**系统目录及数据说明**

1. **主目录**

121564188850567052

APP/ \* 存放所有APP数据

Attribute/\* 存放每个月的用户属性数据。

buy\_phone.csv 一汽丰田已购车数据。

CALL /\* 存放所有语音数据。

Keyword/\* 存放所有关键词数据。

Other/\* 其他数据，包括地铁站、lpoint\_id等相关数据。

qinghua/\* 该路径用于存放所有相关代码及处理后的数据。

SMS/\* 存放所有短信数据。

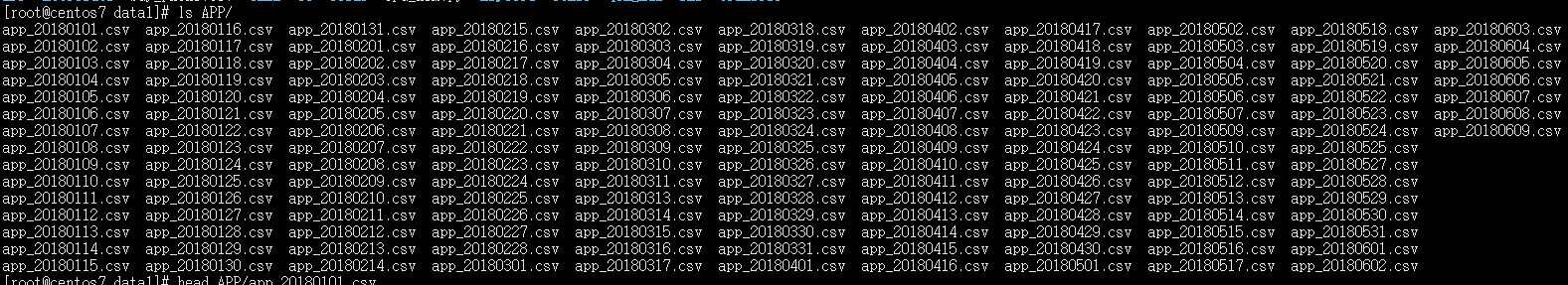
其它为陈程相关路径

1. **APP数据**

APP数据按天存储，文件路径：APP/app\_20180101.csv

单个文件大小：1G-3G

单个文件长度约为：5千万条左右。



字段说明：

phone\_nm：加密号码

app\_name：APP的名称

app\_count：APP访问频次（联通根据APP对应的host统计当天出现的次数。该对应关系为联通内部知识库，无法提供）。

app\_time：数据统计时间，精确到天。



说明：APP所有数据均已经进行初步清洗整理，均可直接使用pandas直接读取。每个文件均包含头。

1. **关键词数据**

关键词数据按天存储，文件路径：Keyword/key\_20180101.csv

单个文件大小：100M以内。

单个文件长度约为：一百多万条左右。



字段说明：

phone\_nm：加密号码。

key\_souce：用户搜索时使用的平台。

keyword：用户搜索的关键词。

key\_time：用户搜索关键词时间，精确到天。



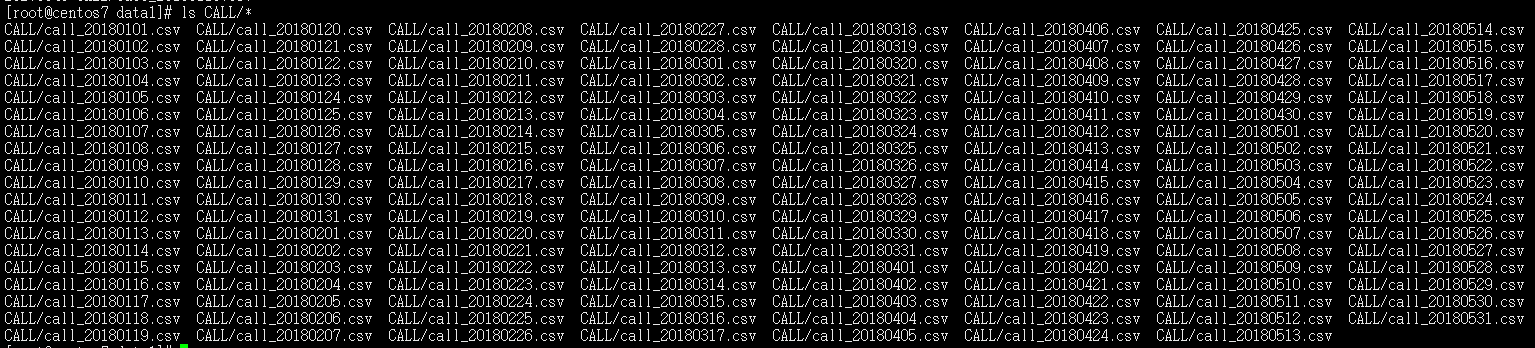
说明：关键词所有数据均已经进行初步清洗整理，均可直接使用pandas直接读取。每个文件均包含头。

