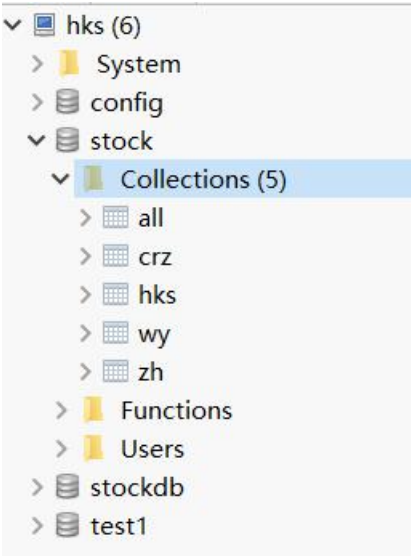


数据库我们采用的是非关系型数据库 mongodb，因为我在云计算课程中用的也是 mongodb，所以感觉很方便也很熟悉。

使用 robo 3T 可视化工具，可以直观的展现数据内容。



数据库名为“stock”，旗下有五张表，表的名字分别为我们组四个人名字的简称和“all”。这样比较能够直观地看到每个人所爬取的数据是哪些。

Crz：（中国证券网）（曹润泽）（stock.py）

股票代码	000001
公司名称	平安银行
上市时间	1991-04-03
公司全称	平安银行股份有限公司
注册资本(万元)	1940591.82
股东列表	[10 elements]
[0]	{ 3 fields }
股东名称	中国平安保险(集团)股份有限公司-集团本级-自有资金
持股数	961854.02
占总股本比例	49.56
[1]	{ 3 fields }
[2]	{ 3 fields }
[3]	{ 3 fields }
[4]	{ 3 fields }
[5]	{ 3 fields }
[6]	{ 3 fields }
[7]	{ 3 fields }
[8]	{ 3 fields }
[9]	{ 3 fields }
总股本(万股)	1940591.82
总流通股本(万股)	1940591.82
每股净资产(元/股)	14.55
主营收入(万元)	3792600.00
净利润(万元)	854800.00
高管列表	[10 elements]
数据源	中国证券网

由上图可以直观地看出，该表主要爬取的是一个公司的基本信息，比如总股本、注册资本、注册时间等等。比较关键的信息是股东列表，该字段能够将不同的公司联系起来，成为构建

知识图谱的关键信息。

Wy: (金融界)(王阳)(jrj.py)

公司全称	平安银行股份有限公司
股票代码	000001
板块	主板
发行价格 (元)	40.0
主营业务	(一)吸收公众存款;(二)发放短期、中期和长期贷款;(三)办理国内外结算;(四)办理...
注册资本 (元)	19,405,918,208
实际控制人	谢永林
股东列表	[10 elements]
行业分类	{ 3 fields }
证监会行业分类	金融业
全球行业分类	银行
申万行业分类	银行
评级	{ 4 fields }
综合评级	5
市盈率(TTM)	8.48
市净率(MRQ)	0.71
市现率(TTM)	--
状态	上市
数据源	金融界

该表其实也主要是静态字段，有些内容是对第一张表的补充，比如主营业务、板块字段等。也有一些是相同的信息如注册资本、股东列表等等，则是为了以防万一，能够相互补充。

Zh: (东方财富网)(赵航)(zh.py)

股票代码	600000
每股收益	{ 9 fields }
2020-03-31	0.5600
2019-12-31	1.9500
2019-09-30	1.6200
2019-06-30	1.0700
2019-03-31	0.5300
2018-12-31	1.8500
2018-09-30	1.4400
2018-06-30	0.9500
2018-03-31	0.4600
每股经营现金流	{ 9 fields }
净资产收益率	{ 9 fields }
总资产	{ 6 fields }
总负债	{ 6 fields }
总利润	{ 6 fields }
净利润	{ 6 fields }
经营现金流	{ 6 fields }
股本变动	{ 2 fields }

可见，该张表爬取的全是实时数据，用来反映一些数据的变化，在项目文档里也说过，该数据是通过简介爬取得到的，所以爬取的数据比较全。为了减少数据冗余以及展示困难，只选取了 9 个时间点的数据。

Hks: (Tushare)(黄恺晟)(tushare-spider.py)

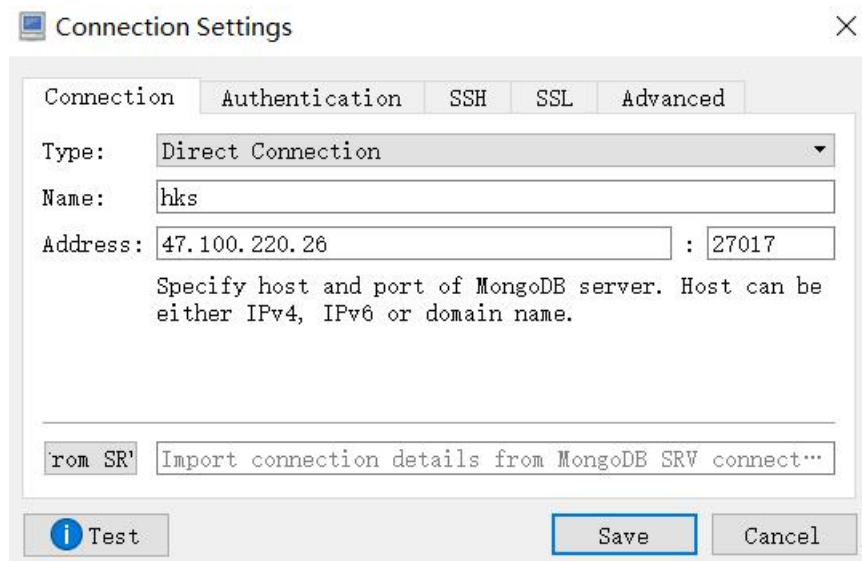
股票代码	600268
交易日期	2019-06-28
开盘价	6.08
最高价	6.17
最低价	5.93
收盘价	5.94
昨日收盘价	6.09
涨跌额	-0.15
成交量(手)	115455.0
成交额(千元)	69875.454

以上三张表大部分都是静态数据，无法反应实际的变化，而 tushare 则提供了一个良好的接口来提供实时变化的数据。

本来这些数据是组员通过 python 爬取到 csv 文件的，共计一年的数据。为了方便存储，又将这些数据存到 mongodb 中，可是一年的数据量太过庞大，考虑到也没有必要将那么多的数据展示出来，所以只存储了 13W 条数据，大概是从 2019-6-28 到 2019-9-11 这段时间的数据。该数据处理为 git 上的 yearData.py。

数据库访问：

我们的 mongodb 是建立在服务器上的，通过 robo3T 连接



如图所示，地址为 47.100.220.26，端口为 27017，账号密码都是 mongoadmin。