# 数据迁移步骤说明V1.0

# 环境搭建

## 1.1环境说明（业务说明）

**版本差异说明 V4.0.3.7 🡪 V4.0.7.15**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 版本 | mysql | ElasticSearch |  |  |
| V4.0.3.7 | 有 | 无 |  |  |
| V4.0.7.15 | 有 | 有 |  |  |

**将4.0.3.7 只有mysqld的坏境升级到 4.0.75，需要做的业务处理有**:

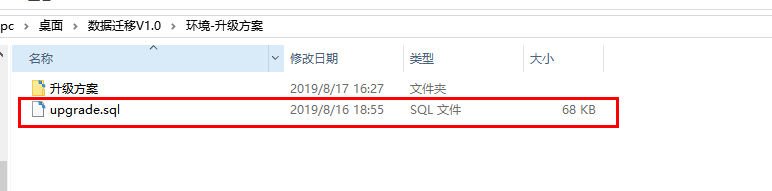
* Mysql里面取出小图图片
* 将图片批量提交到特征提取接口
* Mysql的objext\_result\_ymd数据和 vlpr\_result\_ymd数据存入es数据
* 将提取到的特征值的数据推送到kafka的topic feature
* 消费kafak的topic为featrue的数据 featrue-search 模块建立搜索数据
* 前台页面支持搜索功能

前面标红的数据为此次代码需要实现的功能，后面的为其他模块的业务逻辑，不需要关心，

## 1.2 mysql数据库表字段更新

版本之间的数据库表字段存在差异

因此需要先 将V4.0.3.7的mysql数据通过脚本upgrade.sql 升级到V4.0.7.15版本mysql字段

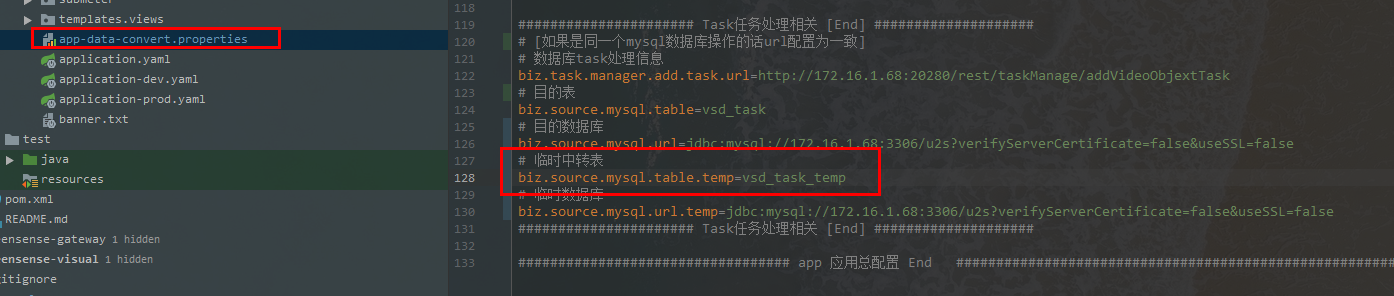


Mysql 由于是同库升级，因此需要先对数据库进行一定的操作在进行处理，vsd\_task表涉及到搜图业务，因此需要对数据进行一次处理，具体的处理逻辑如下

1. 将vsd\_task表备份一份，名字reanme为vsd\_task\_temp



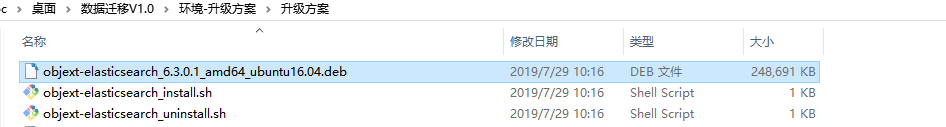
1. 清空表vsd\_task , tb\_analysis\_detail，tb\_analysis\_task 三个表
2. 配置迁移程序 里面的配置文件为对应的数据库信息



