集算器

创新大数据计算引擎

产权交易所解析HTML与计算

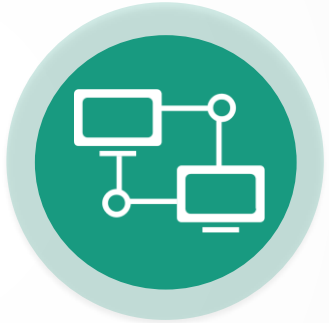
润乾软件出品

项目背景

随着某产权交易所的业务发展，内部产生了大量的交易数据（如：交易的标的、价格、目前交易的阶段等），这对于参与交易的相关单位非常具有参考价值！
同样地，其它地区的交易所也含有很多类似的重要数据，该交易所希望整合这些外部的公开数据，与自身数据相结合，对外提供完整的交易数据服务！

数据采集多

通过爬虫技术手段抓取多个不同地区的交易所网站数据，结构化转换之后，存入数据集市



资源整合难

抓取的外部数据如何存储?与内部数据如何关联?为选合理、有效的方案而犹豫不决！

网站会改版

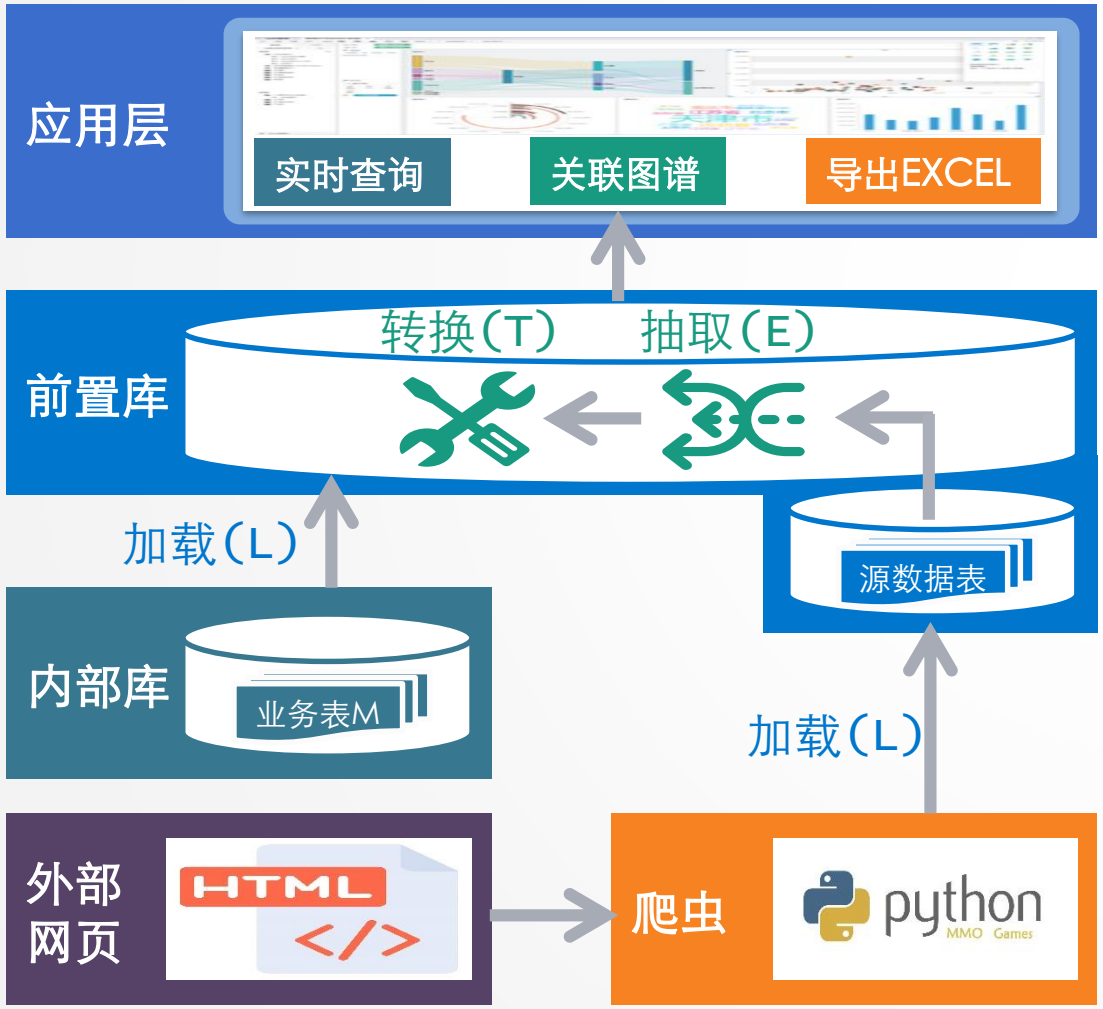
爬虫抓取的网页，有可能会改版，相应的解析程序就要重新修改，工作量大，对人员要求高

