



集算器

创新大数据计算引擎

# 木槿生活数据平台案例

润乾软件出品



# 项目背景



## 报表数量多

- IT部门已负责维护的报表数量达到85+张
- 随着零售行业业务种类的不断丰富，报表数量还在不断增加

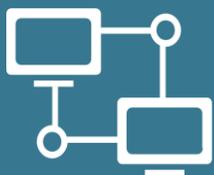
## 报表查询慢

- 20%的报表查询速度是分钟级，5%的报表查询是小时级
- 一些常用报表由于查询慢已影响各部门做业务，如市场营销、数据整理再汇报

## 资源整合难

- 数据分散在不同的业务系统中，彼此互不关联，联合查询困难，造成信息孤岛

# 洞察现象后的本质



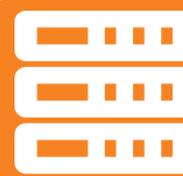
数据量大，交易流水表5亿，6张常用表在4千万至1亿之间，库存表及其他维表在百万到千万不等



数据库承载过多的计算作业，同时满足业务交易和报表查询，负担重，性能到达瓶颈



报表业务稳定性差，开发没完没了；新需求需专业程序员现场开发测试，缺乏长期应对机制

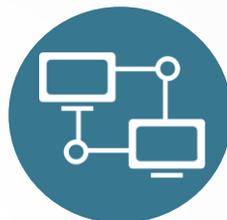
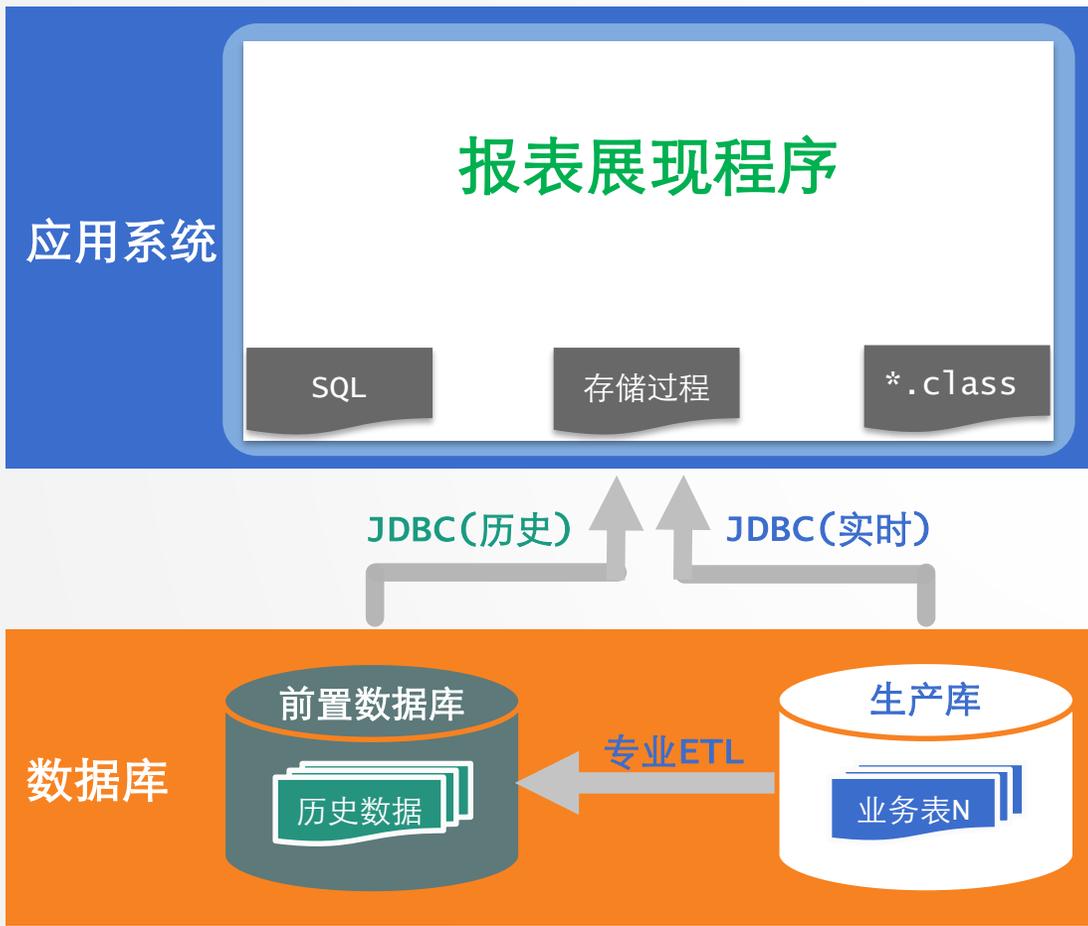