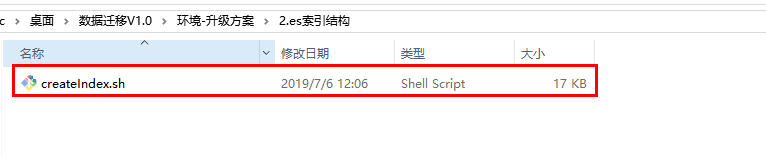
程序会自动将vsd\_task\_temp表里面的数据，以接口的形式将数据，提交到vsd\_task该业务表，实现原有数据的业务逻辑的转换升级

## 1.3装elasticSearch(4.0.3.7只有mysql无es)

将V4.0.3.7升级到V4.0.7.15，因为V4.0.3.7无es，业务数据来源全部通过mysql，所以需要先将V4.0.7.15安装好elasticSearch



创建es的索引结构 执行创建脚本



## 1.4升级到4.0.7.15



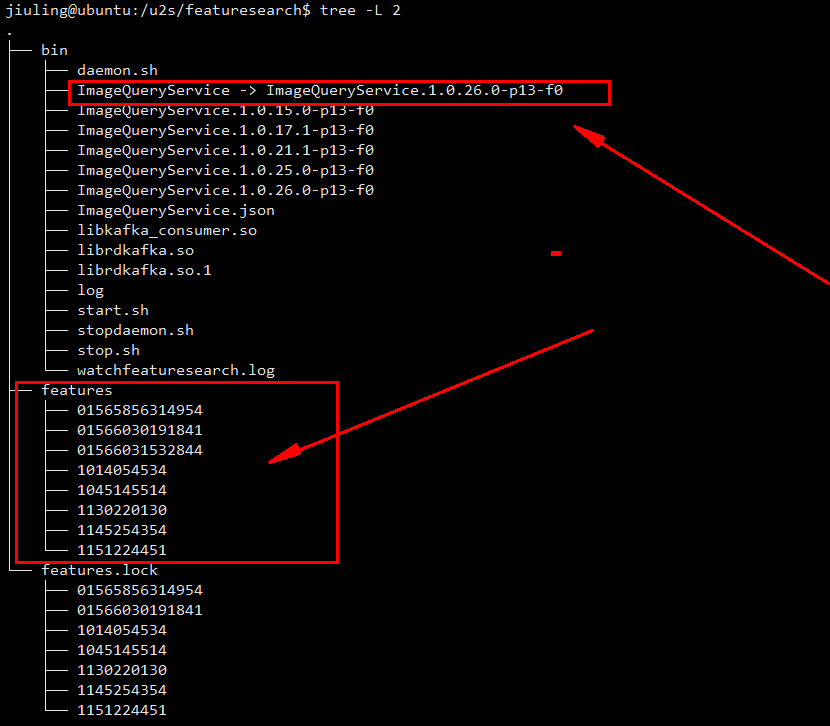
先安装

VideoObject-sdk-4.0.7.15-amd64-ubuntu16.04.tar.gz

再安装

VideoObject-master-4.0.7.15-amd64-ubuntu16.04.tar.gz

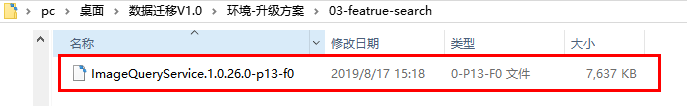
## 1.5 featrue-search 搜图模块处理



1. 首先需要先对搜图模块的service以软链接的形式进行升级

rm -rf ImageQueryService

1. 上传ImageQueryService.1.0.26.0-p13-f0最新的服务



1. 将ImageQueryService.1.0.26.0-p13-f0软链起来

Sudo ln –s ImageQueryService ImageQueryService.1.0.26.0-p13-f0

