部署手册

1. 程序包结构说明：

filebeat-1.3.1-x86\_64.tar.gz

logstash-2.3.4.tar.gz

下载地址：

<https://download.elastic.co/beats/filebeat/filebeat-1.3.1-x86_64.tar.gz>

https://download.elastic.co/logstash/logstash/logstash-2.3.4.tar.gz

filebeat,logstash做为实时流的基础组件，监控文件夹数据变化，并将数据打入kafka以及kafka数据的分流

2. 术语表

Filebeat监控文件夹内文件变化的组建，并可以将数据以TCP接口的方式输出到logstash。

Logstash负责接收filebeat的信息，传输给kafka以及kafka的分流。

3. 环境要求

logcenter部署的机器需要配置filebeat，filebeat对接logstash的内存要求高16g。

部署filebeat的机器，要求spark用户能够访问/var/log/logcenter/目录。

logstash端口占用：5044

4. 部署说明

4.1 初始化

spark用户操作：

解压为如下目录

/app/filebeat-1.3.1-x86\_64

/app/logstash-2.3.4

建立软链：

filebeat -> /app/filebeat-1.3.1-x86\_64

logstash -> /app/logstash-2.3.4

4.2 配置文件说明

对于filebeat的配置,新环境替换掉hosts中的内容。

filebeat:

prospectors:

-

paths:

- /var/log/logcenter/current/\*.log

input\_type: log

close\_older: 30m

scan\_frequency: 2s

harvester\_buffer\_size: 163840

publish\_async: true

idle\_timeout: 1s

output:

logstash:

hosts: ["10.10.2.13:5044","10.10.2.16:5044"]

worker: 16

loadbalance: true

file:

path: "/tmp/filebeat"

filename: filebeat

rotate\_every\_kb: 10000

shipper:

logging:

files:

对于logstash的配置,新环境替换掉kafka组建中bootstrap\_servers的信息。

打入kafka供forest处理的topic id：

medusa-pre-log

helios-pre-log

对接filebeat的logstash配置如下：

input {

beats {

port => "5044"

}

heartbeat {

message => "logstash"

interval => 60

type => "heartbeat"

}

}

filter {

if [type] != "heartbeat" {

metrics {

meter => "events"

add\_tag => "metric"

flush\_interval => 60

}

}

}

output {

if [type] == "heartbeat" {

file {

path => "/tmp/logstash\_heartbeat.log"

}

} else if "metric" in [tags] {

file {

codec => line {

format => "%{[events][rate\_1m]},%{[events][rate\_5m]},%{[events][rate\_15m]}"

}

path => "/tmp/logstash\_metrics.log"

}

}else {

kafka {

bootstrap\_servers => "10.10.2.20:9092,10.10.2.21:9092,10.10.2.22:9092,10.10.2.23:9092,10.10.2.24:9092"

topic\_id => "medusa-pre-log"

compression\_type => "snappy"

codec => plain {

format => "%{message}"

}

workers => 16

}

}

}

kafka分流的logstash配置如下：

需要替换zk\_connect，bootstrap\_servers信息，以及ouput插件的if判断条件来分流。

kafka topic切分条件

广告前播片数据

[logType] == "start\_end" and [actionId] == "ad-vod-pre-play"

广告请求数据

if [logType] == "event" and [eventId] == "ad-vod-request"

会员嘉年华

if [logType] == "play" and [playStatus] == "success" and [accountId] != "" and [vipLevel]!= "非会员" and [duration] <= 10800

配置模版如下：

input {

kafka {

zk\_connect => "bigdata-computing-02-015:2182,bigdata-computing-02-016:2182,bigdata-computing-02-017:2182,bigdata-computing-02-018:2182,bigdata-computing-02-019:2182"

group\_id => "kylin\_medusa\_processed\_log\_consumer\_group"

topic\_id => "medusa-processed-log"

codec => "json"

consumer\_threads => 10

}

heartbeat {

message => "logstash"

interval => 60

type => "heartbeat"

}

}

filter {

if [type] != "heartbeat" {

mutate {

remove\_field => [ "tags" ]

}

metrics {

meter => "events"

add\_tag => "metric"

flush\_interval => 60

}

}

}

output {

if [logType] == "play" and [contentType] == "hot" and [event] == "startplay" {

kafka {

bootstrap\_servers => "10.10.2.20:9092,10.10.2.21:9092,10.10.2.22:9092,10.10.2.23:9092,10.10.2.24:9092"

topic\_id => "kylin\_medusa\_hot\_play\_test"

workers => 5

compression\_type => "snappy"

}

} else if [type] == "heartbeat" {

file {

path => "/tmp/logstash\_heartbeat.log"

}

} else if "metric" in [tags] {

file {

codec => line {

format => "%{[events][rate\_1m]},%{[events][rate\_5m]},%{[events][rate\_15m]}"

}

path => "/tmp/logstash\_metrics.log"

}

}

}

4.3 任务启动

filebeat:

在/opt/filebeat目录创建start\_filebeat.sh脚本，执行start\_filebeat.sh脚本，脚本内容如下：

#!/usr/bin/env bash

current\_bin\_path="`dirname "$0"`"

echo "current\_bin\_path is ${current\_bin\_path}"

cd ${current\_bin\_path}

nohup ./filebeat -c filebeat\_product.yml -e > nohup.log 2>&1 &

logstash：

在/opt/logstash目录创建startup\_with\_filebeat.sh脚本，执行startup\_with\_filebeat.sh脚本，

依赖与具体机器分配，分发不同配置

模版脚本内容如下：

#!/usr/bin/env bash

LOGSTASH\_HOME="$( cd "$( dirname "$0" )" && cd .. && pwd )"

echo ${LOGSTASH\_HOME}

nohup $LOGSTASH\_HOME/bin/logstash -f $LOGSTASH\_HOME/conf/filebeat\_to\_kafka\_only\_message\_product.conf --auto-reload -w 16 -b 2000 > $LOGSTASH\_HOME/logs/filebeat\_to\_kafka\_only\_message\_product.log 2>&1 &

参考：

<http://172.16.17.100:8090/pages/viewpage.action?pageId=3835903>

http://172.16.17.100:8090/pages/viewpage.action?pageId=3835265