

大数据发展现状及趋势

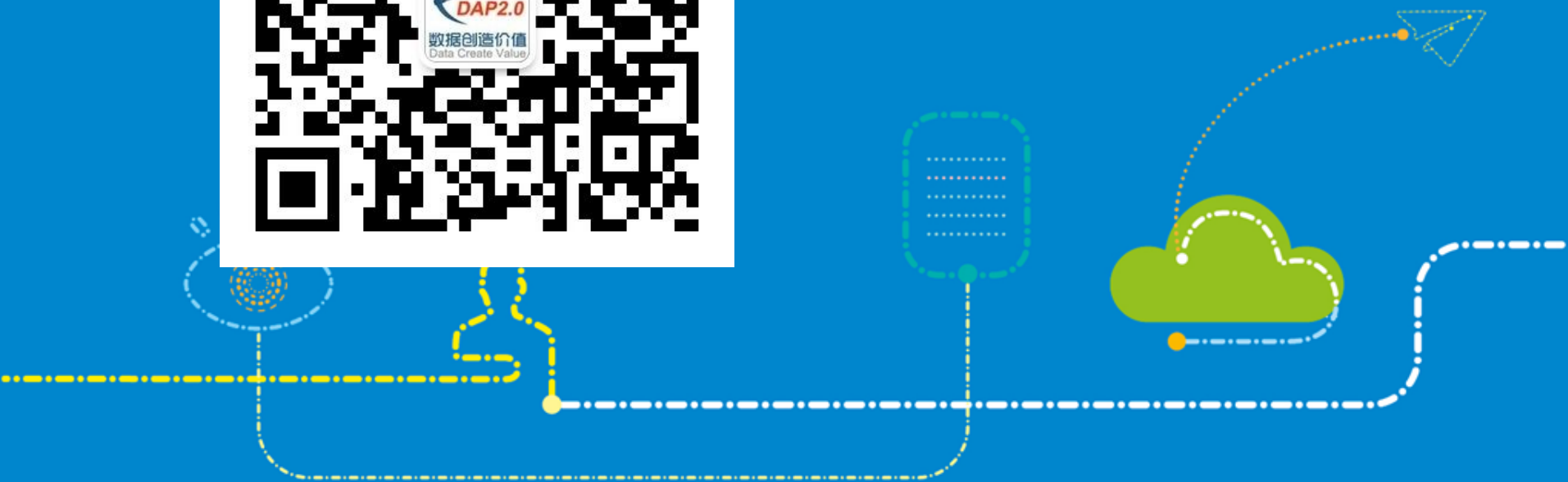
中兴通讯大数据总工 朱科支

ZTE

未来，不等待



ZTE_BigData



Gartner : 2014年大数据技术已经越过炒作顶峰

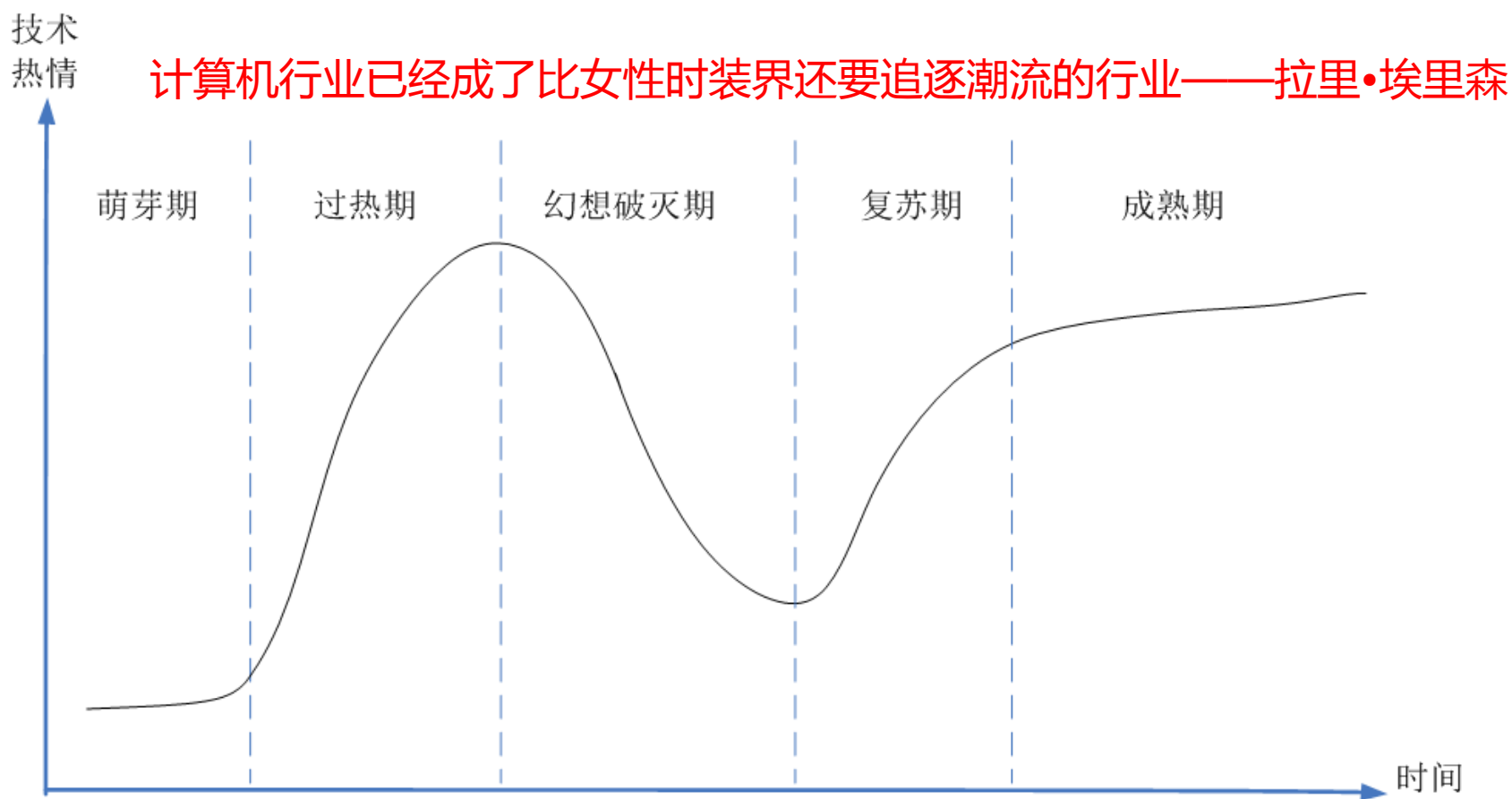
Emerging Technologies of 2014

Source: Gartner (August 2014)



