

创新计算助力政务大数据应用飞翔

北京润乾 黄曜令



集算器助力政务大数据应用飞翔

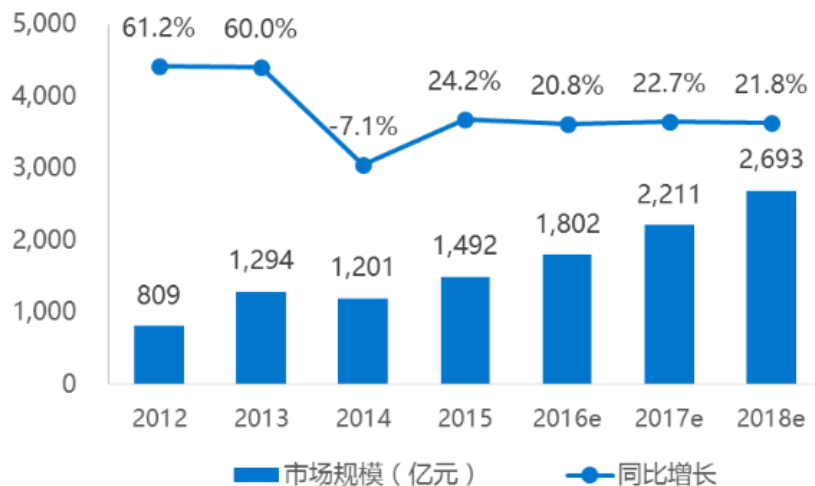
大数据市场方兴未艾

数据成为核心生产资料

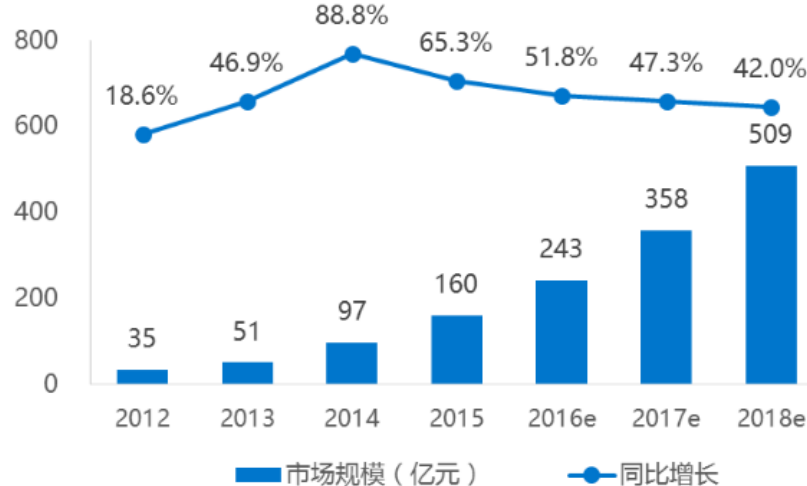
大数据市场潜力巨大

政府成为推动大数据发展的主力

2011-2018年全球大数据市场规模

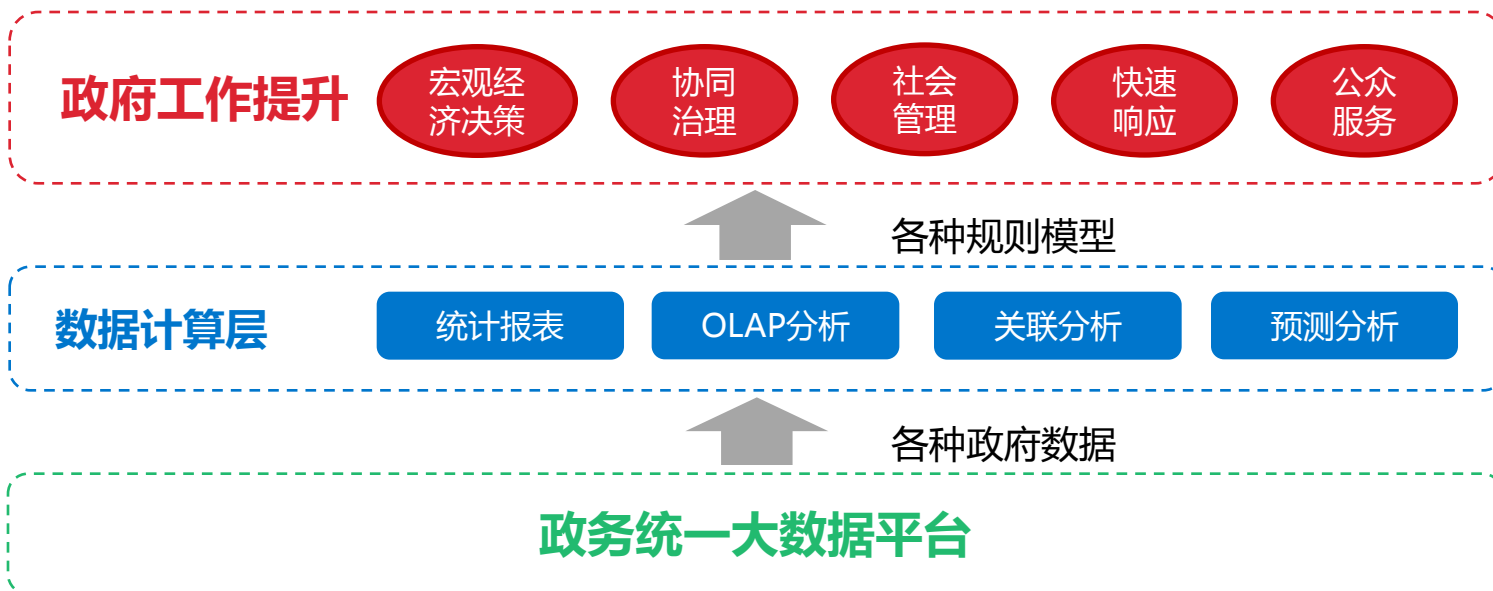


2011-2018年我国大数据市场规模



注：资料来源36氪研究院《大数据服务行业研究报告》

政府大数据的价值



三个难题

在政务大数据共享的推进过程中，有三个难题。

不愿共享 不敢共享

不能共享

政务大数据平台架构



政务大数据中心——难点

一、制定标准规范

二、构建数据资源体系

业务信息目录、共享数据目录、服务目录、文件目录、物联感知目录

三、建立统一平台

- 1、数据治理：数据清洗转换，标准化，规范化
- 2、信息资源目录：数据检索
- 3、数据交换：数据共享交换
- 4、综合运维：系统维护管理监控
- 5、数据管控：数据审核、发布、订阅
- 6、数据服务：数据查询、分析、挖掘、可视化展示
- 7、数据应用：互联网+政务服务、智慧城市、物联网综合平台

四、形成数据模板与业务模型

- 1、几十个政府部门的共享数据模板
- 2、上千甚至更多的数据库表，上万甚至更多的共享数据字段模板
- 3、人口、法人、空间地理、宏观经济、城市资源等多个基础数据的数据模板和数据元
- 4、物联感知数据、信用数据、开放数据、公共服务数据、社会治理网格化数据、电子证照等多个主题数据库的数据模板
- 5、人口、法人、空间地理、宏观经济、物联网业务分析模型

五、满足性能要求

- 1、成百上千个常用数据需求处理作业
- 2、数据量巨大、数据源众多

润乾创新大数据计算引擎 - 集算器是.....

