

量化交易实战培训

基于开源框架vn.py

Day 1/6

李来佳 QQ/WeChat 28888502

- ① 内 容： 较为完整地讲解量化交易体系，并通过vn.py实战量化策略。
- ② 听 众： 具有一定编程基础的量化从业人员。
- ③ 讲 师： 一群爱好交易的程序员。
- ④ 感 谢： vn.py的创始人陈晓优和他的开源社区

交易
体系

数据

基础
库

可视
化分
析

VNPY
框架

深入
VNPY

基于
VNPY
编写
策略

回测
策略

模型
实战

1.1 金融量化交易

Quantitative Trading, 借助现代统计学和数学方法, 利用计算机进行交易的投資方式。

起源于20世纪70年代
美国股票市场



Virtu Financial 2008年成立

高频交易公司, 2015年4月纳斯达克上市 (VIRT)
在2009年至2013年的1238个交易日里, 仅有一天
出现亏损。在2014年的252个交易日里, 又一次创
下“零亏损日”的纪录



AIDYIA香港

人工智能自动化系统在美国证券市场中进行交易。
在它2016年5月上线的第一天, 它就在一个未披露的
资金池中产生了2%的回报。
首席科学家Babak Hodja, 是苹果Siri的开发者之一。



量化比例

美国: 70%

中国: 10%



詹姆斯·西蒙斯

数学家, 量化交易的代表人物
大奖章基金, 1989—2009年间大奖章基金平均年回报率高达35%
依靠数学模型和电脑, 管理着自己旗下的巨额基金, 用数学模型
捕捉市场机会, 由电脑作出交易决策。
“从某种角度看, 我们用的是机器学习”



旧金山创业公司Sentient Technologies

2015年, 获得了1.43亿美元投资,
正在开发和应用使用世界上最强大的分布式人工智能系统构建
的专有量化交易和投资策略。



量化基金经理

2016年5月, 《机构投资者》旗下出版物《阿尔法》公布的
“2016年全球收入最高的对冲基金经理”排行榜显示, 前十位
收入最高的对冲基金经理中, 有八位被归为量化基金经理, 前
25位有一半属于量化分析。

正期望值系统

- $\text{仓位} * \text{盈利概率} * \text{盈利点数} - \text{仓位} * \text{亏损概率} * \text{止损点数} - \text{交易手续费}$

高频交易：仓位低，盈利概率高，盈利点数小，交易手续费高

CTA交易：动态仓位，盈利概率低，盈利点数高
亏损概率高，止损点数低

套利交易：动态仓位，盈利概率高，盈利点数低
亏损概率低，止损点数低

1.2 角色与系统



策略分析师



策略开发员



基金经理、
风控员等

① 商业客户端：文华、金字塔、TB、MC、博易、通达信、交易师..

② 在线分析：RiceQuant、JoinQuant、优矿..

③ 定制类：AlgoTrade、VN.PY...