大数据已经越过炒作顶点，
业界将会越来越悲观

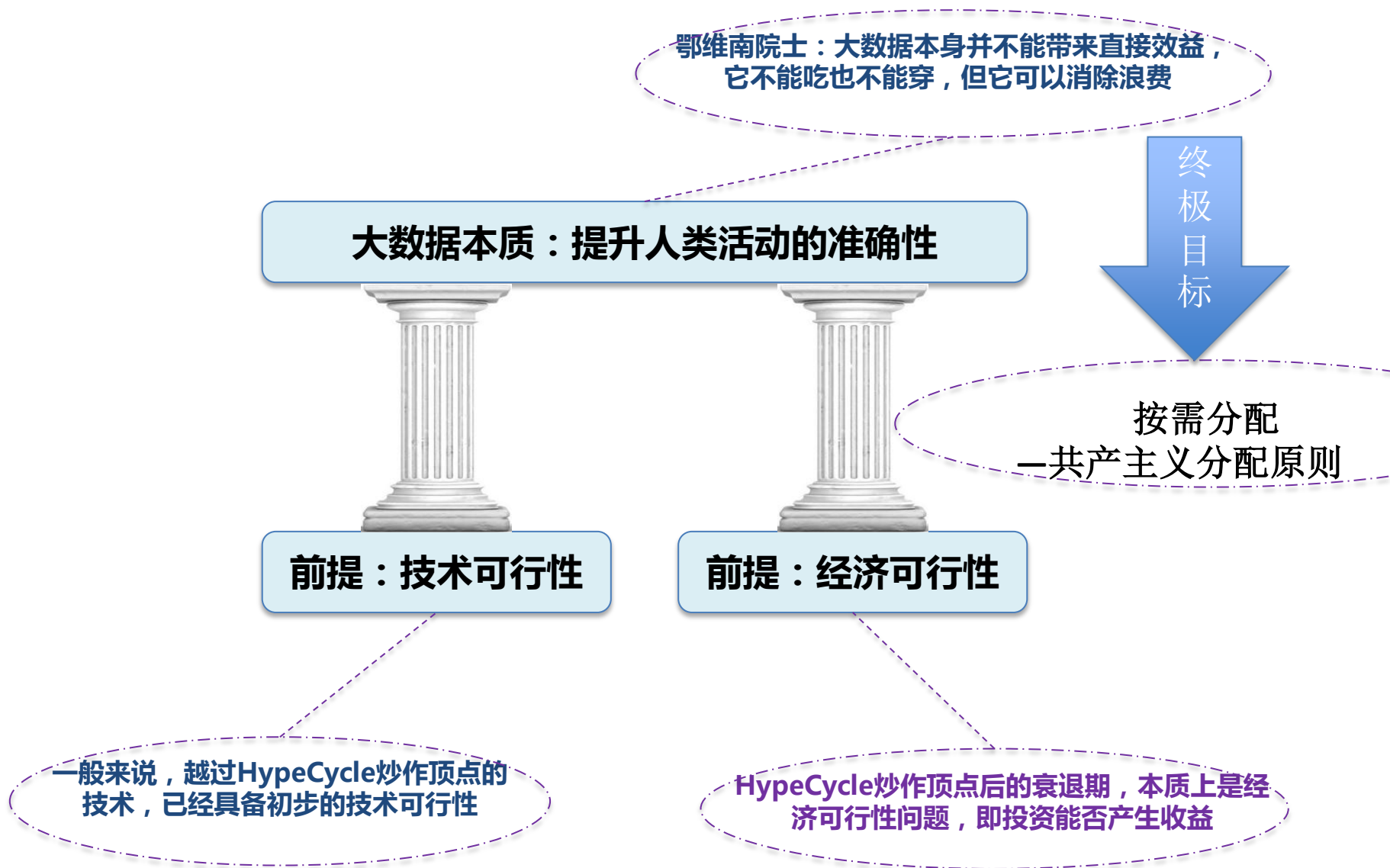
HypeCycle曲线



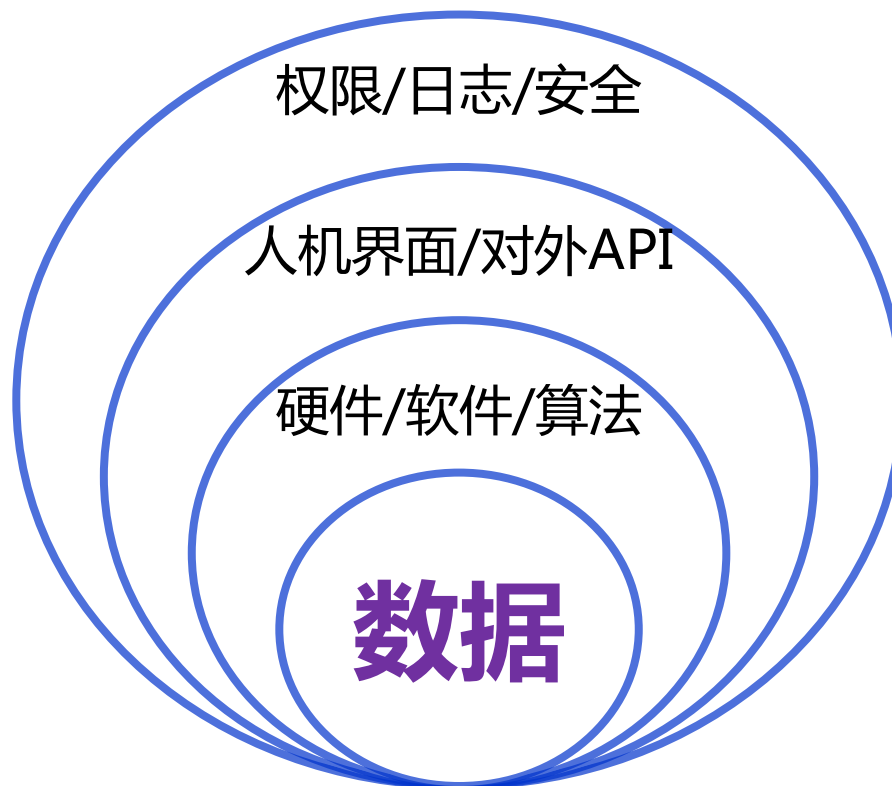
过热期：具备技术可行性

复苏期：具备经济可行性

大数据的本质与实施前提



技术可行性：以数据为中心的架构与部署



- ◆商业部署以数据为中心
- ◆软件设计以数据为中心