1. 重启featrue-search的服务

Service objext-featuresearch restart

1. 备份原有的 u2s/featuresearch/features 目录的搜索数据

## 1.6 elasticSearch字段和mysql字段的匹配

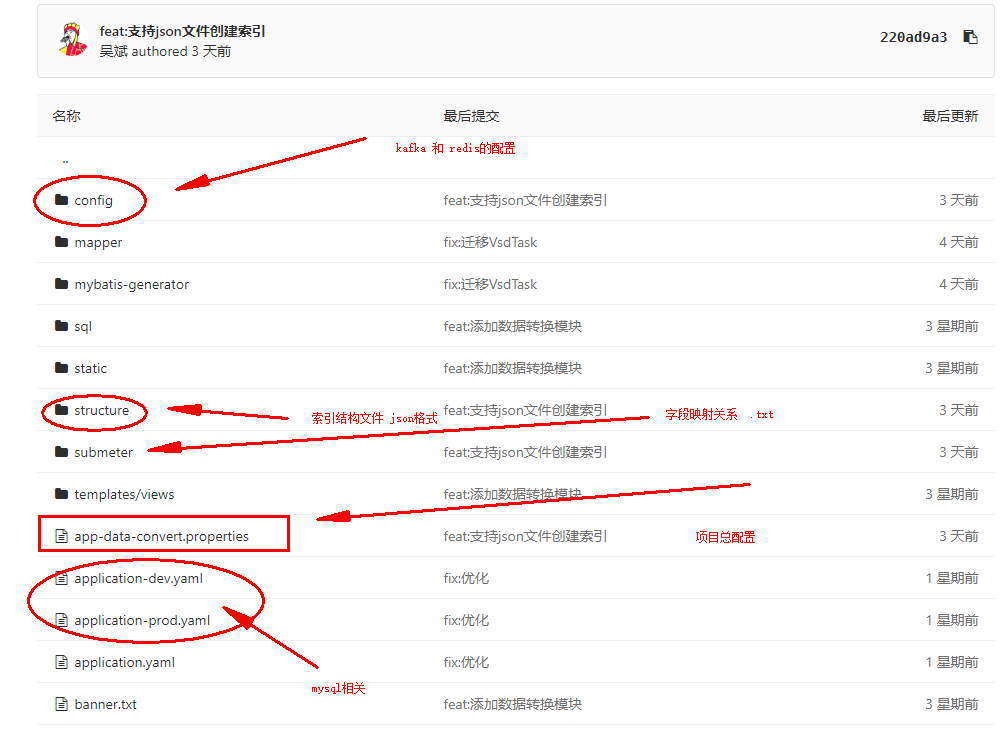
# 程序实现数据转换升级

## 2.1 程序地址

<https://qianshitong.320.io:13085/cuiss/keensense-u2s/tree/master/keensense-dataconvert>

## 2.2程序配置业务说明

1.程序概览

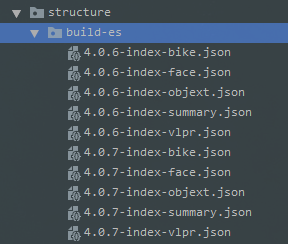


2.基本的配置说明

总配置app-data-convert.properties

# 是否开启索引构建  
sys.switch.case.open.build.es=true  
# 是否开启es数据迁移到es功能  
sys.switch.case.open.es.to.es=false  
# 是否开启mysql数据迁移到es 并且推送到kafka功能  
sys.switch.case.open.mysql.to.es=true  
# 表示mysql迁移到es是否需要转换字段[默认为false]  
sys.mysql.to.es.need.convert.field = true