期货、股市、外汇



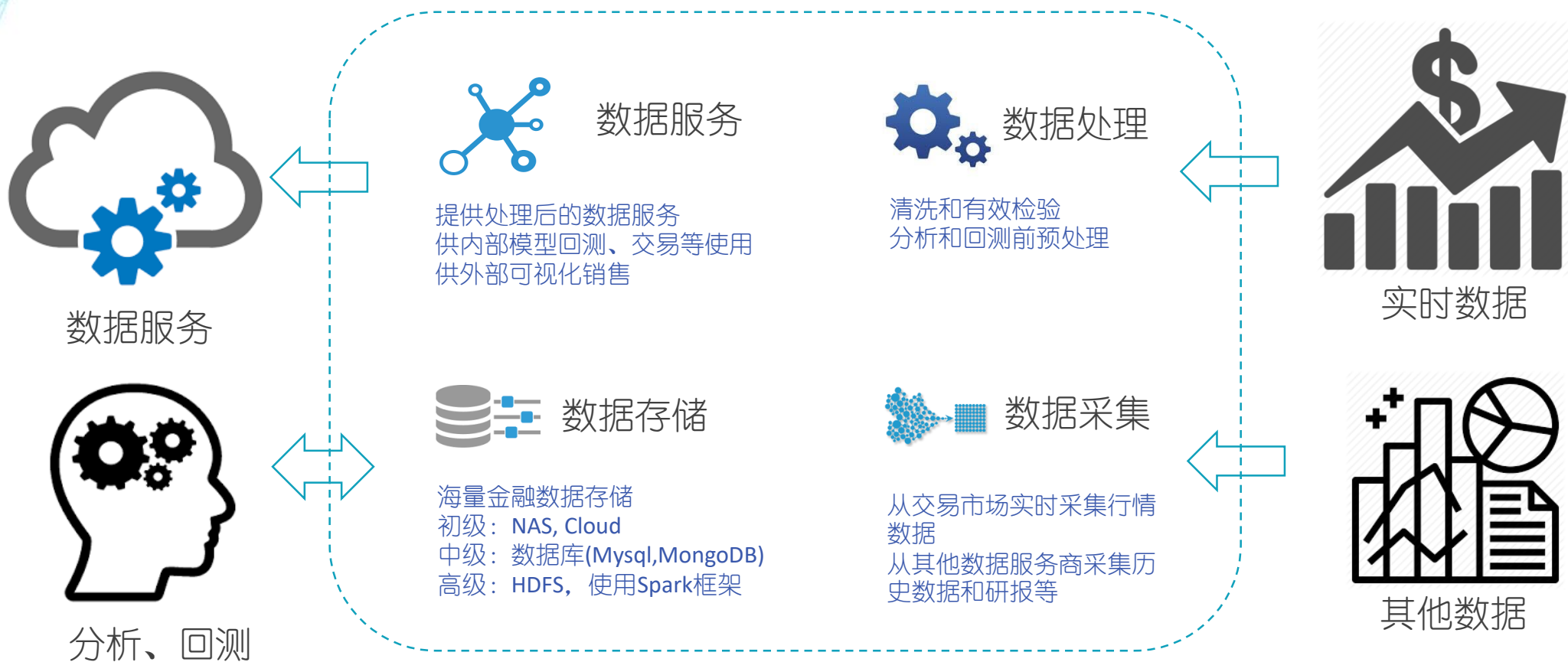
商业数据

1.3 交易体系组成



1.3.1 数据

数据是金融量化和人工智能的根基，海量、稳定、高效。



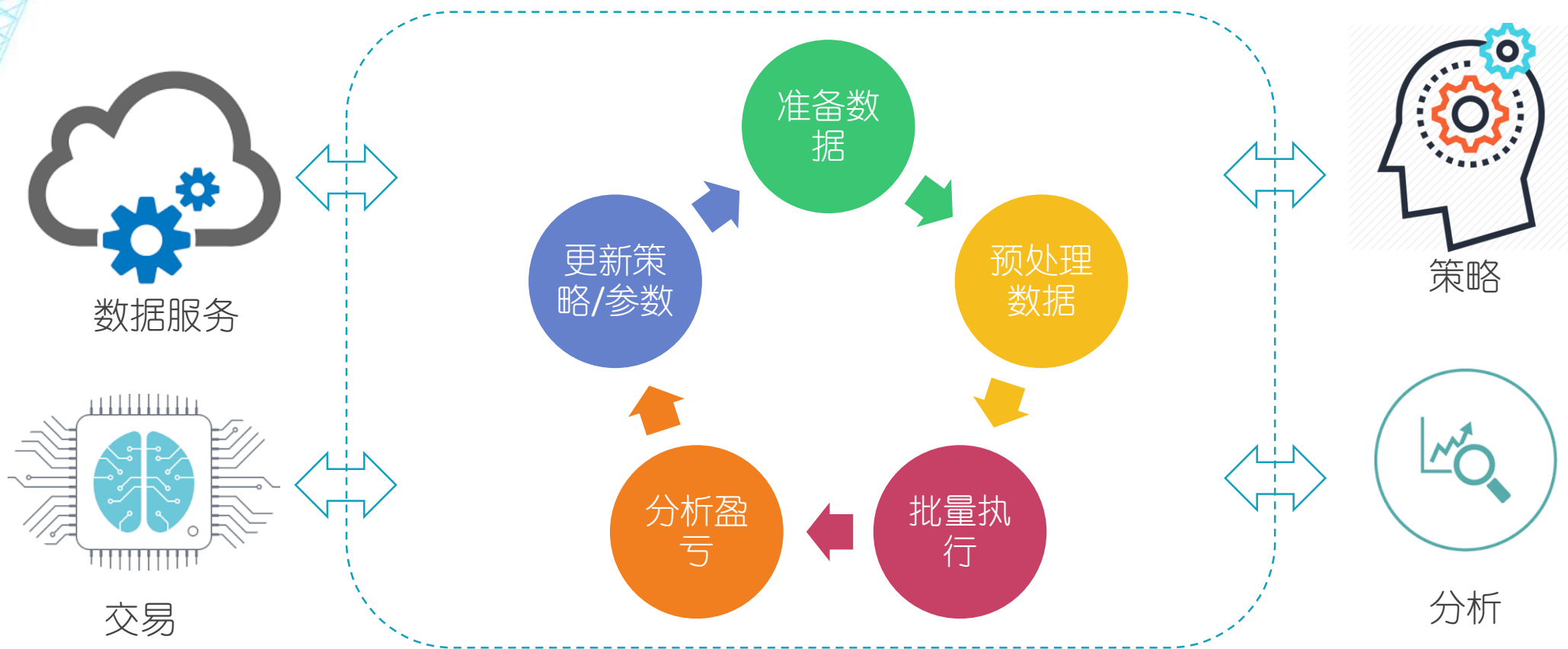
1.3.2 分析

金融领域的分析包括特征挖掘、置信区域、趋势分析等



1.3.3 回测

回测是使用历史数据，最大程度模拟实盘交易验证策略



1.3.4 策略

策略是核心机密资产，安全、有效、持续优化



1.3.5 交易

