

1.在 flume-ng 命令中，如何指定 agent 的配置文件？

- A.-C
- B.-f
- C.-F
- D.--c

2.使用下面命令启动 Flume 的 agent，下面说法错误的是？

```
bin/flume-ng agent --conf conf/ --name a1 --conf-file job/flume-netcat-logger.conf -  
Dflume.root.logger=INFO,console
```

- A.--conf 用来指定 flume 配置文件夹所在的位置
- B.命令中的 conf 目录和 job 目录一定是同级目录
- C.agent 的名字是 a1
- D.如果配置了环境变量，可以在任意目录下执行此命令

3.如果要实时监控收集一个指定日志文件的新增内容，可以使用何种 flume 的 source？

- A.spooldir
- B.avro
- C.hdfs
- D.exec

4.下列 Channels 选项中,哪一个可以保证数据不会丢失。

- A.Memory Channel
- B.File Channel

5.Flume 数据传输的基本单元是？

- A.Source
- B.Event
- C.Sink
- D.Channel

6.Flume 的 spooldir 源可以实时监控文件的变化说法是？

- 对
- 错

7.使用 Flume 监听一个端口，收集该端口数据，并打印到控制台，需要使用什么 Source 实现？

- A.spooldir
- B.exec
- C.netcat
- D.HTTP

8.Flume Sink 的作用是？

- A.缓存通道组件，用于数据传输
- B.数据源组件，用于跟数据源对接，以获取数据

C.获取数据，封装成 Event，写入 channel

D.下沉组件（输出），向最终存储系统传递数据

9.Flume 的 Channel 类型不包括？

A.JDBC Channel

B.Memory Channel

C.File Channel

D.HDFS Channel

10.Flume 的 Agent 就是一个 Java JVM 进程的说法是？

对

错

11.何种 Flume 的 Source 数据源类型可以监控指定目录，一有新文件进入指定监控目录则立刻收集该文件，收集完成后添加后缀标记文件为已处理。

A.exec

B.avro

C.spooldir

D.netcat

12.Flume Agent 的 Source 可以无需 Channel 而直接连接 Sink 的说法是？

对

错

参考答案错

13.在 Flume 配置文件中配置 Agent a1 的 Source r1 为 exec 的语法正确的是？

A.a1.sources.r1.type = exec

B.r1.sources.a1.type = exec

C.a1.source.r1.type = exec

D.r1.source.a1.type = exec

14.下面关于 Flume 的说法正确的是？

A: 一个 Flume 的 Agent 只能有一个 Source

B: Flume 可以将多个 Agent 连接成拓扑结构

A.A 对 B 对

B.A 错 B 错

C.A 错 B 对

D.A 对 B 错

15.以下哪个不属于 Flume 的 Agent 的组件？

A.Channel

B.Sink

C.Source

D.Topology

16.Memory Channel 和 File Channel 相比性能更好速度更快。

对

错

17.下面说法正确的是？

A.Flume 有可靠性、可伸缩、可扩展性

B.flume 有三大组件 source、channel 和 sink，各个组件之间都可以相互组合，各组件间耦合度低。

C.Flume 除了可以收集日志，还能对日志进行简单处理

D.以上都对

18.Flume Agent 组件中用于跟数据源对接，获取数据的是？

A.Topology

B.Source

C.Chanel

D.Sink

19.Flume 一个 Source 可以连接多个 Sink 的说法？

对

错

20.Flume 的 Event 说法正确的是？

A.Event 是数据传输的基本单位

B.以上都对

C.Event 来自外部数据源

D.Event 是携带 Header 信息的字节数组

1.阅读以下陈述：

A. Flume 的数据流可以根据 headers 的信息发送到不同的 channel 中

B. Flume Source 可以通过添加拦截器在运行时对 Event 进行修改或丢弃

A.A 正确 B 错误

B.A 错误 B 正确

C.A 错误 B 错误

D.A 正确 B 正确

2.关于 Flume 的说法错误的是？

A.在大数据分析中 Flume 的作用无可替代

B.可以实现自定义 Sink

C.Flume 的主要作用是收集外部日志到 HDFS 文件系统

D.通过 Flume 可以将数据流收集到本地文件系统

3.以下说法错误的是？

- A.一般的采集需求，通过对 flume 的简单配置即可实现
- B.Flume 可以适用于大部分的日常数据采集场景
- C.Flume 针对特殊场景也具备良好的自定义扩展能力
- D.Flume Agent 中的数据经过 Sink 组件之后就不能再发往其他 Agent 了

