

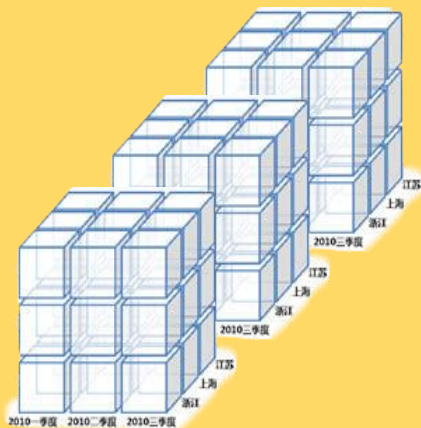
高性能多维分析方案

集算器SPL Base应用场景



加作者，了解更多

多维分析OLAP服务器的常见选择



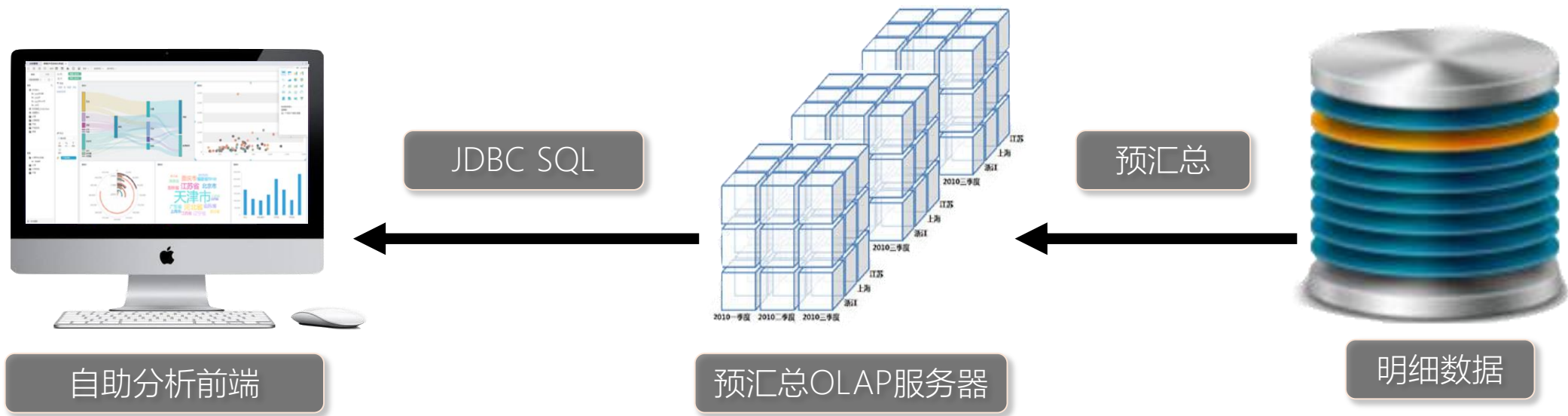
预汇总方案

预先计算统计结果 (cube)
复杂度从 $O(n)$ 变为 $O(1)$

普通数据库方案

用关系数据库做后台
OLAP拖拽SQL查询结果

预汇总OLAP服务器结构



预先计算统计结果

复杂度从 $O(n)$ 变为 $O(1)$

秒级返回结果

300米

250米



几十个维度全量预
汇总空间**几万TB**

1万TB需要存入
1万个1TB硬盘

1万个1TB硬盘
堆在一起**250米高**

预汇总OLAP服务器问题：占用存储空间巨大

扫码阅读



《多维分析预汇总的存储容量》

预汇总OLAP服务器问题：功能盲区

扫码阅读



《多维分析预汇总的功能盲区》

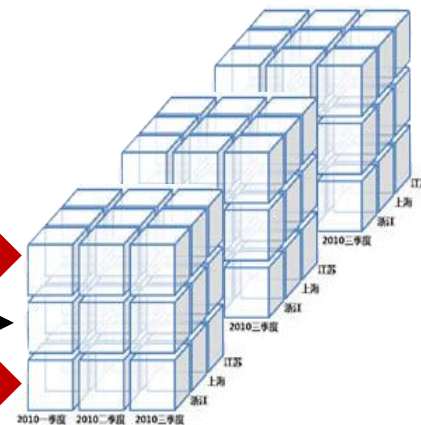
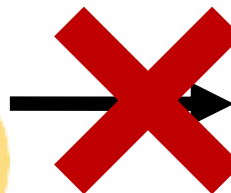


自助分析前端

JDBC
SQL

Countdistinct
组合聚合
中位数 月最大销售额
非常规聚合 唯一计数 方差
条件测度：金额大于100元的订单金额合计
时间段统计：5月8日到6月12日的销售额合计
针对维表的过滤条件
地区人员收入中位数的均值

多种常见查询情况
无法预汇总优化



预汇总OLAP服务器



有条件测度
针对维表的条件



组合聚合
非常规聚合



时间段统计

普通数据库方案问题：性能差，用户等待时间长



行式存储
大宽表查询性能低



压缩率低
大数据量读取慢



无预关联机制
大量JOIN计算慢

突破：集算器用作OLAP服务器



解决预汇总OLAP
引擎存在问题



OLAP提速



硬件要求低

集算器：采用多线程并行机制



基于明细数据实现OLAP查询

明细数据不需要按照各种维度预汇总，不会额外产生大量的cube数据。



多线程并行加速数据文件访问

对于多CPU（多核）的服务器，多线程并行计算可以有效提高明细数据的访问和计算速度。

集算器：压缩列式存储

1
2
3

列式存储

数据文件字段很多，只有少部分参与OLAP计算。
采用列存方式，有效提高数据读取的速度。



压缩存储

采用十倍左右的压缩比，数据文件占用的硬盘空间变小，
读取速度也大大提高。



集算器：预关联加速JOIN

客户	
客户ID	主键
客户名称	
地址	
...	

订单	
订单ID	主键
客户ID	
订购日期	
...	



客户	
客户ID	主键
客户名称	
地址	
...	

订单	
订单ID	主键
客户ID	
订购日期	
...	

指针



数据库JOIN:

查询时临时计算关联

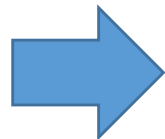


集算器JOIN

预先计算好关联
以各种方式存放在内存或硬盘中
查询时无需计算关联

集算器：针对复杂SQL的优化

```
Select F1, F2, F3,  
  (select FF1 from TABLE1 WHERE...) AS F4,  
  (select min(FF1) from TABLE2 GROUP BY...) AS F5,  
From  
  (select FFF1 from TABLE3 WHERE...) T1  
Left join  
  (select FFFF1 from TABLE4 WHERE...) T2  
On T1.FFF2=T2.FFFF2  
WHERE  
T1.FFF2 in (select min(FFFF1) from TABLE5 GROUP  
BY...)
```



将子查询转成JOIN计算

剥开多层嵌套SQL转成单层计算

...



OLAP前端拖拽的SQL
多层嵌套并包含子查询



OLAP引擎
只能优化简单SQL



集算器对这类SQL
做了深度优化

集算器：提供创新的预汇总能力

扫码阅读



《多维分析预汇总的方案探讨》



部分预汇总，有效权衡
空间与时间的矛盾



对时间段统计给出
切实可行的预汇总手段

- 结束 -

The End

感谢关注，加作者微信了解细节



润乾官微
更多精彩



本文作者
加微信吧

THANK YOU