

# 线下跑批



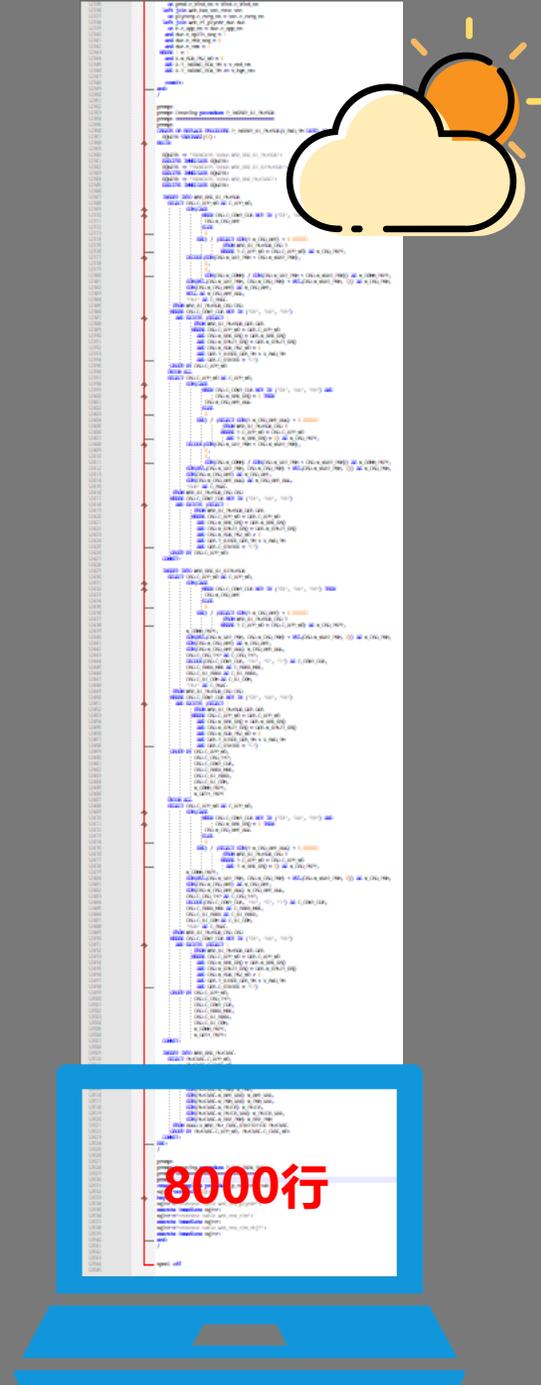
存储过程写了几千行，跑批非常慢！六个多小时才跑完，第二天上班业务受影响！



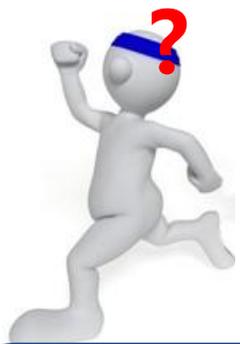
勉强跑得完，提心吊胆怕出错，根本来不及再跑一次！



今年还能对付，但数据量会越来越大，明年不知道还行不行！



# 线下跑批



MySQL



跑批需要外部源数据，到达时间不可控，窗口时间可能被缩小很多



IBM

DB2

时间消耗在数据库的出库入库上，但又只能用数据库跑批，更不敢在生产库上跑批

NO

Informix

业务人员期望能多跑几个月的数据，由于跑批太慢，只能跑几天的数据

# SPL Base为线下跑批提速



## 用SPL更优的算法代替存储过程

SPL支持并行查询

对数据跑批的join、遍历、并行提供优化算法  
计算性能优于存储过程



## 直接基于文件计算，省去出库入库时间

SPL数据文件主要面向数据分析计算

二进制压缩列存、倍增分段、任意并行等技术

充分利用硬件计算能力，提高数据存储、计算性能



跑批提速6.5倍  
扫码了解更多

# 在线查询

关联的数据表太多，报表要等半小时，喝杯茶还没出来，业务人员拍桌子



单个查询没问题，批量查询，并发多了就没戏



为了保证速度只能压制需求，业务部想查一年的，只给一个月

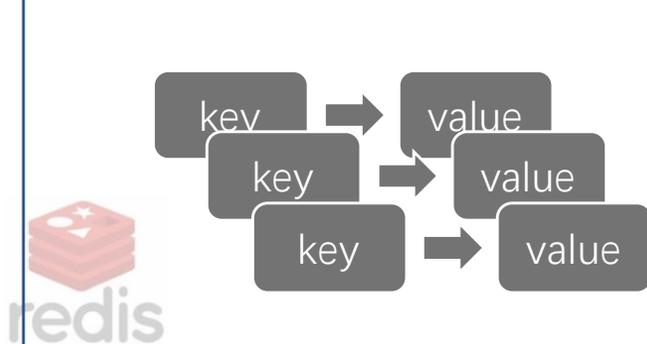


# 在线查询

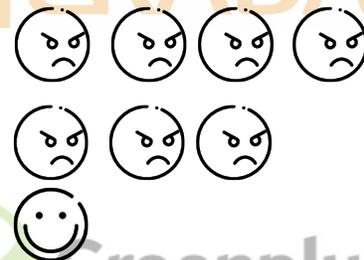


搜索引擎实现查询，性能可接受，但无法实现后续计算，大量写UDF，难以维护

APACHE  
HBASE



KV数据库做查询，单任务可以，批量查询难以支撑，同样无法处理复杂逻辑，UDF满天飞



业务要查询全量数据，压力集中到中央数据仓库上，并发查询撑不住

# SPL Base高性能在线计算方案



有序压缩存储，查询时高速定位、批量读取



自由列存压缩，有效减少磁盘IO时间



SPL代码难度低，容易实现复杂逻辑



刻意分布减少集群节点数据传播



预关联计算技术，将JOIN运算提速到常数时间



3亿条-60并发-单服务器  
查询0.6秒-扫码了解更多



# SPL Base多维分析方案

## 创新的预汇总技术

预汇总部分维度组合  
按更高维度层次预汇总

## 高性能存储和并行计算

有序压缩、自由列存  
多线程并行计算

## 算法优化

预关联加速JOIN  
深度优化复杂SQL



千万数据秒级响应  
扫码了解案例

# 软硬件环境



上线了内存数据库，系统怎么还这么慢？

HANA



内存加到1T了，还是出现内存溢出错误！



MPP节点数已经达到极限，压力还在增加怎么办？

VERTICA



Hadoop已经有几十个节点了，性能还是很差？



一体机这么昂贵，真的管用么？

# SPL Base: 高性能计算数据库



SPL Base应用  
场景扫码了解



## 高性能计算理念

软件无法改变硬件的速度  
高性能计算=高效算法设计+高效  
算法实现



## SPL创新计算体系

关系数据库的SQL象只有加法的算  
术体系，而SPL则发明了乘法



## SPL高性能计算机制

高效遍历、高效关联、高性能存储、  
并行计算