4.何种 Flume Source 类型支持断点续传功能？

- A.netcat source
- B.spooldir source
- C.exec source
- D.taildir source

5.使用 Hive Sink 的前提条件是

- A.以上都正确
- B.Hive 表必须分区分桶
- C.Hive 客户端需要开启事务支持
- D.Hive 表必须是 Acid 表（开启事务支持）

6.Flume 不支持哪种 channel 类型？

- A.File Channel
- B.HDFS Channel
- C.Memory Channel
- D.JDBC Channel

7.下面 Sink 配置表示的含义是\_\_\_\_\_

- a1.sinks.k1.hdfs.round = true
- a1.sinks.k1.hdfs.roundValue = 10
- a1.sinks.k1.hdfs.roundUnit = minute
- A.时间戳会向前一个整 10 分钟取整
- B.时间戳会向后一个整 10 分钟取整
- C.时间戳保持不变
- D.以上都不正确

8.要实现数据流在 Agent 之间传输，可以\_\_\_\_\_

- A.使用 Kafka Source 和 Kafka Sink
- B.使用 Kafka Source 和 Kafka Sink
- C.使用 Thrift Source 和 Thrift Sink
- D.以上都可以

9.以下关于 Flume，说法错误的是

- A.Flume 的数据流由事件（Event）贯穿始终，事件是 Flume 的基本数据单位
- B.Flume 以 agent 为最小的独立运行单位，一个 agent 就是一个 JVM。
- C.Flume agent 由 Source、Sink 和 Channel 三大组件构成
- D.Channel 中 filechannel 可将数据持久化到本地磁盘，但配置较为麻烦，需要配置数据目录和 checkpoint 目录，不同的 file channel 可以配置同一个 checkpoint 目录

10.Kafka Source 本质上就是一个\_\_\_\_

- A.Kafka Producer
- B.Kafka Streaming
- C.Kafka Broker
- D.Kafka Consumer

11.需求是收集 Web Socket 的数据并将数据打印到 linux 的控制台上, 何种 Flume 的 Agent 配置是可行的?

- A.netcat-source + file-channel + logger-sink
- B.netcat-source + memory-channel + console-sink
- C.socket-source + memory-channel + logger-sink
- D.socket-source + file-channel + console-sink

12.Flume Sink 类型不包括

- A.Hadoop Sink
- B.HBase Sink
- C.HDFS Sink
- D.HTTP Sink

13.以下说法正确的是?

- A. Flume 只能单机部署
- B. Flume 是海量日志采集、聚合和传输的系统。
- A.A 错 B 对
- B.A 对 B 对
- C.A 错 B 错
- D.A 对 B 错

1.绑定 Flume Agent a1 的 source r1 和 sink k1 到 channel c1 的配置写法正确的是?

- A.a1.source.r1.channels = c1  
a1.sinks.k1.channels = c1
- B.a1.sources.r1.channels = c1  
a1.sinks.k1.channels = c1
- C.a1.source.r1.channels = c1  
a1.sinks.k1.channel = c1
- D.a1.sources.r1.channels = c1  
a1.sinks.k1.channel = c1

2.如何配置把 channel 里面的 Event 按照轮询方式发送到 sink 各自对应的目的地?

- A.a1.sinkgroup.g1.processor.selector = round\_robin
- B.a1.sinkgroup.g1.processor.selectors = round\_robin
- C.a1.sinkgroups.g1.processor.selectors = round\_robin
- D.a1.sinkgroups.g1.processor.selector = round\_robin

3.Sink 组的默认的 SinkProcessor 是\_\_\_\_\_

- A.DefaultSinkProcessor
- B.FailoverSinkProcessor
- C.LoadBalancingSinkProcessor
- D.CustomSinkProcessor

4.LoadBalancingProcessor 需要和\_\_\_\_\_配合使用实现\_\_\_\_\_。

- A.Sink Group, 负载均衡
- B.Source Group, 负载均衡
- C.Sink, 负载均衡
- D.Sink Group, 高可用

5.FailoverSinkProcessor 可以实现 FLume 集群的\_\_\_\_\_功能。

- A.负载均衡
- B.故障转移
- C.数据一致性
- D.以上都不是

6.\_\_\_\_\_作用就是选出 event 将要被发往哪个 Channel

- A.Channel Processor
- B.Channel Dispatcher
- C.Channel Selector
- D.以上都不是

7.Flume Agent 中的哪一种 Source 组件支持断点续传功能?