说明是否需要 使用该模块创建索引 需要的话 程序会加载

对应的版本还号的索引json格式数据，调用api创建索引到对应的es库。

### 1.2.1 系统转换开关

# 是否开启es数据迁移到es功能  
sys.switch.case.open.es.to.es=false

如果开启的话 可以实现 es库到es库的数据的迁移 {V4.0.3.7 升级到 V4.0.7.15不需要,因此可以置为false

# 是否开启mysql数据迁移到es 并且推送到kafka功能  
sys.switch.case.open.mysql.to.es=true

这是当前V4.0.3.7 🡪 V4.0.7.15 需要做的处理 置为开启状态 也就是要把 mysql的数据迁移到es,提取特征,然后推送到kafka

# 表示mysql迁移到es是否需要转换字段[默认为false]  
sys.mysql.to.es.need.convert.field = true

是否需要将mysql的字段转换为es需要的字段，因为某些字段mysql和es字段对应不上，因此需要做字段的转换操作

### 1.2.2 图片映射关系配置

########### [图片地址映射相关]-[Start] #####################  
# 严格按该方式,127.0.0.1:8082 需要映射的地址  
http.picture.prefix=http://172.16.1.63:8082  
# 需不需要做图片的地址映射 默认为false 即127.0.0.1  
http.picture.is.need.mapping=true  
########### [图片地址映射相关]-[End] #######################

如果是同一台机器做处理 可以不需要做图片地址映射的处理，因此可以将

http.picture.is.need.mapping=true

置为false,表示不作图片映射处理

### 1.2.3 批处理设置

########### [处理调用接口等相关配置]-[Start] #####################  
# [注意] 向上取整 [ pageSize \* n + m(m<pageSize) = batchDeal ]  
# 每次从mysql里面load加载的数据  
biz.deal.mysql.day.load.page.size=1000  
# 每次mysql load 到的数据 每批次处理量  
biz.deal.mysql.day.load.batch.deal=3000  
# 定时处理queue残余的数据[总数据量/处理能力](分钟)  
biz.deal.timing.batch.data.time.minutes=60  
########### [处理调用接口等相关配置]-[End] ########################

biz.deal.mysql.day.load.page.size=1000

mysql数据量比较大,因此需要分页从mysql数据库中拿出数据，这里的设置即是每次mysql表里面取出的记录条数

# 每次mysql load 到的数据 每批次处理量  
biz.deal.mysql.day.load.batch.deal=3000

Mysql数据量很多 不可能一起加载 最后处理 因此可以在这里设置每次 满多少数据 开始处理一批 减轻系统的压力

# 定时处理queue残余的数据[总数据量/处理能力](分钟)  
biz.deal.timing.batch.data.time.minutes=60

因为调用特征提取的接口为batch提交，会出现残留的情况，因此启动了一个定时线程，设定时间后，调用一次，防止 < BatchSize 数量的数据丢失，这里需要根据具体的现场情况来设定一个比较合理的数值

### 1.2.4 特征提取相关

########### [特征处理相关配置]-[Start] ###########################  
# 每次传入 特征提取模块的batchSize  
biz.deal.recog.feature.pic.batch=10  
# 调用特征分析服务 固定线程池大小  
biz.deal.recog.feature.request.size=30  
  
# 下载图片固定线程池大小 [默认为机器的1.5f]  
biz.deal.download.image.thread.size=10  
# 取出imageQueen里面的数据进行处理的线程大小  
biz.deal.queen.take.image.thread.size=10  
  
# 图片下载 queenCapacity  
biz.deal.download.queue.capacity=500  
# 图片http下载连接超时时间 (s)  
biz.deal.download.connect.time.out=10  
# 图片http下载socket通信超时时间 (s)  
biz.deal.download.read.timeout=15  
  
# 处理base64后的图片 call 特征接口的 queenCapacity  
biz.deal.download.queue.request.size=600  
# HttpClient连接超时时间 (s)  
biz.deal.http.client.connect.timeout=10  
# socket 通信超时时间 (s)  
biz.deal.http.client.socket.timeout=15  
########### [特征处理相关配置]-[End] ##############################

# 每次传入 特征提取模块的batchSize  
biz.deal.recog.feature.pic.batch=10

比较重要的参数,表示每次提交到特征接口的数据batchSize，正常情况下 <= 8-10

# 调用特征分析服务 固定线程池大小  
biz.deal.recog.feature.request.size=30

识别接口性能较好的时候可以加大一点，性能不好的时候调小，调太大,会出现很多readtime-out接口响应不过来的情况，建议为 25-30之间，太小处理速度又要慢很多