开源的分布式交易平台VNPY，高速、灵活、风险可控。



1.3.6 风控

执行严谨可靠的风控体系，是基金在金融市场的立命之本。



1.3.7 管理

人、宏观、微观

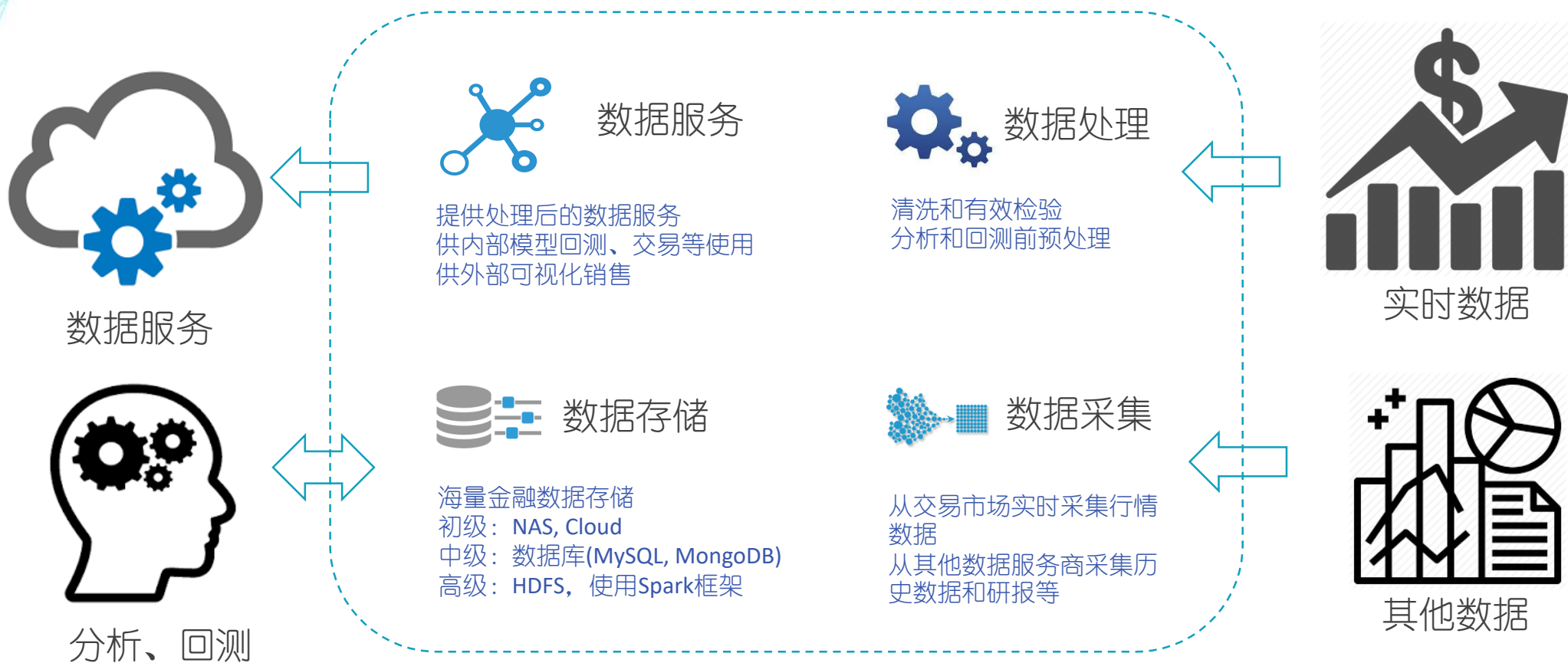




问答时间

2 数据

数据是金融量化和人工智能的根基，海量、稳定、高效。



2.1 数据分类

原始数据

结构化数据

行情类:Tick, 合约
交易类:委托/成交, 账号
资金

非结构化数据

舆情, 公告, 研报

处理数据

行情:分钟/小时/日/周/
月/年
主力合约,财务,流通股,
板块...