- ◆硬件设计以数据为中心

技术可行性：开源生态加速技术成熟

小团队（个人）可完成传统自研模式无法想象的工作

收益

开源软件一般具备良好架构，相较自研，架构失败可能性较小

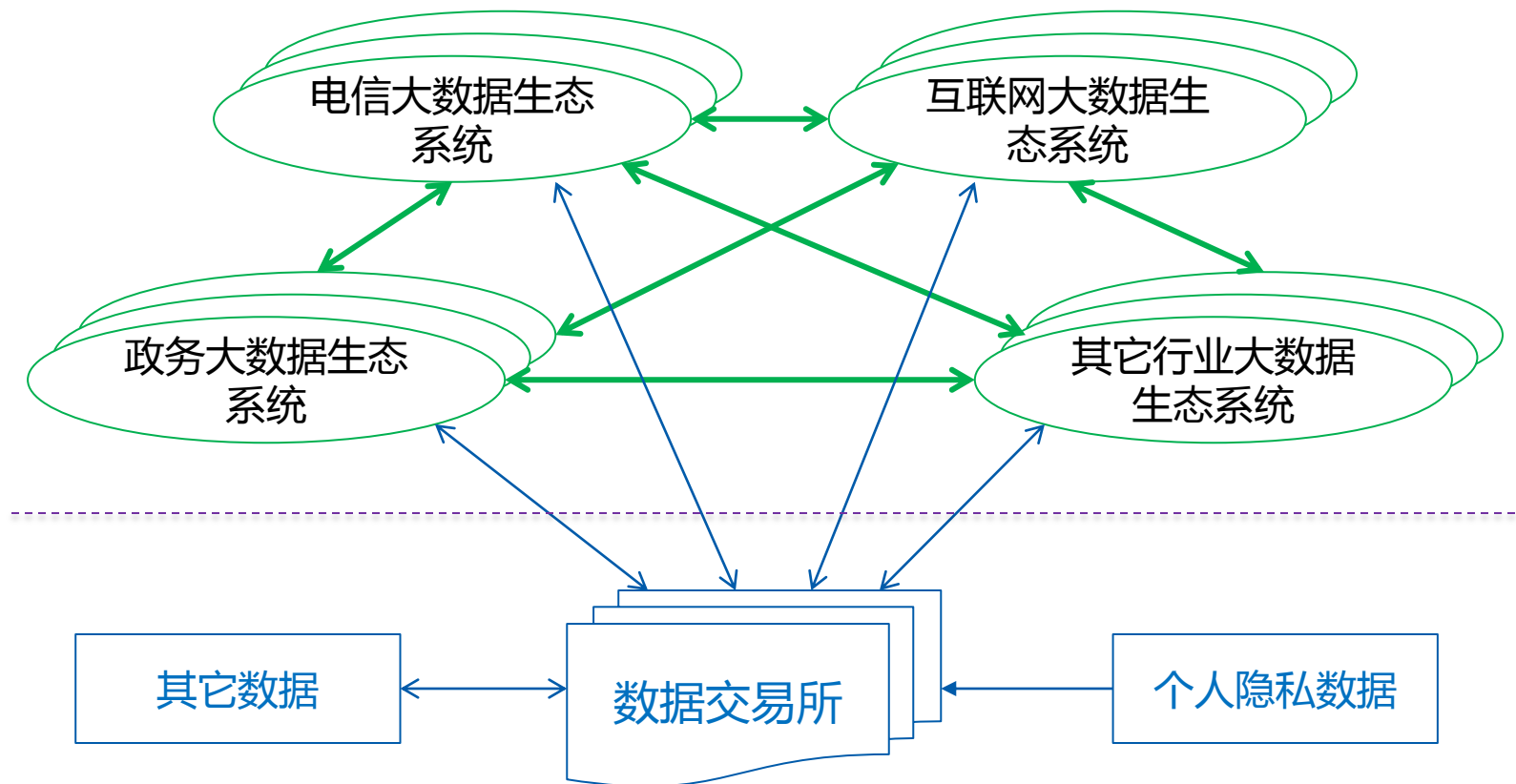


技术门槛降低，相同工作，100人大团队与5人小团队有何差别？

问题

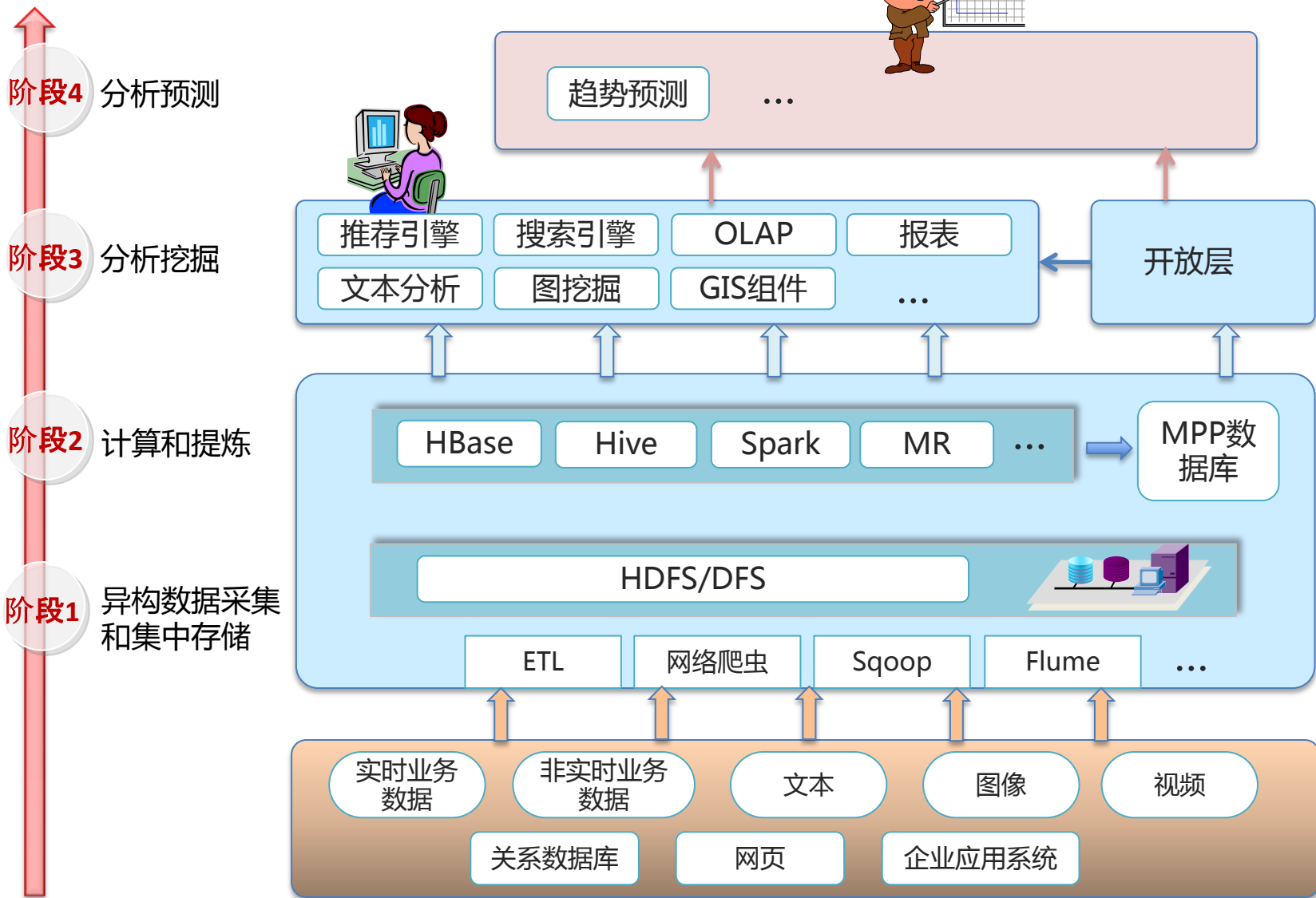