### 1.2.5 EasticSearch相关配置

########## [Elasticsearch 等相关配置] [Start]#####################  
# [特别提醒:indexName,indexTypes,idColumns三者一定要对应上]  
# 需要数据转换处理的es indexNames  
deal.old.es.to.es.index.names=vlpr\_result,objext\_result,bike\_result  
# typeNames  
deal.old.es.to.es.index.types=data,data,data  
# id columns 以xx做id 存入 es库  
deal.old.es.to.es.id.columns=idTest,idObjextTest,idBike  
# es 滚动查询大小 每次查询多少条记录  
deal.old.es.to.es.load.batch.size=10000  
# 版本号汇总[4.0.3.7,4.0.6,4.0.7]  
# 源es配置 [需要处理的es库] 172.16.1.15 172.16.1.29  
source.elasticsearch.version=4.0.7  
source.elasticsearch.host=172.16.1.15  
source.elasticsearch.port=9200  
source.elasticsearch.username=  
source.elasticsearch.password=  
# 目标es源配置 172.16.1.68  
target.elasticsearch.version=4.0.7  
target.elasticsearch.host=172.16.1.68  
target.elasticsearch.port=9200  
target.elasticsearch.username=  
target.elasticsearch.password=  
# mysql 刷新到 es [只用来创建表数据]  
# 默认情况下和target保持一致的配置  
refresh.elasticsearch.version=4.0.7  
refresh.elasticsearch.host=172.16.1.68  
refresh.elasticsearch.port=9200  
refresh.elasticsearch.username=  
refresh.elasticsearch.password=  
########## [Elasticsearch 等相关配置] [End]#####################

### 1.2.6 特征提取相关配置

###################### 系统BIZ业务相关 [Start] #################  
# 选择的识别类型type=[0:视频结构化][1:图片结构化]  
api.recognize.service.chose.type=1  
# [视频服务]识别服务接口地址 [默认]  
api.recognize.video.feature.request.url=http://172.16.1.68:8100/v8/images/objects  
# 是否开启多个识别接口集群轮询选择[默认为false]  
api.recognize.video.feature.cluster.enable=true  
# 提供识别服务的多个接口 [\ 多个换行] [url#权重] [性能好一点的机器权重配高一点]  
api.recognize.video.feature.cluster.urls=http://172.16.1.46:8101/v8/images/objects#1,http://172.16.1.46:8102/v8/images/objects#1,\  
 http://172.16.1.46:8103/v8/images/objects#1,http://172.16.1.46:8104/v8/images/objects#1,\  
 http://172.16.1.63:8100/v8/images/objects#1,http://172.16.1.63:8101/v8/images/objects#1  
  
# [图片服务] 识别图片接口地址 10901  
api.recognize.picture.feature.request.url=http://172.16.1.63:10901/images/recog  
# 是否开启多个识别接口集群轮询选择[默认为false]  
api.recognize.picture.feature.cluster.enable=true  
# 提供识别服务的多个接口 [\ 多个换行] [url#权重] [性能好一点的机器权重配高一点]  
api.recognize.picture.feature.cluster.urls=http://172.16.1.63:10901/images/recog#1,http://172.16.1.63:10902/images/recog#1  
# kafka 主题  
api.kafka.recognize.object.topic=feature-test-820  
# 4.0.3.7 mysql数据处理时间节点  
# [起始时间]/[结束时间] eg: 10天前-10天后 [默认] true 开启  
# 注意: 时间配置有两种方式 true为index方式[index1-index2] eg:[-10,10]  
deal.old.mysql.to.es.date.index.enable=false  
deal.old.mysql.to.es.start.index=-5  
deal.old.mysql.to.es.end.index=0  
# 根据时间来配置[index=false才生效] [20190701-20190801]  
# 如果要跑0726的数据[20190726-20190726]  
deal.old.mysql.to.es.start=20190817  
deal.old.mysql.to.es.end=20190817  
###################### 系统BIZ业务相关 [End] ####################