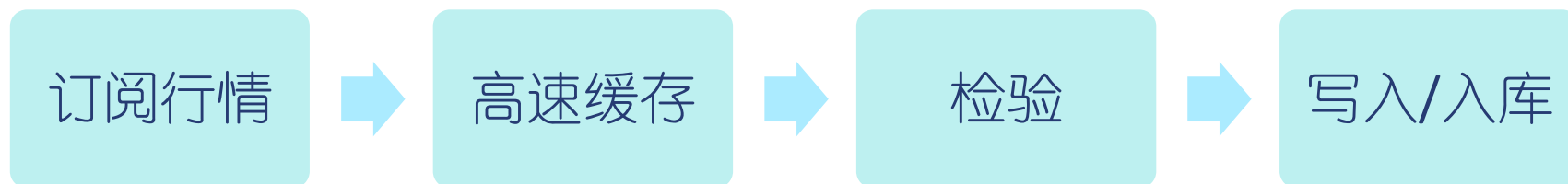
关联度, 热度



2.2 数据采集

- 实时采集
- Tushare使用
- 通联数据 (datayes)
- wind数据
- RiceQuant
- 淘宝购买数据
- 网站共享爬网

2.1.1 数据采集-实时采集



VNPY的实现：（期货为例） `vn.trader/dataRecorder`

- 1、通过单独数据记录模块
- 2、设置多个合约Symbol，通过`gateway.Subscribe()`订阅行情
- 3、注册Tick到达事件，生成tick和Bar数据，置入缓存
- 4、另建线程，写入MongoDB

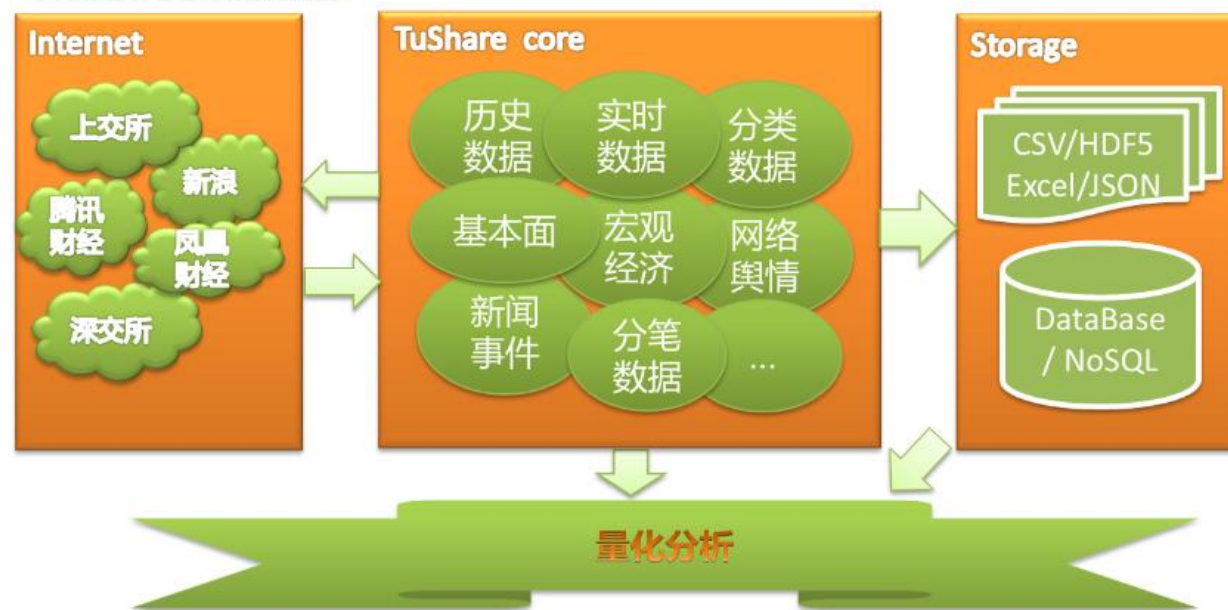
需完善：

- 1、需定期更新合约。 解决方法：通过行情接口获取所有合约
- 2、数据完整性校验。 解决方法：多个行情接口+校验机制

2.1.2 数据采集-TUSHARE

- TuShare是一个免费、开源的python财经数据接口包。主要实现对股票等金融数据从数据采集、清洗加工 到 数据存储的过程。其支持的数据种类比较丰富，API设计也是围绕着pandas来进行的。
- <http://Tushare.org>

TUSHARE 功能概览



- `import tushare as ts`
- `df = ts.get_hist_data('600848')` #一次性获取全部日k线数据
- `df.head(10)`

2.1.3 数据采集-通联数据

- **通联数据** (DataYes) 是国内目前最大的开放金融数据平台，整合了包括股票、基金、期货、期权和港股方面的全品类金融数据。
- Data API返回HTTP JSON和CSV格式数据，任何语言都可以直接调用HTTP API，以获得数据。
- <https://m.datayes.com/>
 - getFutureTickRTIntraDay --获取一只期货在本清算日内某时间段的行情信息
 - getFutureBarRTIntraDay --获取当日期货分钟线
 - getMktFutd --期货日行情(主力以持仓量计算)

