



集算器

创新大数据计算引擎

交易所统一数据集市案例

润乾软件出品



项目背景



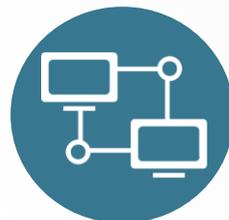
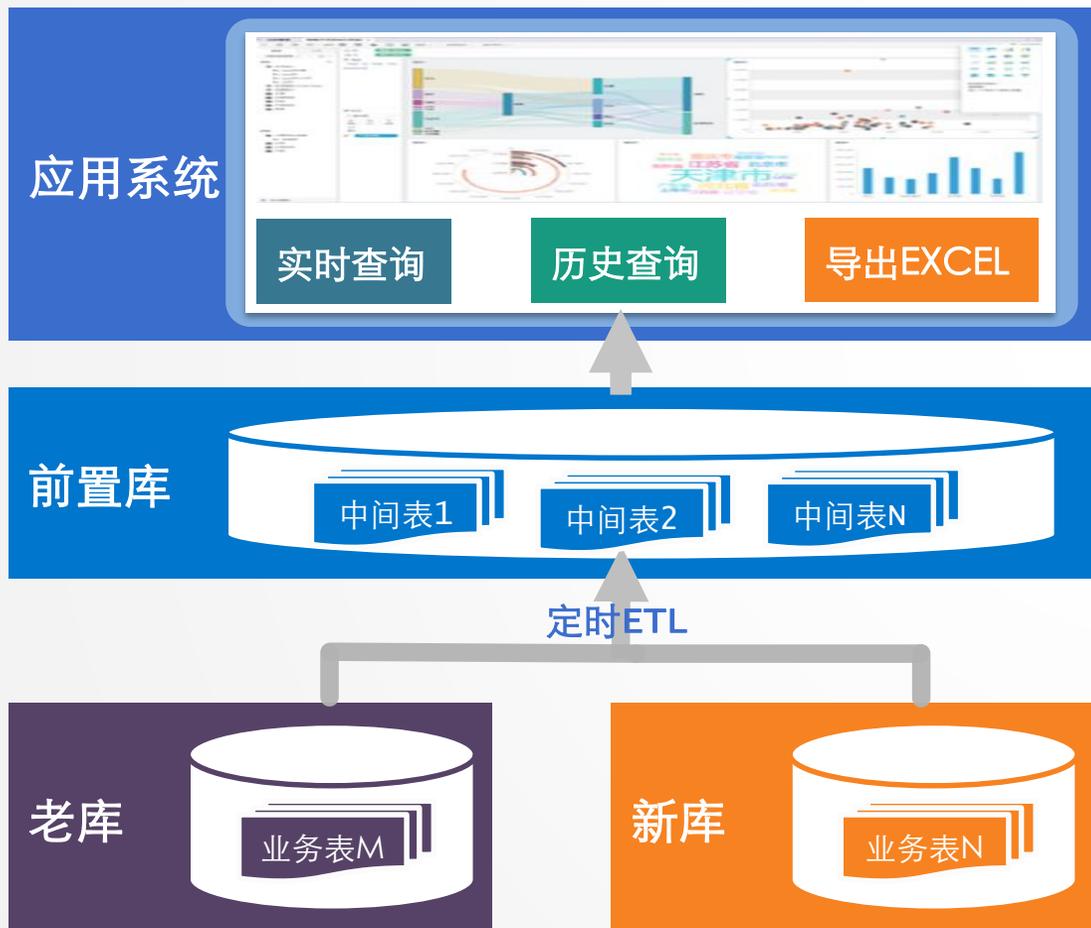
随着交易所业务的发展和IT系统的建设，出现了多个系统同时运行，互不连通的问题。在实现数据分析时，必须要解决多个数据系统之间联合分析的难题！





