# 任务描述

入库合约信息表

# 流程

## 提取合约信息

按时间顺序遍历\\168.36.1.170\share\archives\MD\_\*.rd

文件命名规范: MD\_yyyymmdd\_hhmmss.rd 中文编码为gb18030

其中以"OPLST:01"打头的行,即为一条合约信息

格式: OPLST:01合约编号 hash 后面的内容可以直接通过python eval提取为字典 dict[str, str]

字段可参考c的结构注释

|  |
| --- |
| ///合约  struct CThostFtdcInstrumentField  {  ///合约代码  TThostFtdcInstrumentIDType InstrumentID;  ///交易所代码  TThostFtdcExchangeIDType ExchangeID;  ///合约名称  TThostFtdcInstrumentNameType InstrumentName;  ///合约在交易所的代码  TThostFtdcExchangeInstIDType ExchangeInstID;  ///产品代码  TThostFtdcInstrumentIDType ProductID;  ///产品类型  TThostFtdcProductClassType ProductClass;  ///交割年份  TThostFtdcYearType DeliveryYear;  ///交割月  TThostFtdcMonthType DeliveryMonth;  ///市价单最大下单量  TThostFtdcVolumeType MaxMarketOrderVolume;  ///市价单最小下单量  TThostFtdcVolumeType MinMarketOrderVolume;  ///限价单最大下单量  TThostFtdcVolumeType MaxLimitOrderVolume;  ///限价单最小下单量  TThostFtdcVolumeType MinLimitOrderVolume;  ///合约数量乘数  TThostFtdcVolumeMultipleType VolumeMultiple;  ///最小变动价位  TThostFtdcPriceType PriceTick;  ///创建日  TThostFtdcDateType CreateDate;  ///上市日  TThostFtdcDateType OpenDate;  ///到期日  TThostFtdcDateType ExpireDate;  ///开始交割日  TThostFtdcDateType StartDelivDate;  ///结束交割日  TThostFtdcDateType EndDelivDate;  ///合约生命周期状态  TThostFtdcInstLifePhaseType InstLifePhase;  ///当前是否交易  TThostFtdcBoolType IsTrading;  ///持仓类型  TThostFtdcPositionTypeType PositionType;  ///持仓日期类型  TThostFtdcPositionDateTypeType PositionDateType;  ///多头保证金率  TThostFtdcRatioType LongMarginRatio;  ///空头保证金率  TThostFtdcRatioType ShortMarginRatio;  ///是否使用大额单边保证金算法  TThostFtdcMaxMarginSideAlgorithmType MaxMarginSideAlgorithm;  ///基础商品代码  TThostFtdcInstrumentIDType UnderlyingInstrID;  ///执行价  TThostFtdcPriceType StrikePrice;  ///期权类型  TThostFtdcOptionsTypeType OptionsType;  ///合约基础商品乘数  TThostFtdcUnderlyingMultipleType UnderlyingMultiple;  ///组合类型  TThostFtdcCombinationTypeType CombinationType;  ///最小买下单单位  TThostFtdcVolumeType MinBuyVolume;  ///最小卖下单单位  TThostFtdcVolumeType MinSellVolume;  ///合约标识码  TThostFtdcInstrumentCodeType InstrumentCode;  }; |

e.g.

|  |
| --- |
| {'CombinationType': '0',  'CreateDate': '20190328',  'DeliveryMonth': '9',  'DeliveryYear': '2019',  'EndDelivDate': '20190925',  'ExchangeID': 'SSE',  'ExchangeInstID': '10001745',  'ExpireDate': '20190925',  'InstLifePhase': '1',  'InstrumentCode': '510050C1909M02950',  'InstrumentID': '10001745',  'InstrumentName': '50ETF购9月2950',  'IsTrading': '0',  'LongMarginRatio':'1.7976931348623157e+308',  'MaxLimitOrderVolume': '30',  'MaxMarginSideAlgorithm': '0',  'MaxMarketOrderVolume': '10',  'MinBuyVolume': '1',  'MinLimitOrderVolume': '1',  'MinMarketOrderVolume': '1',  'MinSellVolume': '1',  'OpenDate': '20190226',  'OptionsType': '1',  'PositionDateType': '2',  'PositionType': '2',  'PriceTick': '0.0001',  'ProductClass': '7',  'ProductID': 'ETF\_O',  'ShortMarginRatio': '1.7976931348623157e+308',  'StartDelivDate': '20190925',  'StrikePrice': '2.95',  'UnderlyingInstrID': '510050',  'UnderlyingMultiple': '1.0',  'VolumeMultiple': '10000'} |

## 入库

create table archive\_oplst\_raw ......

''ExchangeInstID'' + date

create table archive\_oplst ......