2.1.4 数据采集-WIND数据

- Wind Datafeed服务
- 中国市场的精准金融数据服务供应商，为量化投资与各类金融业务系统提供准确、及时、完整的落地数据，内容涵盖股票、债券、基金、衍生品、指数、宏观行业等各类金融市场数据
- 提供了python的API
 - `w.wsd` 获取历史序列数据
 - `w.wsi` 获取分钟数据
 - `w.wst` 获取日内tick级别数据
 - `w.wss` 获取历史界面数据（针对股票为主）
 - `w.wsq` 获取和订阅实时行情数据（需要指定回调函数）

2.1.4 数据采集-WIND数据

■ 代码生成器

分钟序列数据向导 - 步骤 1/3 - 证券选择

范围: 期货 待选: 总数: 11 选中: 总数: 1

选中: RB1710.SHF [螺纹钢1710]

注: 分钟数据只支持国内6大交易所产品(上交所、深交所、中金所、上期所、大商所、郑商所)

输入代码:

上一步 下一步 确定 取消

分钟序列数据向导 - 步骤 2/3 - 指标选择

可选指标:

- 基本行情
 - 开盘价
 - 最高价
 - 最低价
 - 收盘价
 - 成交量
 - 成交额
 - 涨跌
 - 涨跌幅
 - 持仓量
- 技术指标
 - BIAS乖离率
 - BOLL布林带
 - DMI趋向标准
 - EXPMA指数平滑移动平均
 - KD随机指标
 - MA简单移动平均
 - MACD指数平滑异同平均
 - RSI相对强弱指标

已选指标:

指标名称	字段	参数	编辑
开盘价	open		
最高价	high		
最低价	low		
收盘价	close		
成交量	volume		
成交额	amt		
涨跌	chg		
涨跌幅	pct_chg		
持仓量	oi		
BOLL布林带	BOLL	26;2;MID	

支持函数名称、指标中文名称和拼音查找

删除 全部删除

上一步 下一步 确定 取消

分钟序列数据向导 - 步骤 3/3 - 参数设置

周期: 1 分钟(M)

时间设置

起始日期: 2017- 3-17 9:00:00 截止日期: 2017- 3-17 15:00:00 ☐ 当前时间

☐ 指定时间段 起始时间: 9:00:00 结束时间: 16:04:27

非交易间隔处理

☐ 沿用之前数据 ☒ 空值

复权方式

☒ 不复权 ☐ 前复权 ☐ 后复权 ☐ 指定复权点: 2017- 3-17

运行设置

☒ 不运行 ☐ 直接运行

数据列表 证券代码列表 字段列表 时间列表 错误ID 请求ID

输出变量名: w_wsi_data w_wsi_codes w_wsi_fields w_wsi_times w_wsi_errorid w_wsi_reqid

上一步 下一步 确定 取消

w.wsi("RB1710.SHF",
"open,high,low,close,volume,amt,chg,pct_chg,oi,BOLL",
"2017-03-17 09:00:00", "2017-03-17 15:00:00",
"BOLL_N=26;BOLL_Width=2;BOLL_IO=1")

数据采集-RICEQUANT

- RiceQuant是一家提供在线回测的量化科技公司，近期，推出了RQData和RQAlpha
- RQDataC，供用户本地快速调用远程数据
 - 提供主力连续合约(如IF88)和指数合约(如IF99)
 - `get_price(order_book_id, start_date=None, end_date=None, frequency='1d', fields=None, adjust_type='pre', skip_suspended=False, country='cn')` 历史行情
 - `order_book_id`, 1~多个合约
 - `Frequency`, 1d:1天; 1m: 1分钟, 28m: 28分钟
 - `get_dominant_future` 获取主力合约
 - `get_future_contracts` - 获取期货可交易合约列表
 - 未来支持Tick数据

