需求分析：

采集需求：比如业务系统使用 log4j 生成的日志，日志内容不断增加，需要把追加到日志文件中的数据实时采集到 hdfs

根据需求，首先定义以下 3 大要素
- 采集源，即 source——监控文件内容更新 ： exec　'tail -F file'
- 下沉目标，即 sink——HDFS 文件系统 ： hdfs sink
- Source 和 sink 之间的传递通道——channel，可用 file channel 也可以用 内存 channel

定义 flume 的配置文件

node03 开发配置文件

```
cd /export/servers/apache-flume-1.6.0-cdh5.14.0-bin/conf
vim tail-file.conf
```

配置文件内容

```
agent1.sources = source1
agent1.sinks = sink1
agent1.channels = channel1

# Describe/configure tail -F source1
agent1.sources.source1.type = exec
agent1.sources.source1.command = tail -F /export/servers/taillogs/access_log
agent1.sources.source1.channels = channel1

#configure host for source
#agent1.sources.source1.interceptors = i1
#agent1.sources.source1.interceptors.i1.type = host
#agent1.sources.source1.interceptors.i1.hostHeader = hostname

# Describe sink1
agent1.sinks.sink1.type = hdfs
#agent1.sinks.sink1.channel =    channel1
agent1.sinks.sink1.hdfs.path          =          hdfs://192.168.52.100:8020/weblog/flume-collection/%y-%m-%d/%H-%M
agent1.sinks.sink1.hdfs.filePrefix = access_log
agent1.sinks.sink1.hdfs.maxOpenFiles = 5000
agent1.sinks.sink1.hdfs.batchSize= 100
agent1.sinks.sink1.hdfs.fileType = DataStream
agent1.sinks.sink1.hdfs.writeFormat =Text
```

```
agent1.sinks.sink1.hdfs.rollSize = 102400
agent1.sinks.sink1.hdfs.rollCount = 1000000
agent1.sinks.sink1.hdfs.rollInterval = 60
agent1.sinks.sink1.hdfs.round = true
agent1.sinks.sink1.hdfs.roundValue = 10
agent1.sinks.sink1.hdfs.roundUnit = minute
agent1.sinks.sink1.hdfs.useLocalTimeStamp = true

# Use a channel which buffers events in memory
agent1.channels.channel1.type = memory
agent1.channels.channel1.keep-alive = 120
agent1.channels.channel1.capacity = 500000
agent1.channels.channel1.transactionCapacity = 600

# Bind the source and sink to the channel
agent1.sources.source1.channels = channel1
agent1.sinks.sink1.channel = channel1
```

创建文件夹
```
mkdir -p /export/servers/taillogs
```

启动 flume

```
cd    /export/servers/apache-flume-1.6.0-cdh5.14.0-bin
```

```
bin/flume-ng    agent    -c    conf    -f    conf/tail-file.conf    -n    agent1    -
Dflume.root.logger=INFO,console
```

开发 shell 脚本定时追加文件内容

```
mkdir -p /export/servers/shells/
cd    /export/servers/shells/
vim tail-file.sh
```

```
#!/bin/bash
while true
do
  date >> /export/servers/taillogs/access_log;
    sleep 0.5;
```

```
done
```

启动脚本

```
sh /export/servers/shells/tail-file.sh
```

```
tail -f access_log
```