注重时效性

关键数据一旦有变化或更新，能在第一时间被消化、利用，保证其时效性，实现数据驱动运营

常见方案的不足

增加前置数据库，外部数据利用Python爬虫技术采集，内部数据定时同步到前置库，在库内做数据的清洗、转换、所有数据查询都基于前置数据库，从而达到统一查询的目标！



架构成本高

做一轮同库来试图解决跨源混合运算，增加了人员开发工作量、数据库、etl等软硬件成本

开发效率低

Python语法不是专为结构化数据计算设计，解析多种HTML，开发周期长，应对困难

实时响应差

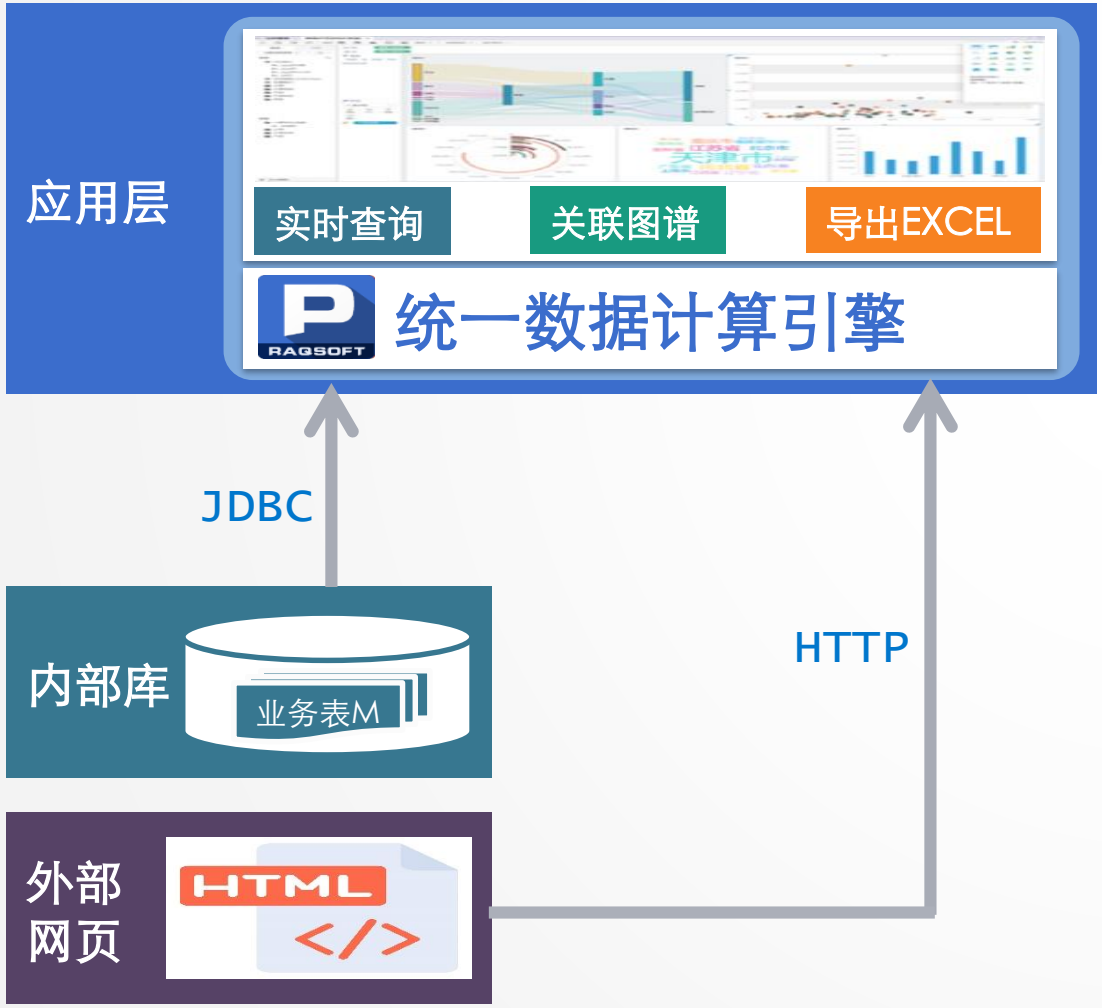
总是定时将HTML数据加载到前置库中，再经存储过程计算后，才能得出有效指标

适应能力弱

当网站改版时，Python解析程序要重新修改、联调测试、上线，运维成本高

引入数据计算层

不变动基础层架构，在应用层集成数据计算引擎中间件，适配各类SQL、NOSQL，使用一致的结构化计算模型，为前端应用提供统一计算服务！



架构更精简

独立的计算引擎，可轻松实现跨源混合计算，无须多余的数据库和etl组件

开发更简单

精心设计的丰富库函数和一致性语法，比python更易掌握并且性能更好

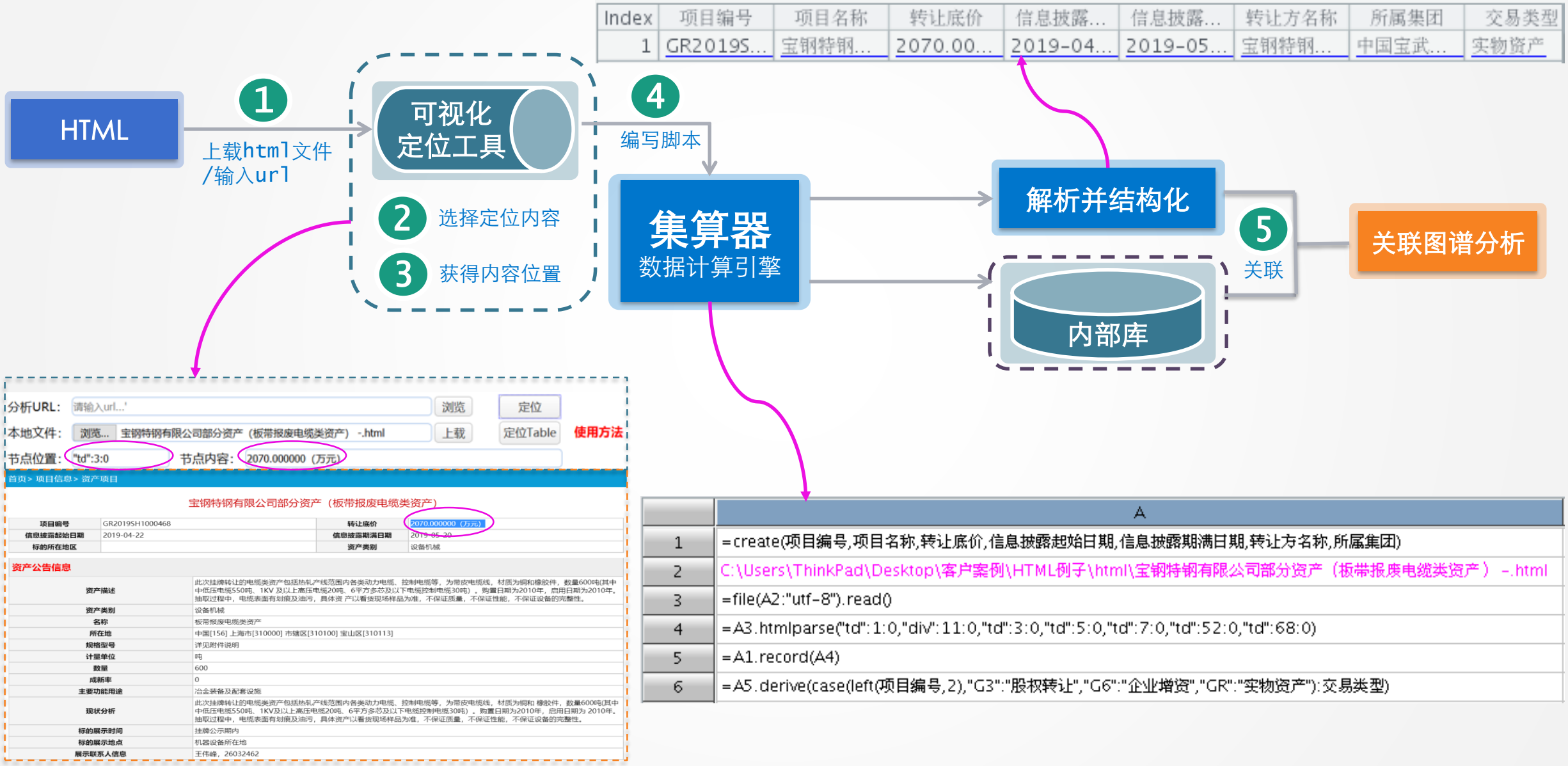
指标实时算

