

代码操作说明

此脚本目前针对本地数据库，但是由于数据量过大，后期可能选择云数据库（开会讨论）

操作注意点：

第六行记得改成本机用户名和密码

```
6 conn = connect(host='localhost', port=3306, user='root', password='root1234', charset='utf8')
```

第45行记得修改文件路径

```
45 path = "/Users/allen/Desktop/py流数据/data" # 文件夹目录
```

表

本次设计了两个表：user和login

user表用于记录用户的行为

user相关字段

```
CREATE TABLE IF NOT EXISTS `user` (  
    `key` int(11) NOT NULL AUTO_INCREMENT, # 由于其他字段均存在重复，因此采用自增字段作主键  
    `sessionID` varchar(40) NOT NULL,  
    `time` datetime NOT NULL, # 格式为yyyy-mm-dd hh-mm-ss  
    `action` varchar(20) NOT NULL, # 根据url将action 分成了getDetail, cart, favor, buy  
    `userID` varchar(15) NOT NULL,  
    `itemID` varchar(15) NOT NULL,  
    `categoryID` varchar(15) NOT NULL,  
    `isSecondKill` int(1) default NULL, # 仅针对购买行为，即buy,其他行为此字段为null, 购买为  
0/1  
    `success` int(1) default NULL, # 同上，仅针对上条字段为1的情况，值为1/0，其他情况下为null  
    PRIMARY KEY (`key`)  
) ENGINE=InnoDB DEFAULT CHARSET=utf8 AUTO_INCREMENT=0
```

login用于记录登陆行为

大体字段意义同上

```
CREATE TABLE IF NOT EXISTS `login` (  
  `key` int(11) NOT NULL AUTO_INCREMENT,  
  `ipAddress` varchar(20) NOT NULL, # 记录ip地址  
  `sessionID` varchar(40) NOT NULL,  
  `time` datetime NOT NULL,  
  `userID` varchar(15) NOT NULL, # 用户名  
  `pwd` varchar(40) NOT NULL, # 密码  
  `auth` varchar(40) NOT NULL, # 授权码  
  `success` int(1) NOT NULL, # 登陆是否成功, 非空, 0/1  
  PRIMARY KEY (`key`)  
) ENGINE=InnoDB DEFAULT CHARSET=utf8 AUTO_INCREMENT=0
```

每读完一个文件会commit一次，并在控制台进行输出

有任何问题请及时联系