**深运平台数据质量说明文档**

1. **输入字段**
2. **ajm\_idcard**

|-- serverReceiveTimestamp: long

|-- name: string

|-- sex: string

|-- nation: string

|-- birth: string

|-- addr: string

|-- idno: string

|-- authority: string

|-- validBegin: string

|-- validEnd: string

|-- face: string

|-- capTime: string

|-- deviceNum: string

|-- deviceSn: string

|-- gbNo: string

|-- cameraOneFace: string

|-- cameraTwoFace: string

|-- deviceAddr: string

1. **ap\_raw**

|-- serverReceiveTimestamp: long

|-- gbNo: string

|-- sn: string

|-- deveice\_mac: string

|-- MACs: array

| |-- element: struct (containsNull = true)

| | |-- timestamp: long

| | |-- mac: string

| | |-- channel: integer

| | |-- signal: integer

1. **ifaas\_face**

|-- serverReceiveTimestamp: long

|-- gbNo: string

|-- timestamp: long

|-- capturePic: string

|-- thumbPic: string

|-- cameraId: long

1. **ifaas\_warning**

|-- serverReceiveTimestamp: long

|-- gbNo: string

|-- alarmId: long

|-- personId: long

|-- repoId: long

|-- taskId: long

|-- blackId: long

|-- faceId: long

|-- confidence: double

|-- timestamp: long

|-- alarmType: integer

|-- serverId: long

|-- realName: string

|-- idcardNo: string

|-- capturePic: string

|-- thumbPic: string

|-- blackPic: string

|-- cameraId: long

1. **rzx\_device**

|-- serverReceiveTimestamp: long

|-- EQUIPMENT\_NUM: string

|-- EQUIPMENT\_NAME: string

|-- MAC: string

|-- SERVICE\_CODE: string

|-- PROVINCE\_CODE: string

|-- CITY\_CODE: string

|-- AREA\_CODE: string

|-- EQUIPMENT\_CODE: string

|-- LONGITUDE: string

|-- LATITUDE: string

|-- SUBWAY\_STATION: string

|-- SUBWAY\_VEHICLE\_INFO: string

|-- SUBWAY\_LINE\_INFO: string

|-- SUBWAY\_COMPARTMENT\_NUM: string

|-- CAR\_CODE: string

|-- UPLOAD\_TIME\_INTERVAL: long

|-- COLLECTION\_RADIUS: long

|-- CREATE\_TIME: string

|-- gbNo: string

1. **rzx\_feature**

|-- serverReceiveTimestamp: long

|-- ACCOUNT: string

|-- AP\_CHANNEL: string

|-- AP\_ENCRYTYPE: string

|-- BSSID: string

|-- COMPANY\_ID: string

|-- CONSULT\_XPOINT: string

|-- CONSULT\_YPOINT: string

|-- DEVICENUM: string

|-- DEVMAC: string

|-- END\_TIME: long

|-- ESSID: string

|-- HISTORY\_ESSID: string

|-- IMEI: string

|-- IMSI: string

|-- MAC: string

|-- MODEL: string

|-- OS\_VERSION: string

|-- PHONE: string

|-- POWER: string

|-- PROTOCOL\_TYPE: string

|-- SERVICECODE: string

|-- START\_TIME: long

|-- STATION: string

|-- TYPE: integer

|-- URL: string

|-- XPOINT: string

|-- YPOINT: string

|-- gbNo: string

1. **rzx\_location**

|-- serverReceiveTimestamp: long

|-- SERVICE\_CODE: string

|-- SERVICE\_NAME: string

|-- STATUS: integer

|-- SERVICE\_TYPE: integer

|-- PROVINCE\_CODE: string

