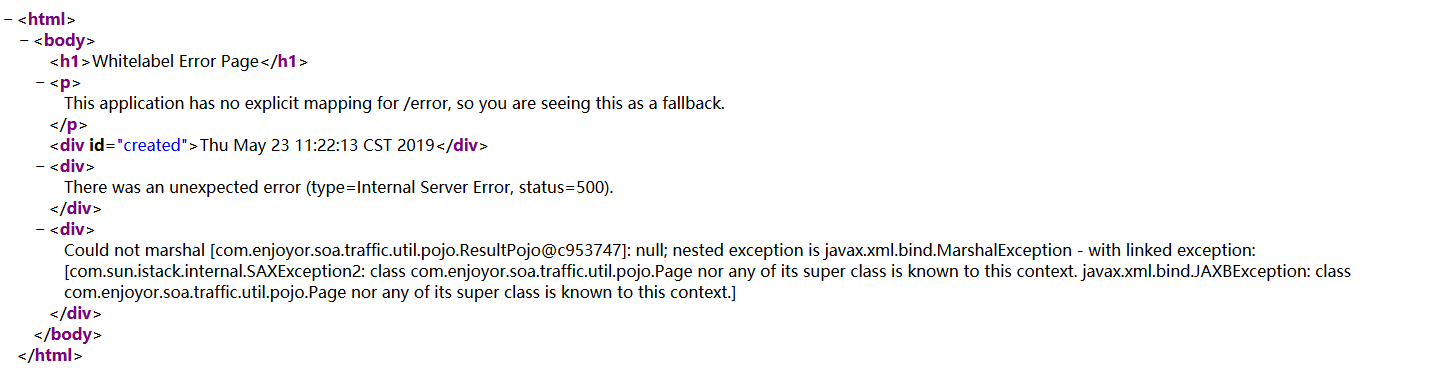
# Solr 代码优化异常汇总

## 1:rest服务在url查询出问题,swagger测试正常



现象:

rest接口在swagger测试正常,使用url查询出现

抛Could not marshal [com.enjoyor.soa.traffic.util.pojo.ResultPojo@c953747]: null; nested exception is javax.xml.bind.MarshalException - with linked exception: [com.sun.istack.internal.SAXException2: class com.enjoyor.soa.traffic.util.pojo.Page nor any of its super class is known to this context. javax.xml.bind.JAXBException: class com.enjoyor.soa.traffic.util.pojo.Page nor any of its super class is known to this context.]

定位代码:



发现接口中已经将对象转成JSon类型 ,这里又将代码转成XML格式 出错 ,代码注释问题解决

## 2:盐城单机Solr对Schema识别问题

在盐城发布rest查询时,一直提示pojo跟查询的结果在bean转换时字段类型不一样

盐城单机版本的string+multiValued 识别是string

集群是string+multiValued 识别是ArrayList

但是公司搭建的单机Solr并没有出现这个问题!!!!

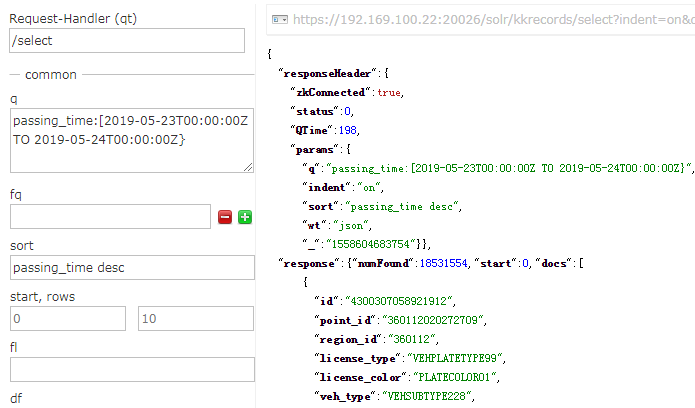
**难道是安装包的问题**

盐城装时候不能联网, 使用的之前的

公司单机安装时联网下的

wget http://archive.apache.org/dist/lucene/solr/6.2.0/solr-6.2.0.tgz

## 3: Solr查询优化



数据量大的情况下 先用时间固定一批数据加载到内存,再从在这批数据中过滤

原因 q:先拿到数据 fq再从这批数据做筛选

## 4: SolrException: Error opening new searcher

一般发生在增加/修改字段之后,没有及时进行reload

对应异常:

Solr入库程序会出现id :XXXXXXX 不能正常入库,程序正常运行,数据没有入库

解决方式:reload 该collection 或者 重启solr服务

## 5: 3 字节的 UTF-8 序列的字节 3 无效

Main.xml 文件修改后,数据入库抛这个异常

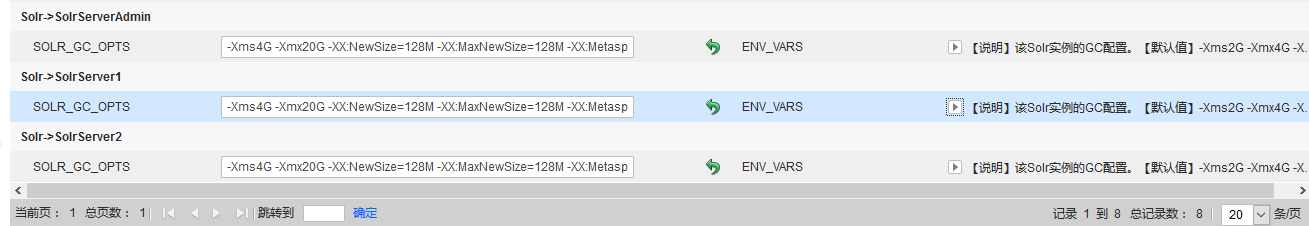
原因是xml配置文件中出现中文或中文空格 将中文注释删除掉

是使用记事本编译时,格式出现问题

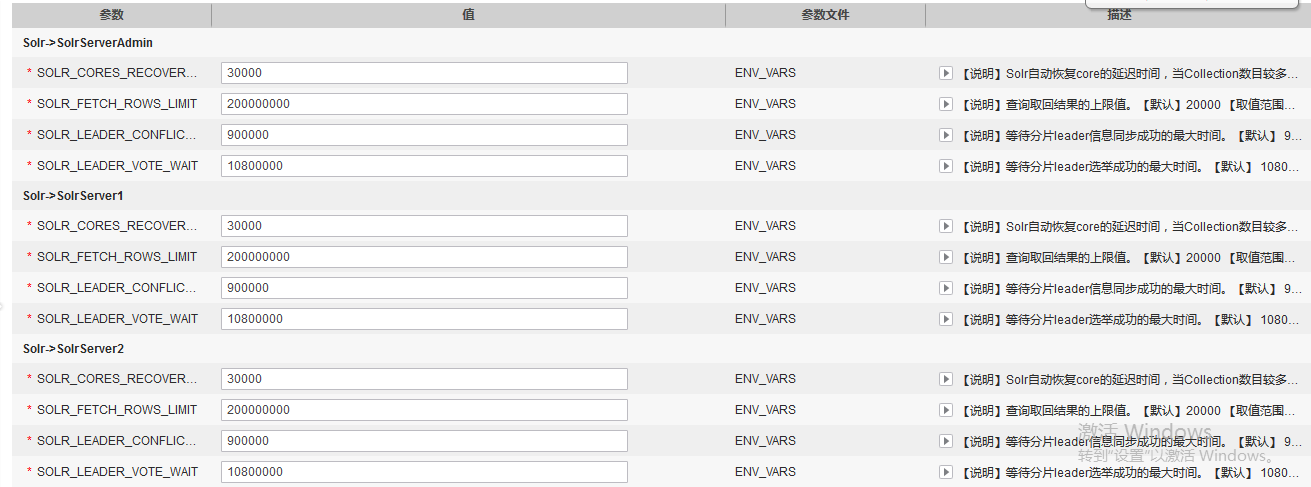
## 6:想看最后一页

返回限制改到2亿条

内存加大到7节点/10G 3个节点/20G



最大限制的修改成7天的返回2亿条

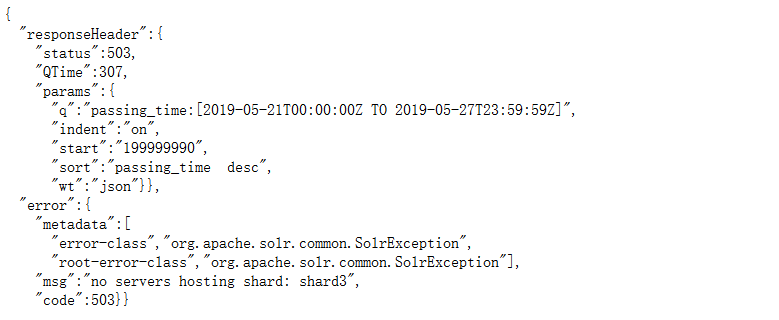


在CPU99%的情况下 重启Solr 服务?