自主可控

自主研发、完全自主知识产权的新一代数据计算引擎

计算体系

不依赖关系型数据库的、完备的计算体系

理论基础

离散数据集模型

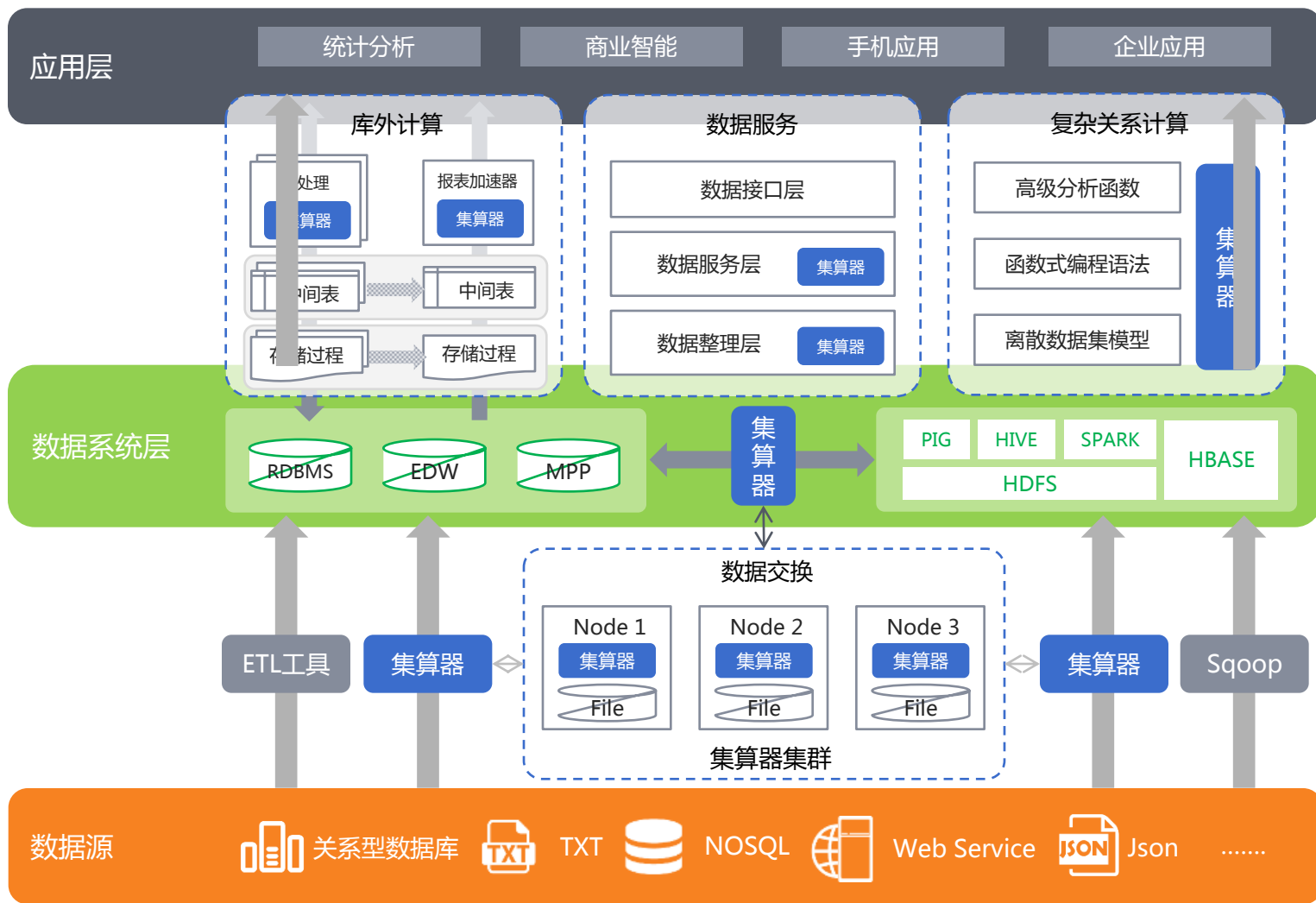
自有语法体系

集合SQL和其他高级语言的优点，让描述计算更简单、更贴近人的思路

计算效率更高

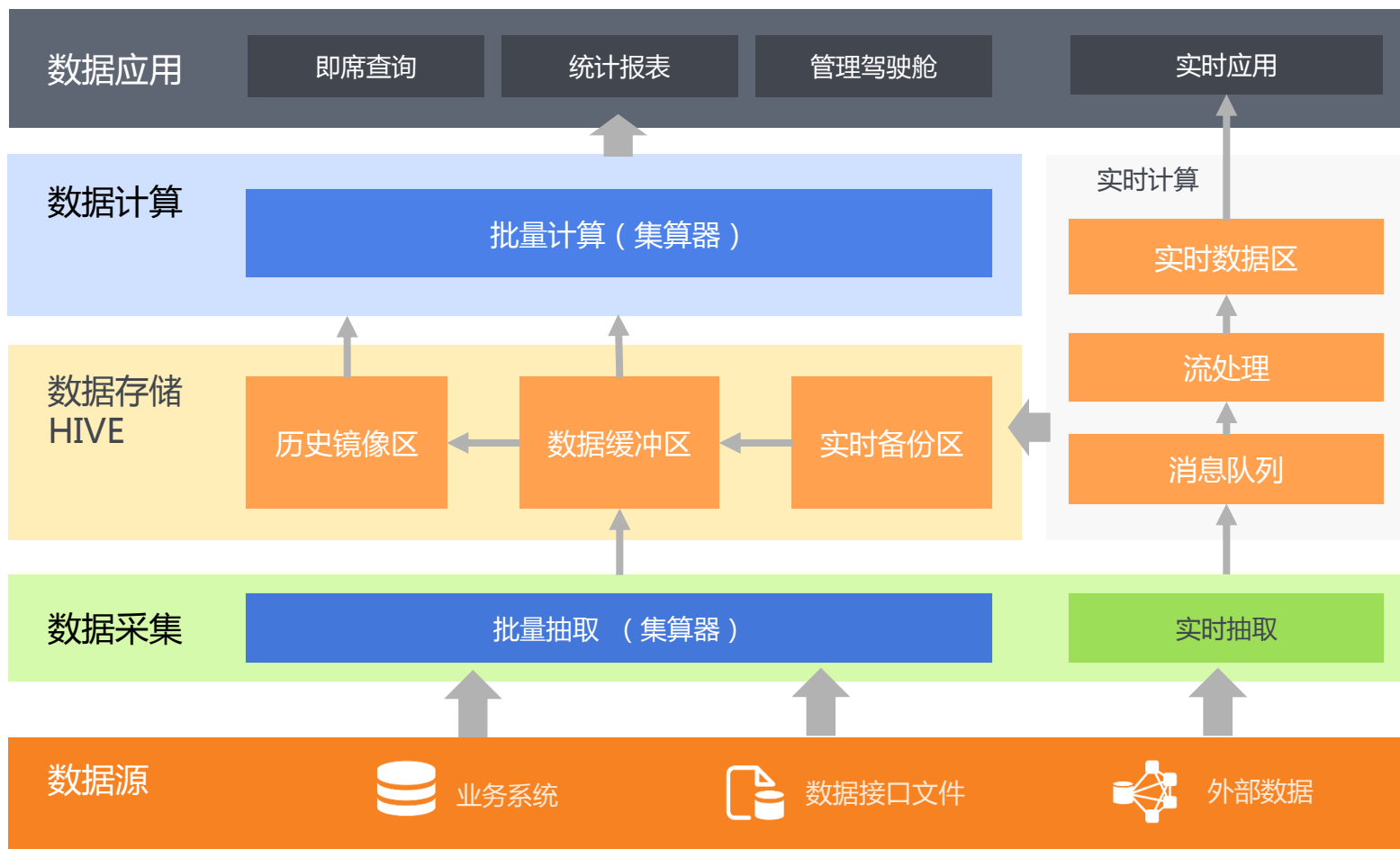
数据和算法模型不限制高效算法实现

计算无处不在！



CASE 1

基于集算器与Hadoop共同构建的大数据平台



CASE 1

该政府委办局大数据平台-集算器应用点

集算器主要完成以下几方面工作

ETL 将各类异构数据源ETL到大数据平台

数据计算 将Hive数据通过Hive CLI接口读入集算器进行计算

数据回写 某些集算器计算结果仍然通过CLI接口写回到Hive中