|-- CITY\_CODE: string

|-- AREA\_CODE: string

|-- XPOINT: string

|-- YPOINT: string

|-- gbNo: string

1. **sensordoor\_face**

|-- serverReceiveTimestamp: long

|-- gbNo: string

|-- mac: string

|-- genId: long

|-- time: string

|-- captureFaceList: array

| |-- element: struct

| | |-- face: string

| | |-- info: integer

1. **sensordoor\_idcard**

|-- serverReceiveTimestamp: long

|-- gbNo: string

|-- mac: string

|-- genId: long

|-- time: string

|-- idno: string

|-- name: string

|-- sexCode: string

|-- nationCode: string

|-- birth: string

|-- addr: string

|-- authority: string

|-- validBegin: string

|-- validEnd: string

1. **st\_alarm**

|-- serverReceiveTimestamp: long

|-- cameraId: string

|-- cameraName: string

|-- chaosDensity: integer

|-- crowdDensity: integer

|-- crowdNumberAllImage: integer

|-- crowdNumberWithEvent: integer

|-- dateTime: string

|-- image: string

|-- eventType: string

|-- startTime: string

|-- endTime: string

|-- lastTime: integer

|-- gbNo: string

|-- regionId: string

1. **st\_status**

|-- serverReceiveTimestamp: long

|-- cameraId: string

|-- cameraName: string

|-- dateTime: string

|-- intensity: double

|-- peopleCount: integer

|-- safetyIndex: integer

|-- gbNo: string

1. **ty\_imsi**

|-- serverReceiveTimestamp: long

|-- gbNo: string

|-- deviceId: long

|-- capTime: long

|-- imsi: string

|-- imei: string

|-- location: string

1. **basic\_device\_info(txt) (仅在需求‘1.1 设备量统计’中用到)**

|-- gb\_code: String

|-- dev\_id: String

|-- dev\_type: String

|-- dev\_name: String

|-- dev\_mac: String

|-- dev\_ip: String

|-- network\_type: String

|-- comp\_name: String

|-- org\_name: String

|-- project\_desc: String

|-- dev\_addr: String

|-- px: String

|-- py: String

|-- station\_id: String

|-- station\_name: String

|-- camera\_type: String

|-- camera\_quality: String

|-- dev\_status: String

|-- dev\_grid: String

|-- dev\_gridlist\_control: String

|-- create\_time: String

|-- update\_time: String

|-- line\_name: String

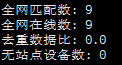
1. **输出字段(所有输出均为csv文件)**
2. **全网设备量统计**

全网匹配数:

全网在线数:

去重数据比:

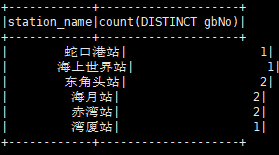
无站点设备数:



1. **各个站点匹配数**

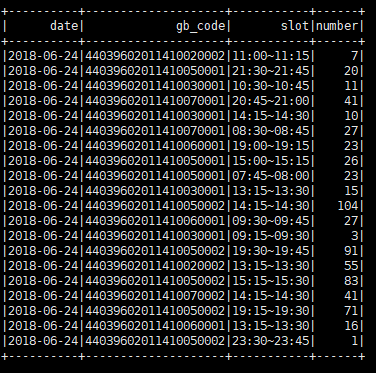
|  |  |
| --- | --- |
| station\_name | count(DISTINCT gbNo) |

注：station\_name从basic\_device\_info中取得，count(DISTINCT gbNo)表示对应站点的匹配数。

****

1. **单个设备15分钟粒度的数据采集数**

|  |  |  |  |
| --- | --- | --- | --- |
| date | gb\_code | slot | number |



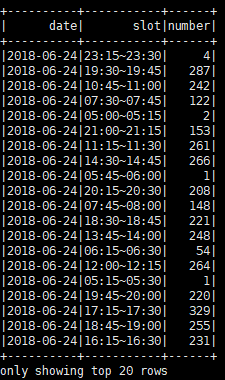
注：date指serverReceiveTimestamp所对应的日期，形式为："yyyy-MM-dd HH:mm:ss"；

slot指所处时段，分别可取：

"00:00~00:15","00:15~00:30","00:30~00:45", ……,"23:30~23:45","23:45~00:00"