方案与架构严重同质化，与同等竞争对手相比，有何优势？

- 开源生态环境已是无可抵挡的潮流
- 不可能退回到封闭自研的生态环境

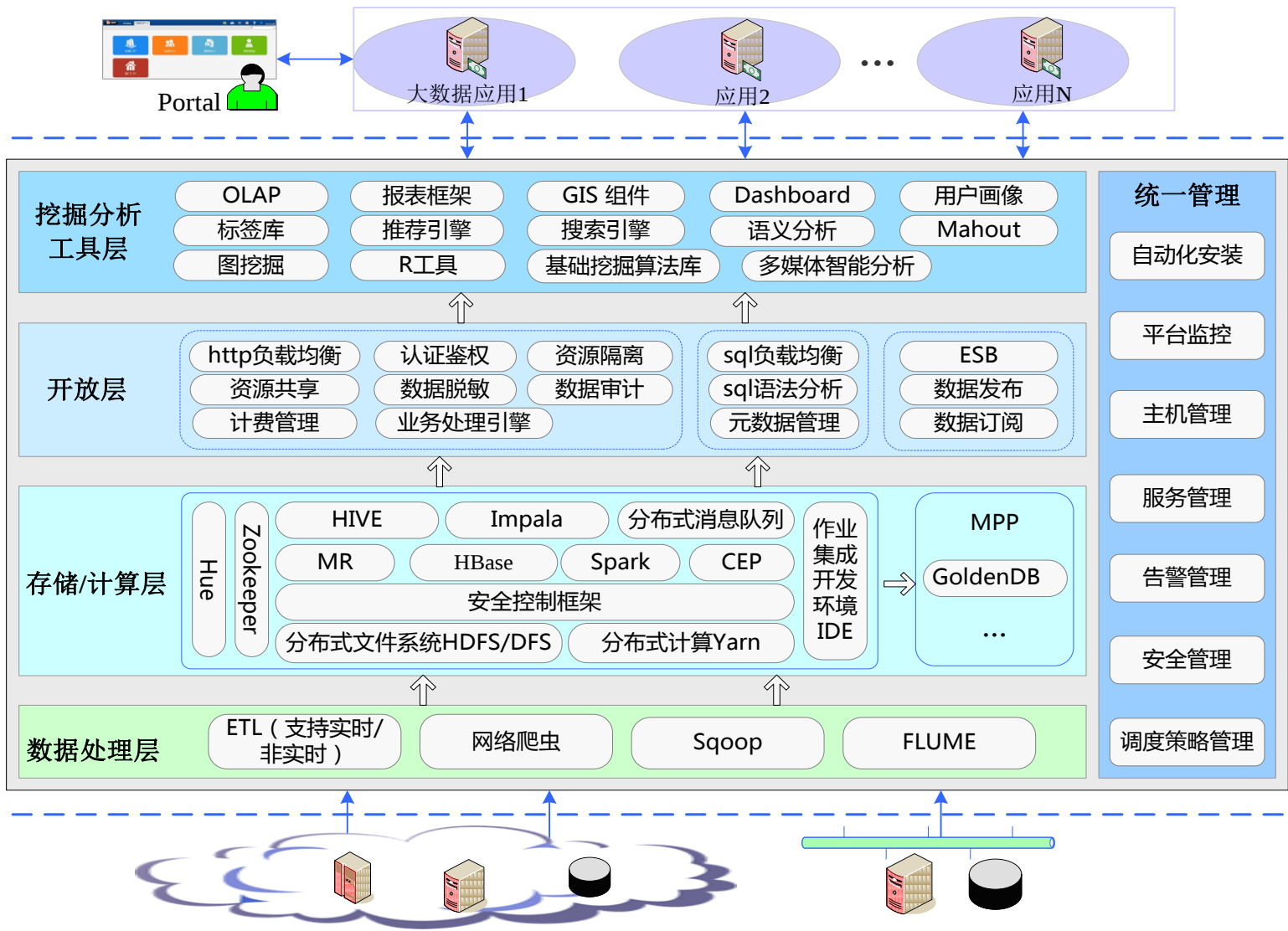


- ◆ 不同的行业内，将产生强势的生态系统（1个或多个）
- ◆ 生态系统间将按照互利原则进行等价数据交换
- ◆ 数据交易所是生态系统的有益补充，解决数据交换的长尾问题

