

# Q：使用Kafka做日志收集

---

需要收集的信息：

- 1、用户ID (user\_id)
- 2、时间 (act\_time)
- 3、操作 (action, 可以是：点击：click, 收藏：job\_collect, 投简历：cv\_send, 上传简历：cv\_upload)
- 4、对方企业编码 (job\_code)

说明：

- 1、HTML可以理解为拉勾的职位浏览页面
- 2、Nginx用于收集用户的点击数据流，记录日志access.log
- 3、将Nginx收集的日志数据发送到Kafka主题：tp\_individual

架构：

HTML+Nginx+[ngx\\_kafka\\_module](#)+Kafka

提示：

学员需要自己下载nginx，配置nginx的ngx\_kafka\_module，自定义一个html页面，能做到点击连接就收集用户动作数据即可。

作业需提交：

- 1、html的截图+搭建的过程+结果截图以文档或视频演示形式提供。
- 2、作业实现过程的代码。

(注意：1、需要把搭建过程用文档写清楚，2、运行效果：在html上点击相应的连接或按钮，应该在Kafka中看到有消息流进来，效果最好以视频形式演示。)

## A: 实现

---

### 1. 安装依赖

---

```
# 1.安装git
yum install -y git
# 2.安装相关依赖
yum install -y gcc gcc-c++ zlib zlib-devel openssl openssl-devel pcre pcre-devel
# 3.将kafka的客户端源码clone到本地
cd /opt/lagou/software
git clone https://github.com/edenhill/librdkafka
# 4.编译
./configure
make && make install
```

## 2. 下载nginx并解压

```
wget http://nginx.org/download/nginx-1.18.0.tar.gz
tar -zxf nginx-1.18.0.tar.gz
```

## 3. 下载ngx\_kafka\_module，并编译

```
cd /opt/lagou/software
git clone https://github.com/brg-liuwei/ngx_kafka_module
# 进入到nginx的源码包目录下    (编译nginx, 然后将插件同时编译)
cd /opt/lagou/software/nginx-1.18.0
./configure --add-module=/opt/lagou/software/ngx_kafka_module/
make && make install
```

## 4. 启动zookeeper和kafka，提前创建收集日志的topic，并启动消费者

```
kafka-topics.sh --zookeeper linux121:2181/myKafka --create --topic
tp_individual --partitions 3 --replication-factor 2
kafka-console-consumer.sh --bootstrap-server linux121:9092 --topic
tp_individual --from-beginning
```

## 5. 修改nginx.conf，启动nginx

```
vim /usr/local/nginx/conf/nginx.conf
# 在http标签下新增
kafka;
kafka_broker_list linux122:9092 linux122:9093;

# 在server标签下修改
listen      8675;
```

```
server_name linux123;

location /kafka/test {
    #root    html;
    #index   index.html index.htm;
    kafka_topic tp;
}
# 启动nginx
cd /usr/local/nginx/sbin
./nginx
```

## 6. 启动kafka，拉数据

---

```
kafka-topics.sh --zookeeper linux122:2181/myKafka --create --topic tp --
partitions 1 --replication-factor 1
kafka-console-consumer.sh --bootstrap-server linux122:9092 --topic tp --from-
beginning
```

## 7. 发送请求

---

```
curl linux123:8675/kafka/test -d "user1 click" -v
curl linux123:8675/kafka/test -d "user1 collect_job" -v
curl linux123:8675/kafka/test -d "user1 send_cv" -v
curl linux123:8675/kafka/test -d "user1 upload_cv" -v
```

## 结果验证

---

```
root@linux122:~
```

```
▶ kafka-console-consumer.sh --bootstrap-server linux122:9092 --topic tp --from-beginning
```

```
test message kafka
```

```
test message kafka
```

```
user1 click iphone
```

```
user1 click
```

```
user1 collect_job
```

```
user1 send_cv
```

```
user1 upload_cv
```

```
client_loop: send disconnect: Broken pipe
```