直接计算HTML数据得出最终指标，无须借助中间表二次过渡，实时性好

易迭代升级

解析算法归属于应用，网站改版时，修改计算逻辑不会影响系统其它部分

数据计算流程详解

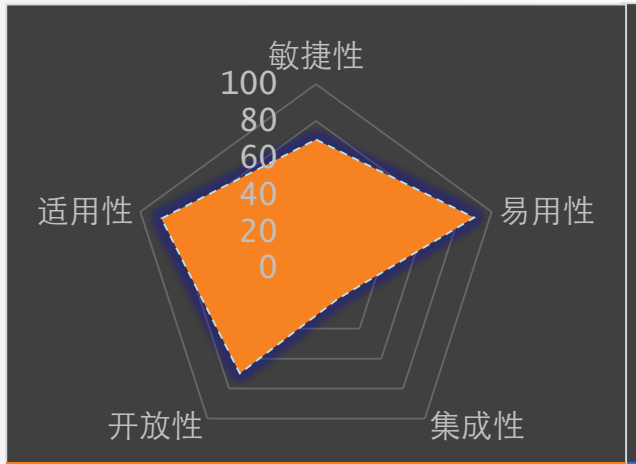


Index	项目编号	项目名称	转让底价	信息披露...	信息披露...	转让方名称	所属集团	交易类型
1	GR2019S...	宝钢特钢...	2070.00...	2019-04...	2019-05...	宝钢特钢...	中国宝武...	实物资产

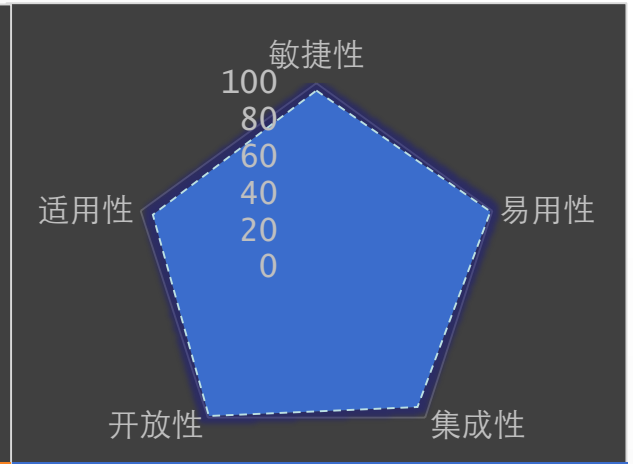
	A
1	=create(项目编号,项目名称,转让底价,信息披露起始日期,信息披露期满日期,转让方名称,所属集团)
2	C:\Users\ThinkPad\Desktop\客户案例\HTML例子\html\宝钢特钢有限公司部分资产（板带报废电缆类资产） -.html
3	=file(A2:"utf-8").read()
4	=A3.htmlparse("td":1:0,"div":11:0,"td":3:0,"td":5:0,"td":7:0,"td":52:0,"td":68:0)
5	=A1.record(A4)
6	=A5.derive(case(left(项目编号,2),"G3":"股权转让","G6":"企业增资","GR":"实物资产"):交易类型)