- A.spooldir
- B.taildir
- C.exec
- D.avro

8. 可通过写 Unix command 的方式组织数据, 最常用的就是 tail -F [file]的 Source 组件类型是\_\_\_\_\_。

- A.spooldir source
- B.taildir source
- C.netcat source
- D.exec source

9.Flume Agent 之间的通信一般使用\_\_\_\_\_和\_\_\_\_\_来实现。

- A.Avro Source, Avro Channel
- B.Avro Channel, Avro Sink
- C.Avro Source, Avro Sink
- D.以上都不对

10.Flume Agent 通过在 source 和 sink 中间增加 channel 的设计，使 Event 从 source 到 sink 的传输是一个\_\_\_\_的过程

- A.同步
- B.异步
- C.并行
- D.并发

11.如果数据源是文本文件，对应的 Event 通常是\_\_\_\_

- A.一个字符
- B.一行记录中的一个字段
- C.一行记录
- D.一个文件

12.Channel: 连接\_\_\_\_和\_\_\_\_，用于中转\_\_\_\_的一个\_\_\_\_存储。

- A.source sink event 永久
- B.source sink agent 永久
- C.sink source event 临时
- D.sink source agent 临时

13.Source: 数据源，用来消费传递到该组件的 Event，每个 Agent\_\_\_\_数据源；

- A.只能有一个
- B.必须有多多个
- C.可以有一个或多个
- D.可以没有

14.Event 从 Source 流向 Channel，再到 Sink，本身为一个\_\_\_\_数组。

- A.bit
- B.byte
- C.binary
- D.boolean

15.Event: 一个存储数据的单元，带有一个可选的\_\_\_\_，可以是日志记录、avro 对象等；

- A.header
- B.body
- C.footer
- D.tail

16.在 Flume 中，最重要的抽象是\_\_\_\_，描述了数据从产生、传输、处理并最终写入目标的一条路径。

- A.data block
- B.data flow
- C.data chain
- D.data path

17.阅读以下陈述：

- A. Flume 的数据流可以根据 headers 的信息发送到不同的 channel 中
- B. Flume Source 可以通过添加拦截器在运行时对 Event 进行修改或丢弃

A.A 正确 B 错误

B.A 错误 B 正确

C.A 错误 B 错误

D.A 正确 B 正确

18.Flume 不支持哪种 channel 类型？

A.Memory Channel

B.File Channel

C.JDBC Channel

D.HDFS Channel

19.以下关于 Flume，说法错误的是

A.Flume 以 agent 为最小的独立运行单位，一个 agent 就是一个 JVM。

B.Flume 的数据流由事件（Event）贯穿始终，事件是 Flume 的基本数据单位

C.Channel 中 filechannel 可将数据持久化到本地磁盘，但配置较为麻烦，需要配置数据目录和 checkpoint 目录，不同的 file channel 可以配置同一个 checkpoint 目录

D.Flume agent 由 Source、Sink 和 Channel 三大组件构成

20.何种 Flume Source 类型支持断点续传功能？

A.exec source

B.spooldir source

C.taildir source

D.netcat source

21.需求是收集 Web Socket 的数据并将数据打印到 linux 的控制台上，何种 Flume 的 Agent 配置是可行的？

A.netcat-source + memory-channel + console-sink

B.netcat-source + file-channel + logger-sink

C.socket-source + file-channel + console-sink

D.socket-source + memory-channel + logger-sink

22.以下说法错误的是？

A.Flume 可以适用于大部分的日常数据采集场景

B.一般的采集需求，通过对 flume 的简单配置即可实现

C.Flume 针对特殊场景也具备良好的自定义扩展能力

D.Flume Agent 中的数据经过 Sink 组件之后就不能再发往其他 Agent 了



23.以下说法正确的是？

- A. Flume 只能单机部署
- B. Flume 是海量日志采集、聚合和传输的系统。
- A.A 对 B 错
- B.A 错 B 对
- C.A 对 B 对
- D.A 错 B 错

24.Kafka Source 本质上就是一个\_\_\_\_

- A.Kafka Consumer
- B.Kafka Producer
- C.Kafka Broker
- D.Kafka Streaming

25.关于 Flume 的说法错误的是？

- A.通过 Flume 可以将数据流收集到本地文件系统
- B.可以实现自定义 Sink
- C.Flume 的主要作用是收集外部日志到 HDFS 文件系统
- D.在大数据分析中 Flume 的作用无可替代

26.要实现数据流在 Agent 之间传输，可以\_\_\_\_\_

- A.使用 Kafka Source 和 Kafka Sink
- B.使用 Thrift Source 和 Thrift Sink
- C.使用 Avro Source 和 Avro Sink
- D.以上都可以

