# 简述

Java源代码地址：

http://git.minrow.com/lipengbo/autodatacheck

自动化检查工具只做2件事情：

1/离线和实时数据比对，发现异常需要告警（电话，短信，邮件）；

2/对数据下载功能所需要的数据进行检查，发现异常需要告警（电话，短信，邮件）；

工具的代码执行逻辑：

1/并行执行检查点任务（并行度为10）。跟据5.1表配置信息，比对实时离线指标，将比对结果写入tkio\_data\_check\_result表

2/对于上一步执行出错的任务重新执行（并行度为1）

3/并行执行分指标统计（并行度为5），对tkio\_data\_check\_result表进行处理写入tkio\_data\_check\_result\_app表，然后再对app表进行汇总写入tkio\_data\_check\_result\_indicator。

在此步骤里加入差异过滤规则：（指标值大于等于10且差异大于等于10%的数据条数/总数居条数）得到的比值这里这里写入rows\_all\_rate字段。

4/对于上一步执行出错的任务重新执行（并行度为1）

5/对第3步得到的结果进行检查，如果rows\_all\_rate字段大于5.2.2表内相应的报警值则会出发报警。如果所有指标都小于等于报警值，则发送邮件显示tkio\_data\_check\_result\_indicator表详细结果。

6/检查日志表tkio\_data\_check\_log，如果发现前4步有报错也会电话断行发邮件告警。

7/删除以下四张表15天以前的数据：

tkio\_data\_check\_log，tkio\_data\_check\_result，tkio\_data\_check\_result\_app，tkio\_data\_check\_result\_indicator

8/下载数据检查：访问相应的hive表，得到各个appid对应的数据条数，遍历所有appid看是否在s3对应目录里有数据，如果某个app在s3上找不到数据，则会告警

# 2．调用

sh /mnt/engine\_general/autodatacheck/autodatacheck.sh >/mnt/engine\_general/autodatacheck/autodatacheck.log 2>&1 &

|  |  |  |
| --- | --- | --- |
| 参数名称 | 默认值 | 备注 |
| exectype | All | all（检查所有），check（只做实时离线数据比对），download（只做下载数据检查） |
| ds | 前一天 | yyyy-MM-dd数据检查日期 |
| taskno | Null | 检查点编号，如果有多个则用逗号分隔 |

# 3．参数文件配置

存放位置：/mnt/engine\_general/autodatacheck/checkconf/settings.properties

|  |  |
| --- | --- |
| 参数名 | 备注 |
| mySQLConfConnectionURL | 自动化检查工具配置信息数据库 |
| mySQLDataConnectionURL | 实时离线数据存放数据库 |
| hiveTkioConnectionURL | Hive连接地址 |
| tkio\_mid\_install\_day | 数据下载-激活 |
| tkio\_mid\_reg\_day | 数据下载-注册 |
| tkio\_mid\_pay\_day | 数据下载-支付 |
| tkio\_mid\_clk\_detail\_day\_down | 数据下载-在表名单里的黑名单点击 |

# 4．告警信息

## 4.1离线和实时数据比对

如果正常，每天会有一封邮件，内容是按指标统计的详细信息。

如果异常，会打电话发短信，短信内容是邮件标题。邮件内容是按app统计的详细信息，和按指标统计的错误信息。

如果检查点检查任务执行报错，会打电话发短信并邮件告知，内容是报错信息。

如果分指标统计任务执行报错，会打电话发短信并邮件告知，内容是报错信息。

## 4.2 下载数据检查

如果正常则没有告警信息。

如果异常会打电话发短信并邮件通知

# 5．数据库配置

## 5.1检查点配置

|  |  |  |
| --- | --- | --- |
| dqc\_data\_check\_config  数据检查点-配置表 | | |
| 字段名 | 示例 | 描述 |
| row\_id | 1 | 自增主键 |
| task\_no | 10001 | 检查点编号（必须以1开头） |
| task\_desc | 推广活动详情 | 检查点描述 |
| table\_type | T | table\_a和table\_b字段值是表则填T,是SQL语句则填V） |
| table\_a | tkio.tkio\_rpt\_source\_analysis\_day | a表名称，或SQL语句 |
| table\_b | tkio\_stream.tkio\_rpt\_source\_analysis\_day | b表名称，或SQL语句 |
| dim\_name | appid|campaignid|cid|ds | 维度字段列表，以|分隔（必须再table\_a,table\_b中同时存在）） |
| fact\_name | amt\_income\_new\_user|dupnum\_click\_all|dupnum\_click\_day|dupnum\_reged\_day|num\_black|num\_channel\_fake\_click|num\_click|num\_click\_fake|num\_install|num\_install\_fake|num\_payer\_new\_user | 指标字段列表，以|分隔（必须再table\_a,table\_b中同时存在） |
| filter\_where | and ds=dt | 跟在表后边的过滤条件（table\_type='T'时有效） |
| filter\_having | \N | 保留字段，暂无用 |
| notify\_type | 0 | 保留字段，暂无用 |
| dim\_desc | \N | dim\_name的字段描述 |
| fact\_desc | amt\_income\_new\_user|dupnum\_click\_all|dupnum\_click\_day|dupnum\_reged\_day|num\_black|num\_channel\_fake\_click|num\_click|num\_click\_fake|num\_install|num\_install\_fake|num\_payer\_new\_user | fact\_name的字段描述 |
| is\_delete | 0 | 是否删除1是0否 |
| create\_time | ####### |  |
| update\_time | ####### |  |
|  |  |  |
|  |  |  |
|  |  |  |

## 5.2 分指标统计配置

### 5.2.1 tkio\_data\_check\_indicator

|  |  |  |
| --- | --- | --- |
| dqc\_data\_check\_indicator  数据检查点-指标清单（必须是5.1中fact\_name字段内包含的指标） | | |
| 字段名 | 示例 | 描述 |
| row\_id | 1 | 自增主键 |
| indicator\_name | amt\_income\_new\_user | 指标名称 |
| indicator\_desc | 新增付费金额 | 指标描述 |
|  |  |  |

### 5.2.2 tkio\_data\_check\_indicator\_value

|  |  |  |
| --- | --- | --- |
| dqc\_data\_check\_indicator\_value  数据检查点-指标差异报警上限值（对各个检查点各个指标分别设置上限值） | | |
| 字段名 | 示例 | 描述 |
| indicator\_name | amt\_income\_new\_user | 指标名称 |
| task\_no | 10001 | 检查点编号 |
| maxrate | 0.363 | 报警值rows\_all\_rate |
| avgrate | 0.2511 | 无用字段 |
| minrate | 0.15 | 无用字段 |
| days | 10 | 无用字段 |
| create\_time | ####### |  |
| update\_time | ####### |  |

# 6．日志查询

|  |  |  |
| --- | --- | --- |
| dqc\_data\_check\_log  tkio数据检查点-执行日志 | | |
| 字段名 | 示例 | 描述 |
| row\_id | 66 | 自增主键 |
| ds | 2018-02-27 | 日期yyyy-MM-dd |
| task\_no | dupnum\_click\_day | 检查点编号，如果是指标此处会写指标名称 |
| exec\_sql |  | 执行SQL |
| exec\_state | 0 | 执行状态0成功-1失败 |
| error\_info |  | 错误信息 |
| create\_time | 2018-02-28 23:16:49 | 创建时间 |