1. **python环境**

请在电脑上安装anaconda和pycharm

* 1. anaconda，集成了绝大多数你需要的python packages，我们使用python2.7版本  
     <https://www.continuum.io/downloads/>
  2. pycharm，python ide，下载community版本并安装

http://www.jetbrains.com/pycharm/download/#section=windows

* 1. 将pycharm的python解释器切换成anaconda下python解释器所在的位置  
     一般是installed\_file/anaconda/bin/python

1. **代码规范**

编写python代码时，请参照以下规范

* 1. Python官方pep8 <https://www.python.org/dev/peps/pep-0008/>

如果打不开，请用https://my.oschina.net/u/1433482/blog/464444?p=1

2.2 google python code style

http://zh-google-styleguide.readthedocs.io/en/latest/google-python-styleguide/contents/

**3 pandas库相关**

<http://pandas.pydata.org/pandas-docs/stable/index.html>

**3.1 10 Minutes to pandas**

**3.2 Tutorials**

**3.3 Intro to Data Structures**

**3.4 [Indexing and Selecting Data](http://pandas.pydata.org/pandas-docs/stable/indexing.html)**

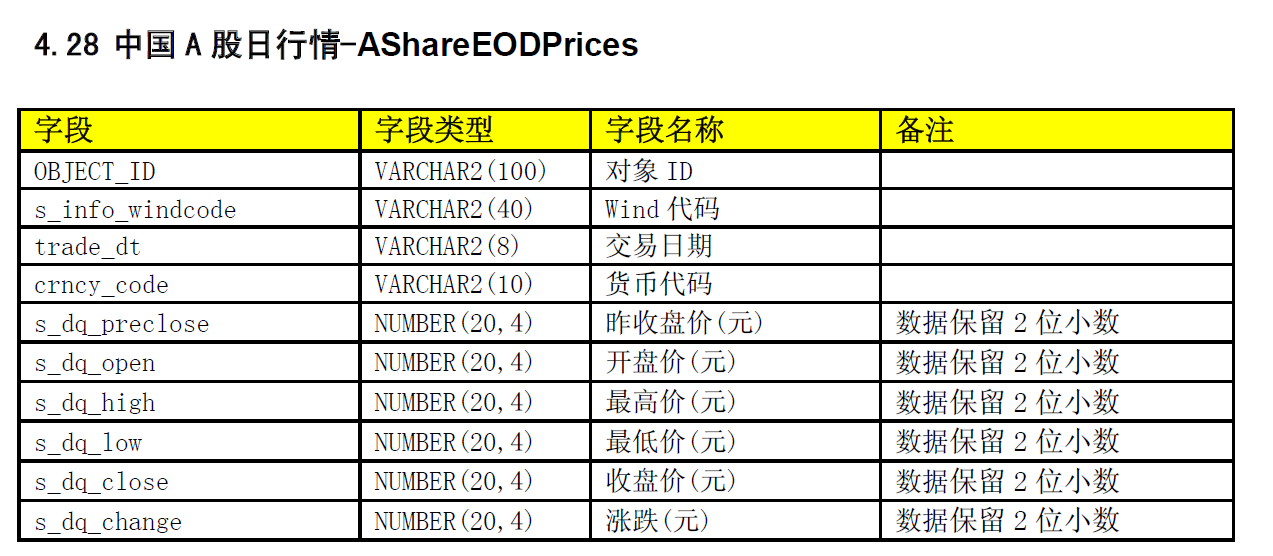
**3.5 IO Tools(read\_csv, to\_csv)**

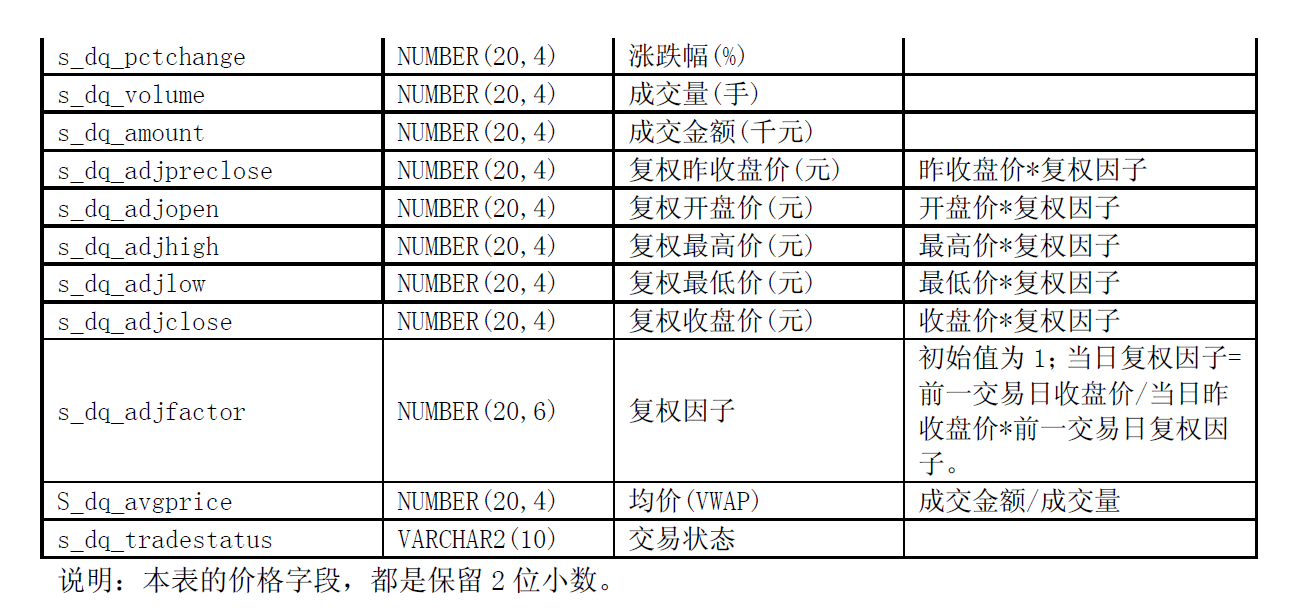
阅读以上章节，完成：

E1 你可以看到一些csv文件，每个csv文件是股票的日频数据，以s\_info\_windcode命名，

**注意：a, 以下所说的股票代码是指股票在交易所的代码，即s\_info\_windcode的前6位。SZ表示深圳，SH表示上海。**

**b, 以下用symbol代表股票代码，symbols表示股票代码列表。  
 c，以下用date表示trade\_dt，turnover代表s\_dq\_amout, 其余对应去掉s\_dq。**



  