实测：集算器在各方面的工作量评估

对比指标	Python+SQL	集算器（SPL）	提升
代码量（单一场景）	96行	17行	5.6倍
工作量	30人天	6人天	5倍
再优化	无	容易	质变
耦合度	高	低	模块化、工具化
可维护性	较差	好	易调试、易集成



Python



SPL

SPL技术优势特点

- **敏捷性**-即装即用，环境配置简单，无须额外插件
- **适用性**-完善的类库和过程计算，适应复杂数据处理
- **易用性**-语法简单直观，调试方便，无须专业程序员
- **开放性**-内置多种数据源访问接口，直接计算
- **集成性**-无缝嵌入应用系统，易于转为日常计算

为什么集算器能如此提高效率

独立计算引擎

不依赖于数据库的计算能力，适配各类SQL、NOSQL混合运算

敏捷语法体系

语法简洁，实现同样算法，只需更少的代码，更短的开发周期

应用部署更易

无框架，标准接口易于实现应用嵌入式集成

多种数据接口

直接计算，无须入库，体系结构更精简，减少中间表降低开发量

创新技术
推动应用进步！


各设备故障分析
近7天设备的总故障率
近7天
近1天
近6小时
1.53%
0.95%
2.40%
近7天故障情况
全年故障综合分析
近7天故障次数前5的设备
服务器CPU使用率
今日
昨日
今天硬盘情况