数据治理的几个层次（阶段）



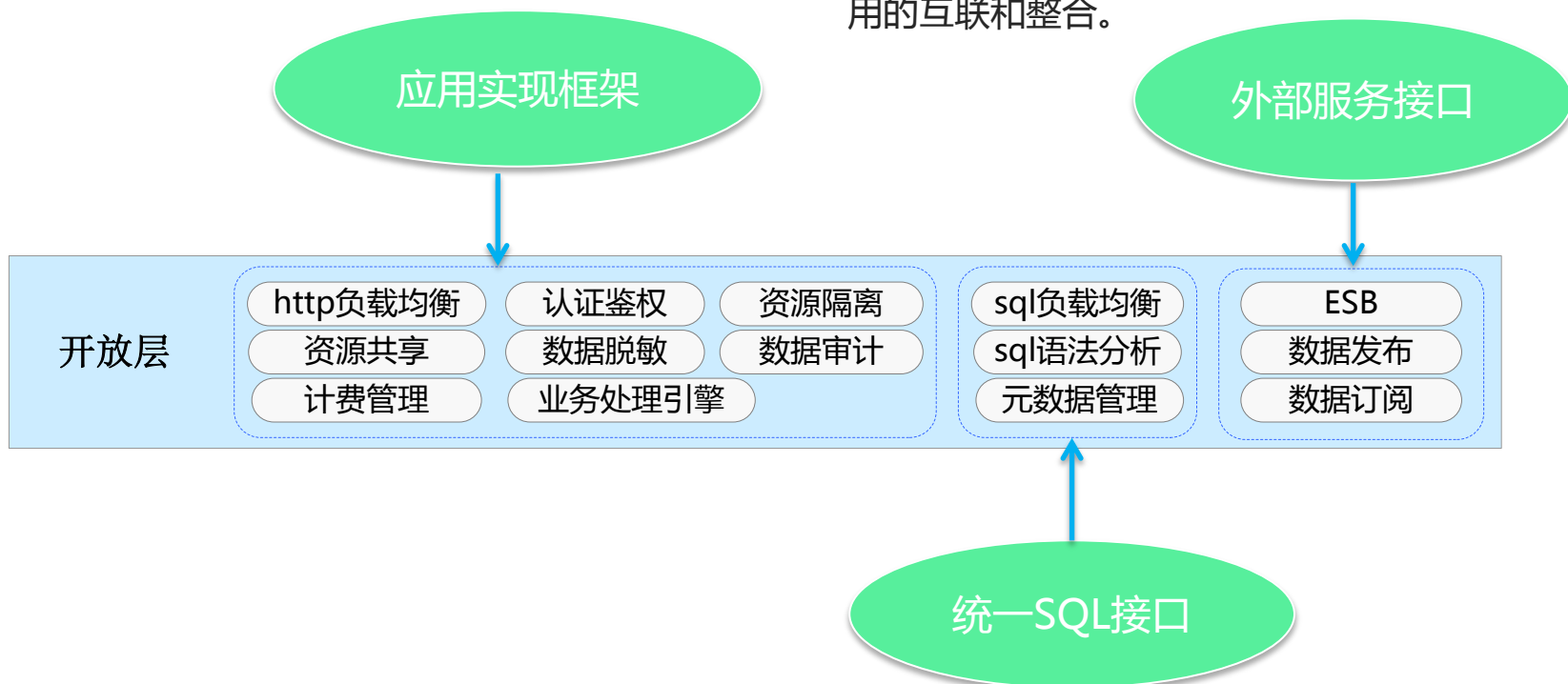
DAP大数据平台架构



DAP大数据开放层三大功能和技术范围

在大数据平台之上架设友好的使用环境，包括数据存储、查询、导入导出；企业数据隔离、共享服务；管理权限，输出计费数据等。

ESB（Enterprise Service Bus企业服务总线）、数据发布和订阅功能为企业内部或外部的其它应用系统提供通用的访问和交互接口，实现多个应用的互联和整合。