跨库混合计算、需硬编码实现、实施难度太高



# 常见方案的不足

增加前置数据库，利用ETL工具定时从生产库中提取数据，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，从而和生产库分离！



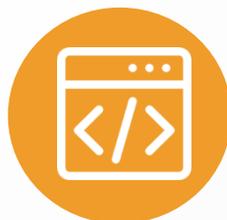
## 性能一般

前置数据库的IO效率低，而报表性能又严重依赖于取数环节，没能从根子上解决



## 改变了原有架构

数据存储层变化较大，增加数据库提升软件成本，而且管理和维护更为复杂



## 报表开发难

不支持跨库，混合运算的压力需扔给报表环节，依然面临开发难的窘迫



## 升级成本高

提高性能需庞大的人员开发工作量、数据库、专业ETL工具等软硬件扩容成本

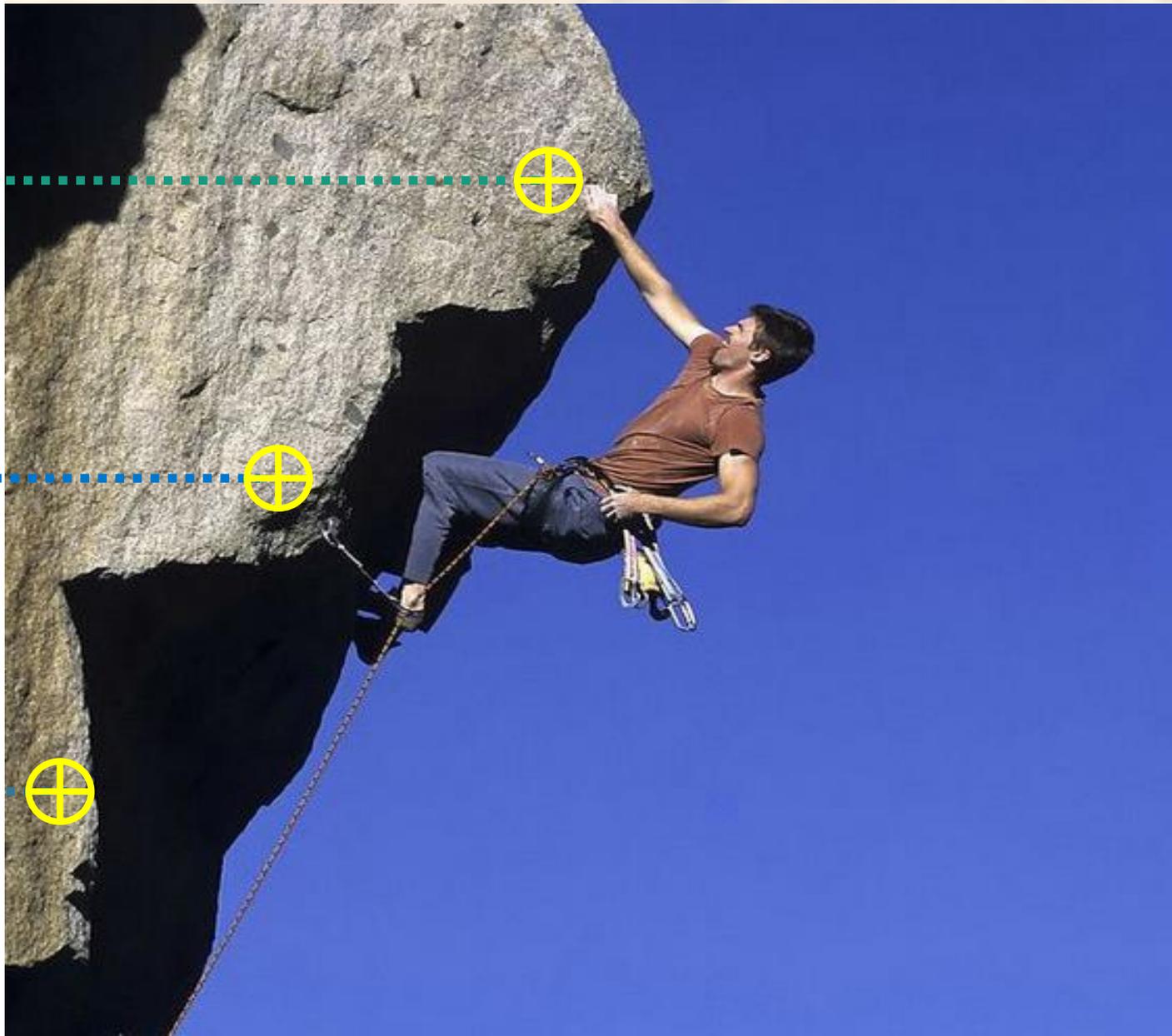
# 用户期望

这些能做到吗?

大数据量查询不超过1分钟  
按参数条件任意组合查询?

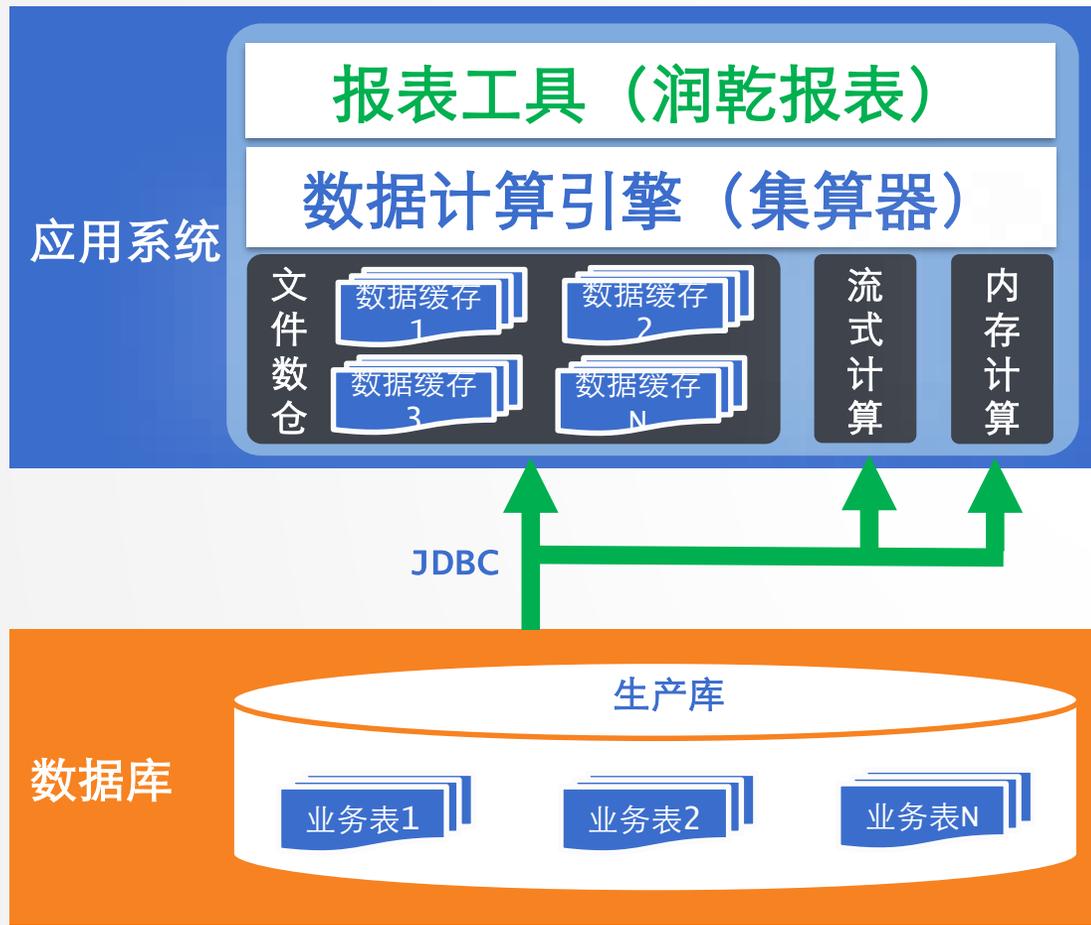
不变动现有基础层架构  
可在应用层面做努力?

