

查询慢、跑批慢、性能低怎么办？

查询一次等的花儿都谢了
跑批跑不完，第二天被业务追杀
数据库慢如牛，我也没办法唉

你是否也曾如此崩溃?



查询与报表慢!!!



- 清单大报表数据量大, 呈现太慢
- 关联运算太多, SQL嵌套太深, 数据库优化不起作用
- T+0全量查询数据量大, 耗用资源过多, 影响生产系统
-

- 存储过程步骤多, 大量使用临时表, 性能低下
- 涉及库外数据, 要先同库再跑批, 浪费时间
- 数据库IO效率太低, 大量数据迁移耗时太长
-



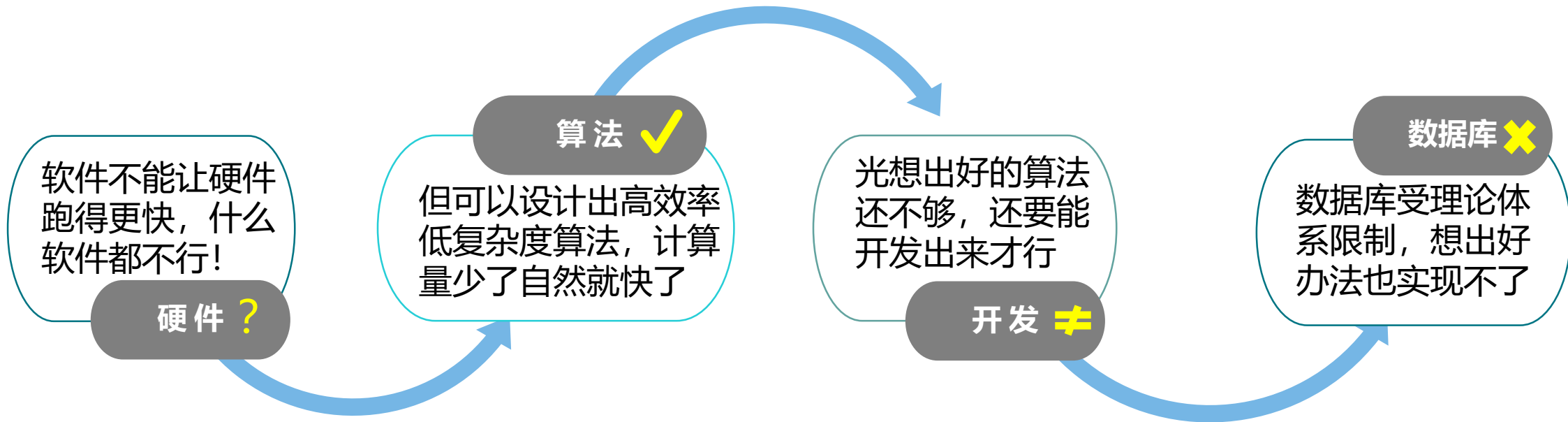
跑批慢!!!

数据库压力大!!!



- 数据仓库的容量已达到极限, 扩容成本高昂
- 数据仓库支撑过多应用, 并发过多导致性能不可控
- 冗余的汇总中间表太多, 无端消耗数据库资源
-

这是怎么回事呢?



那咋办呢?

往后看!

哦, 原来是这样

对咯, 说破了不神奇

那找程序员去做呗

没有这么容易滴

那不是只能干瞪眼吗?

嘿嘿, 绝大多数情况就是这样滴

因此

高性能计算

=

算法设计

+

算法实现

→ 成为制约高性能计算的瓶颈

SQL, NoSQL, NewSQL, 还有Hadoop, 都会限制算法实现哦

那还能怎么办?



引入新的**计算体系**及开发工具! 易于实现高性能、低复杂度算法

该技术应具备的特点

低复杂度

能实现还不够, 算法实现要相对简单

高性能

算法运行高效

使用灵活

可以根据计算和数据特征灵活设计算法

易扩展

支持分布式计算



解决问题
一定要从根本入手

集算器 - 性能优化专家



集算器采用新的理论模型和SPL语法，可以很方便实现低复杂度、高性能算法

The screenshot shows the Jisuanqi V2018 software interface. The main window displays a script editor with the following code:

```
1 =B5.group(clerk_name;~.sum(sale_amt).total_amt)
2
3
4 for A1
5   =A2.select(sale_date>=A4.start_date && sale_date<=A4.end_date)
6   =B5.group(clerk_name;~.sum(sale_amt).total_amt)
7   =B6.maxp(total_amt)
8   >A3.insert(0,A4.promo_name,B7.clerk_name)
9 =A3
10
```