正常重启

性能 查询反而变慢了

70G内存 太大了 导致内存超过域值 ,Solr节点出问题



改成20G,限制2亿NCJJ07 起不来

限制返回改成3000w,服务正常,但是翻页仍然不行

{

"appCode": "9",

"dataBuffer": "未知",

"resultList": "No live SolrServers available to handle this request:[https://192.169.101.4:21107/solr/kkrecords, https://192.169.101.8:21107/solr/kkrecords, https://192.169.101.5:21107/solr/kkrecords, https://192.169.101.7:21104/solr/kkrecords, https://192.169.101.6:21104/solr/kkrecords, https://192.169.101.9:21104/solr/kkrecords, https://192.169.101.4:21104/solr/kkrecords, https://192.169.101.7:21107/solr/kkrecords, https://192.169.101.6:21107/solr/kkrecords, https://192.169.101.1:21101/solr/kkrecords]"

}

返回3000w条 一查2990w条数据 就卡死了



## 7:新增加lane\_number和二次识别的字段作为查询条件

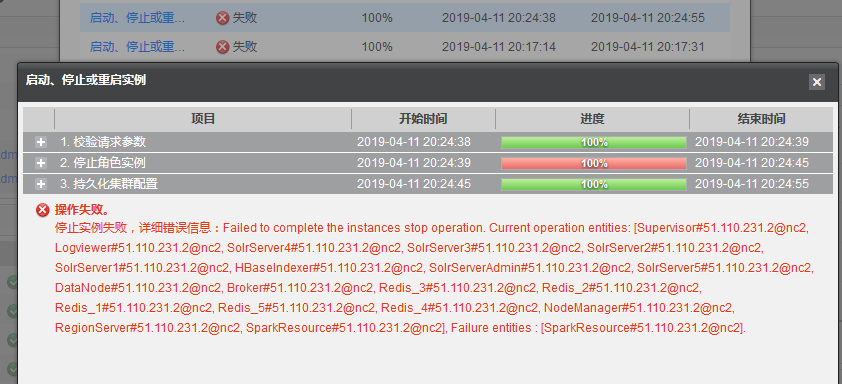
单机上测试字段的时候 ,将lane\_number字段删除,string变成tlong,copy to \_text ;

出现查询异常

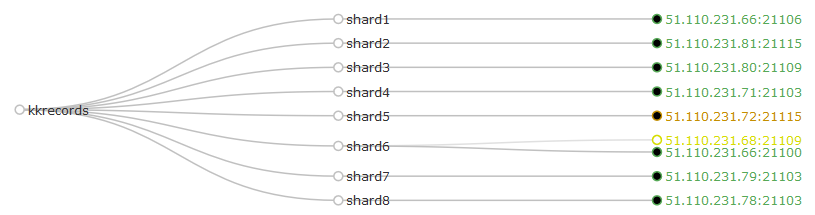
最终定位原因是

修改字段前必须先关闭数据写入,否则会出现这种情况

## 8:Solr数据节点挂掉,导致往这里存的数据失败 少数据

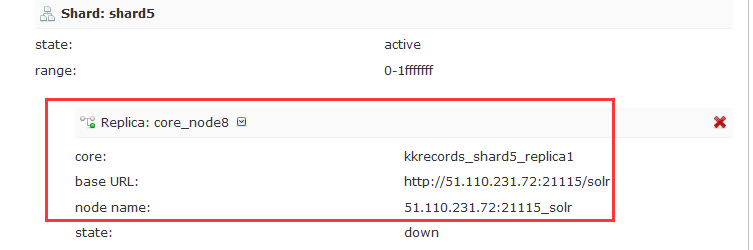


入库程序关闭,重启服务:



定位到这台机子nc8 231.72:





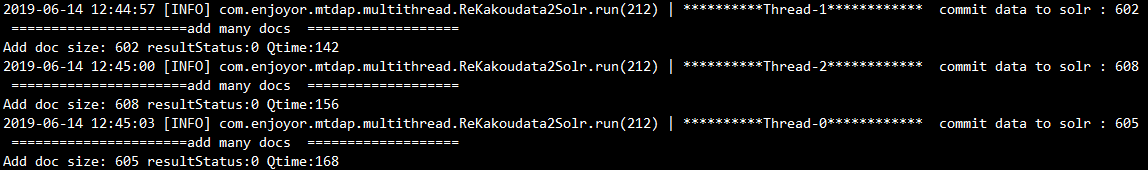


将出问题的shard5 删除完,自动分配

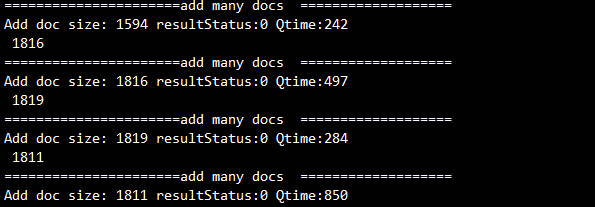
## 9:Solr入库程序消费不及时

南昌数据量3000w 之前的程序 性能突然下降了一半1800降到700+) 导致现在数据滞后

改写成多线程消费,效率如下

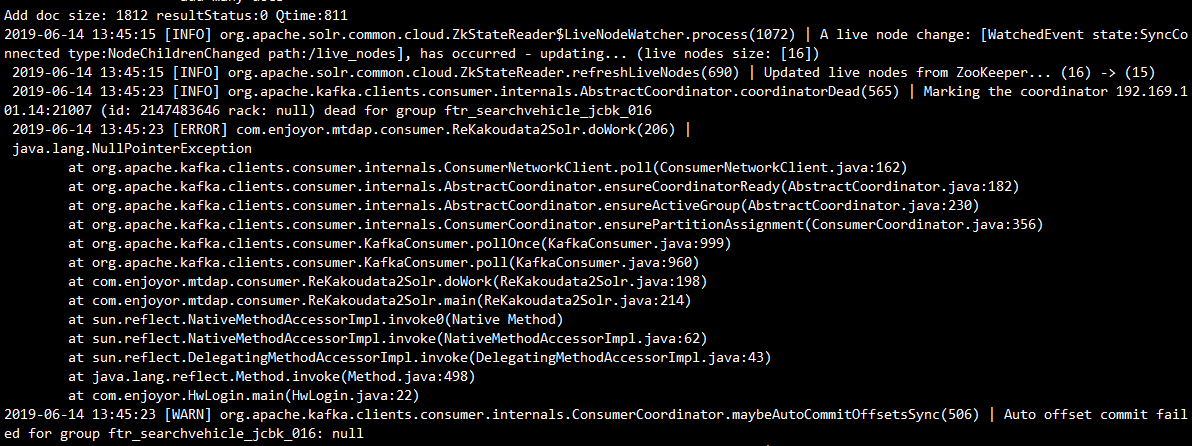


现象:消息滞后 我打印输出是600+ 多线程程序关闭后又变成1800(程序没改变)



原程序测试: 40分钟 滞后3分钟

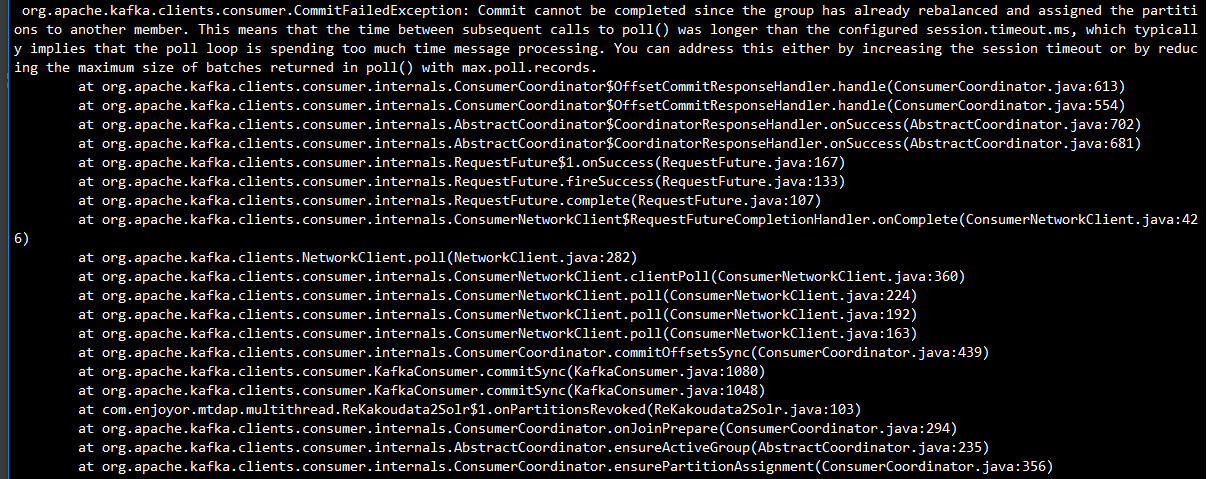
老程序挂掉:



多线程修改:

Topic的分区6 多线程加到6,去掉client\_id ,poll(500) 测试消费速度

出现这个bug ,时间加大:



这个bug是poll()时间太短

Bug:多线程的问题(每次恒定拿数据600+)



不加线程,取数据 有段时间也是600+ 后来多线程执行一次后 ,单线程的老程序变成1800

原因:就是同时发布多个消费者程序:一个程序1800+/次,然后开一个消费者抢夺2个变成600+,关掉消费者,原来的程序还是600+,解决方式:原来程序kill掉,重新发布.

使用多线程并没有增加消费能力的原因是,多个线程存取和一个线程存取数据是一样的 ,应该是一个线程负责取数据到BlockingQueue,另外几个线程负责从BlockingQueue消费数据

## 10:南昌solr分表(历史表,实时表)

详细:

南昌问题:出现严重的数据滞后,添加实例,找不到配置集 重建索引字段 写不进数据 (在于错误操作:添加副本,数据锁定在备份)

原因:数据量存3个月25亿条 过车数据实时创建索引 页面滞后明显

操作:添加实例(2->4) 数据写入失败,找不到索引字段(配置集从confWithCommon --> confWithSchema 原因未知)

-->错误操作: 先添加备份 导致数据写不进去(此时数据在备份)

-->在confWithSchema添加索引

-->删除Recovering状态下的副本 问题解决

南昌solr问题优化方式:

在分片的基础上分表 历史表(全部) 实时表(最近3个月)

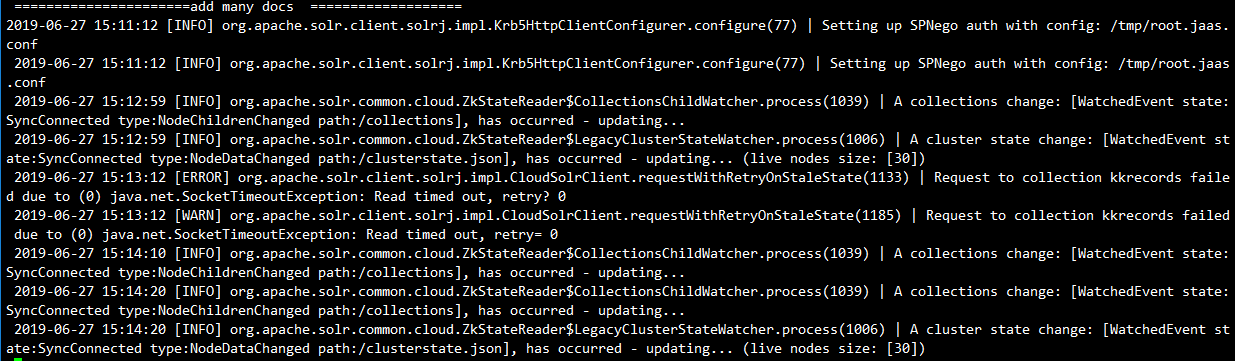
rest接口做成可配置的 判断查询的过车开始时间绝地查3个月实时表,或全部的历史表

(九江2年前查历史表(空的),南昌3个月前查历史表)

kkrecord:作为实时表(28shard) 数据同步(hive-->solr)

实时接入kafka数据 ,历史数据导入3个月 4->6月

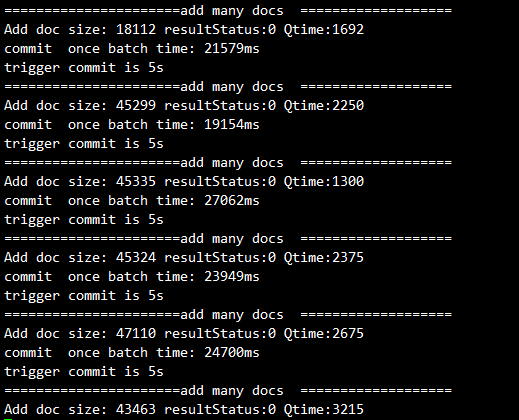
Bug如下:添加实例,数据写入抛:



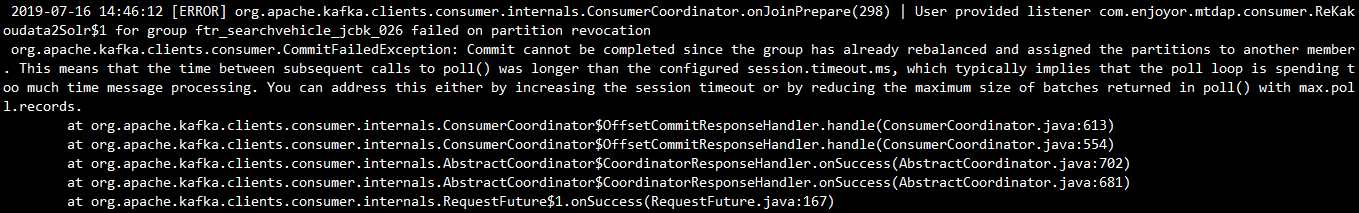
原因是创建一个备份,数据在同步 锁死了 将备份删掉了 可以

## 11:Solr数据入库时间加长

数据库存28亿条 日增3300w

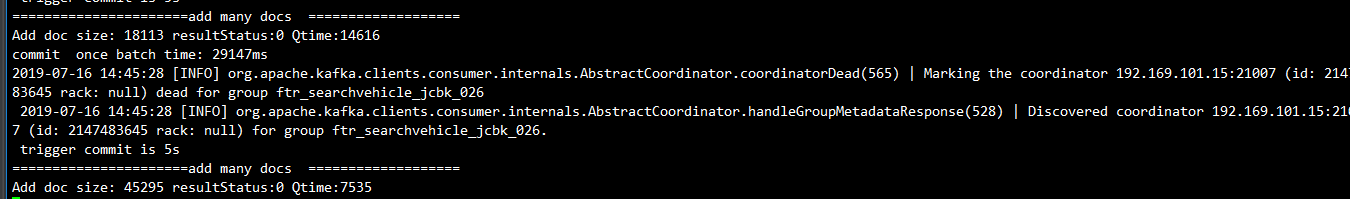


当入库时间超过30s时,会出现



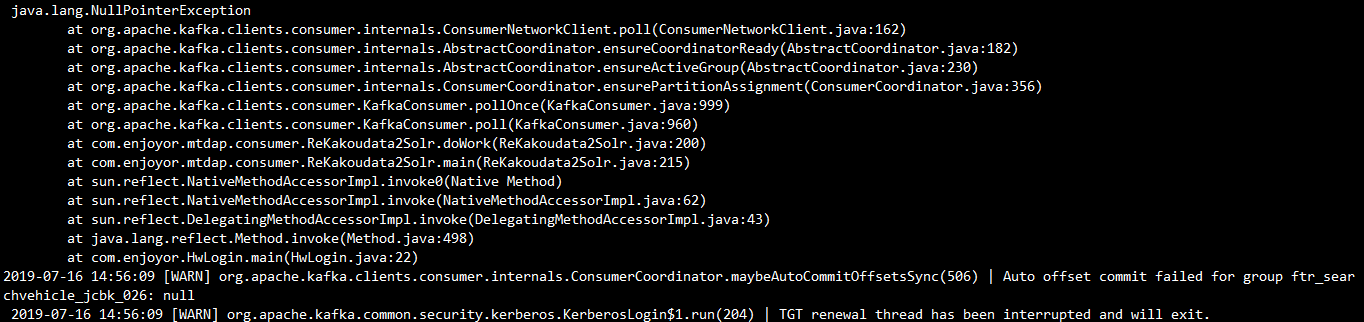
但是不会导致程序挂掉.

入库时间之前23亿条是7s 28亿现在15-20多s



出现rebalance 因为入库根据主键更新所以不会影响solr存储

有时会出程序异常退出:



在poll这一行添加try-catch{} 同时减小consumer.poll(n)中的参数n为100

数据正常:

