



全国产高性能计算数据库

润乾软件出品

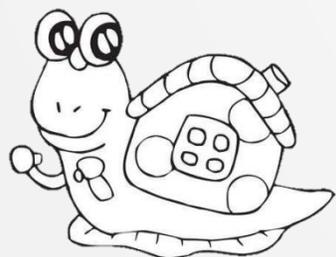


全国产技术的尴尬



- 90%以上的高端芯片依赖进口
- 国产芯片单机没人能跑得过intel
- 经常要靠增加硬件数量来获得速度上的提升

- 目前都是欧美企业的天下
- 国内的基本都是拿开源改
- 只有靠大集群，加大成本和管理复杂度才可以提升速度



慢芯片 + 慢数据库 → 慢上加慢

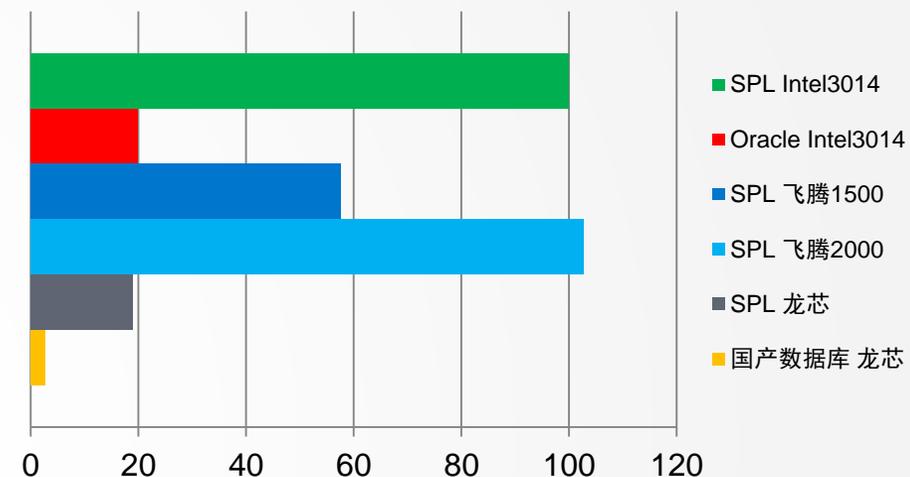
大多只能应用于边缘业务（OA，邮件…）

一份性能测试报告

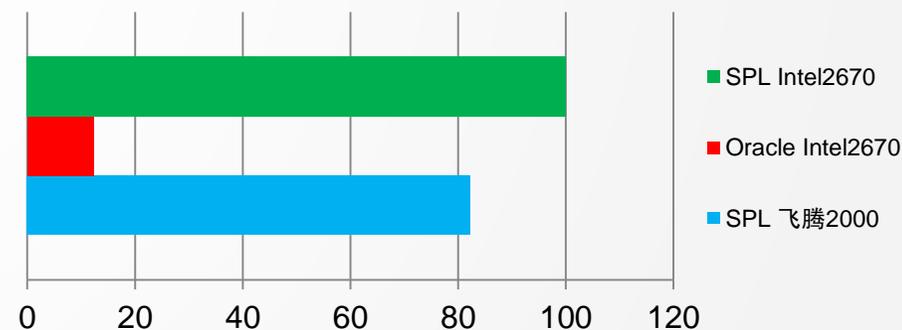


TPCH (单位: 秒)	100G						200G		
	Intel 12核		飞腾16核	飞腾64核	龙芯8核		Intel 16核		飞腾64核
	SPL	Oracle	SPL	SPL	SPL	国产数据库	SPL	Oracle	SPL
1	38	131	62	19	275	507	40	325	36
2	4	27	8	6	18	247	8	73	13
3	16	222	33	22	97	4451	23	582	35
4	12	207	27	18	89	1790	21	454	43
5	20	225	36	24	72	1761	25	463	45
6	9	135	22	6	60	757	11	352	12
7	16	184	32	20	91	700	22	496	30
8	29	192	46	48	93	1611	29	485	80
9	68	234	125	65	517	1066	85	636	135
10	23	215	35	22	99	1634	34	493	42
11	5	33	12	6	29	165	9	63	11
12	25	184	72	38	173	647	52	464	55
13	57	37	114	85	335	2209	135	103	135
14	22	157	65	12	142	500	65	368	38
15	18	155	60	26	103	506	61	358	46
16	10	13	19	12	53	105	14	71	22
17	21	165	48	9	100	963	40	349	19
18	21	344	35	13	163	2382	25	966	26
19	23	154	65	12	137	518	60	345	29
20	18	175	57	11	110	594	55	442	20
21	233	326	222	190	901	3349	191	790	398
22	22	48	37	27	99	139	30	99	49
合计	710	3563	1232	691	3756	26601	1035	8777	1319

TPCH 100G



TPCH 200G



Intel3014 1.7G/12核/64G内存

飞腾FT1500/16核/32G内存

龙芯/8核/64G内存

Intel2670 2.6G/16核/128G内存

飞腾FT2000/64核/256G内存

TPCH是什么？



TPCH

- **数据库性能测试的国际标准，具有充分的权威性和广泛的被认可度，以免厂商选择对自己有利的测试用例**
- **主要评价数据库执行查询分析计算的性能**
- **22个题，涉及内容全面，规避性能测试的片面性，可反映数据库的整体综合性能**



SPL是什么？



创新的数据库计算语言

— SPL (Structured Process Language)

100% 理论创新，独创离散数据集模型

已经实现成可商用软件产品

100% 产品创新，每一行代码均自主研发，安全可控

SPL的优势来源于创新代数体系，关系数据库的SQL就像只有加法的简单算术体系，而SPL则像发明了乘法后的高级运算体系，所以能获得更高性能！

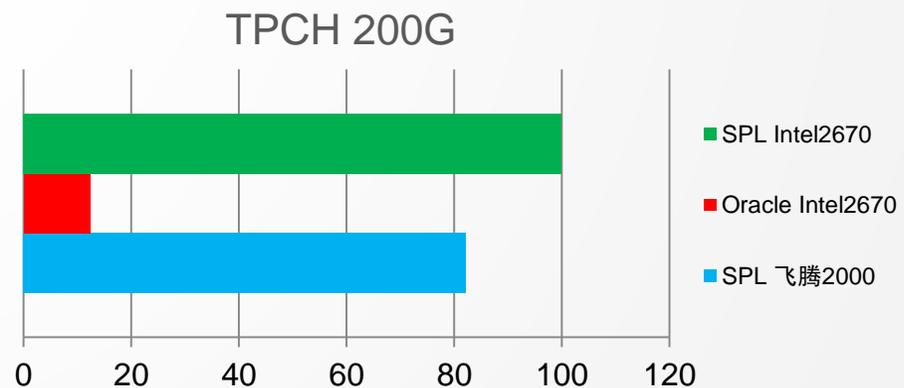
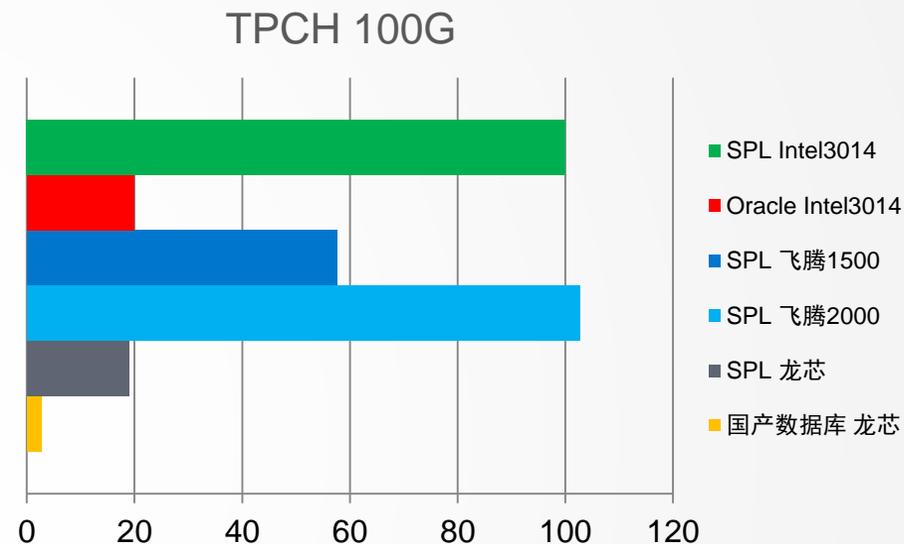




这份测试数据说明了什么

在SPL的支持下:

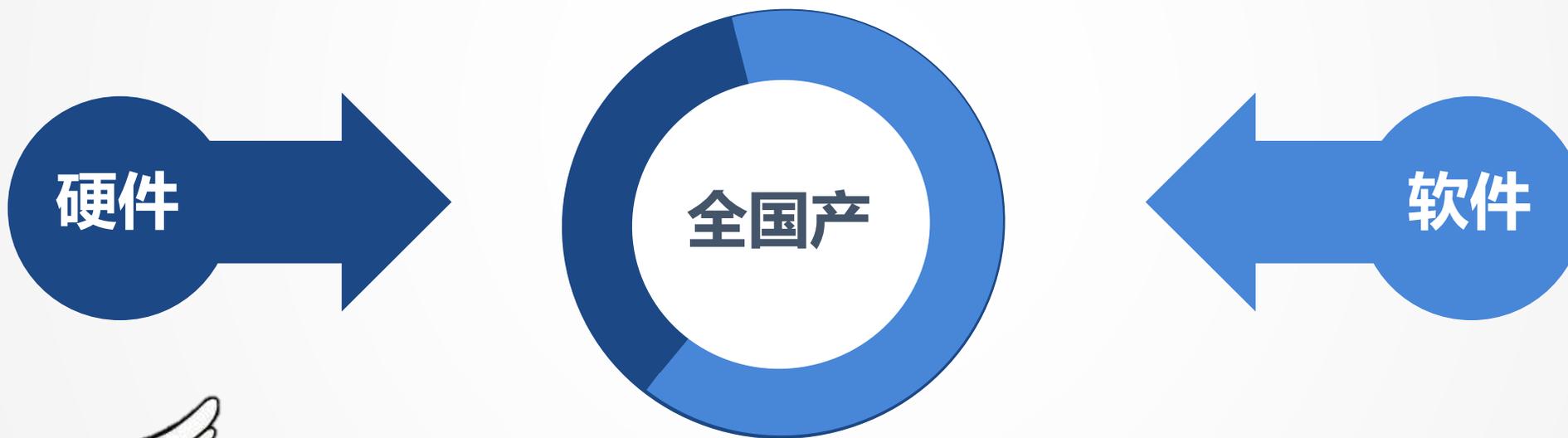
- 飞腾芯片跑出超越Intel上Oracle数倍（**2-5倍**）的性能
- 龙芯也能与Intel上Oracle的性能基本相当（**接近95%**）





这又意味着什么？

- 软件弥补硬件，全国产技术以后不必只在边缘业务中打转，不必依靠大集群取胜，同样规模的硬件就可以让关键支撑业务跑起来！



I can fly, you see me in the sky!



高性能计算的理念



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, Hadoop, 都会限制算法实现



高性能来自于创新计算体系

【类比】 计算 $1+2+3+\dots+100=?$

普通人这么算

$1+2=3$
 $3+3=6$
 $6+4=10$
 $10+5=15$
 $15+6=21$
 $21+7=28$
...

高斯这么算

$1+100=101$
 $2+99=101$
...
一共有50个101
 $50*101=5050$

高斯很聪明，想到了高效的算法，但更关键的是：那时人们已经发明了**乘法**！

思考题： 1亿行数据取前10名在SQL下会怎么做？

SQL理论上就是大排序后取前10个，效率很低，程序员都知道有不必要大排序的办法实现这个运算，却无法用SQL表达，只能用指望数据库引擎自动优化，但复杂情况时数据库并不会优化！

关系数据库的SQL就象只有**加法**的算术体系，而SPL则发明了**乘法**！

SPL中还有更多**乘法**（高性能计算和存储库），人人都能成为**高斯**（快速实现高性能算法）。

部分高性能计算机制



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



高性能算法举例

聚合理解

	A	
1	<code>=file("data.ctx").create().cursor()</code>	
2	<code>=A1.groups(:top(10,amount))</code>	金额在前10名的订单
3	<code>=A1.groups(area;top(10,amount))</code>	每个地区金额在前10名的订单

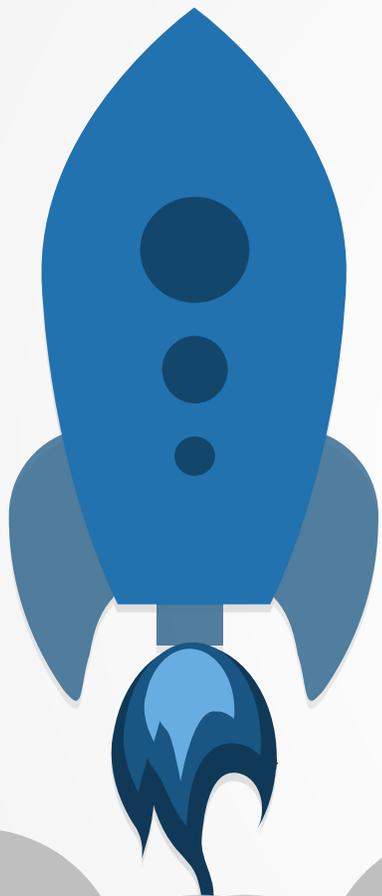
高复杂度的排序转换为低复杂度的聚合

遍历复用

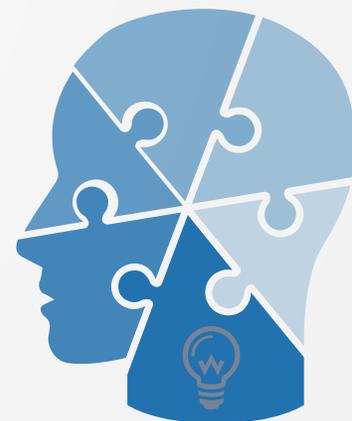
	A	
1	<code>=file("order.ctx").create().cursor()</code>	准备遍历
2	<code>=channel(A1).groups(product;count(1):N)</code>	配置复用计算
3	<code>=A1.groups(area;sum(amount):amount)</code>	遍历, 并获得分组结果
4	<code>=A2.result()</code>	取出复用运算的结果

一次遍历可返回多个结果集

我们还能跑得更快



- 目前是Java写的（为了可移植性和开发成本），如果用C语言改写，...
- 对操作系统内核做定制改造，设计更适合的存储机制



全国产高性能计算数据库架构





1. 新型算法引擎 – SPL语法



离散化与集合性的结合



内存、外存游标、集群运算



分步计算与完善的调试功能



300多个结构化计算函数



循环，分支适应过程式计算



特色网格式代码层次清晰

	A	B	C
1	<code>=file("../demo\\zh\\bt\\Promotion.bt").import@t()</code>		
2	<code>=file("../demo\\zh\\bt\\SalesRecord.bt").import@t()</code>		
3	<code>=create(promo_name,best_sale)</code>		
4	<code>for A1</code>		
5		<code>=A2.select(sale_date>=A4.start_date && sale_date<=A4.end_date)</code>	
6		<code>=B5.group(clerk_name;~.sum(sale_amt):total_amt)</code>	
7		<code>=B6.maxp(total_amt)</code>	
8		<code>>A3.insert(0,A4.promo_name,B7.clerk_name)</code>	
9	<code>=A3</code>		



2.数据组织

高性能数据存储

私有数据存储格式，集文件、组表

文件系统存储形式

支持以树状目录方式按业务分类存储数据

集文件

倍增分段方式支持任意数量并行
自有高效压缩编码（减少空间；CPU占用少；安全）
泛型存储，允许集合数据



组表

行列混合存储
有序存储提高压缩率和定位性能
高效智能索引
倍增分段方式支持任意数量并行
主子表合一减少存储与关联
排号键值实现高效定位关联





3. 多源混合

开放灵活的计算体系

支持各类数据源；

外部数据无需导入，直接与系统内数据混合计算

多源混合汇总计算

内



外





4. 并行计算

数据容错和计算容错

提供内外存两种数据容错机制，外存冗余式容错，内存备胎式容错



外存冗余式容错



内存备胎式容错

支持计算容错，节点故障时自动将该节点计算任务迁移到其他节点继续完成

可控数据分布

用户可根据数据和计算任务的特点灵活定制数据分布及冗余方案，有效减少节点间数据传输量，以获得更高性能

无中心架构，避免单点失效

集群没有永久的中心主控节点，程序员用代码控制参与计算的节点

负载均衡能力

根据每个节点空闲程度（线程数量）决定是否分配任务，实现负担和资源的有效平衡



5.应用接口

SQL+

在传统SQL的基础上，额外加入新的标记语法，用来描述数据特征等关键信息，从而弥补传统SQL描述能力不足的缺陷，大幅提高计算性能，并有效降低学习成本

SPL

面向多步骤的复杂计算，采用过程式编程，借助丰富的集合运算类库、离散性、深度集合化等特性，极大改善传统SQL或JAVA在实现复杂计算时难实现、难维护、性能低等情况



6.可集成计算引擎



计算引擎可嵌入
应用增强应用端
数据处理能力

利于结构分层

降低耦合性

计算

存储

国产硬件平台



7.数据固化

数据从热到冷转换固化

使用SPL实现

语言级控制粒度更加精细

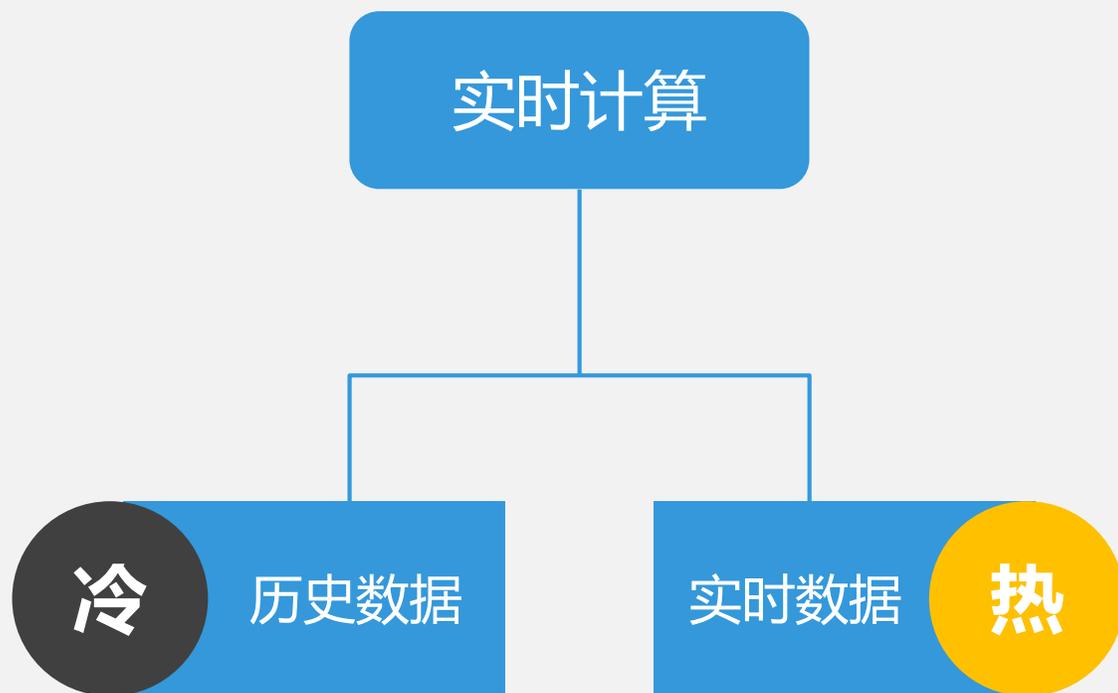
开发人员使用更加灵活

非常适应复杂数据处理过程





8.支持实时计算



支持ETL后的历史数据和当期实时数据进行混合计算，实现前端实时T+0查询

9.安全可控



全部国产 · 完全自主 · 安全可控

国产高性能计算数据库优势总结



完全自主可控



理论模型/接口/计算/数据库
/存储/硬件全部自主研发

性能卓越



万亿级数据处理能力，比传统
方案平均提高1至2个数量级

5大

优势总结

弥补国产芯片不足



模型创新提供高效算法
算法弥补硬件计算性能不足

计算体系开放灵活



多源混合计算
点对点计算前置

提升开发效率



语法过程化，符合自然思维
丰富类库和函数

基于全国产高性能计算数据库的应用方案

全国产商业智能解决方案

什么是商业智能？



商业智能（Business Intelligence，简称：BI），指用现代数据仓库技术、线上分析处理技术、数据挖掘和数据展现技术进行数据分析以实现商业价值

广泛性

商业智能是数据分析的第一步，**所有行业**只要积累了一定数据都会用到

必要性

商业智能已经是**关键性**支撑业务，在企业机构的决策支持过程中**不可或缺**

强计算性

商业智能涉及数据量较大，计算强度大，当前成熟方案几乎全部要**基于国外芯片**运行

为什么缺乏基于国产芯片的商业智能解决方案？



国产商业智能 (BI)

低性能国产数据库

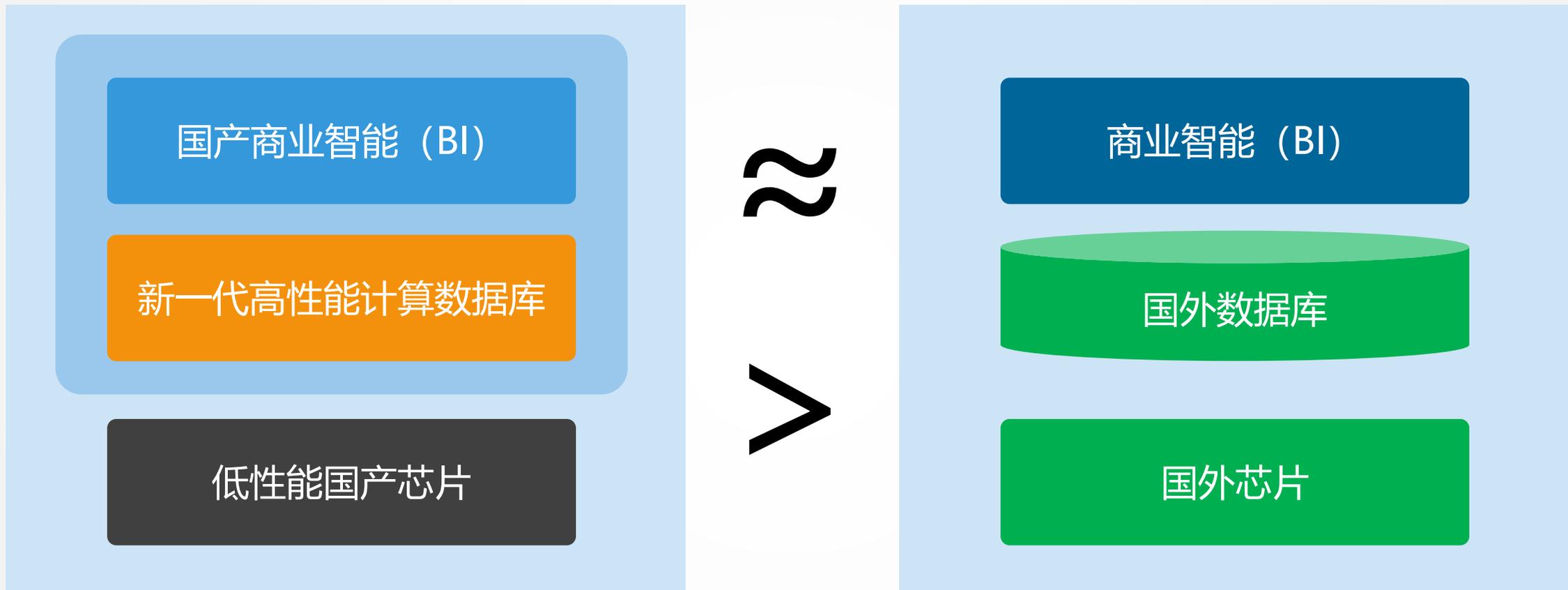
低性能国产芯片



慢到没法用



现在有了新武器



通过全国产高性能计算数据库，**弥补国产芯片的性能不足**。在确保安全可控的前提下，使全国产商业智能产品的整体性能与基于国外软硬件的商业智能产品相当甚至超越

全国产商业智能系统架构



酷炫的大屏展现



A data table with multiple columns and rows, likely representing a list of items or transactions. The table has a light blue header and a white body with alternating row colors.



A data table with multiple columns and rows, similar to the one above. The table has a light blue header and a white body with alternating row colors.



A data table with multiple columns and rows, similar to the one above. The table has a light blue header and a white body with alternating row colors.



A data table with multiple columns and rows, similar to the one above. The table has a light blue header and a white body with alternating row colors.



拖拖拽拽做分析



R 分析
保存
打开
数据文件
横向平铺
纵向平铺
重叠显示
数据集

订单分析

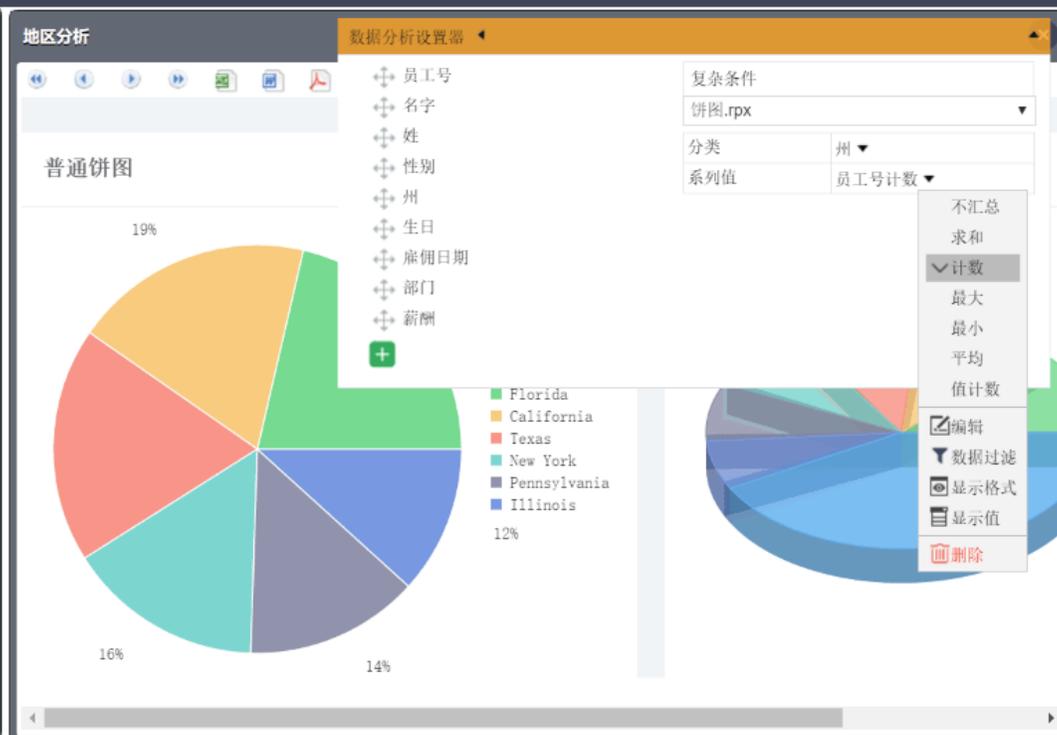
商品类别: 商品		饮料							
省:市		苹果汁	牛奶	汽水	啤酒	蜜桃汁	绿茶	运动饮料	柳橙汁
辽宁	大连	810.00	0.00	90.00	280.00	918.00	12939.60	1080.00	322.00
	总计	810.00	0.00	90.00	280.00	918.00	12939.60	1080.00	322.00
上海	上海	0.00	152.00	0.00	0.00				
	总计	0.00	152.00	0.00	0.00				
山东	青岛	0.00	0.00	72.00	329.00				
	总计	0.00	0.00	72.00	329.00				
江苏	南京	1660.50	2303.75	106.00	0.00				
	常州	54.00	874.00	112.00	1095.00				
总计	1714.50	3177.75	218.00	1095.00					
江西	南昌	0.00	0.00	266.00	0.00				
	总计	0.00	0.00	266.00	0.00				
浙江	温州	324.00	649.80	90.00	0.00				
	总计	324.00	649.80	90.00	0.00				
北京	北京	324.00	0.00	140.00	0.00				
	总计	324.00	0.00	140.00	0.00				

数据分析设置器

- 客户
- 雇员
 - 雇员姓名
 - 职务
 - 出生日期
 - 雇佣日期
 - 市

商品类别
商品

R 分析
保存
打开
数据文件
横向平铺
纵向平铺
重叠显示
数据集
数据集[员工分析]已加载完毕, 共 500 行数据



独创的强关联分析



DQL分组分析 x +

localhost:6868/demo/raqsoft/guide/jsp/olap.jsp#

分析 保存 打开 数据文件 横向平铺 纵向平铺 重叠显示 数据集

报表名称

省:客户	订单数量	订单金额求和	订单数量-总计	订单金额求和-总计	
辽宁	国皓	26	¥21821.55	26	¥21821.55
	威航货运有限公司	11	¥1780.00	11	¥1780.00
	三捷实业	2	¥126.00	2	¥126.00
	五金机械	9	¥3262.00	9	¥3262.00
	华科	32	¥32690.70	32	¥32690.70
	总计	80	¥59680.25	80	¥59680.25
上海	业兴	19	¥8624.60	19	¥8624.60
	总计	19	¥8624.60	19	¥8624.60
山东	悦海	17	¥9650.90	17	¥9650.90
	东帝望	14	¥5386.10	14	¥5386.10
	总计	31	¥15037.00	31	¥15037.00
通恒机械	52	¥26033.60	52	¥26033.60	
迈策船舶	9	¥1674.00	9	¥1674.00	
五洲信托	45	¥30115.50	45	¥30115.50	
国银贸易	6	¥3153.00	6	¥3153.00	

数据分析设置器

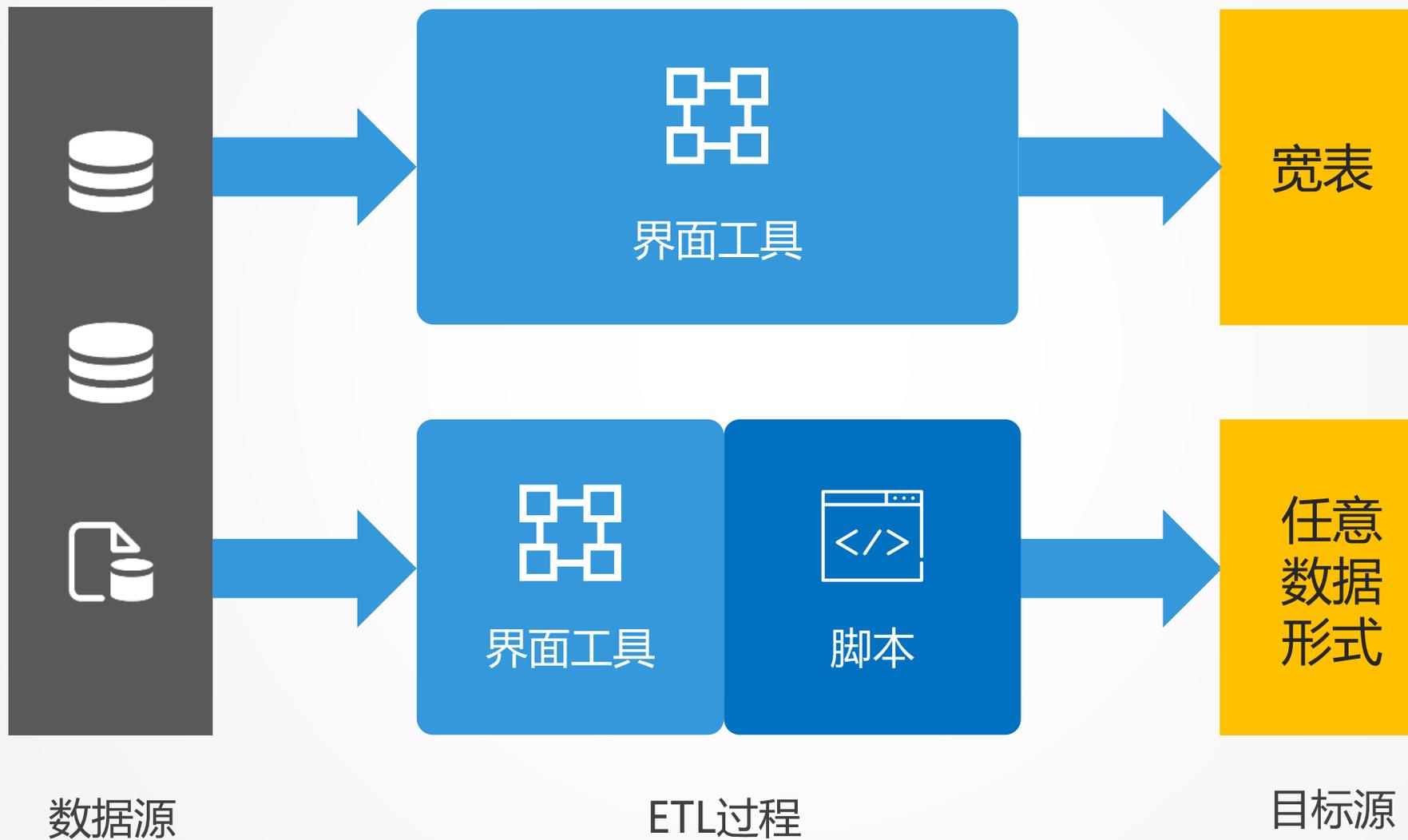
报表列表 添加报表

折扣
明细金额
订单
客户
市
省
名称
人口
VIP级别
VIP折扣率
VIP起始时间
VIP结束时间
雇员
签单日期
发货日期
入库单

来源订单表
指标标题
订单数量
订单金额求和
来源订单明细表
来源地区表
省
客户
来源客户表

多表（多层维表）关联查询

自助式ETL工具

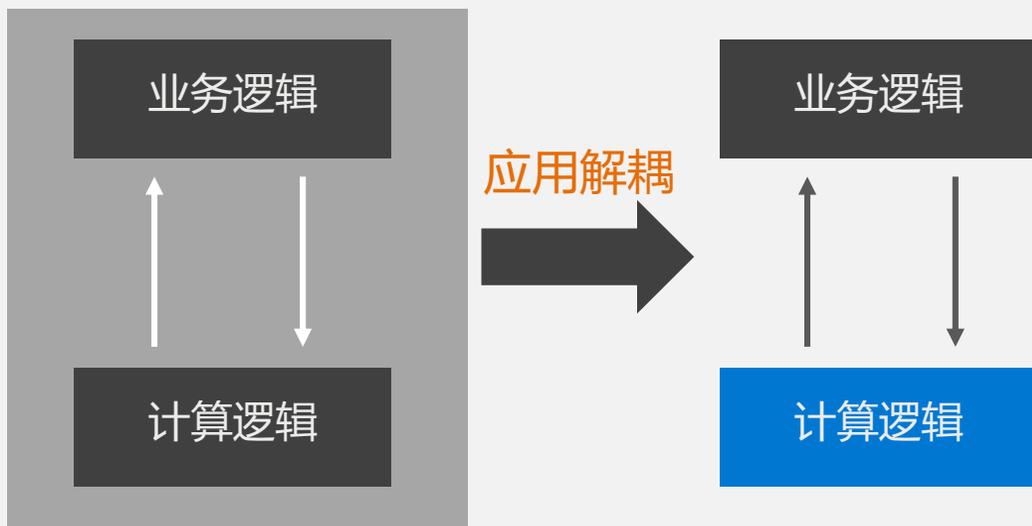




计算引擎带来低耦合与热切换

低耦合

计算逻辑脚本单独维护，方便模块化
将计算与业务解耦、计算与数据源解耦，易于应用扩展



热切换

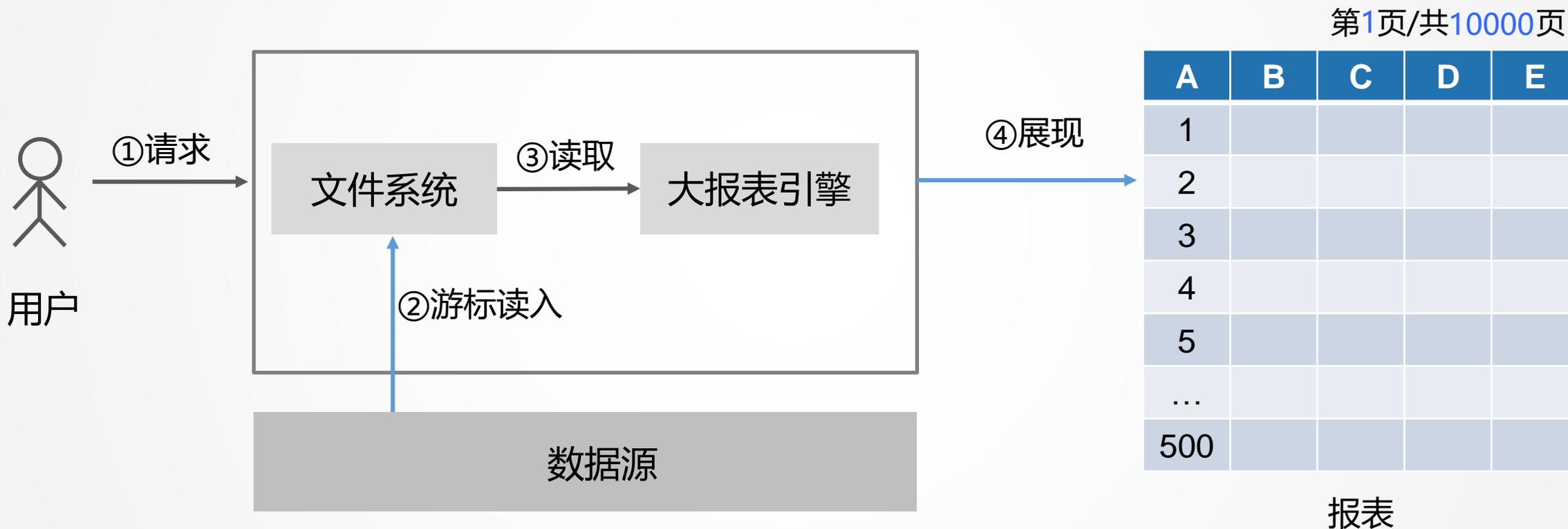
算法解释执行
支持不停机热切换





百万级大报表查询秒级响应

海量清单数据查询秒级响应，海量数据导出Excel和打印



除清单查询，还支持海量大数据分组查询

全面支持移动端



中国联通 73B/s 43% 下午4:18

WEB预览报表

第1页/共1页

中科软 www.sinosoft.com

规模: 500人 项目: 15 (合作) 7 (潜)

联系人: 李瑶瑶 电话: 155 1234 5678

邮箱: liyaoyao@126.com

产品分类构成

过去三年销售金额(万元)

ID	产品	订单时间	套数	销售代表
1	V4	2012-03-05	400	王磊

中国联通 97B/s 77% 傍晚6:24