数据输出 将计算结果通过集算器应用接口为应用实时输出，节省关系数据库中转

对大数据平台的优化

支持多样数据源

可提高计算性能

并降低复杂算法工作量

应用效果

在hadoop平台内应用集算器后，使得整个**项目周期缩短至2个月**，平台搭建由于只选用了HDFS、Hive和集算器，只花费一周时间。

使用集算器**实现业务逻辑**相比Hive+MR**提升了至少2倍开发效率**。

集算器直接为应用输出有效计算结果，解决了需要通过关系数据库中转的难题。

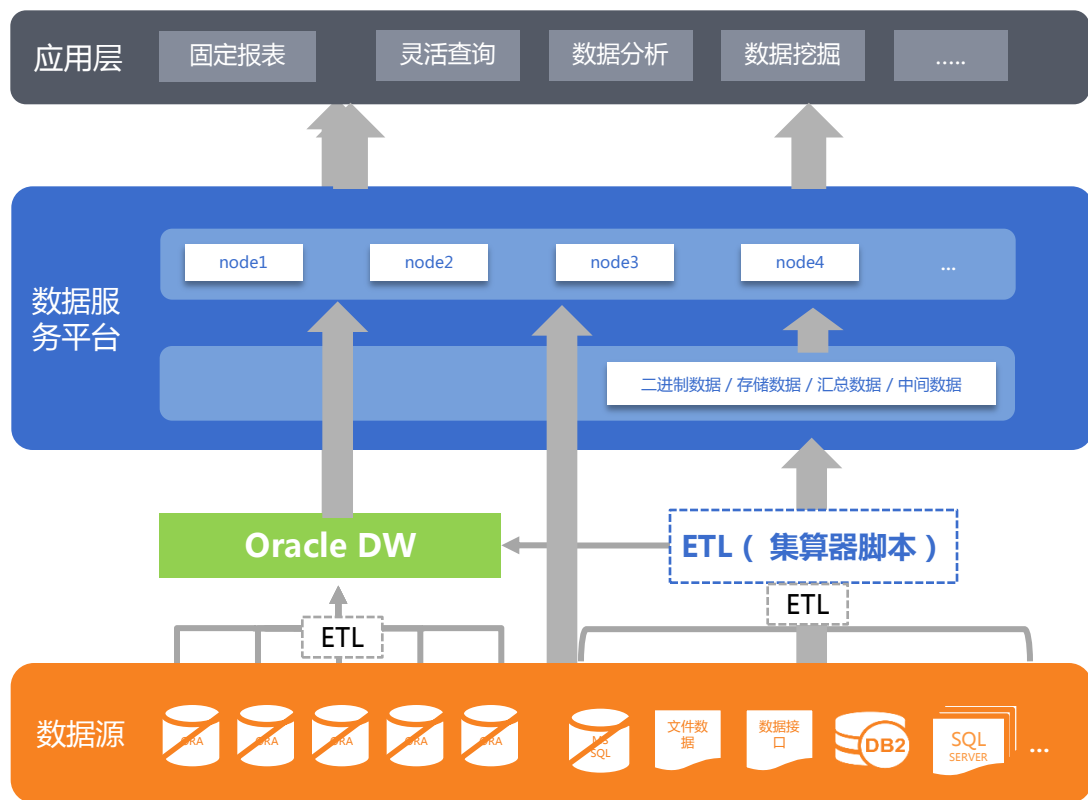
整个平台**实施周期短、快速上线、运维成本更低、更加快捷**。

CASE 2

基于润乾集算器构建的数据服务平台

使用集算器作为统一数据服务接口，通过标准JDBC接口为上层应用提供数据输出服务

面向分析系统，提供了固定报表查询和为业务人员提供的灵活查询



集算器作为统一数据服务平台带来的效果：

- 1、应用直接基于集算器开发，屏蔽底层数据源之间的差异，规范了数据管理
- 2、实现多数据源混合计算
- 3、实现历史数据归档，并基于归档数据做分析
- 4、提供多线程并行计算机制，加速数据计算性能
- 5、提供分布式计算，支持横向扩容 (scale out) ，实现弹性计算框架

CASE 2

快速提供数据服务

业务需求

每月需要将各个部门提供的kpi明细数据的excel文件整合到一起形成一张供领导查看的9个统计图的kpi报表；

面临挑战

数据来源分散在多个子系统中，无法直接于数据库中计算

痛点

采用excel工具实现
实现周期长；可复用性差；统计图不美观；类似的业务多

解决方案

使用集算器连接多个业务子系统
在线计算kpi报表

客户收益

节约制作成本；
统计图比excel美观，领导满意；
集算器构建的数据服务平台（数据上传）实现应用（0.5天）
集算器+润乾报表快速实现数据呈现（0.5天），快速响应领导数据服务需求

