通用kpi说明书

1. kpi计算原始数据生成
2. 先创建五张原始数据存储表

使用auto\_cretb.sh中的建表语句进行五张表的创建，五张表的表名分别为：doss\_table,claim\_table,part\_table,order\_table,cust\_table

创建操作如下：

sh auto\_cretb.sh cre\_tb\_five\_fact

1. 五张原始表的数据导入

将十张表的数据通过pyspark文件将十张表的数据导入到五张数据表中。使用的py文件为：claim.py,cust\_order.py,part.py,doss.py,order.py运行命令见任务提交脚本，文本存放在任务提交命令文件夹下。

注意事项：

1. 将py文件中的参数传入进行修改，需要将参数的个数，及格式是否正确进行相关代码的修改
2. 看代码的能否进行优化，看数据能否进行暂时的缓存
3. 创建表的时候需要注意表的存放路径以及数据库，不需要参数传入
4. auto\_cretb.sh文件存放在auto项目bin目录下
5. 按规定要求需要的用户的家目录下建立目录：genal/bin,data,log
6. 所有的可执行文件都存放在bin
7. 数据导入后需要查看数据是否导入成功，数据的存放目录是否正确
8. 数据去重

1.去重规则

数据去重主要是对原始数据表中的数据进行去重操作。

规则是含有a，b，c三个字段都相同的若干数据为重复  
以下是每张表中对应的a，b，c字段对应的下标

claim=  
doss=7,12 VIN,DEALER\_NUMBER  
customer=0,1,10,46 VIN,LNUMBER,CH\_CODE,OWNER\_CODE  
order=3,27 ORDER\_NUMBER,ASC\_CODE  
part=

1. 去重操作

去重的mr程序全部是在uniq-data项目下，所有去重操作均为一套逻辑代码

去重操作的脚本是auto\_mr.sh,存放在脚本目录下

执行的命令如下：

sh auto\_mr.sh uniq\_cus <参数：时间>

sh auto\_mr.sh uniq\_doss <参数：时间>

sh auto\_mr.sh uniq\_order <参数：时间>

sh auto\_mr.sh uniq\_claim <参数：时间>

注意事项：

1. 在程序运行的时候需要按要求更改数据的输入路径和文件的输出路径，如果当文件不需要后续操作时，可以将文件的输出路径当作是外部表的存储路径，否则的话进行下步操作。
2. 在参数传值的时候格式如下：

参数： mon=2017-05

3.参数传入时应当和原始数据表中的分区格式保持一致

1. 打标签

目前需要对order表进行打标签，doss表在生成五张数据表的时候已经在打上标签，其写在py文件中。

1.标签类型

(1)是否首保 是否索赔  
 需要关联claim表 索赔单号 4s  
 (2)维修类别1(maintType)  
 需要关联part表 asc\_code,工单号，结算日期，vin

2.执行操作：

sh auto\_mr.sh delete\_useless <参数：时间>

sh auto\_mr.sh label\_fm <参数：时间>

sh auto\_mr.sh label\_fix <参数：时间>

sh auto\_mr.sh label\_other <参数：时间>

1. 创建最终原始数据存储表

将打完标签以后的数据的存储地址作为最终kpi计算的原始表的存储路径。

需要使用命令如下：

sh auto\_cretb.sh cre\_tb\_fact

4.给数据表添加分区

使用命令如下：

sh auto\_mr.sh add\_part

注意事项：

1.order表已经在最后一列打入一列“删除”标签，字段数 +1,这一步结束后会在order中再加入两列，第一列是否首保，第二列是否索赔  
2.打完标签后order表会有40个字段

3.在创建最终数据表的时候里面的tb\_fact的路径需要使用打完标签以后数据的存储路径

4.必须在打完标签创建好表以后给表添加分区

1. Kpi计算

dim表中的数据是否需要更新。kpi计算需要执行四个py文件，分别是zjxj.py,zhanbi.py,fjp,py,jpl.py,py文件的执行是有顺序的，执行顺序是直接相加类，占比类，非基盘类，基盘类

1. 创建dim表，命令如下：

sh auto\_cretb.sh cre\_tb\_dim

sh auto\_cretb.sh cre\_flow\_dim

1. 创建kpi结果表

sh auto\_cretb.sh cre\_tb\_kpi\_tmp 8areas #占比类的数据存储表

sh auto\_cretb.sh cre\_tb\_kpi\_tmp 8areas\_k1 #直接相加的数据存储表

sh auto\_cretb.sh cre\_tb\_kpi\_other 8areas #基盘类和非基盘类

1. dim 表数据的导入

dim表数据需要我们手动导入，导入命令如下：

sh auto\_cretb.sh load\_data\_dim

4.执行kpi计算脚本

执行命令见提交命令

注意事项：

1. 使用原始sql文件转py文件时，一定要注意文件转换时使用的方法
2. 在使用文件是需要修改文件的变量，文件的路径
3. 文件在转换的时候使用的是create-table包下面的类进行文件的转换，py文件转换使用的是createshellpy类。
4. 需要注意的是dim表中一些表带有分区，需要对有分区的表进行添加分区的操作。
5. 所需要的脚本可以在工程的bin目录下找到
6. 每个py文件执行的时候传入的参数是不相同的
7. 通用平台使用的是spark-submit2
8. 本地新建目录

general/data

general/log

geneeal/bin

1. hive放文件

将建表语句的脚本与mr的脚本先放到bin目录下。