livershui1(Ocen

liyonghui160com

博客

微博

收藏

留言

关于我

 \cap



liyonghui160com

浏览: 8423 次

性别: 💣

来自: 北京



最近访客 <u>客>></u> 更多访





fullstar

dylinshi126





angke93

<u>lyg5623</u>

文章分类

- 全部博客 (123)
- java (20)
- windows系统 (4)
- mongodb (0)
- Highcharts (6)
- **c**# (6)
- SqlServer (1)
- <u>jquery (6)</u>
- javascript (2)
- <u>ibatis (1)</u>
- IIS (1)
- mybatis (4)
- mysql (4)
- thrift (2)
- ubuntu (8)

<u>Flume-ng1.5安装配置</u>

博客分类: flume-ng

Flume-ng1.5安装配置

本人亲自执行操作,希望可以帮到想学Flume的你

1) 简介

Flume是一个分布式、可靠、和高可用的海量日志聚合的系统,支持在系统中定制各类数据发送方,用于收集数据;同时,Flume提供对数据进行简单处理,并写到各种数据接受方(可定制)的能力。

设计目标:

(1) 可靠性

当节点出现故障时,日志能够被传送到其他节点上而不会丢失。Flume提供了三种级别的可靠性保障,从强到弱依次分别为: end-to-end(收到数据agent首先将event写到磁盘上,当数据传送成功后,再删除;如果数据发送失败,可以重新发送。),Store on failure(当数据接收方crash时,将数据写到本地,待恢复后,继续发送),Best effort(数据发送到接收方后,不会进行确认)。

(2) 可扩展性

Flume采用了三层架构,分别为agent,collector和storage,每一层均可以水平扩展。其中,所有agent和collector由master统一管理,这使得系统容易监控和维护,且master允许有多个(使用ZooKeeper进行管理和负载均衡),这就避免了单点故障问题。

(3) 可管理性

所有agent和colletor由master统一管理,这使得系统便于维护。多master情况,Flume利用ZooKeeper和gossip,保证动态配置数据的一致性。用户可以在master上查看各个数据源或者数据流执行情况,且可以对各个数据源配置和动态加载。Flume提供了web 和shell script command两种形式对数据流进行管理。

(4) 功能可扩展性

用户可以根据需要添加自己的agent, collector或者storage。此外, Flume自带了很多组件,包括各种agent (file, syslog等), collector和storage (File, HDFS, HBase等)。

- storm (16)
- tomcat (3)
- hadoop (6)
- 文件同步 (2)
- 推荐&统计系统 (14)
- zookeeper (5)
- memcache (4)
- Kafka (10)
- mevan (2)
- nginx (3)
- flume-ng (2)
- Redis (4)
- <u>hbase (2)</u>

社区版块

- 我的资讯 (0)
- 我的论坛 (0)
- 我的问答(0)

存档分类

- **2014-09** (2)
- **2014-08** (24)
- **2014-07** (32)
- 更多存档...

最新评论

2) 配置

之前配置过Hadoop和hbase,所以需要先将hadoop和hbase启动,才能将文件写入hdfs和hbase。hadoop-2.2.0和hbase-0.96.0的配置分别参考《Ubuntu和CentOS中分布式配置Hadoop-2.2.0》

http://www.21ops.com/front-tech/10782.html 和《CentOS分布式环境安装HBase-0.96.0》

http://www.21ops.com/front-tech/10788.html 。

本次配置环境为两台装有centos 的测试集群。主机名为master的机器负责收集日志,主机名为node的机器负责日志的写入,本次配置的写入方式有三种:写入普通目录,写入hdfs。

首先下载flume-ng的二进制压缩文件。地址: http://flume.apache.org/download.html。下载好后,解压文件。首先编辑/etc/profile文件,在其中添加如下几行:

Xml代码 ☆

```
    export FLUME_HOME=/home/aaron/apache-flume-1.4.0-bin
    export FLUME_CONF_DIR=$FLUME_HOME/conf
    export PATH=$PATH:$FLUME_HOME/bin
```

添加好之后记得运行\$ souce /etc/profile命令使修改生效。

在master的flume文件夹的conf目录中,新建一个flume-master.conf文件,内容如下:

Xml代码 😭

```
agent.sources = seqGenSrc
      agent.channels = memoryChannel
 2.
 3.
      agent.sinks = remoteSink
 4.
      # For each one of the sources, the type is defined
      agent.sources.seqGenSrc.type = exec
 5.
      agent.sources.seqGenSrc.command = tail -F /home/aaron/test
 6.
 7.
      # The channel can be defined as follows.
8.
      agent.sources.seqGenSrc.channels = memoryChannel
9.
      # Each sink's type must be defined
10.
      agent.sinks.loggerSink.type = logger
      #Specify the channel the sink should use
11.
12.
      agent.sinks.loggerSink.channel = memoryChannel
13.
      # Each channel's type is defined.
14.
      agent.channels.memoryChannel.type = memory
      # Other config values specific to each type of channel(sink or source)
15.
16.
      # can be defined as well
      # In this case, it specifies the capacity of the memory channel
17.
18.
      agent.channels.memoryChannel.capacity = 100
19.
      agent.channels.memoryChannel.keep-alive = 100
20.
      agent.sinks.remoteSink.type = avro
21.
      agent.sinks.remoteSink.max.message.size=1000000
      agent.sinks.remoteSink.hostname = node
23.
      agent.sinks.remoteSink.port = 23004
24.
      agent.sinks.remoteSink.channel = memoryChannel
```

在node机器上也将/etc/profile文件添加上面的配置。然后,在conf中新建一个flume-node.conf文件,修改如下:

Xm1代码

```
agent.sources = seqGenSrc1
      agent.channels = memoryChannel
 3.
      #agent.sinks = fileSink
      agent.sinks = <SPANstyleSPANstyle="FONT-FAMILY: Arial, Helvetica, sans-serif">fileSink</SPAN>
 5.
      # For each one of the sources, the type is defined
 6.
      agent.sources.seqGenSrc1.type = avro
 7.
      agent.sources.seqGenSrc1.bind = node
 8.
      agent.sources.seqGenSrc1.port = 23004
9.
      # The channel can be defined as follows.
10.
      agent.sources.seqGenSrc1.channels = memoryChannel
11.
      # Each sink's type must be defined
12.
      agent.sinks.loggerSink.type = logger
      #Specify the channel the sink should use
13.
14.
      agent.sinks.loggerSink.channel = memoryChannel
15.
      # Each channel's type is defined.
16.
      agent.channels.memoryChannel.type = memory
      # Other config values specific to each type of channel(sink or source)
17.
18.
      # can be defined as well
19.
      # In this case, it specifies the capacity of the memory channel
20.
      agent.channels.memoryChannel.capacity = 100
      agent.channels.memoryChannel.keep-alive = 100
21.
22.
      agent.sources.flieSink.type = avro
      agent.sources.flieSink.max.message.size=1000000
23.
24.
      agent.sources.fileSink.channel = memoryChannel
25.
      agent.sources.fileSink.sink.directory = /home/aaron/
26.
      agent.sources.fileSink.serializer.appendNewline = true
```

#命令参数说明

- -c conf 指定配置目录为conf
- -f conf/example.conf 指定配置文件为conf/example.conf
- -n agent1 指定agent名字为agent1,需要与example.conf中的一致(这里不一致,可能会一直停在那里,请参考笔记中后面的错误全集Flume部分,那里介绍了错误的分析,原因,解决)
- -Dflume.root.logger=INFO,console 指定DEBUF模式在console输出INFO信息

在master上面运行命令:

Xml代码 ☆

1. [root@rs229apache-flume-1.5.0-bin]# ./bin/flume-ng agent -c conf/ -f conf/flume-maste.conf -n agent -Dflume.root.logger=INFO,console

在node上运行命令:

Xm1代码 🛣

1. [root@rs229apache-flume-1.5.0-bin]# ./bin/flume-ng agent -c conf/ -f conf/flume-node.conf -n agent -Dflume.root.logger=INFO,console

启动之后,就可以发现两者之间可以相互通信,master上面的文件就能发送到node上,修改master上的test文件,在后面追加内容时,node也可以接收到。

如果想要将内容写入hadoop,可以将node中的flume-node.conf文件做如下修改:

Xml代码 ☆

```
    agent.sinks = k2
    agent.sinks.k2.type = hdfs
    agent.sinks.k2.channel = memoryChannel
    agent.sinks.k2.hdfs.path = hdfs://master:8089/hbase
    agent.sinks.k2.hdfs.fileType = DataStream
    agent.sinks.k2.hdfs.writeFormat = Text
```

其中,hdfs://master:8089/hbase为hadoop的hdfs文件路径。

Java代码 😭

```
1.
      # producer config
 3.
      4.
      #agent section
 5.
      producer.sources = s
 6.
 7.
      producer.channels = c
 8.
      producer.sinks = r
9.
10.
      #producer.sources.s.type = seq
11.
      producer.sources.s.channels = c
12.
      producer.sources.s.type = exec
13.
      producer.sources.s.command=tail -f -n+1 /usr/aboutyun.log
14.
      #fks.sources.source1.type=spooldir
15.
      #fks.sources.source1.spoolDir=/usr/local/yting/flume/tdata/tdir1
16.
      producer.sources.s.deletePolicy=never
17.
      #producer.sources.s.type = avro
      #producer.sources.s.bind = localhost
18.
19.
      #producer.sources.s.port = 10000
20.
      # Each sink's type must be defined
21.
      producer.sinks.r.type = org.jd.clickstream.sink.KafkaSink
22.
23.
      producer.sinks.r.metadata.broker.list=127.0.0.1:9092,127.0.0.1:9093,127.0.0.1:9094
24.
      producer.sinks.r.partition.key=0
25.
      producer.sinks.r.partitioner.class=org.jd.clickstream.partition.TypePartitioner
      producer.sinks.r.serializer.class=kafka.serializer.StringEncoder
26.
27.
      producer.sinks.r.request.required.acks=0
28.
      producer.sinks.r.max.message.size=1000000
29.
      producer.sinks.r.producer.type=sync
30.
      producer.sinks.r.custom.encoding=UTF-8
31.
      #Specify the channel the sink should use
      producer.sinks.r.channel = c
32.
33.
34.
      # Each channel's type is defined.
35.
      producer.channels.c.type = memory
36.
      producer.channels.c.capacity = 1000
37.
38.
      #producer.channels.c.type=file
39.
      #producer.channels.c.checkpointDir=/usr/local/yting/flume/checkpointdir/tcpdir/example_agent1_0
40.
      #producer.channels.c.dataDirs=/usr/local/yting/flume/datadirs/tddirs/example_agent1_001
```

1 _000000

3 ______

5 农村平房设计

2 ______

4 冲压不锈钢水箱

6 <u>儿童教育培训</u> 📽

分享到:
评论
发表评论
您还没有登录、请您登录后再发表评论

声明: ITeye文章版权属于作者,受法律保护。没有作者书面许可不得转载。若作者同意转载,必须以超链接形式标明文章原始出处和作者。 © 2003-2014 ITeye.com. All rights reserved. [京ICP证110151号 京公网安备110105010620]