http://192.168.0.141:6868/demo/raqsoft/

应用

手机销售各区占比 & 各月对比

地区	占比
华北	~45%
华东	~35%
华南	~20%

中国联通 48B/s 44% 下午4:20

WEB预览报表

第1页/共1页

选择地区: 华东 查询

选择城市: 下拉联动

- 上海
- 南京
- 南昌
- 常州
- 济南
- 温州
- 烟台
- 青岛

城市	运货	订单金额
北京	急速快递	¥2974.76
	统一包裹	¥10117.76
	联邦货运	2 ¥1455.80

完整的平台管理功能



- 润乾报表DEMO
- Dashboard
- 基本报表
- 高级报表
- 交互报表
- 集算器相关
- 统计图形
- 金融行业
- 交通运输
- 能源行业
- 更多报表
- 数据分析
 - DQL样例
 - SQL分析
 - 填报分析

节点属性

节点类型

DQL分组分析

节点名称

DQL样例

分析类型:

dql表

填报文件

数据源+查询语句

数据文件

dfx文件

dfx脚本

需要分析的文件:

analyseFiles/DQL

提交

浏览权限

机构浏览权限

用户浏览权限

访客

金商百货批发公司

金商百货批发公司-管理层

金商百货批发公司-销售部

raq_visitor

王伟

郑建杰

张颖

金融行业落地



采用高性能计算数据库的商业智能解决方案，可实现实时联机分析，随时调整查看口径，支持超多标签和大数据量



动态实时选择标签