2.3 数据处理

- 如何有效地清洗，处理数据

- Tick数据训练（去重）

- $(Ask1+Bid1)/2$ ，若只关注价格，将连续不变的价格tick去除，提高训练速度和回测速度。

- Bar数据回测（主连数据/指数数据）

- 需要人工/或程序，产生某一主连合约的换月日期。例如：
 - 策略中回测时，根据实际情况处理
 - 切换合约，或者强制平仓

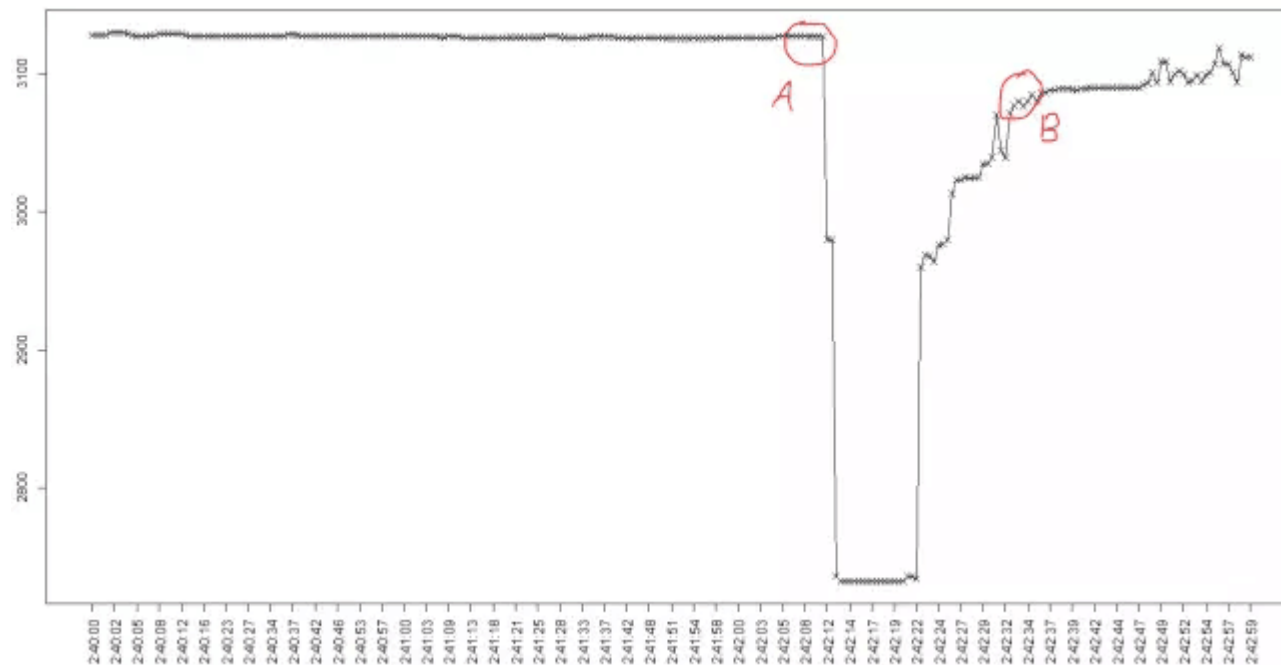
ContractNam	StartDate	EndDate	StartPric	EndPrice
RB1105	2010-10-14	2011-2-1	4515	5022
RB1110	2011-2-9	2011-8-4	5124	4980
RB1201	2011-8-5	2011-11-2	4884	4065
RB1205	2011-11-3	2012-3-1	4020	4277
RB1210	2012-3-2	2012-7-10	4294	3995
RB1301	2012-7-11	2012-10-22	3929	3679
RB1305	2012-10-23	2013-2-8	3733	4127

- 套利数据（价差/比率/方差）

- 暂未发现数据商提供套利数据，需自己生成
 - 利用套利公式和2腿合约tick，生成套利合约tick
 - 使用套利合约tick，进行套利回测。
 - 或回测模块中，同时读取两腿合约，实时计算套利合约tick，提交给策略

2.3 数据处理

2016.5.31 10:42:11左右, IF1606价格运行正常, 维持在3126.8附近。在10:42:12时, 最新价突变到2980, 一个tick的时间, 价格变化-146.8点, 价格维持了两个tick, 第四个tick即10:42:13.500时价格就到了跌停板2732.4, 价格距离3126.8变化为-394.4, 从正常价格到跌停板总共耗时2秒, 共接收到4个tick。价格在跌停板持续运行了7秒, 开始反弹, 一直反弹至10:42:37 价格为3088时, 价格趋于正常稳定相对3126.8价格变化为-38.8。



乌龙指过滤器:

将连续两个tick的价格变化不能超过30跳, 则IF为60点定义为异常行情出现。
如果当前tick与前一tick满足此条件时, 则本tick无效, 并且随后1分钟以内的行情均定义为异常行情, 不进入模型。

