# DataReceiver数据接收器

## 功能

1. GetInstruments.exe，登录交易账号，下载合约列表。此工具需要添加到计划任务当中。每天早上9点前和晚上9点前都要启动一次。
2. DataReceiver.exe，登录行情账号，根据合约列表进行订阅并生成pd0格式的文件。早上8点45左右和晚上8点45左右最好重新连接并订阅一次，因为连接得过快，登录过早，有可能服务器没有记录下来，导致开盘时没有行情推送。
3. ArchiveData.exe，将数据文件按交易日，按合约名进行归档。也是需要添加到计划任务中。可以归档成h5格式。

## 工作流程

项目运行需要XAPI的各种库，可以将【QuantBox统一接口完整版】的XAPI目录复制出来，放在DataReceiver同目录下。然后对每个exe进行配置，可能工作流程比较复杂，一般人很难理解。所以建议直接安装【QuantBox统一接口完整版】，XAPI的库就不使用复制到DataReceiver下的了，而是使用【QuantBox统一接口完整版】给OpenQuant2014安装到C盘的库即可。

即，请与【QuantBox统一接口完整版】配合使用。

### GetInstruments.exe

1. GetInstruments.exe.config中设置了配置文件的路径ConfigPath，ConnectionConfigFileName是交易账号信息。
2. 添加到计划任务后，每次启动自动登录，查询合约列表，查到了就写入InstrumentInfoListFileName文件中，然后自动退出。

### DataReceiver.exe

1. DataReceiver.exe.config中设置了配置文件路径ConfigPath和数据存储路径DataPath。
2. 软件先根据行情地址与账号信息ConnectionConfigListFileName，登录N\*SessionLimit个行情连接。注意：这里的json是存的是对象List,而GetInstruments.exe.config用的ConnectionConfigFileName用的是一个对象。
3. 然后，软件加载InstrumentInfoListFileName、IncludeFilterListFileName、ExcludeFilterListFileName，过滤出要订阅的合约，进行订阅。
4. 如果在软件正在运行中对InstrumentInfoListFileName、IncludeFilterListFileName、ExcludeFilterListFileName三个文件中所对应的内容进行修改，将会自动进行订阅。这样每天GetInstrument.exe运行时出现新上市的合约都能得到订阅。

### ArchiveData.exe

1. ArchiveData.exe.config中设置了数据文件的一个输入DataPath与两个输出OutputPath\_TradingDay(按交易日)、OutputPath\_Instrument(按合约名)
2. SevenZipExePath设置成7-zip软件的安装目录，按合约名的压缩后分目录由它实现。按交易目生成目录后的压缩也依赖于它。
3. Clear\_DataPath是删除原始数据，只在OutputPath\_TradingDay完成后才删除，慎用
4. Clear\_OutputPath\_TradingDay，删除按交易生成的目录，即将中间过程的文件删除
5. DefaultExchange在接收CTP行情时可能没有交易所信息，所以提供一个默认目录，无法提取交易所的文件名就存在此目录下。
6. 请在接收完结算价后再执行此功能，实盘一般是3点半左右已经收到结算价了，但仿真行情就很不一定的，出于稳妥考虑，可晚一点做此工作，比如5点后。
7. 按交易日打包时会同时打包一个h5格式的文件，h5文件可以通过Python或MATLAB进行访问，demo文件夹下提供了读取的示例和工具。

## 其它说明

1. GetInstruments.exe.config和DataReceiver.exe.config中的InstrumentInfoListFileName设为同一文件，这样正好可以每天更新
2. InstrumentInfoListFileName中记录了合约的基本信息，TickSize/Factor都是文件记录时必需的参数。
3. 每个连接可以订阅的合约数到一定量时就会不稳定，所以不要给每个连接订阅过多合约，比如CTP在互联网下一个连接订阅20个。而LTS是服务器做了限制每个连接只能订阅200个，每个服务器只能同时支持2个会话，一共有3台实盘行情服务器，即最多只能订阅200\*2\*3=1200个。
4. 所以行情接收时需要选择自己比较关心的合约，可以用IncludeFilterListFileName和ExcludeFilterListFileName,其中的IncludeFilterListFileName中的Time\_ssf\_Diff是用来记录数据时减少时间所占空间的功能，股指期货500ms所以设成5，间隔1s就设成10，股票由于两笔之间可能时间完全一样，所以设成0，如果不知道设成多少就使用默认的0
5. IncludeFilterListFileName和ExcludeFilterListFileName中都是用的正则表达式。是由交易接口中返回的Symbol来决定。目前期货是{0}，而证券是{0}.{1}
6. ArchiveData可以复制多份，在15点30后做一次按交易日打包，但不删除原数据。这样用户可直接下载一个文件即可。在17点再做一次，这样结算价也已经收全了。也可以等第二个交易日，时将上一个交易日的数据做次完整的归档。
7. 为何不一开始就保存成h5格式，而是pd0格式，等归档时才转h5？pd0格式由于是专门设计的，压缩率更高，同时是增量格式，新行情来时，只要接着后面写入即可，效率更高。

## 计划任务

有用户使用操作系统中的计划任务，在定时执行GetInstruments和ArchiveData，我们也提供了计划任务的功能。DataReceiver.exe.config中的ScheduleTasksListFileName指定了计划任务的配置文件，参考进行设置即可。

支持定时执行函数的功能，现在支持的是Connect/Disconnect/ReSubscribe/SaveData。

## 常用正则表达式

这里要区分一下，有一个订阅用Symbol和一个行情接收Symbol。行情接收一定是{0}.{1}格式。在这下面都是用的订阅用Symbol。

1. 所有合约："SymbolRegex": "."
2. 所有的上交所的期权："SymbolRegex": "\\d{8}\\.SSE"
3. 所有的组合合约SP/SPD/SPC："SymbolRegex": "^SP"
4. 所有的组合合约IPS："SymbolRegex": "^IPS"
5. 期转现efp："SymbolRegex": "efp"
6. 所有的SR期货："SymbolRegex": "SR\\d{3}$"
7. 所有的SR期权："SymbolRegex": "SR\\d{3}[CP]"
8. 上交所6开头的股票："SymbolRegex": "^6\\d{5}\\.SSE"
9. 深交所0开头的股票："SymbolRegex": "^0\\d{5}\\.SZE"