对于大数据平台中的结构化、非结构化数据库访问接口进行封装，提供统一的类似SQL语言接口，简化二次开发。

开放层的客户价值

1

- 统一的数据访问接口，对用户和开发者屏蔽关系库和大数据，并在数据增大的时候实现无缝切换

2

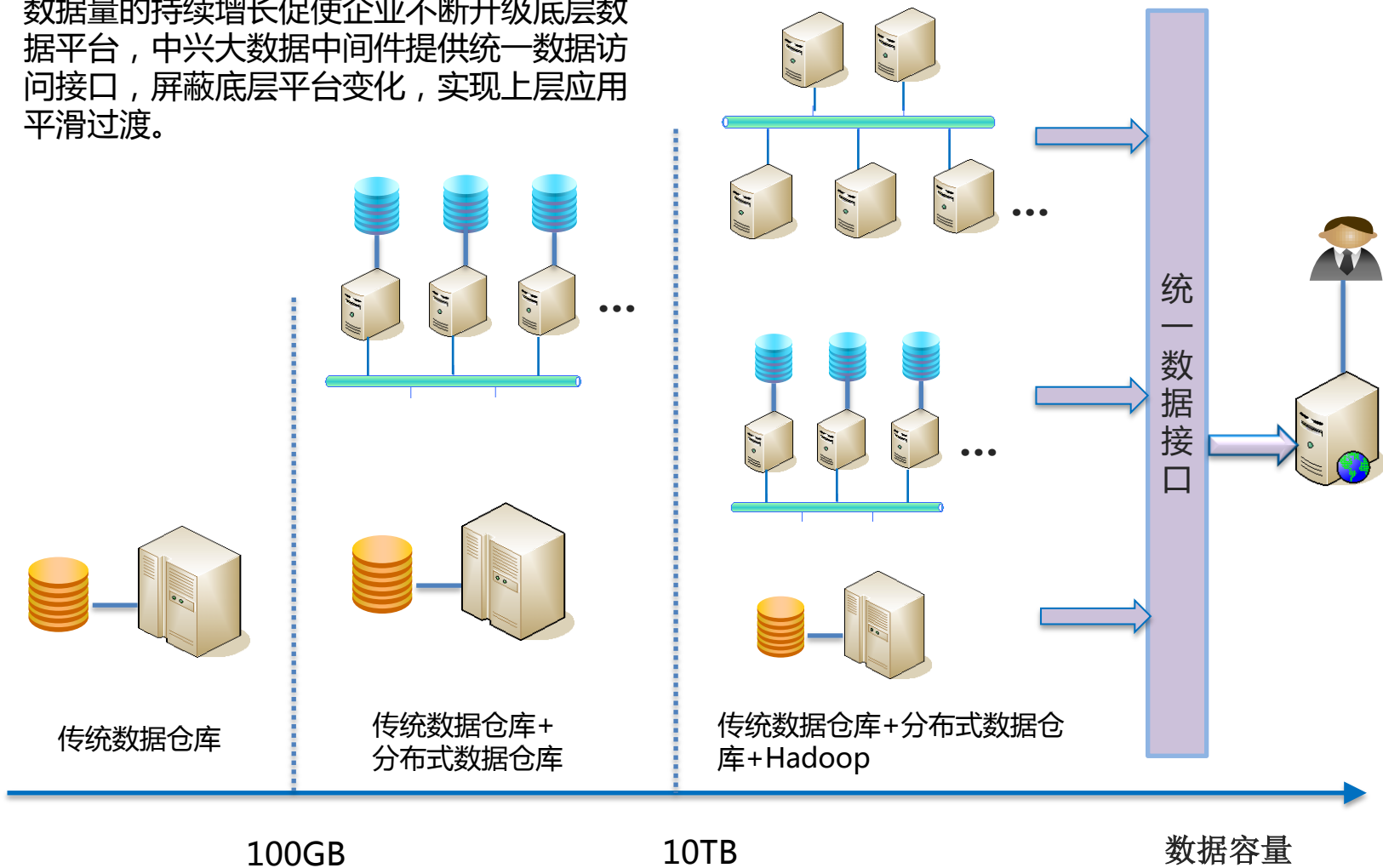
- 支持多租户的方式使用大数据平台，减少企业大数据投入

3

- 支持IT基础架构与数据的无缝演进与切换

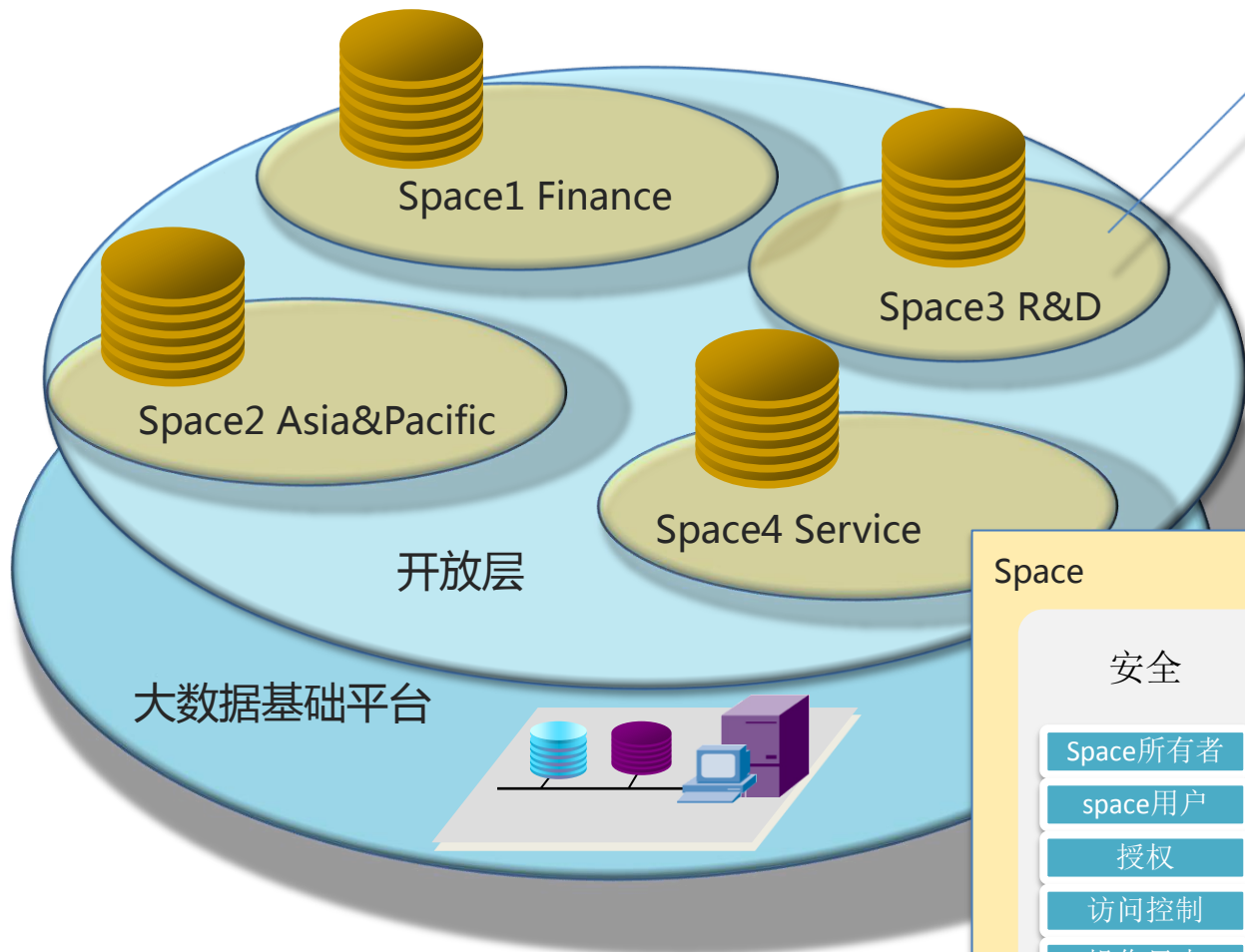
开放层保证企业应用平滑升级

数据量的持续增长促使企业不断升级底层数据平台，中兴大数据中间件提供统一数据访问接口，屏蔽底层平台变化，实现上层应用平滑过渡。



开放层实现多租户使用方式--实施隔离后：

支持多租户的方式使用大数据平台，减少企业重复投入



