数据模块技术需求

V0.2

# 版本管理

|  |  |  |  |
| --- | --- | --- | --- |
| 版本号 | 修改内容 | 撰写人 | 提交时间 |
| V0.1 |  | 吕平;张恒 | 20191009 |
| V0.2 | 主力连续TICK数据的复权  主力连续K线的持仓量和成交量 | 吕平;张恒 | 20191010 |

# 数据模块需求

数据模块处理目前实盘交易中使用的数据，数据的落地需求包括历史数据的入库、数据的每日更新和维度、数据的清洗等工作。

数据按照生成方式可分为：

* 指数数据、主力连续数据、WLC指数复权数据、跨期价差数据

数据按照周期可分为：

* TICK、1M、5M、15M、Day数据

数据按照品种可分为：

* 商品期货数据、股指期货数据、现货数据(目前为股票指数数据)、后续可增添股票基本面数据和商品基本面数据

现阶段主要的工作目标为实现一代平台中的策略历史回测和实盘交易数据，包括指数数据、主力连续数据、WLC指数复权数据；数据涵盖商品期货数据、股指期货数据、股指指数数据。具体需求见后续章节。

## 指数数据

周期需求：1M、5M、15M、Day；1M为存储的基本数据。其他周期可由此数据进行衍生。具体指数数据生成方法如下：

1. 合约选取
   1. 按照昨日持仓量排序，选择昨日持仓量最大的三个合约
   2. 若选中合约中包含临近到期日的合约（即合约到期日与交易日属于同一个自然月），则跳过该合约，按昨日持仓量顺延选中下一个合约
   3. 在同一个交易日内，所选择合约不改变
   4. 股指期货不需回避交割月，商品需要回避交割月；
2. 指数生成TICK
   1. 开盘价：以三个合约的开盘价按昨日持仓量加权平均后得出；如某个合约没有开盘价，则以昨日收盘价代替 注：按开盘时的持仓量计算，如无则用昨持仓量
   2. 收盘价：以三个合约的收盘价按昨日持仓量加权平均后得出；如某个合约没有收盘价，则以昨日收盘价代替 注： 按收盘时的持仓量计算
   3. 盘中价：收到任一合约的行情tick后，以三个合约的最新价格，按照**当前**持仓量加权平均后得出。
   4. 成交量：收到任一合约的行情tick后，**以该tick的成交量作为指数tick的成交量**
   5. **持仓量**：**三个合约当前持仓量的累加和**
3. K线生成
   1. 以左闭右开区间作为分钟K线的计算区间，以1分钟K线为例，09:00的计算区间是09:00:00.000~09:00:59.999，09:01的计算区间是09:01:00~09:01:59.999，其他类同
   2. 价格：按照指数生成K线的开、高、低、收各价格
   3. 成交量：**区间内所有tick成交量的累加和**
   4. **持仓量：区间内最后一个tick的当前持仓量**

## 主力连续数据

### TICK数据

周期需求: TICK数据，该部分工作主要由张恒负责核对和使用。

1. 合约选取
   1. 持仓量最大的三个合约的tick数据。全字段
   2. 有LEVER2行情更佳
   3. 过滤非交易时间段数据；如开盘前和收盘后有冗余数据

### 主力连续TICK数据

1. 合约选取
   1. 持仓量最大的合约的tick数据。全字段。 注：指交易日收盘时的持仓量
   2. 有LEVER2行情更佳
   3. 过滤非交易时间段数据；如开盘前和收盘后有冗余数据
   4. 主力合约切换后，不进行切回近月合约。举例：1901合约切换到1905合约，即使下一交易日1901合约持仓量大于1905合约持仓量，也不进行切回。
   5. 在同一个交易日内，所选择合约不改变
2. 合约拼接
   1. 持仓量最大的合约的tick数据直接按照时间拼接
   2. 加入数据列字段：当前主力合约代码和前一个主力合约代码
3. 特别说明：
   1. 主力切换时tick数据是否也应该记录复权因子数据：
      1. 主力连续TICK为按照合约切换日期直接进行主力合约拼接完成的数据，不进行任何处理。复权因子数据仅在生成主力连续K线数据时进行存储。
   2. 主力切换时，tick的成交量、持仓量如何定义？
      1. 主力切换时，存储的为主力合约的全字段数据
      2. 成交量：该tick的成交量
      3. 持仓量：当前主力合约的TICK对应的持仓量

### 主力连续K线数据

周期需求： 1M、5M、15M、Day；基本数据存储为1M数据，1M数据可由主力连续TICK数据生成，1M以上周期的数据由1M数据生成。

1. 合约选取：
2. 持仓量最大的合约
3. 主力合约切换后，不进行切回近月合约。举例：1901合约切换到1905合约，即使下一交易日1901合约持仓量大于1905合约持仓量，也不进行切回。
4. 在同一个交易日内，所选择合约不改变
5. 主力连续数据K线生成
6. 由TICK数据生成1M的K线数据
7. 成交量：即为原K线生成机制中的成交量，即**区间内所有tick成交量的累加和**
8. 持仓量：即为原K线生成机制的最新持仓量，即**区间内最后一个tick的当前持仓量**
9. 复权因子数据
10. 合约发生切换时，生成复权因子数据；如7月14日收盘后，合约由IF1907切换到IF1908合约； K=CLOSE(IF1908)/ CLOSE (IF1907)；Date = 20190714
11. 存储字段：日期、K，当前主力合约代码、前一个主力合约代码
12. 主力连续复权数据(该部分功能由策略模块的数据部分进行处理)
13. 复权方式：向前复权
14. 策略模块使用主力连续复权数据时，调用主力连续数据K线和复权因子数据进行合成主力连续复权数据

## WLC指数复权数据

### WLC\_Trade商品期货数据

周期需求：5M和15M；目前实盘使用的周期为15M，建议以5M为周期生成，便于后续策略开发。

### WLC\_IF股指期货数据

周期需求：5M和15M；目前实盘使用的周期为15M，建议以5M为周期生成，便于后续策略开发。