设计StockData类，  
E1.1 StockData.\_\_init\_\_(path)

初始化函数，path为csv数据文件所在路径

E1.2 StockData.read(symbols)  
 symbols是股票代码列表，包括1~n个股票代码，函数从对应symbol的文件中读取相  
 数据并选取合适结构保存在类中  
E1.3 StockData.get\_data\_by\_symbol(symbol, start\_date, end\_date)  
 获取某个symbol从start\_date 到end\_date之间的所有日频数据  
E1.4 StockData.get\_data\_by\_date(adate, symbols):  
 获取某一天中，对应symbols的所有日频数据

E1.5 StockData.get\_data\_by\_filed(field, symbols):  
 获取symbols在某个field上所有的所有交易日的数据

其中E1.3，E1.4，E1.5返回pandas.DataFrame结构，

E1.3 dataframe格式如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| date | open | high | low | close |
| 20170320 |  |  |  |  |
| 20170321 |  |  |  |  |
| 20170322 |  |  |  |  |
| 20170323 |  |  |  |  |
| 2017032 |  |  |  |  |

E1.4 dataframe格式如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| symbols | open | high | low | close |
| 600000 |  |  |  |  |
| 600036 |  |  |  |  |
| 000001 |  |  |  |  |
| 000002 |  |  |  |  |
| 000004 |  |  |  |  |

E1.5 dataframe格式如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| date | 600000 | 600036 | 000001 | 000004 |
| 20170320 |  |  |  |  |
| 20170321 |  |  |  |  |
| 20170322 |  |  |  |  |
| 20170323 |  |  |  |  |
| 2017032 |  |  |  |  |