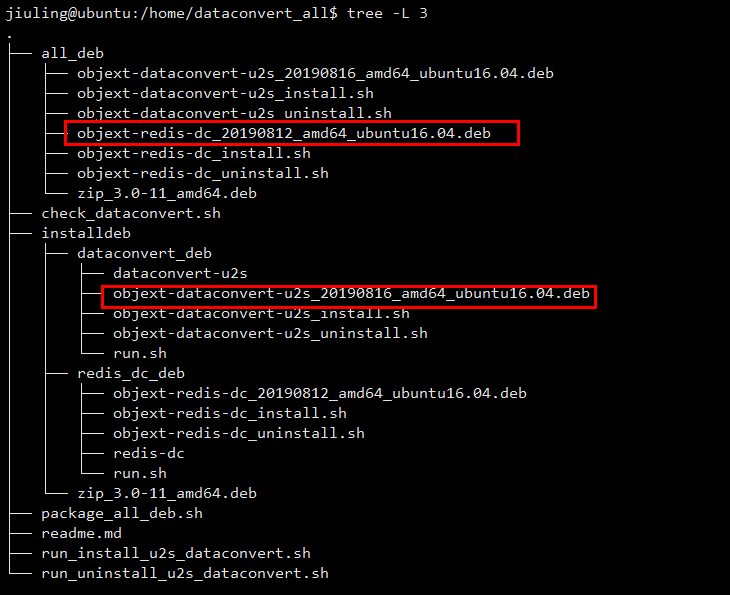
可以支持单个接口提供特征识别功能，也可以实现多个接口，然后轮询调用处理数据，提取特征，具体根据实际的情况来.

### 1.2.7 vsd\_task\_temp 🡪 vsd\_task

###################### Task任务处理相关 [End] ####################  
# [如果是同一个mysql数据库操作的话url配置为一致]  
# 数据库task处理信息  
biz.task.manager.add.task.url=http://172.16.1.68:20280/rest/taskManage/addVideoObjextTask  
# 目的表  
biz.source.mysql.table=vsd\_task  
# 目的数据库  
biz.source.mysql.url=jdbc:mysql://172.16.1.68:3306/u2s?verifyServerCertificate=false&useSSL=false  
# 临时中转表  
biz.source.mysql.table.temp=vsd\_task\_temp  
# 临时数据库  
biz.source.mysql.url.temp=jdbc:mysql://172.16.1.68:3306/u2s?verifyServerCertificate=false&useSSL=false  
###################### Task任务处理相关 [End] ####################