1. **语音数据**

语音数据按天存储，文件路径：CALL/call\_20180101.csv

单个文件大小：2G以内。

单个文件长度约为：一两千万条左右。



字段说明：

phone\_nm1：加密号码1。（全为联通）

phone\_nm2：加密号码2。（可能联通，可能其他）

call\_type：呼叫类型，分为主叫和别叫。（主叫为号码1打出去；被叫为号码1接到电话）

call\_start\_time：呼叫时间，精确到秒。

call\_duration：通话时长。



说明：语音所有数据均已经进行初步清洗整理，均可直接使用pandas直接读取。每个文件均包含头。

**五、短信数据**

短信数据按天存储，文件路径：SMS/smsl\_20180101.csv

单个文件大小：4G左右。

单个文件长度约为：四千万条左右。

767920678270826785

字段说明：

phone\_nm1：加密号码1（联通和其他均由）

sms\_type：类型，分为短信和彩信；SMS为短信，MMS为彩信。

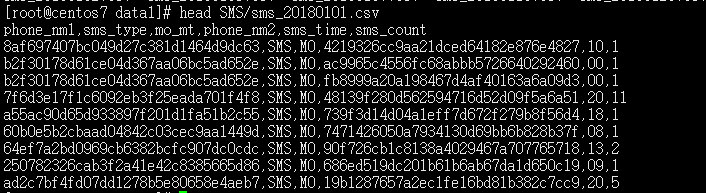
mo\_mt：MO上行（即发出）；MT下行（收到）

phone\_nm2：加密号码（联通和其他均由）

sms\_time：发送时间段

sms\_count：发送次数

备注：统计联通号码短信情况时，使用phone\_nm1。例如，统计联通A号码某天发短信次数，则在加密号码1中匹配A，不考虑A在号码2中出现的情况。



说明：短信所有数据均已经进行初步清洗整理，均可直接使用pandas直接读取。每个文件均包含头。

**六、用户属性数据**

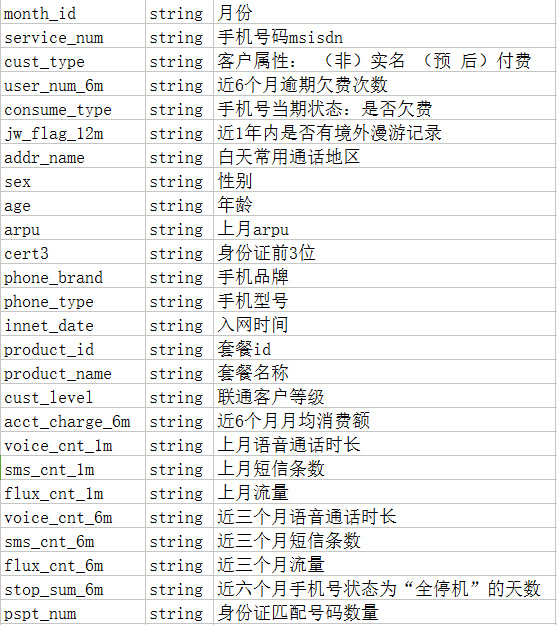
用户属性数据按月存储，文件路径：Attribute/attr201805.csv