CASE 3

政府大数据场景——综合治税

偷漏税企业特征分析

制造型企业的缴税额与产量通常会存在正相关关系

产量与用电量往往也存在很大相关性

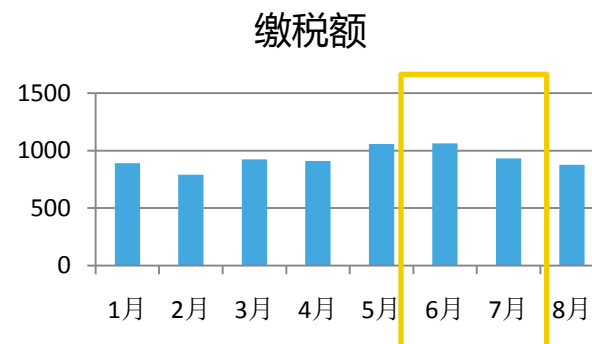
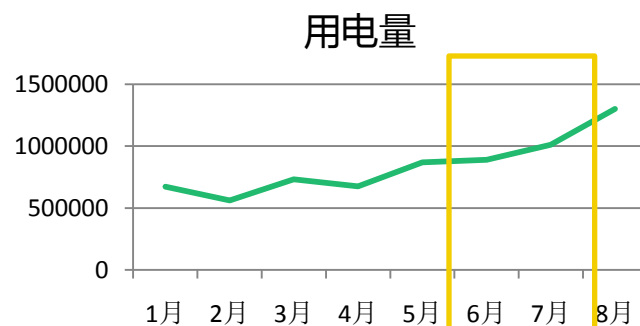
缴税额与用电量存在一定的线性关系

润乾集算器

快速集合对位运算，迅速发现偏离大的月份

基于位置的运算，迅速发现税额和用电量趋势相反的月份

迅速发现可疑月份



CASE 3 示例代码比较 (SQL)

从allpower表和alltax表中找出出现连续两个月用电量上升同时税额下降的企业编号 (eid)

1	select distinct(tpp.eid) from	SQL
2	(select eid,month from	
3	(select t1.eid,t1.month,t1.power,t2.power as power1,t3.power as power2	
4	from allpower t1,allpower t2,allpower t3	
5	where t1.eid=t2.eid and t1.eid=t3.eid and t1.month=t2.month+1 and t1.month=t3.month+2) tp	
6	where power>power1 and power1>power2) tpp,	
7	(select eid,month from	
8	(select t1.eid,t1.month,t1.tax,t2.tax as tax1,t3.tax as tax2	
9	from alltax t1,alltax t2,alltax t3	
10	where t1.eid=t2.eid and t1.eid=t3.eid and t1.month=t2.month+1 and t1.month=t3.month+2) tt	
11	where tax<tax1 and tax1<tax2) ttt	
12	where tpp.eid=ttt.eid and tpp.month=ttt.month	

CASE 3 示例代码比较（集算器脚本）

从allpower表和alltax表中找出出现连续两个月用电量上升同时税额下降的企业编号（eid）

	A
1	=allpower.select(eid==eid[1] && power>power[-1] && power<power[1]).new(eid,month)
2	=alltax.select(eid==eid[1] && tax<tax[-1] && tax>tax[1]).new(eid,month)
3	=[A1,A2].merge@i().id@o(eid)