1. **所有设备15分钟粒度的数据采集数**

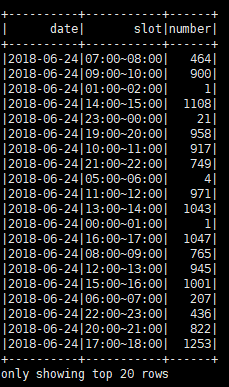
|  |  |  |
| --- | --- | --- |
| date | slot | number |

****

1. **所有设备每小时粒度的数据采集数**

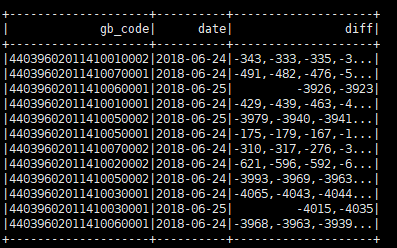
|  |  |  |
| --- | --- | --- |
| date | slot | number |

注：这里的时段分别取：”00:00~01:00”,”01:00~02:00”,……,”22:00~23:00”,”23:00~00:00”

****

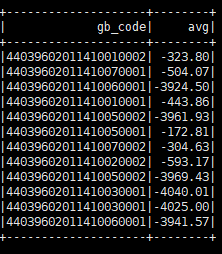
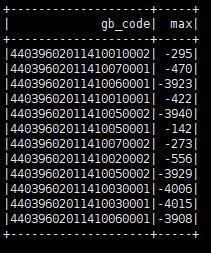
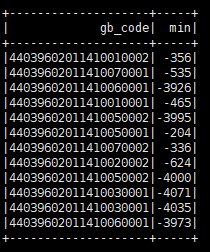
1. **单个设备分天计算延时分布**

|  |  |  |
| --- | --- | --- |
| gb\_code | date | diff |

****

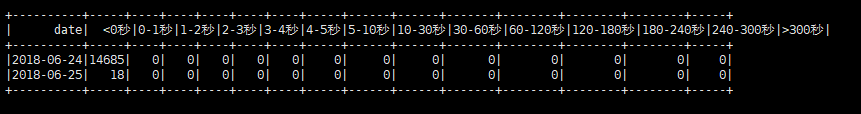
1. **单个设备平均/最大/最小延时**

|  |  |
| --- | --- |
| gb\_code | avg |
| gb\_code | max |
| gb\_code | min |

**  **

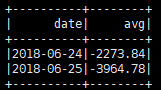
1. **所有设备分天计算分布**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| date | <0秒 | 0-1秒 | 1-2秒 | 2-3秒 | 3-4秒 | 4-5秒 | 5-10秒 |
| 10-30秒 | 30-60秒 | 60-120秒 | 120-180秒 | 180-240秒 | 240-300秒 | >300秒 |  |

****

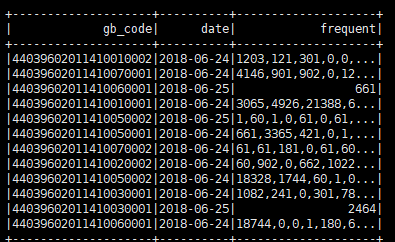
1. **所有设备延时日均值**

|  |  |
| --- | --- |
| date | avg |

****

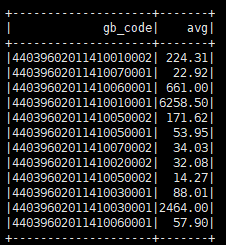
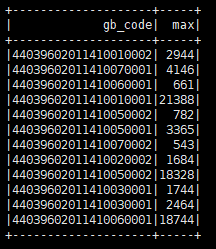
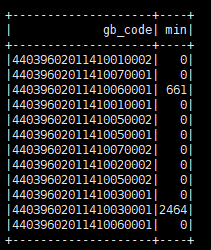
1. **单个设备分天计算发送频率**