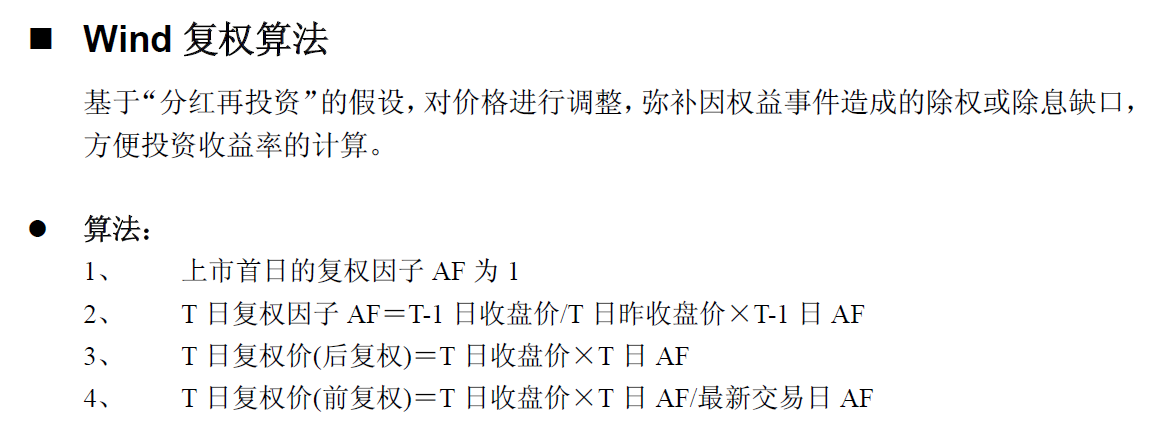
注：停牌或者缺失数据可以不处理，日期范围最大选取近十年的。

**3.6 Time Series/Date functionally**

**3.7 Reshaping and Pivot Tables  
3.8 Visualization**

阅读以上章节，完成  
E2，csv数据均为日频数据，你需要做一些采样和转换  
E2.1 StockData.format\_date(symbol)  
 symbol的数据中date在读入时，可能是int或者string类型，请将这一列中的每个date  
 转化为datetime类型（或者pandas.Timestamp类型）

E2.2 StockData.plot(symbol, field)   
 做出symbol关于field的走势图，field取open，high，low，close, vwap中的某一个。  
 走势图为折线图，取volume或者turnover中的某一个时，走势图为柱状图。  
 简单画一下就行，类似于一般股票网站上价量的图。

E2.3 StockData.adjust\_data(symbol)  
 对代表价格（open, high, low, close）进行（从后向）前复权。计算方法如下：  
 

对于每一天的价格均要复权。

E2.4 StockData.resample(symbol, freq)  
 将symbol的日频数据进行sample（采样），freq表示天数。  
 采样只需完成open，high，low，close，volume，turnover，vwap这些field即可  
 例如：StockData data

df =data.resample(“600000”, 5)  
 表示对600000的日频数据进行5天的采样，其中

open，close表示这5天第一日开盘价和最后一日的收盘价

high，low表示这5天的最高价和最低价  
 volume，turnover 表示这5天的总成交量，总成交额

vwap表示这5天的成交量加权平均价格

E2.1和E2.3的函数直接改变了StockData存储的数据。对于每个symbol而言，就是改变了存储这个symbol日频数据的dataframe中相应field的值。

E2.4返回dataframe，格式如E1.3

**3.9 Computational Tools  
3.10 excel文件中的说明**

阅读以上章节，完成：

E3，你需要根据3.10的excel中的说明，做一些计算，所用到的数据，是在E2.3中前复权过的数据，缺失或者停牌数据不处理。

E3.1 StockData.moving\_average(symbol，field，window)  
 计算symbol对应field的移动平均， window表示滑动窗口。

E3.2 StockData.ema(symbol, params)  
 计算symbol日频数据的ema指标  
E3.3 StockData.atr(symbol, params)

计算symbol日频数据的atr指标  
E3.4 StockData.rsi(symbol, params)

计算symbol日频数据的rsi指标  
E3.5 StockData.macd(symbol, params)

计算symbol日频数据的macd指标

注意：

a E3.1 ~ E3.5均返回一个pandas.Series对象，index是交易日（date）

b E3.2 ~ E3.5 中的params表示计算用的参数，有几个参数，分别表示什么意思，自己定义。这里只是一个参数说明。

c E3.1 ~ E3.5其中涉及到滑动窗口（或叫频率，周期）的参数，请分别测试当窗口为5，20，60的时候的计算结果，可以用E2.2中的plot函数画图看下结果。

d 指标计算方法在excel中如果没有，请网上自行查找。

e 很多指标由于涉及滑动窗口，需要做一些rolling的计算，请参考pandas相关文档。