代码短小的好处：

- ✓实现简单，更贴近自然思维
- ✓容易排错，可在一屏内写出大部分算法
- ✓易于维护，维护成本大大降低

北京润乾

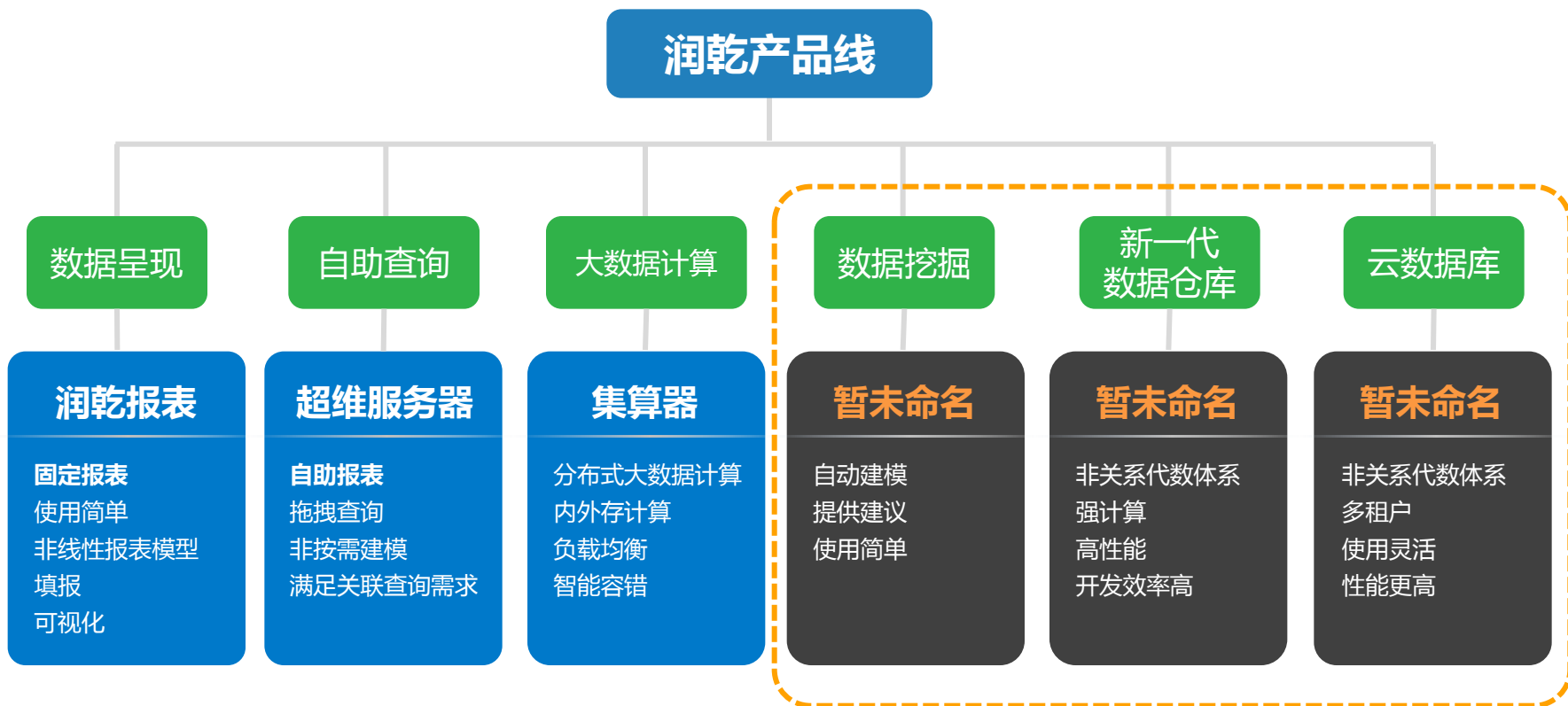
北京润乾成立于2000年6月，是北京市海淀高科技园区内的**高新技术企业**、国内知名软件企业，拥有一支实力强劲的产品研发团队。

北京润乾**十七年**始终坚持**创新技术推动应用进步**的经营理念，是中国最成熟的**数据计算与数据呈现基础软件提供商**，为客户提供国际一流的**通用软件产品**，我们的产品已经得到了广大开发商用户的认可和应用，在金融、电信、能源、政府、军工等多个行业，已积累了20000多成功案例。

- ◆ 2015福布斯中国非上市潜力企业100强
- ◆ 2016年中国软件和信息服务业·十大领军人物（蒋步星）
- ◆ 2016年度中国大数据 - 数据计算领域最佳产品奖（集算器）
- ◆ 2016年度 - 大数据领域创新产品奖（集算器）
- ◆ 2017年4月（集算器）被中国软件评测中心测定为“优秀大数据基础设施类产品”
- ◆ 2017年6月（集算器）获“航天云网”软件应用创新大赛一等奖，润乾报表V5.0获二等奖



坚持数据计算策略，专注基础软件



创新技术 推动应用进步
www.raqsoft.com.cn

谢谢
大家