|  |  |  |
| --- | --- | --- |
| gb\_code | date | frequent |

****

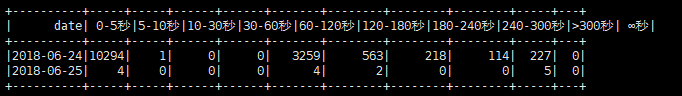
1. **单个设备平均/最大/最小发送频率**

|  |  |
| --- | --- |
| gb\_code | avg |
| gb\_code | max |
| gb\_code | min |

**  **

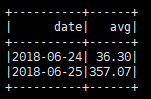
1. **所有设备分天所有频率分布**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| date | 0-5秒 | 5-10秒 | 10-30秒 | 30-60秒 | 60-120秒 |
| 120-180秒 | 180-240秒 | 240-300秒 | >300秒 | ∞秒 |  |

****

1. **所有设备频率日均值**

|  |  |
| --- | --- |
| date | avg |

****

1. **补充说明**
2. **ajm\_idcard**

延时：serverReceiveTimestamp -capTime

频率：serverReceiveTimestamp

1. **ap\_raw**

延时：serverReceiveTimestamp - timestamp

频率：serverReceiveTimestamp

1. **ifaas\_face**

延时：serverReceiveTimestamp - timestamp

频率：serverReceiveTimestamp

1. **ifaas\_warning**

延时：serverReceiveTimestamp - timestamp

频率：serverReceiveTimestamp

1. **rzx\_device**

延时：serverReceiveTimestamp - CREATE\_TIME

频率：serverReceiveTimestamp

延时结果较大，CREATE\_TIME可能不适合作为延时处理

1. **rzx\_feature**

延时：serverReceiveTimestamp - START\_TIME

频率：serverReceiveTimestamp

1. **rzx\_location**

只有receivetime,只能算频率，不能算延时

频率：serverReceiveTimestamp

1. **sensordoor\_face**

延时：serverReceiveTimestamp - time

频率：serverReceiveTimestamp

1. **sensordoor\_idcard**

延时：serverReceiveTimestamp - time

频率：serverReceiveTimestamp

1. **st\_alarm**

延时：serverReceiveTimestamp - dateTime

频率：serverReceiveTimestamp

1. **st\_status**

延时：serverReceiveTimestamp - dateTime

频率：serverReceiveTimestamp

1. **ty\_imsi**

延时：serverReceiveTimestamp - capTime

频率：serverReceiveTimestamp

1. **basic\_device\_info**

station\_name均取自此表，所以只能先匹配再查是否存在站点，故不予输出设备各个站点在线数

1. **问题说明**

① 程序执行和数据存储说明

1. 数据源: /user/hadoop/GongAnV2

2. 存储地址：/user/sibat/GongAn\_analyze

3. Jar包和脚本存放地址：/home/hadoop/GongAn\_2018

4. 每天凌晨四点运行一次

5. 执行sh时，命令行传入一个参数，参数名为日期文件夹名，例如：sh ajm\_idcard.sh 20180815

② 细则

1. 在公安测试20180815的所有数据成功；

2. 注意事项：

1） rzx\_feature，字段内容有“，”，不能用“，”作为分隔符，字段较多空值，报错数组越界，经排查，已解决；

2） rzx\_feature和ap\_raw中有异常数据文件，代码中将文件夹下文件一个个遍历读取，将异常文件跳过，需catch Exception、FileNotFoundException两个异常，问题解决。

3）sh脚本配置如下：

|  |
| --- |
| /share/apps/spark2.0/bin/spark-submit  --class com.beidou.shenyundataaseess.ajm\_idcard\_lzw  --master spark://190.176.35.102:7077  --total-executor-cores 32  --executor-memory 4G  --driver-memory 4G  ajm\_idcard.jar $1 |