**数据更新—20210810**

**更新了大连，上期，能源，郑州，中金的数据，改动如下：**

1，将不规则命名的文件全部删除：例如DYa009，xx0000

2，更新了某些文件里的数据：例如2008年1月的a0801合约，基础数据(open, high等)数据改动

3，添加了漏掉的合约：例如黄大豆1号(a) 在2008年1月缺少的a0809,a0811等合约补齐

4，合约名形式的更改：例如DYa0801🡪a0801

5，最新数据日期延长到2021年6月

**最新数据在25服务器中的位置：**

D:\CTA\highfrequencydata\内地商品期货1分钟-20210730

D:\CTA\highfrequencydata\中金所期货1分钟-20210730

(highfrequencydata中有两个\_db结尾的文件夹可以不用管)

**现有数据库表格：**

HFBase：

BusinessDay—从1990年到今后两年的所有交易日，剔除了节假日

HFData2：

HFData2\_DL—最新的大连数据

HFData2\_SQ—最新的上期数据

HFData2\_IN—最新的能源数据

HFData2\_ZZ—最新的郑州数据

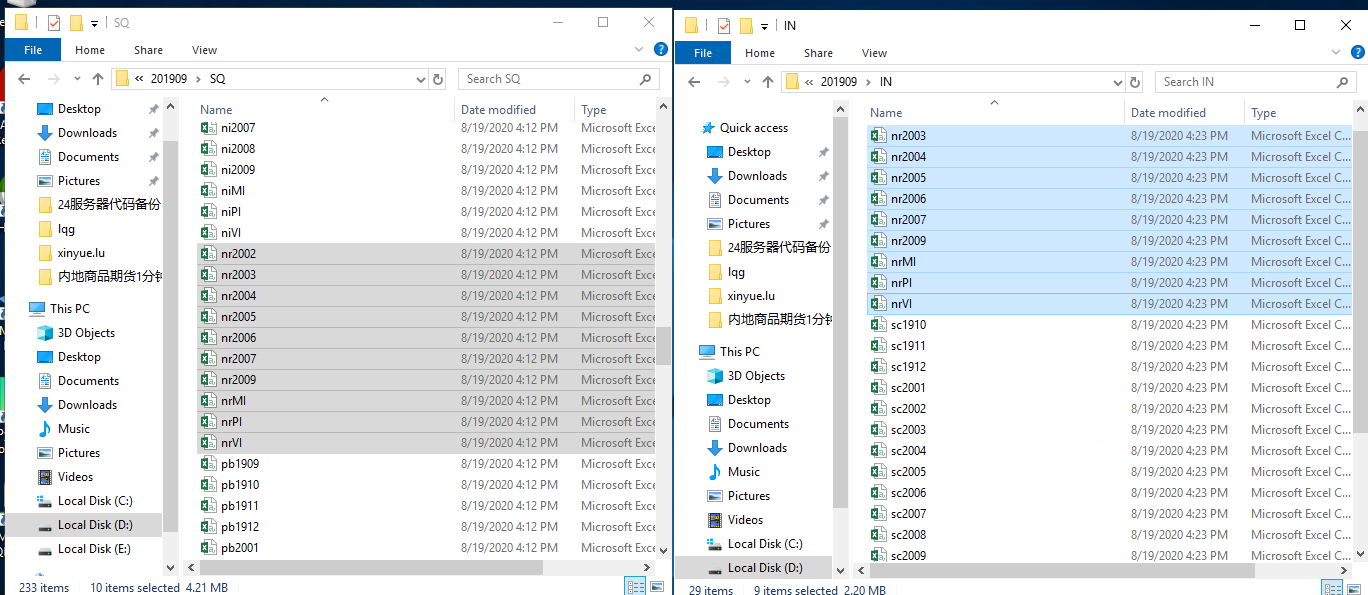
HFData2\_ZJ—最新的中金数据

neidi\_1min\_updated20210810—内地期货1min在2021年8月10日更新完的数据

**需要注意的问题⚠️：**

从2019年9月开始，能源期货交易所和上海期货交易所存在共同拥有的期货

目前只有20号胶(nr)和国际铜(bc)存在这个问题(在sql里面关键字搜索数据的时候注意一下就好，数据方面没有什么问题)



**更新数据涉及的代码位于25服务器中：**

C:\Users\Public\lmj\py存储期货数据

**代码大概介绍：**

* 订单打印的格式—log.py
* 筛选更新过的文件(这里的依据是文件修改日期也就是date modified)并且将csv文件改成和sql server数据表里一样的格式(在每个筛选出来的csv中增加symbol, futures\_market和filename列)—db\_format.py
* 检索每个更新过的csv的旧数据，删除并且导入新数据，double check新数据是否导入完毕—db\_sql.py
* 整体流程的汇总与运行：先转换csv文件格式再导入—main.py

⚠️：db\_format.py和db\_sql.py有自己单独的main，可以单独运行；main.py是所有流程的运行