27.使用 Hive Sink 的前提条件是

- A.Hive 客户端需要开启事务支持
- B.Hive 表必须分区分桶
- C.Hive 表必须是 Acid 表（开启事务支持）
- D.以上都正确

28.下面 Sink 配置表示的含义是\_\_\_\_\_

- a1.sinks.k1.hdfs.round = true
- a1.sinks.k1.hdfs.roundValue = 10
- a1.sinks.k1.hdfs.roundUnit = minute
- A.时间戳保持不变
- B.时间戳会向前一个整 10 分钟取整
- C.时间戳会向后一个整 10 分钟取整
- D.以上都不正确

29.Flume Sink 类型不包括

- A.HDFS Sink
- B.HTTP Sink
- C.HBase Sink
- D.Hadoop Sink

30.在 flume-ng 命令中，如何指定 agent 的配置文件？

- A.--c
- B.-C
- C.-f
- D.-F

31.Flume 的 Event 说法正确的是？

- A.Event 是数据传输的基本单位
- B.Event 是携带 Header 信息的字节数组
- C.Event 来自外部数据源
- D.以上都对

32.下面说法正确的是？

- A.flume 有三大组件 source、channel 和 sink，各个组件之间都可以相互组合，各组件间耦合度低。
- B.Flume 除了可以收集日志，还能对日志进行简单处理
- C.Flume 有可靠性、可伸缩、可扩展性
- D.以上都对

33.在 Flume 配置文件中配置 Agent a1 的 Source r1 为 exec 的语法正确的是？

- A.r1.source.a1.type = exec
- B.r1.sources.a1.type = exec
- C.a1.source.r1.type = exec
- D.a1.sources.r1.type = exec

34.使用下面命令启动 Flume 的 agent，下面说法错误的是？

bin/flume-ng agent --conf conf/ --name a1 --conf-file job/flume-netcat-logger.conf -  
Dflume.root.logger=INFO,console

- A.agent 的名字是 a1
- B.--conf 用来指定 flume 配置文件夹所在的位置
- C.命令中的 conf 目录和 job 目录一定是同级目录
- D.如果配置了环境变量，可以在任意目录下执行此命令

35.使用 Flume 监听一个端口，收集该端口数据，并打印到控制台，需要使用什么 Source 实现？

- A.exec
- B.spooldir
- C.netcat
- D.HTTP

36. 【多选题】在 Agent 中事务发生在\_\_\_\_\_和\_\_\_\_\_

A.Source 从外部消息源接受消息

B.Source 推送消息到 channel

C.Sink 从 channel 拉取消息

D.Sink 发送消息到 Storage

37. 【多选题】Flume Agent 中的哪些类型 Source 组件支持对日志文件的实时收集功能?

A.spooldir

B.taildir

C.exec

D.netcat

38. 【多选题】启动 Flume Agent 时用于指定 Flume Agent 的配置文件的参数是\_\_\_\_或\_\_\_\_

A.--conf-file

B.--config-file

C.-F

D.-f

39. 【多选题】启动 Flume Agent 时用于指定 Flume 配置文件父目录的参数是\_\_\_\_或\_\_\_\_

A.--config

B.--conf

C.-c

D.-C

40. 【多选题】启动 Flume Agent 时用于指定 agent 名称的参数是\_\_\_\_或\_\_\_\_

A.-n

B.-a

C.--name

D.--agent

41. 【判断题】使用 Flume SpoolDir 源监控/spool 目录时, 拷贝到/spool 目录下的文件可以再打开编辑, 这样新编辑增加的内容也会被同样收集。

对

错

42. 【判断题】Flume 可以实现转移不同来源的大量日志数据到中央数据仓库中。

对

错

43. 【判断题】Flume 是一个分布式、可靠和高可用的海量日志采集、聚合和传输的系统。

对

错

44. 【判断题】 Flume 提供对数据进行简单处理，并写到各种 storage。

对

错

45. 【判断题】 Flume 是日志收集系统，支持在日志系统中定制各类数据发送方，用于收集数据。

对

错

46. 【判断题】 Memory Channel 和 File Channel 相比性能更好速度更快。

对

错

47. 【判断题】 Flume 的 spooldir 源可以实时监控文件的变化说法是？

对

错

48. 【判断题】 Flume Agent 的 Source 可以无需 Channel 而直接连接 Sink 的说法是？

对

错

49. 【判断题】 Flume 的 Agent 就是一个 Java JVM 进程的说法是？

对

错

50. 【判断题】 Flume 一个 Source 可以连接多个 Sink 的说法？

对

错