报表开发过程工具化  
自己可解决90%报表难题?





# 引入数据计算层



不变动基础层架构，在应用层构建文件型数据集市，让文件拥有计算能力，且比数据库更高IO性能，解决大数据量慢的难题



独立的计算层可轻松实现跨源混合运算，全面T+0查询等等，为展现层(报表)提供服务



降低了应用与数据库的耦合，当需求变更时修改此模块不影响系统其他部分，完全模块化



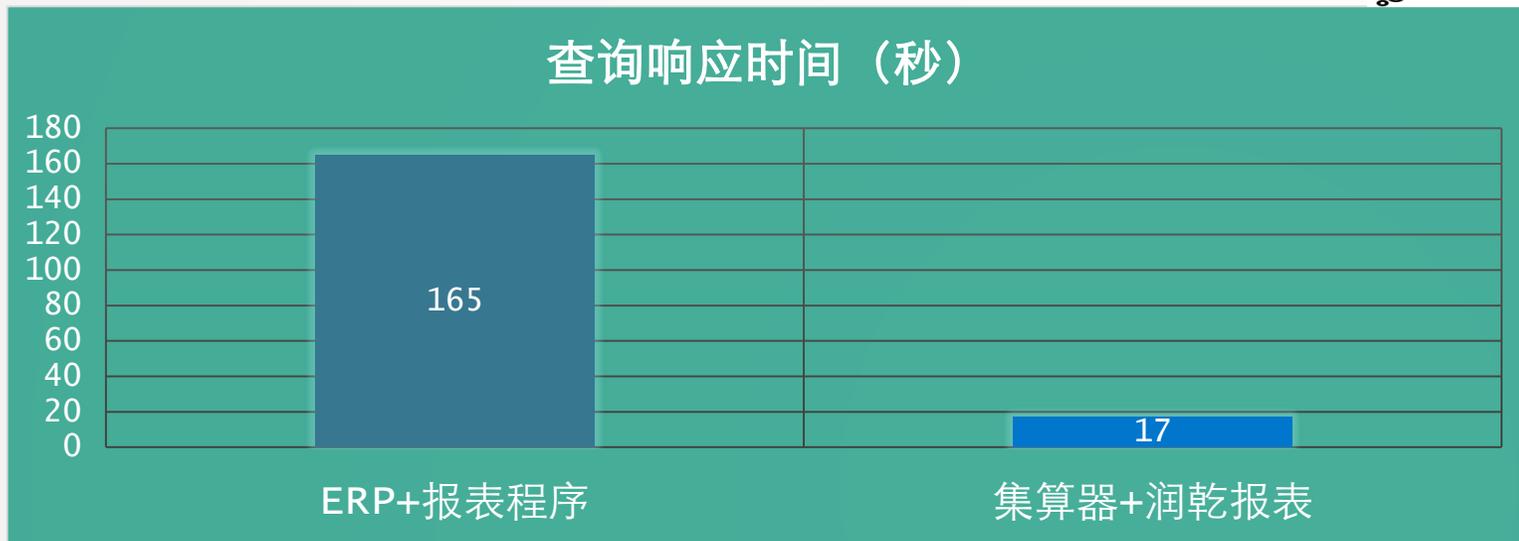
无须复杂环境配置与应用层代码引用，报表开发过程全面工具化，应对没完没了的报表难题，降低对开发人员要求



