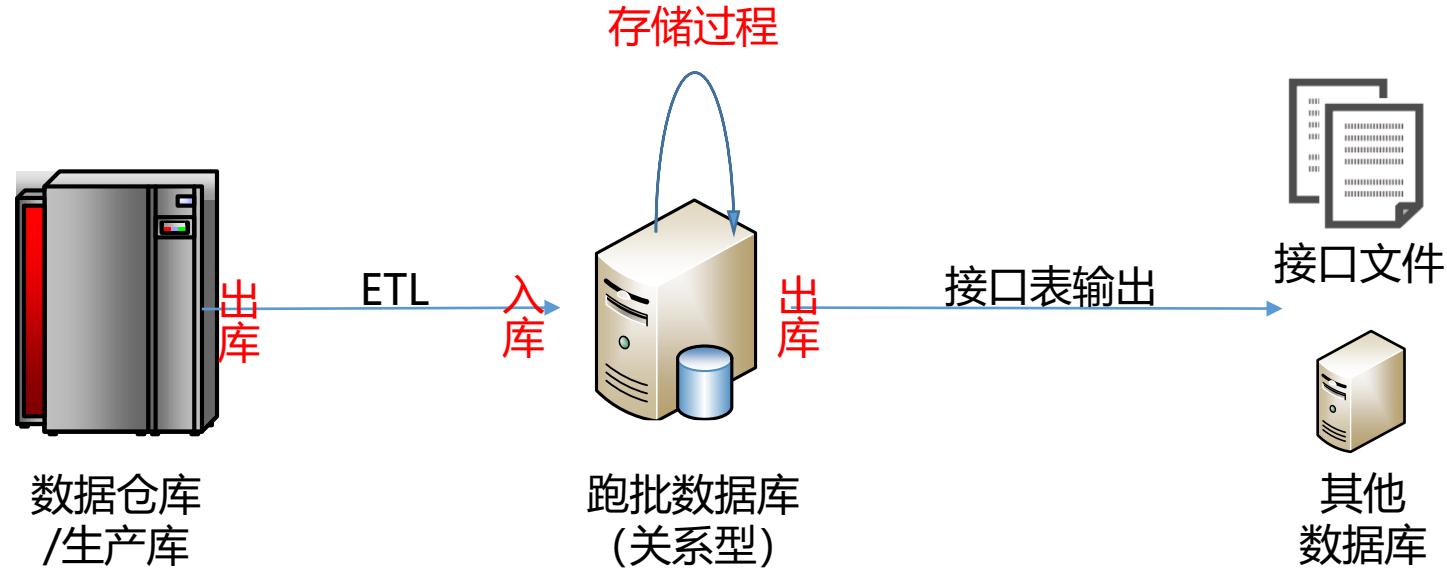


# 高性能跑批（批量计算任务）方案

集算器SPL Base应用方案

# 现状：数据库跑批速度太慢、耗时过长



 数据仓库/生产库

Oracle, Teradata、Hadoop等

 跑批数据库

Oracle、DB2、Mysql等

 接口文件/其他数据库

文本文件：其他应用  
BI数据库、其他应用数据库

# 原因：跑批问题主要原因分析



## 关系数据库出入库太慢

数据库的存储和计算能力封闭，数据进出要做的检查和处理过多，大量数据导入导出耗时非常长



## 存储过程性能差

SQL语法限制，很多高效算法不支持  
算法复杂度高，游标遍历计算很慢  
临时表写表速度慢

# 分析：数据库慢，为什么还要用跑批？

困 境 !!!

目前只有数据库有足够的计算能力，

即使慢也只能用数据库实现跑批计算。

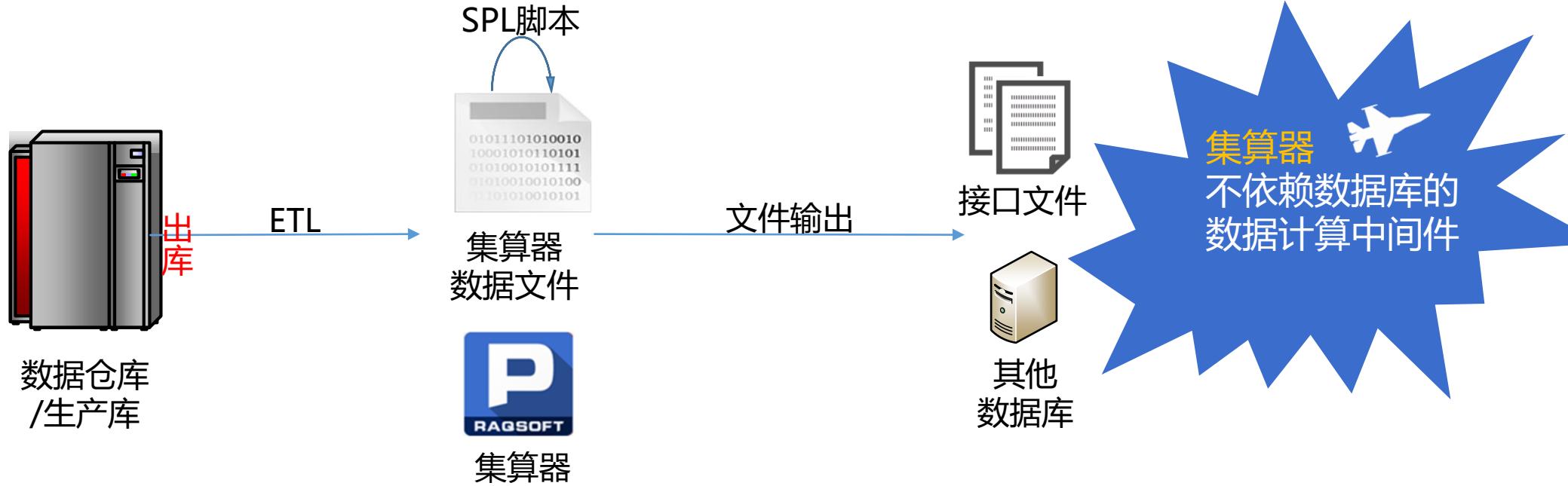
# 分析：如何破局？



不依赖数据库，提供直接文件系统计算、  
并实现更优算法的能力！



# 新方案：集算器新系统结构



ETL  
ETL调用  
集算器SPL脚本



集算器/数据文件  
用集算器数据文件  
实现高性能存储



接口文件/其他数据库

文本文件：其他应用  
BI数据库、其他应用数据库

# 解析：集算器提速的主要方法



## 用SPL更优的算法代替存储过程

集算器支持并行查询

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

集算器数据文件主要面向数据分析计算

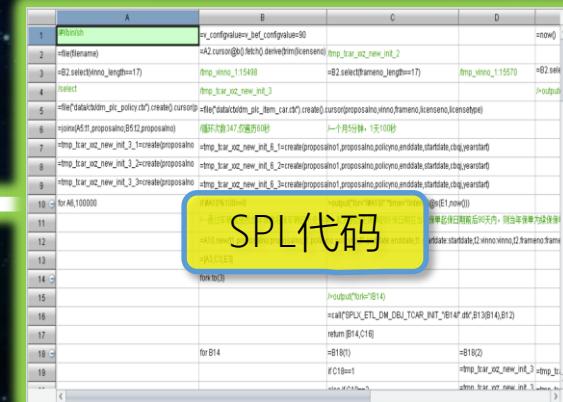
二进制、压缩、列存、倍增分段、任意并行等技术

充分利用硬件计算能力，提高数据存储、计算性能

# 自动转换：集算器自动将SQL自动转换成SPL



```
程序信息：初始化车险新机动车辆并找到上三年保单！  
工作名：ETL_DM_DBJ_TCAR_INIT  
用户名：OP  
密文名：V1_01  
目标表：DM_DBJ_TCAR_XXZ_NEW!  
  
DEFINE v_error_code integer;  
DEFINE v_errormsg varchar;  
define v_loaddate date;  
  
-----  
程序信息：初始化车险新机动车辆(3年)并找到上三年保单！  
工作名：ETL_DM_DBJ_TCAR_INIT  
用户名：OP  
密文名：V1_01  
目标表：DM_DBJ_TCAR_XXZ_NEW!  
  
DEFINE v_error_code integer;  
DEFINE v_errormsg varchar;  
define v_loaddate date;  
  
-----  
程序信息：初始化车险新机动车辆(3年)并找到上三年保单！  
工作名：ETL_DM_DBJ_TCAR_INIT  
用户名：OP  
密文名：V1_01  
目标表：DM_DBJ_TCAR_XXZ_NEW!  
  
DEFINE v_error_code integer;  
DEFINE v_errormsg varchar;  
define v_loaddate date;
```



```
A B C D  
1 printish =>configvalues_tet_configvalue=90  
2 =>dbframe  
3 =>B2.cursor[0].length0=denebIm(cursorIm);tmp_xar_xz_new_int_2  
4 =>B2.selectframe[0].length0=17) tmp_xeno_115570 =>B2.read  
5 =>B2.selectframe  
6 =>B2.cursor[0].length0=17) tmp_xar_xz_new_int_3  
7 =>B2.cursor[0].length0=17) tmp_xar_xz_new_int_3->createproposal  
8 =>B2.cursor[0].length0=17) tmp_xar_xz_new_int_3->createproposal  
9 =>B2.cursor[0].length0=17) tmp_xar_xz_new_int_3->createproposal  
10 =>B2.cursor[0].length0=17) tmp_xar_xz_new_int_3->createproposal  
11  
12  
13  
14 for B14  
15  
16  
17  
18 for B14  
19  
C
```



程序员根据计算和数据特征，对SQL做标记



集算器将标记后的SQL语句自动转换成SPL



集算器自动完成90%转换工作，减少人力成本

# 成功案例：某大型保险公司跑批项目

## 优化需求

跑批采用informix存储过程计算，运行时间过长，急需优化

## 数据规模

保单表：0.35亿；保单明细表：1.23亿

集算器提速 **52.9** 倍

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

# 成功案例：某银行跑批项目

## 优化需求

对公贷款协议明细包含48个SQL，耗时1.5小时，成为跑批任务的瓶颈，DB2优化困难

## 代码量规模

SQL: 3300行；集算器SPL: 500格

集算器提速 **8 . 5** 倍

场景	优化前	优化后	提升倍数
对公贷款协议明细	85分钟	10分钟	<b>8.5倍</b>

- 结束 -

The End

想要了解更多 请联系我们



润乾官微  
更多精彩



本文作者  
加微信吧

THANK YOU