Mysql 数据库里面业务表数据的迁移，前提环境！

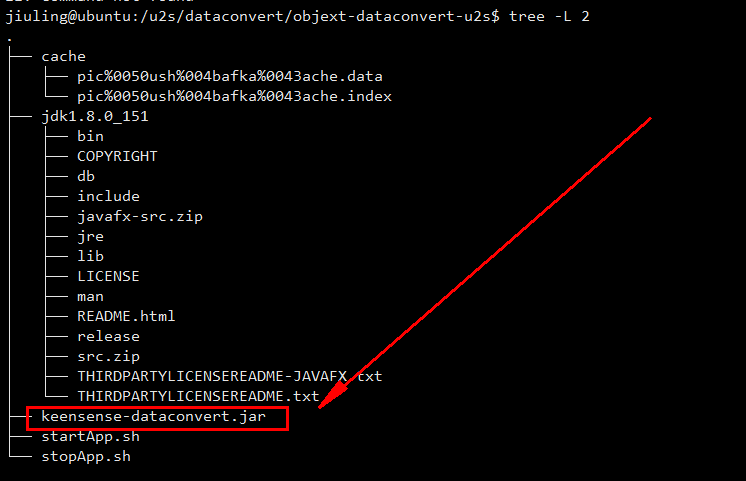
## 2.3程序安装配置



解压安装 data-convert对应的deb包 安装，主要依赖了一个redis，redis用于异常收集，处理处理异常的数据，统计数据的处理量

程序的安装目录

/u2s/dataconvert/objext-dataconvert-u2s



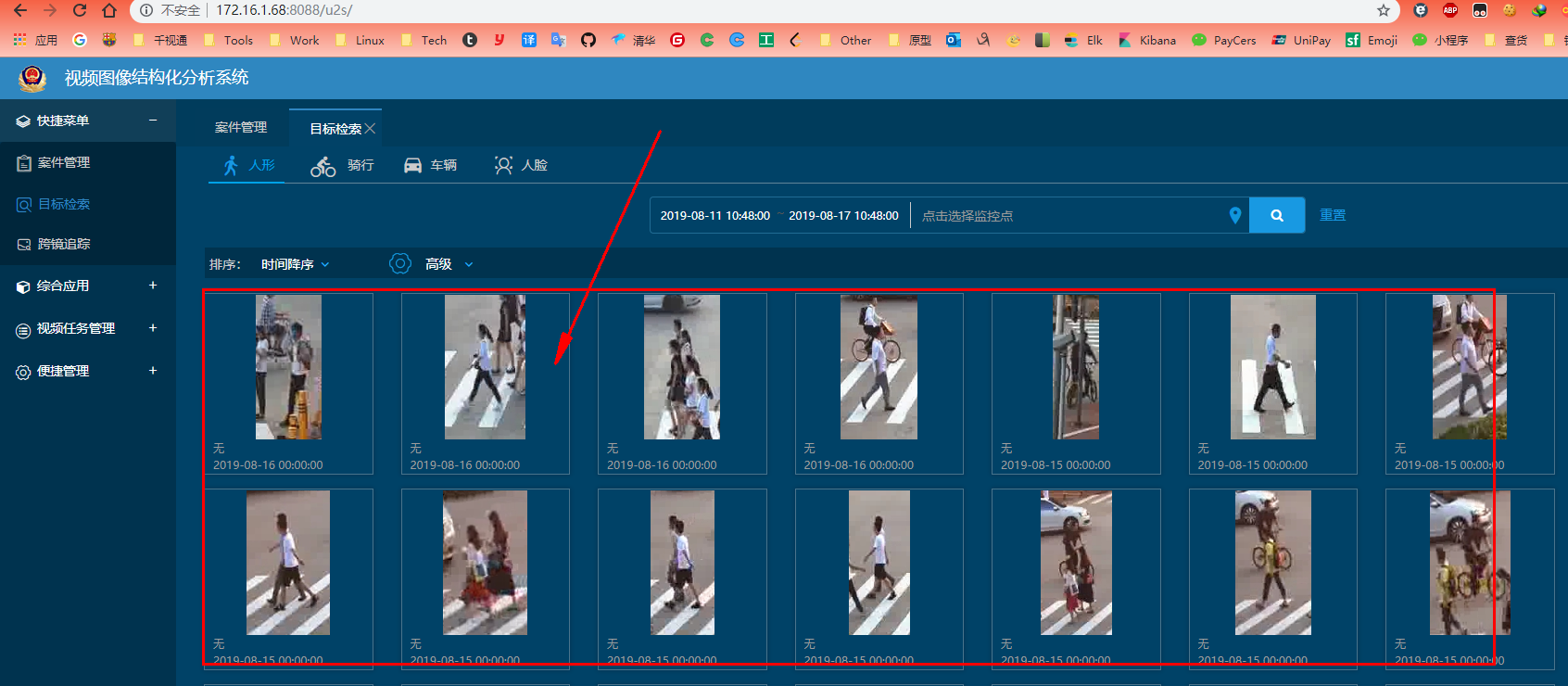