◆各部门、区域在平台上开辟自己的空间，数据放在各自的空间内。

◆各space数据、应用由本space管理。

◆系统管理员负责系统的运维，保证系统的正常运行。

Space

安全

Space所有者

space用户

授权

访问控制

操作日志

数据

Hdfs数据文件

Spark数据表

Hbase数据表

计费数据

计算

MR任务

Spark任务

SQL访问

视频个性化推荐

应用场景

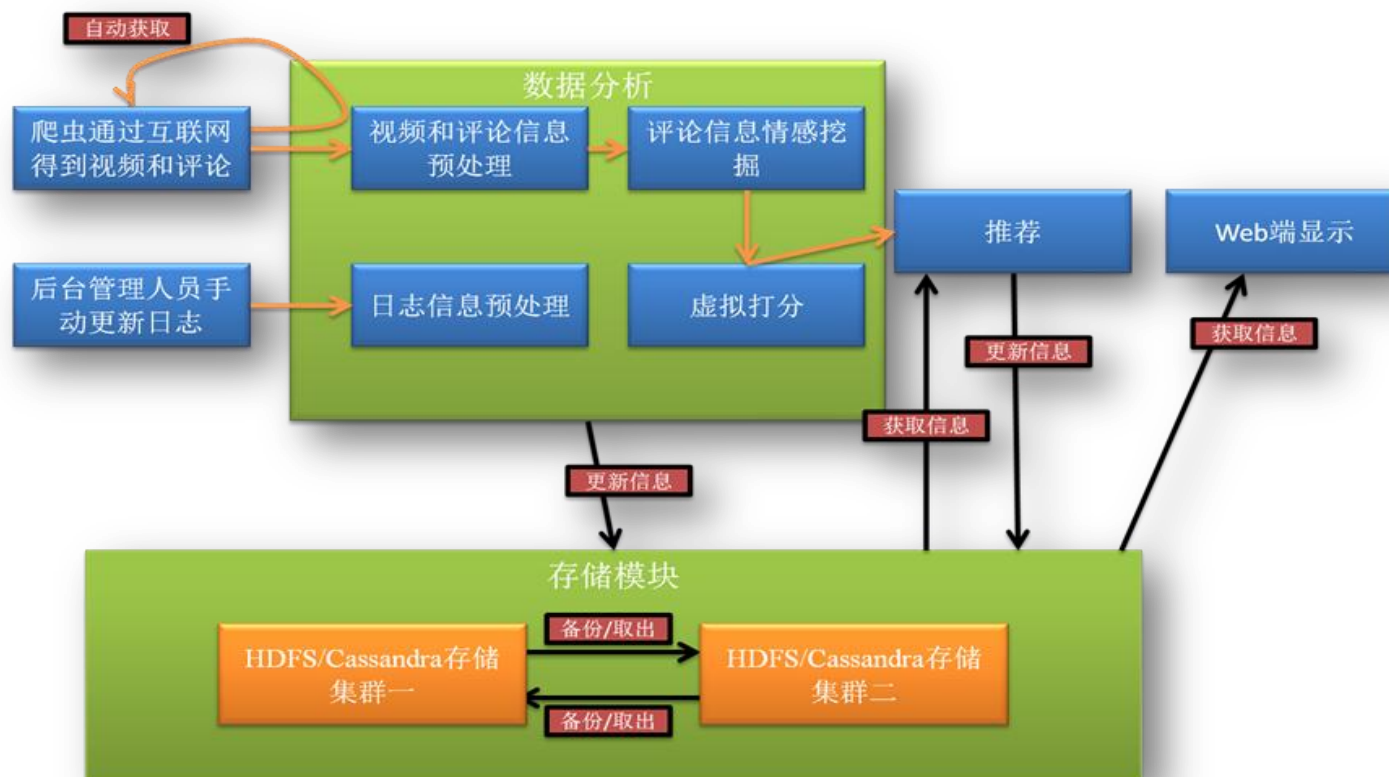
- 结合视频底层特征和用户评论包含的隐式打分，丰富用户稀疏的打分矩阵，为用户提供精准视频推荐。

解决方案

- 基于视频信息，抽取视频底层特征，抽取视频之间隐含的关联；
- 根据用户评论，判断情感倾向，丰富用户打分矩阵；
- 分别从用户行为、评论和内容关联，预测用户打分，进行视频推荐。