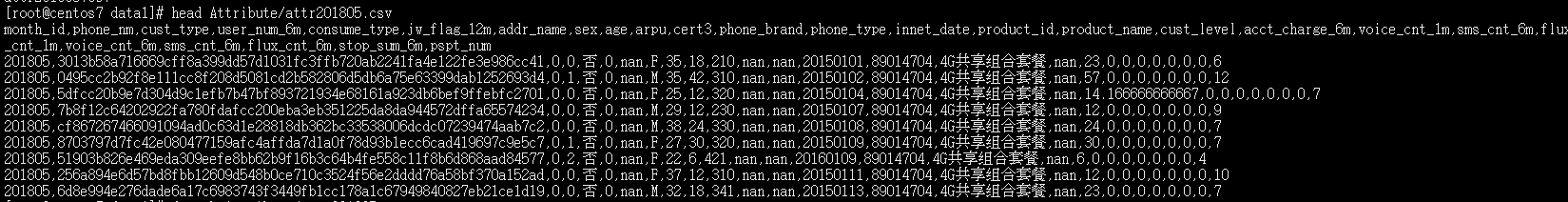
单个文件大小：2G左右。

单个文件长度约为：一千万条左右。

105209602148483330

字段：





说明：用户属性数据均已经进行初步清洗整理，均可直接使用pandas直接读取。每个文件均包含头，空值使用nan。

按月保存，每月一份。

**七、一汽丰田购车数据**

文件路径：buy\_phone.csv

字段：

buy\_phone：购车号码加密。

buy\_time：购买时间。



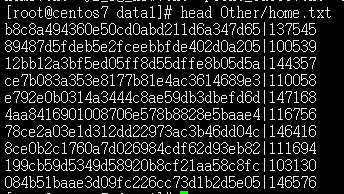
说明：文件包含头。已经经过数据清洗处理，可直接使用pandas直接读取。

**八、其他数据**

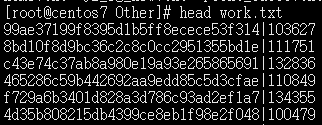
文件路径：Other/\*

430588569546207521

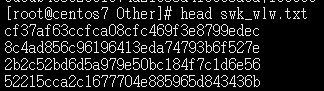
Other/home.txt



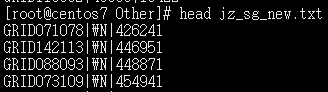
Other/work.txt



Other/swk\_wlw.txt

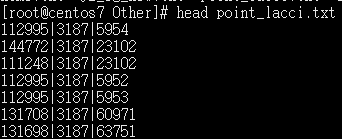


Other/jz\_sg\_new.txt

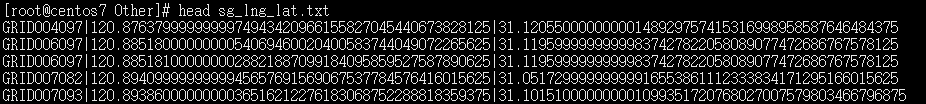


181143769243220717

Other/point\_lacci.txt



Other/sg\_lng\_lat.txt



**九、最新当前数据**

|  |  |  |  |
| --- | --- | --- | --- |
| 数据类型 | 主要字段 | 保存方式 | 大小 |
| APP | 加密电话、app名称、次数、时间 | 按天保存 | 约3G左右 |
| 关键词 | 加密电话、关键词、来源、时间 | 按天保存 | 约3G左右 |
| 语音 | 双方加密号码、通话时长 | 按天保存 | 约几G |
| 短信 | 双方加密号码、上下行 | 按天保存 | 约几G |
| 用户属性 | 略，详见用户属性表 | 按月保存 | 约2G |
| 工作地、居住地 | 其它字段对应删格 | 单份 | 1G以内 |
| 附近地铁站 | 居住地和工作地附近地铁站名、工作地与居住地距离 | 按月保存 | 1G以内 |
| 行为轨迹 | 可对应到相应的删格 | 每天多个文件 | 每天100G左右 |