注意,入库中文编码为utf-8

设置unique主键为 'InstrumentCode'

创建索引: 'InstrumentID' + 'OpenDate'

创建索引: 'InstrumentID' + 'ExpireDate'

字段类型:

可以先jsonb格式写入单一字段

如有可能,可结构化上述所有字段,

# 索引性能比较

## 10000次查询

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 编号 | InstrumentID+OpenDate | InstrumentID+ExpireDate | 无索引 | Json |
| 虚拟列 | 4.989991903305054 | 4.934611797332764 | 5.134729385375977 | 4.972288370132446 |
| 常规字段 | 5.323497772216797 | 5.540904998779297 | 5.587999582290649 |  |

## 单次查询

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 编号 | InstrumentID+OpenDate | InstrumentID+ExpireDate | 无索引 | Json |
| 虚拟列 |  |  |  |  |
| 常规字段 |  |  |  |  |

# 入库代码

## archive\_oplst\_raw

|  |
| --- |
| 参见work/src/intodb/getdata.py |
| *#-\*- coding:utf-8 -\*-* **import** os **import** re **import** psycopg2 **import** chardet  path = **r"\\168.36.1.170\share\archives"** conn = psycopg2.connect(database=**"postgres"**,user=**"postgres"**,password=**"..."**, host=**"192.168.40.129"**, port=**"5432"**) cur = conn.cursor()  dirs = os.listdir(path) file\_type = **".rd"** file\_list = list() **for** i **in** dirs:  **if** re.match(**r'MD\_\*'**, os.path.splitext(i)[0]) **and** os.path.splitext(i)[1] == **'.rd'**:  file\_list.append(os.path.splitext(i)[0])   **for** fileName **in** file\_list:  print(fileName)  file = open(path+**r"/"**+ fileName+file\_type,encoding=**'gb18030'**)  date = fileName[3:11]  time = fileName[-6:]  **if not** time.startswith(**"15"**):  **continue** *# print(date)  # print(time)* **for** line **in** file.readlines():  **if** re.match(**r"OPLST:01\*"**,line):  inf = eval(line[22:])  exchangeInstID = inf[**'ExchangeInstID'**]  inf\_str = str(inf).replace(**"\'"**,**"\""**)  *# inf\_str = inf\_str.encode("utf-8")  # print(chardet.detect(inf\_str.encode()))  # print(exchangeInstID,date)* cur.execute(**"insert into archive\_oplst\_raw values(\'"** + exchangeInstID +**"\',\'"** + date + **"\',\'"** + inf\_str + **"\');"**) cur.execute(**"select \* from archive\_oplst\_raw where date = \'20190506\';"**) rows = cur.fetchall() print(rows) conn.commit() conn.close() |

## archive\_oplst (共504条)

|  |
| --- |
| 参见work/src/intodb/createdb.py |
| *#-\*- coding:utf-8 -\*-* **import** os **import** re **import** psycopg2 **import** chardet  path = **r"\\168.36.1.170\share\archives"** conn = psycopg2.connect(database=**"postgres"**,user=**"postgres"**,password=**"..."**, host=**"192.168.40.129"**, port=**"5432"**) cur = conn.cursor()  dirs = os.listdir(path) file\_type = **".rd"** file\_list = list() **for** i **in** dirs:  **if** re.match(**r'MD\_\*'**, os.path.splitext(i)[0]) **and** os.path.splitext(i)[1] == **'.rd'**:  file\_list.append(os.path.splitext(i)[0])  s = set() **for** fileName **in** file\_list:  print(fileName)  file = open(path+**r"/"**+ fileName+file\_type,encoding=**'gb18030'**)  date = fileName[3:11]  time = fileName[-6:]  **if not** time.startswith(**"15"**):  **continue** *# print(date)  # print(time)* **for** line **in** file.readlines():  **if** re.match(**r"OPLST:01\*"**,line) :  inf = eval(line[22:])   exchangeInstID = inf[**'ExchangeInstID'**]  inf\_str = str(inf).replace(**"\'"**,**"\""**)   *# inf\_str = inf\_str.encode("utf-8")  # print(chardet.detect(inf\_str.encode()))  # print(exchangeInstID,date)* **if** inf[**'InstrumentCode'**] **not in** s:  c = inf[**'InstrumentCode'**]  od = inf[**'OpenDate'**]  ed = inf[**'ExpireDate'**]  id = inf[**"InstrumentID"**]  cur.execute(**"insert into archive\_oplst values(\'"**+ c +**"\',\'"**+od+**"\',\'"**+ed+**"\',\'"**+id+**"\',\'"**+inf\_str+**"\');"**)  s.add(inf[**'InstrumentCode'**])  print(c,od,ed,id) cur.execute(**"select \* from archive\_oplst where InstrumentCode = \'510050P1909M02450\';"**) rows = cur.fetchall() print(rows) conn.commit() conn.close() |

## 比较InstrumentCode

|  |
| --- |
| 参见work/src/intodb/createdb.py |
| **import** os **import** re **import** psycopg2  path = **r"\\168.36.1.170\share\archives"** dirs = os.listdir(path) file\_type = **".rd"** file\_list = list() **for** i **in** dirs:  **if** re.match(**r'MD\_\*'**, os.path.splitext(i)[0]) **and** os.path.splitext(i)[1] == **'.rd'**:  file\_list.append(os.path.splitext(i)[0])  s = set() **for** fileName **in** file\_list:  print(fileName)  file = open(path+**r"/"**+ fileName+file\_type,encoding=**'gb18030'**)  date = fileName[3:11]  time = fileName[-6:]  **if not** time.startswith(**"15"**):  **continue** *# print(date)  # print(time)* **for** line **in** file.readlines():  **if** re.match(**r"OPLST:01\*"**,line):  inf = eval(line[22:])  exchangeInstID = inf[**'ExchangeInstID'**]  d = inf[**'InstrumentCode'**]  print(d,fileName,date)  **if** d **in** s:  print(**"InsturmentCode 不是唯一的"**)  print(s)  **else**:  s.add(d) print(**"InstrumentCode是唯一的"**) |