常见方案的不足

增加前置数据库，利用ETL工具定时从新老生产库中提取数据，清洗后再导入到前置数据库中，所有的历史报表查询都基于前置数据库，从而达到统一查询的目标！



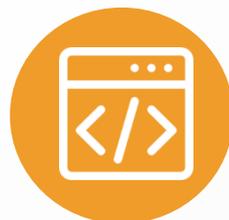
改造成本高

需要庞大的人员开发工作量、数据库、专业ETL工具等软硬件成本



结构变化大

数据存储层变化较大，增加数据库提升软件成本，让管理和维护更为复杂



实时性不足

实时跨库混合计算、仍需程序语言硬编码实现、实施难度高



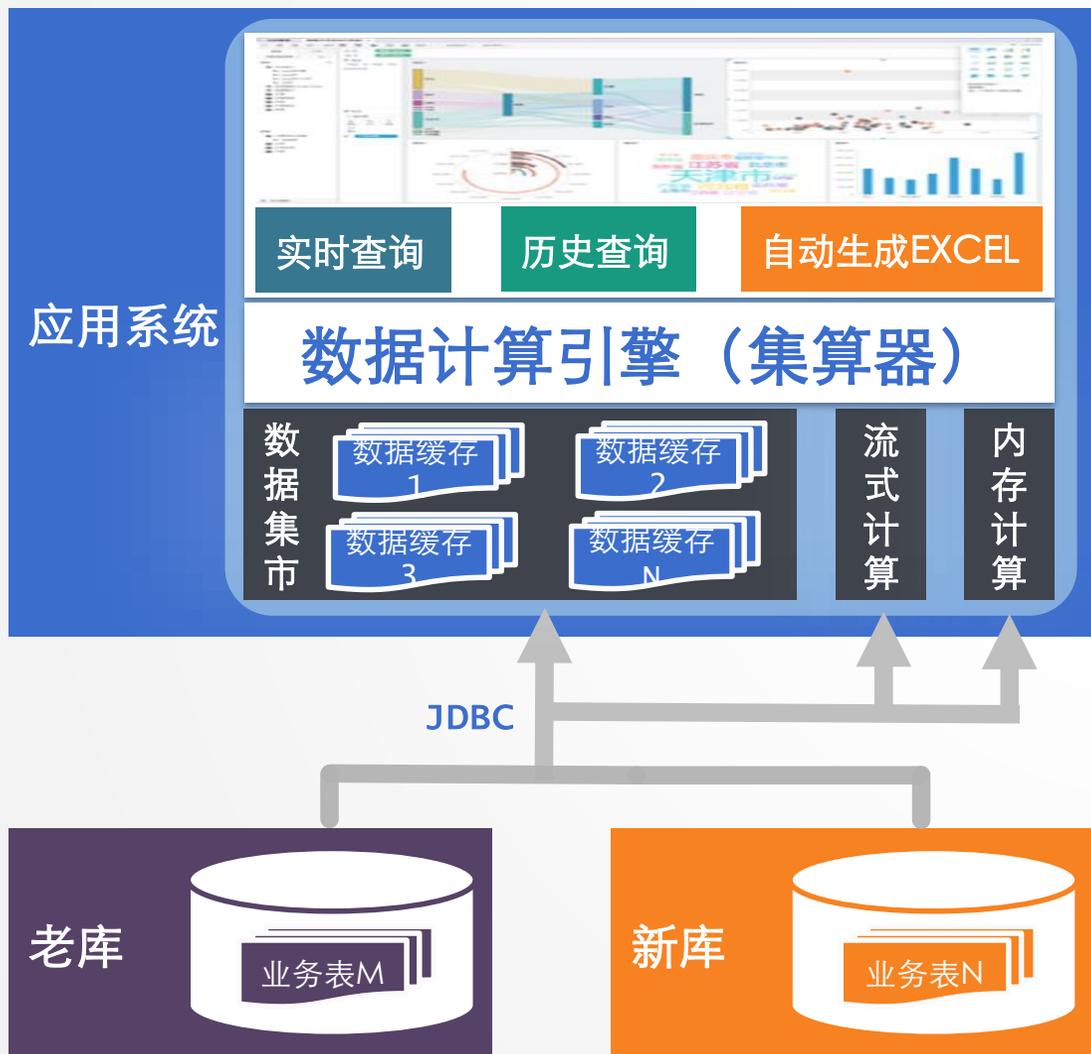
性能很一般

相对于压缩文件，前置数据库的IO效率低，当数据量大时，会面临性能的挑战



引入数据计算层

经过需求分析和产品选型，最终确定采用润乾报表加上润乾集算器的方式来解决数据统一存储、统一计算的问题！



统一数据集市

在应用层构建文件型数据集市，将不同数据源中，不同规范和格式的数据计算整合在一起



实时跨源计算

独立的计算层可轻松实现实时跨源混合运算，从而很容易实现全面T+0等查询，且代码书写简单



更低成本存储

压缩数据文件可以简单地存储在廉价硬盘上，且IO性能更高；无需购买昂贵的数据库专用软硬件



优化体系结构

脚本算法完全归属于应用本身，降低了应用系统与数据库的耦合



实测：集算器开发效率表现优秀 - 举例

一个简单的异构数据合并场景，SQL本身没办法跨库计算，一般借助JAVA等高级语言进行二次封装再返回结果；大概需要200行代码；而集算器只需4行(含类型转换)！

计算结果

Index	项目编号	转让方	交易日期	受让方	交易金额
147	10394	HUNGC	2012-12-25	徐先生	442.0
148	10395	HILAA	2012-12-26	王先生	2122.91999...
149	10396	FRANK	2012-12-27	余小姐	1903.80000...
150	10397	PRINI	2012-12-27	鍾彩瑜	716.719994...
151	10398	SAVEA	2012-12-30	苏先生	2505.59999...
152	10399	VAFFE	2012-12-31	方先生	1765.6
153	10400	EASTC	2013-01-01	谢小姐	3063.0
154	10401	RATTC	2013-01-01	王先生	3868.6
155	10402	ERNSH	2013-01-02	王先生	2713.5
156	10403	ERNSH	2013-01-03	王先生	855.014994...
157	10404	MAGAA	2013-01-03	王炫皓	1591.24999...