The right-hand pane shows a data table with the following content:

序号	clerk_name	total_amt
1	Baker	25184
2	Bill	20776
3	Dow	26332
4	Harry	18103
5	Jenny	22198
6	Johnson	24459
7	Larry	17097
8	Mary	24775
9	Steven	24474
10	Tom	25092

A blue banner at the bottom of the screenshot contains the text: 集算器SPL采用面向过程计算，分步骤实现算法



终于有自己的语言了!
快给朕带上来瞧瞧!

集算器提供的部分高性能计算技术



遍历技术

延迟游标

聚合理解

有序游标

遍历复用

预过滤遍历

高效关联

外键指针化

外键序号化

有序归并

主子同维表一体化

单边HASH连接

高性能存储

有序压缩存储

自由列式存储

层次序号式定位

片状索引及缓存

倍增自由分段并行

分布式计算

抢先式负载均衡

Fork-reduce

外存冗余式容错

内存备胎式容错

集群维表

别搜了，这里大多数技术都是集算器发明的，找不到！



集算器性能表现



*数据规模：100亿行；集群数量：4

测试结果（时间单位：秒）

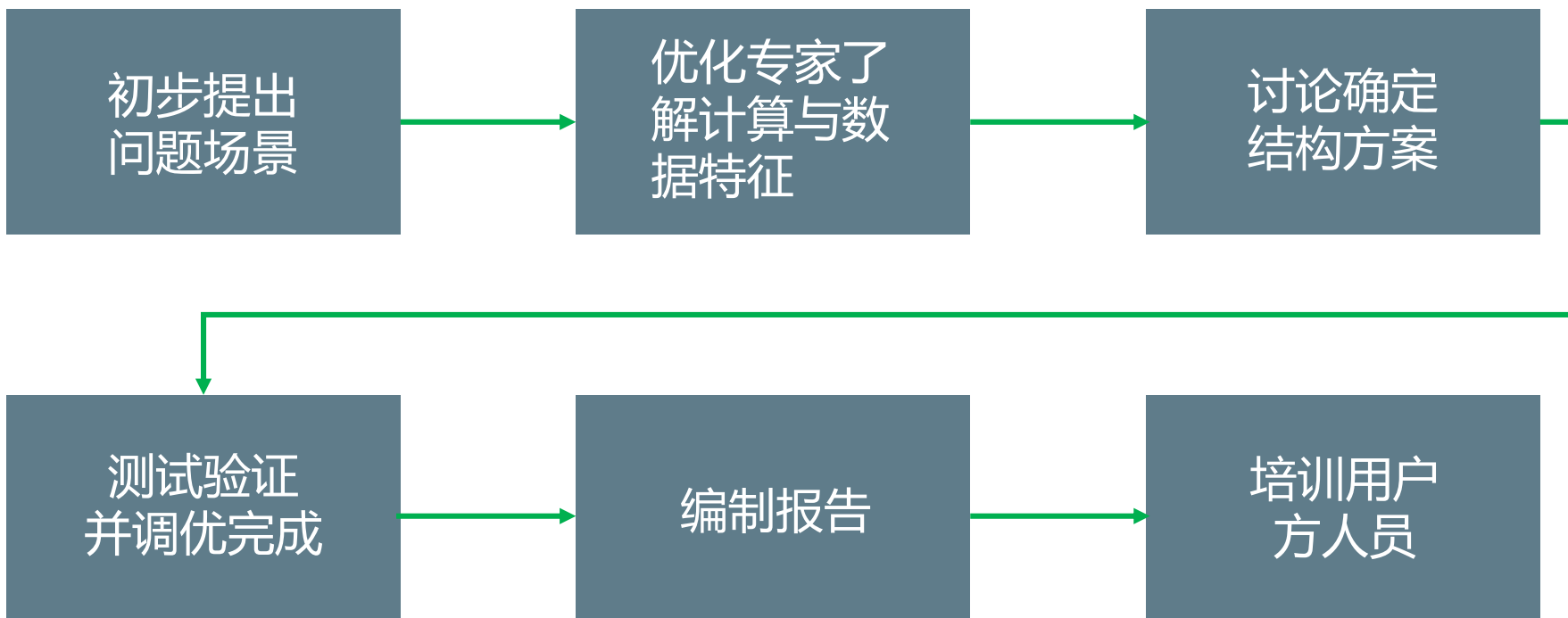
测试用例	Intel X86芯片			国产飞腾芯片	
	SPL读文件计算	SPL读数据库计算	数据库中SQL计算	SPL读文件计算	SPL读数据库计算
连接后并集	-	3.8	>1小时	-	-
连接后交集	-	3.9	>1小时	-	-
多对多连接遍历	69	103	>1小时	93	268
有序分组遍历	100	647	>1小时	102	2037
多步过程计算	272	848	>1小时	377	>1小时
大分组	39	155	2573	56	2493
大表关联分组	111	566	>1小时	178	2106
批量键值查询	15	>1小时	>1小时	15	>1小时

