代码使用

数据表6个

0 t\_finance\_history\_report 深沪公告表（300万）

1 t\_finance\_pre\_disclosure\_report 预披露公告表（几千）

2 t\_finance\_neeq\_company\_report 股份转让表（）

3 t\_finance\_hke\_report 香港公告表（）

4 t\_finance\_fund\_report 基金公告表（）

5 t\_finance\_regulator\_report 监管机制公告表（）

源码存在目录

/home/user1/python/common/code/juchao

如何爬取

在根目录下输入命令

python common\_main.py 2017 1

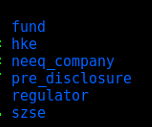
两个参数，年份和序号

年份是采集哪一年的，序号对应上面的6种（0,1,2,3,4,5）

上面命令含义为采集2017年的预披露公告

文件存储目录

/home/user1/python/common/file



每个目录下按年份分子目录



数据库操作：

登录mysql

查看有哪些表

show tables;

查询总记录

select count(\*) from db\_finance.t\_finance\_history\_report;

查询详细信息（limit 后面控制显示的数量）

select \* from db\_finance.t\_finance\_history\_report limit 10;

Linux操作

文件查看数量

进入文件的根目录：

ls -l | grep "^-" | wc -l

统计目录大小：

在目录上层

du -h --max-depth=1