对spingboot的jar包进行替换（如果有代码变动的话！

启动后程序要是进行数据处理

# 观察迁移结果

### app应用页面数据可以查看到

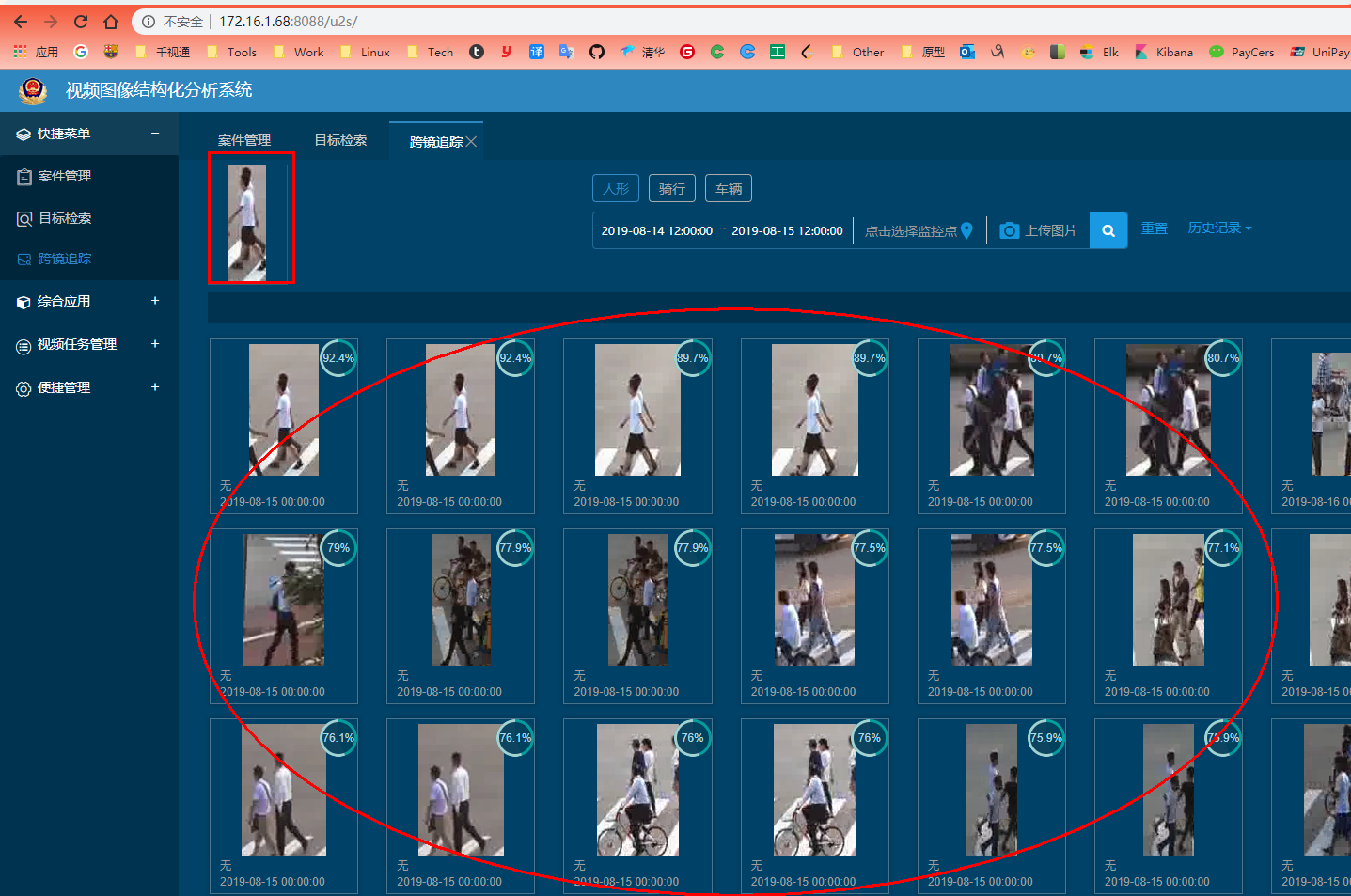
<http://172.16.1.68:8088/u2s/>



数据来源为mysql 页面就可以展示

### 3.2 app应用页面数据可以进行搜索

<http://172.16.1.68:8088/u2s/>



导入的数据 可以在页面进行搜索到