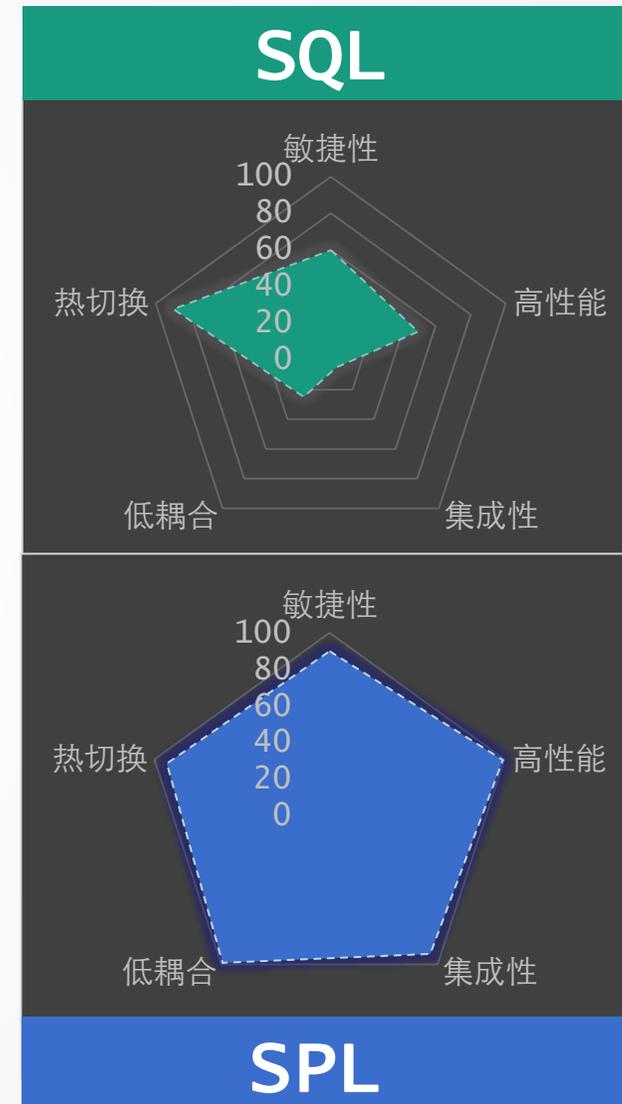


# 取得成果 — 性能优势

部分示例



对比指标	SQL/存储过程	集算器	变化
代码行	253	26	代码量减至10%
开发周期	5人天	1人天	工作量减至20%
执行性能	2分45秒	17秒	平均性能提升10倍
应用耦合	高	低	易管理、热切换
维护性	较差	好	维护成本低



# 集算器在项目中的价值 - 总结



高性能

无需构建数仓解决大数据量慢的难题



轻量级

嵌入应用集成,不改原有基础构架



ETL

搭配定时任务,实现ETL,一机多用



低耦合

报表与算法一体化,与数据库解耦



工具化

开发过程工具化,IT人员即可解决



扩展性

架构高度易用,为项目扩建奠定基础

# 创新技术 推动应用进步！