2.4 数据存储与服务

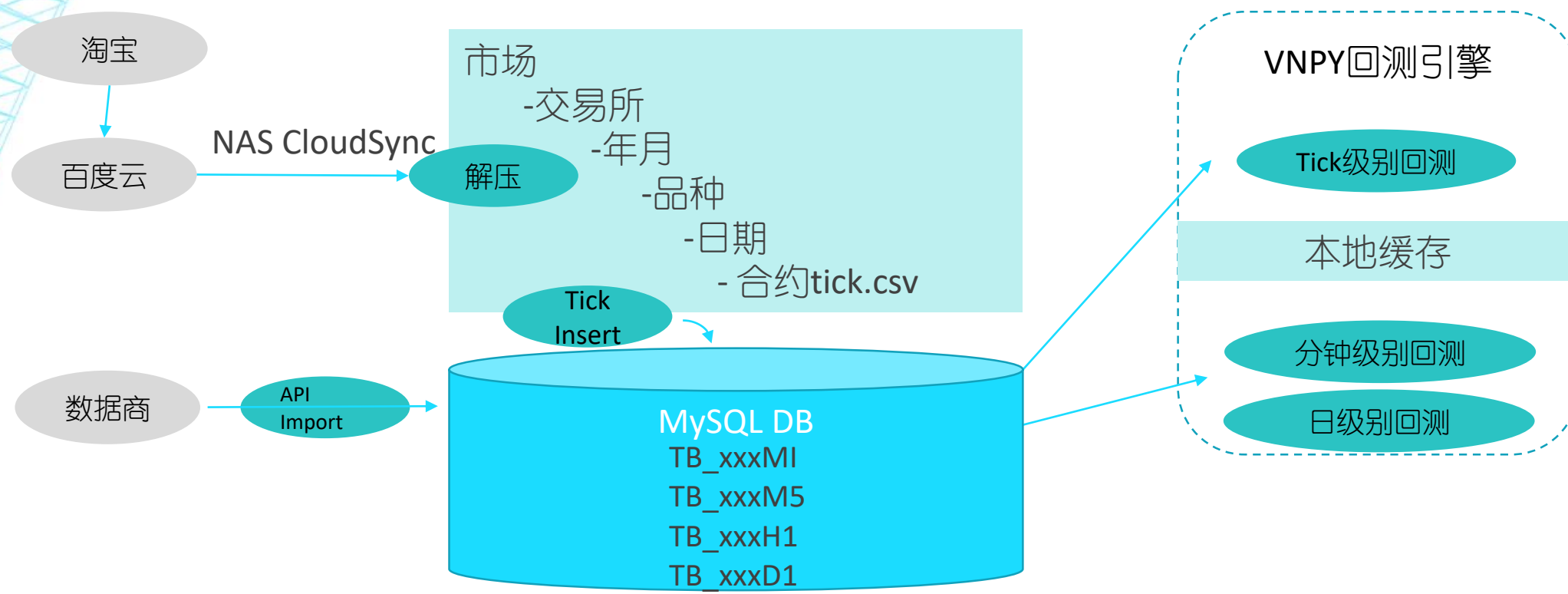
- 本地文件存储
 - 淘宝购买Tick数据=>百度云=> NAS Cloud Sync (本地) => 解压
 - 回测时，自编写数据访问，根据规则读取相应csv文件加载
 - 通过数据服务商接口读取分钟数据，生成csv文件供回测时读取



