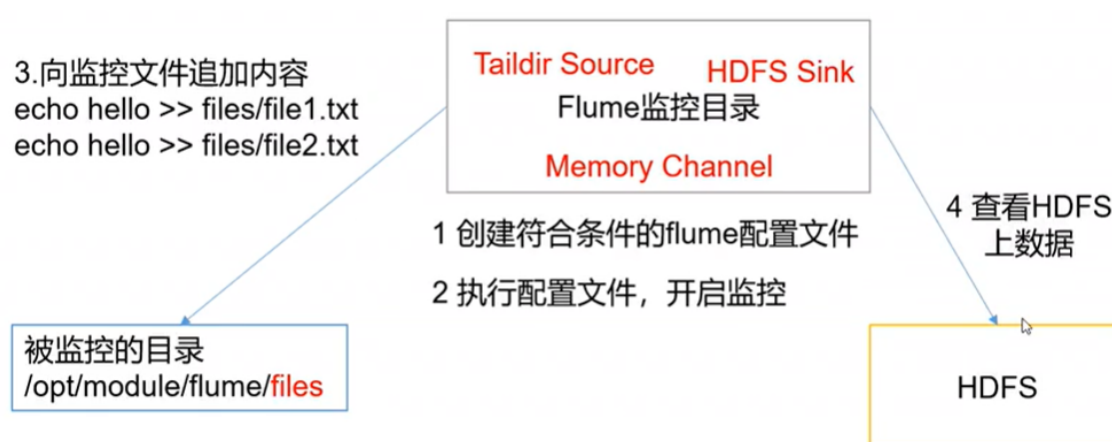
# Flume6 实时监控目录下的多个追加文件

Exec source 适用于监控一个实时追加的文件，但不能保证数据不丢失；Spooldir Source 能够保证数据不丢失，且能够实现断点续传，但延迟比较高，不能实时监控；而Taildir Source既能实现断点续传，又可以保证数据不丢失，还能够进行实时监控。

## 案例需求

使用Flume监听整个目录实时追加文件，并上传至 HDFS

## 需求分析



## 实验步骤

- 建配置文件 flume-file-logger.conf

创建一个文件

```
[atguigu@hadoop102 job]$ touch flume-file-logger.conf
```

打开文件

```
[atguigu@hadoop102 job]$ vim flume-file-logger.conf
```

添加如下内容

```
# Name the components on this agent
a1.sources = r1
a1.sinks = k1
a1.channels = c1

# Describe/configure the source
a1.sources.r1.type = TAILDIR
a1.sources.r1.filegroups = f1,f2
a1.sources.r1.filegroups.f1 = /opt/module/flume/files/file1.txt
```

```

a1.sources.r1.filegroups.f2 = /opt/module/flume/files/file2.txt
a1.sources.r1.positionFile = /opt/module/flume/position/position.json

# Describe the sink
a1.sinks.k1.type = logger

# Use a channel which buffers events in memory
a1.channels.c1.type = memory
a1.channels.c1.capacity = 1000
a1.channels.c1.transactionCapacity = 100

# Bind the source and sink to the channel
a1.sources.r1.channels = c1
a1.sinks.k1.channel = c1

```

- 启动监控文件夹命令

```

[atguigu@hadoop102 flume]$ bin/flume-ng agent -c conf/ -f job/flume-file-logger.conf -n a1 -Dflume.root.logger=INFO,console

```

- 向 files 文件夹中添加文件

在/opt/module/flume 目录下创建 upload 文件夹

```

[atguigu@hadoop102 flume]$ mkdir files

```

向 files 文件夹中的文件添加内容

```

[atguigu@hadoop102 files]$ echo "hello" >> file1.txt
[atguigu@hadoop102 files]$ echo "lili" >> file2.txt

```

- 将flume下线，然后追加其他数据

```

[atguigu@hadoop102 files]$ echo "hello2" >> file1.txt
[atguigu@hadoop102 files]$ echo "lili2" >> file2.txt

```

- 然后重新上线

```

[atguigu@hadoop102 flume]$ bin/flume-ng agent -c conf/ -f job/flume-file-logger.conf -n a1 -Dflume.root.logger=INFO,console

```