158	10405	LINOD	2013-01-06	黄雅玲	400.0

提升开发效率
40倍!

跨库合并

	A	B
1	=MYSQL.query("select * from old_deal")	
2	=ORA.query("select * from new_deal")	
3	=A2.run(交易日期=date(string(交易日期),"yyyyMMdd"))	
4	=A1.rename(成交日期:交易日期,成交金额:交易金额)	
5	=A4 A3	

SPL代码

数据源

Index	项目编号	转让方	成交日期	受让方	成交金额
147	10394	HUNGC	2012-12-25	徐先生	442.0
148	10395	HILAA	2012-12-26	王先生	2122.91999...
149	10396	FRANK	2012-12-27	余小姐	1903.80000...
150	10397	PRINI	2012-12-27	鍾彩瑜	716.719994...
151	10398	SAVEA	2012-12-30	苏先生	2505.59999...
152	10399	VAFFE	2012-12-31	方先生	1765.6

MySQL table:old_deal(部分字段)

Index	项目编号	转让方	交易日期	受让方	交易金额
1	10400	EASTC	20130101	谢小姐	3063.0
2	10401	RATTC	20130101	王先生	3868.6
3	10402	ERNSH	20130102	王先生	2713.5
4	10403	ERNSH	20130103	王先生	855.014994...
5	10404	MAGAA	20130103	王炫皓	1591.24999...
6	10405	LINOD	20130106	黄雅玲	400.0

ORACLE table:new_deal(部分字段)

为什么集算器能如此提高效率



敏捷语法体系

降低算法实现复杂度，极大地提高开发效率



高效函数类库

内置计算引擎，不依赖于数据库计算，轻松实现跨源混算



多样性数据源

丰富的多样性数据源接口，直接取数计算

集算器在项目中的价值 - 总结



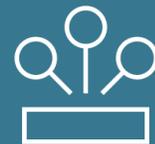
集算器

简称: DCM
(Data Computing
Middleware)



易开发

容易跨库混合运算, 不限是否同构



低成本

让文件拥有计算能力, 无需构建数仓



低耦合

算法归属于应用, 与数据库解耦



开放性

内置多种数据源接口, 直接计算

创新技术 推动应用进步!