2.4 数据存储与服务

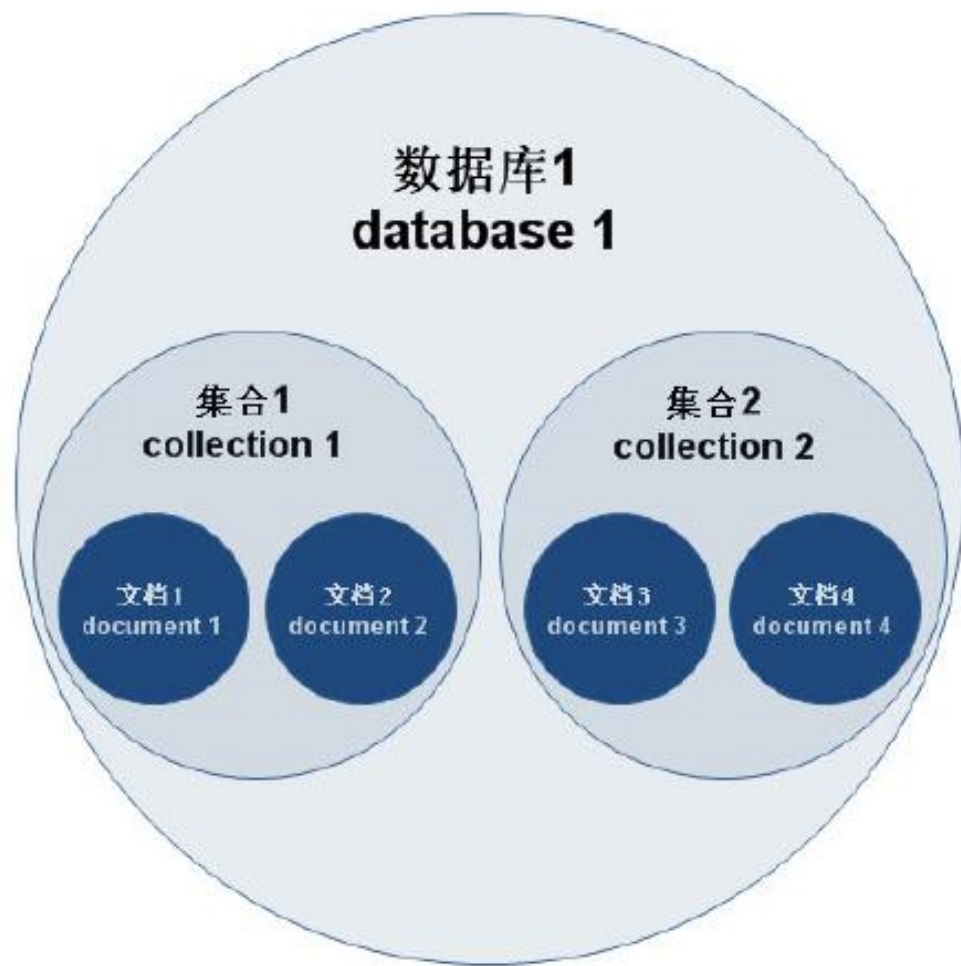
MySQL数据库

- 自编程序，通过数据服务商接口/文件读取并写入MySQL
- 自编tick数据处理程序，写入MySQL
- 回测时，通过mySQL语句读取相应日期数据



2.4 数据存储与服务

- MongoDB数据库

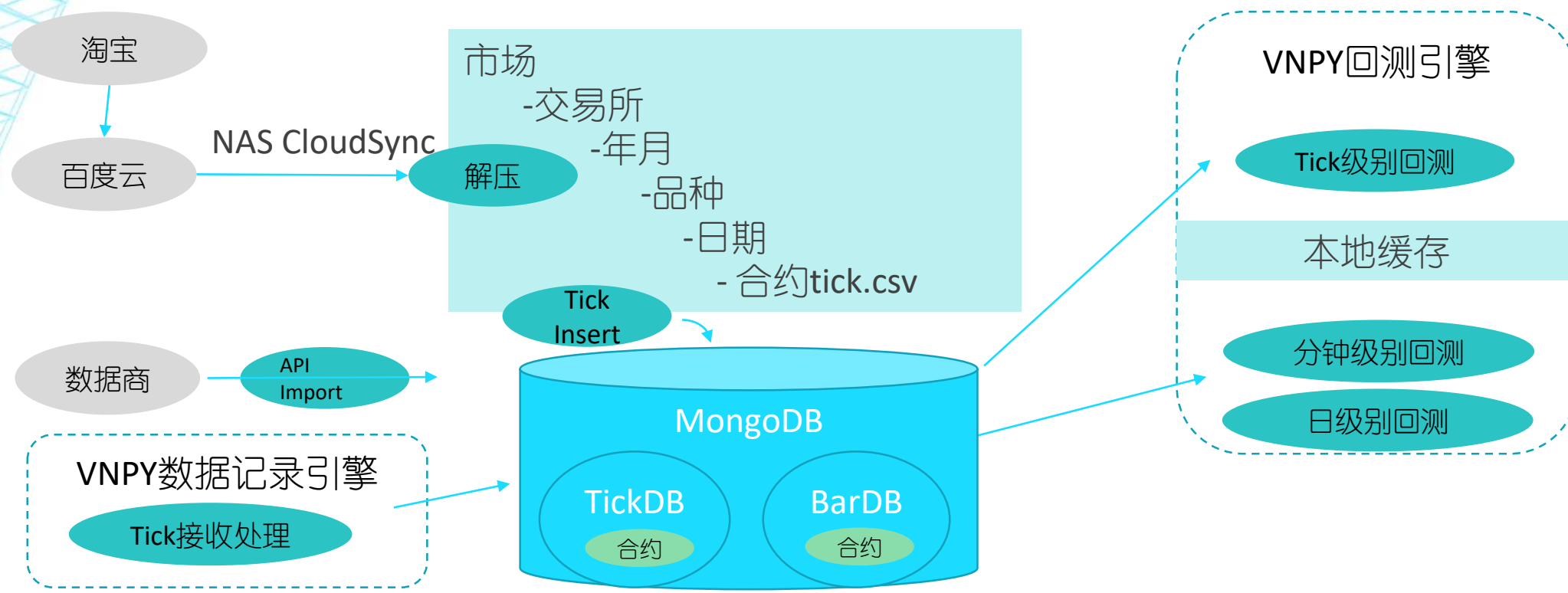


逻辑结构对比	
MongoDB	关系型数据库
文档(document)	行(row)
集合(collection)	表(table)
数据库(database)	数据库(database)

2.4 数据存储与服务

▪ MongoDB数据库

- 自编程序，通过数据服务商接口/文件，读取并写入MongoDB
- VNPY数据记录模块，实时接收并写入MongoDB
- 回测时，通过pyMongo模块获取数据





课后作业

- 读取一个月的文件tick数据，存入MongoDB
- 生成1分钟Bar，存入MongoDB
- 查询两个日期之间的1分钟Bar

THANKS

微信：28888502