【注】SPL是润乾集算器采用的程序设计语言；SQL是关系数据库采用的程序设计语言

国产飞腾芯片上运行的润乾集算器可以超越Intel芯片上分布式数据库的性能



这大概就是骄傲吧！



授人以鱼不如授人以渔!



交出你的方法

某大型保险公司 - 存储过程优化



【案例背景】 定报价风险保费通过informix存储过程计算，运行1小时，效率低下影响使用；初始化车险新规车辆归并,找到上三年保单，数据量大（30天）时，运行将近2小时，急需优化

【数据规模】 保单表：0.35亿；保单明细表：1.23亿

【优化效果】

场景	优化前	优化后	提升倍数
定报价风险保费	3600秒	68秒	52.9倍
查找上三年保单	6672秒	1020秒	6.5倍

52.9 倍

【使用的集算器技术】

压缩列存：使用集算器组表的列式压缩存储，空间小，并且减少了参与计算的列数（约1/7）

有序存储：数据按照保险单号顺序存放，这样可以使使用有序关联提升计算效率

遍历复用：原来需要遍历多次完成多表关联，现在只需遍历一次即可完成多表关联

并行技术：集算器提供多线程并行计算，在多核环境下可以加速计算



我的天呐

【案例背景】 该金融机构需要维护的报表超过7500张，其中20%查询速度为分钟级，5%的查询速度为小时级，这5%非常慢的查询SQL代码行数有几百行，存储过程更是达到几千行，关联运算多。用通俗的语言描述报表慢的原因主要是数据量大，业务逻辑复杂。

【数据规模】 POS交易情况统计表运行时间为3700秒，4个SQL数据集，涉及6个数据库表的关联运算，其中3个表的数据量为千万级（1708万-4886万）。

【优化效果】 不改变硬件环境的条件下，报表查询速度从3700秒降到105秒。

【使用的集算器技术】

高效关联：将原来报表的顺序关联改为更高效的外键指针式关联方式

并行计算：将原有串行计算的4个SQL，改为4线程并行加速计算

不得了 不得了



35.2 倍

【案例背景】月财务结算时需要各省市分级按科目上报财务报表，但由于集团下属分子公司使用了不同的财务软件（SAP和用友），上报的数据存在财务账套无法完全匹配，不同科目需要关联替换等业务数据准备工作需要财务人员手工完成，使用Excel进行报表统计十分低效。

【数据规模】千万级

【优化效果】以费用/利润统计表为例，查询时间由原来317秒，优化到13秒

【使用的集算器技术】

高性能存储：集算器提供了私有二进制压缩存储，提供保存数据类型、泛型存储、任意分段等机制

热切换：集算器SPL采用解释执行机制，报表变更时直接修改SPL实施热部署



24.4 倍

某大型零售企业 - 库龄计算



【案例背景】 库龄计算是零售行业的常见场景，以往采用Oracle存储过程计算效率低下，长时间占用数据库资源得不到释放，影响其他业务

【数据规模】 3亿

【优化效果】 9分钟优化到10秒

【使用的集算器技术】

高性能存储： 集算器提供了私有二进制压缩存储，提供保存数据类型、泛型存储、任意分段等机制

外存游标： 支持数据量大时采用外存游标分批加载到内存进行计算，减少中间数据落地

并行计算： 集算器可实施单机多线程和多机分布式并行计算



再次我的天呐

54倍

想要了解更多 请联系我们



官方认证
如假包换



男神女神，或者~
看你运气



THANK YOU