相关技术要求

- 网页的爬取技术
- 底层视频特征抽取技术
- 用户评论的情感分类
- 协同推荐技术



智能垃圾短信分析系统应用案例

应用场景

- 对短消息中心的实时短信进行分析，根据短信内容，发短信人的行为特征等识别出垃圾短信，进行过滤。

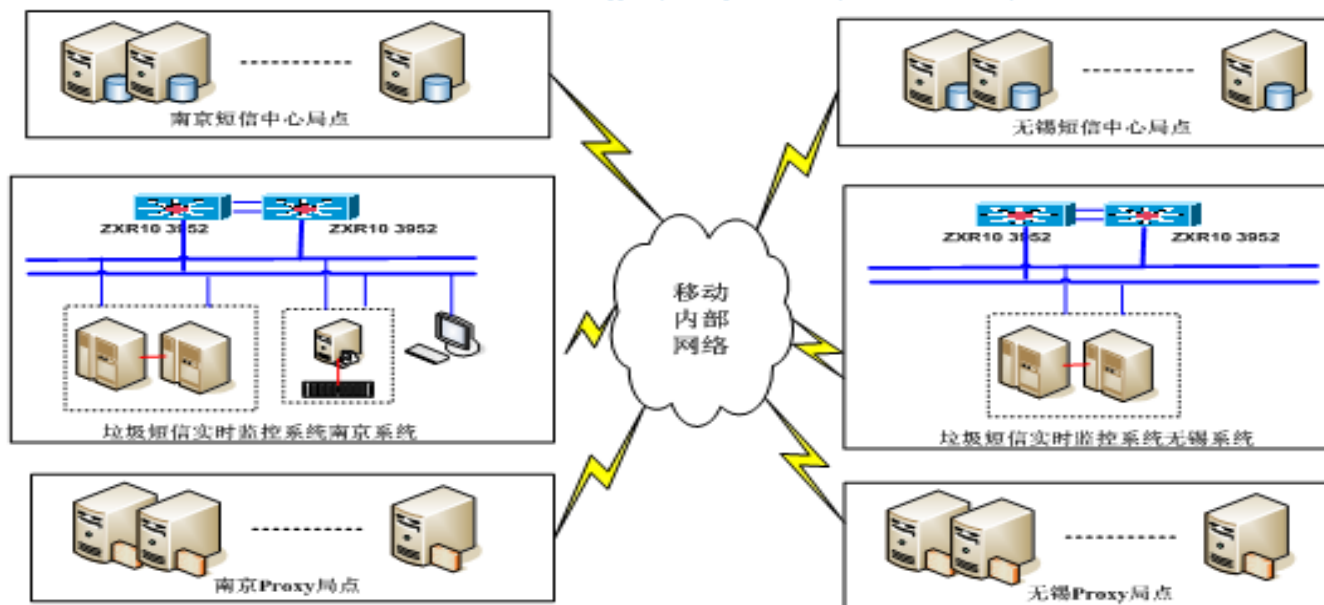
主要数据源

- 省各MO短消息中心的实时短消息。

相关技术要求

- 总处理能力为40000条/秒；

垃圾短信监控组网图



“数据银行” 大数据项目

应用场景

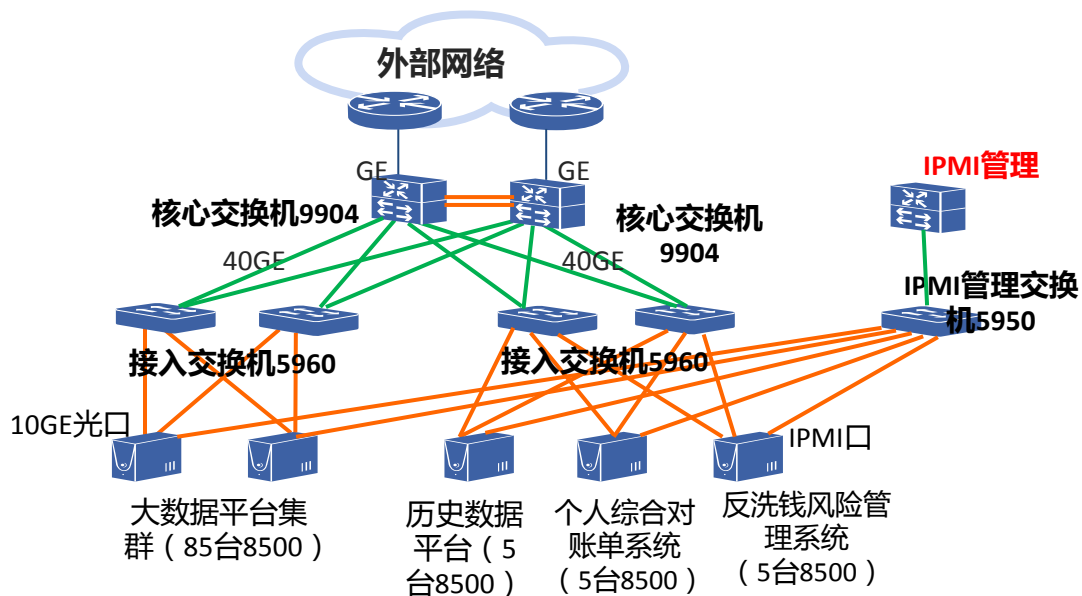
- 主要用于行内海量数据存储、查询和分析；
- 第三方托管应用数据分析及销售；

主要数据源

- 行内ODS系统数据；
- 第三方托管应用数据

相关技术要求

- 提供作业IDE开发、编排和发布工具
- 多租户隔离，计算和数据分离
- 支持数据异地容灾；



日志详单大数据项目

应用场景

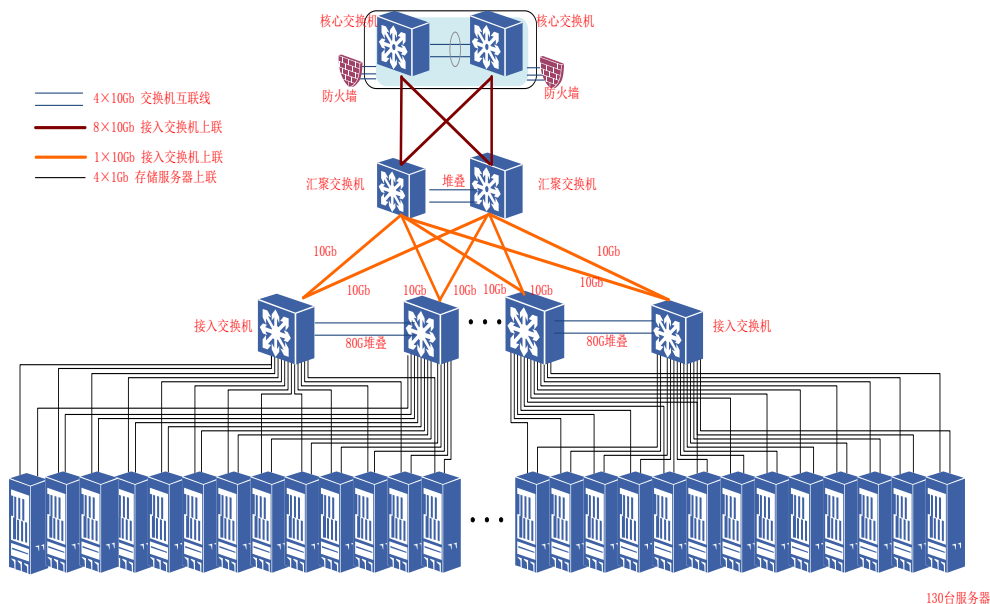
- 主要用于海量数据存储及查询分析
- 在非OLTP场景下逐步替代Oracle、Sybase等数据库服务

主要数据源

- wap网关日志、点击流、计费结算详单、信令等一次写入、无追加及修改要求的数据。

相关技术要求

- 接口要求：HDFS接口、MR接口、标准JDBC/SQL接口等
- 多租户隔离，计算和数据分离



Winner

ZTE

统计模块大数据平台改造

应用场景

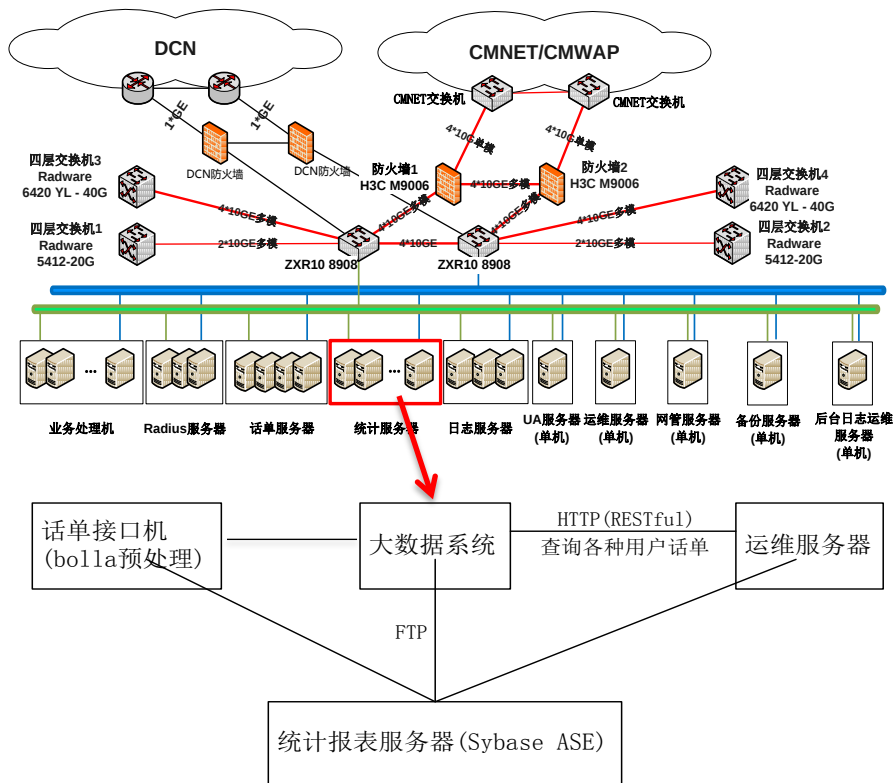
- 数据业务快速发展，数据网络流量在大于5Gbps情况下，原有Sybase IQ性能与价格已经不能满足系统需求，需要采用新技术

系统特征

- 满足数据业务量发展需求
- 降低数据库外购件成本
- 硬件设备利旧：在磁阵+服务器硬件设备上安装部署大数据平台与应用

相关技术要求

- 存储综合网关各种话单，预计最大日话单70亿条，保存3个月
- 提供HTTP(RESTful)接口供运维服务器做话单实时查询使用
- 定时生成统计报表



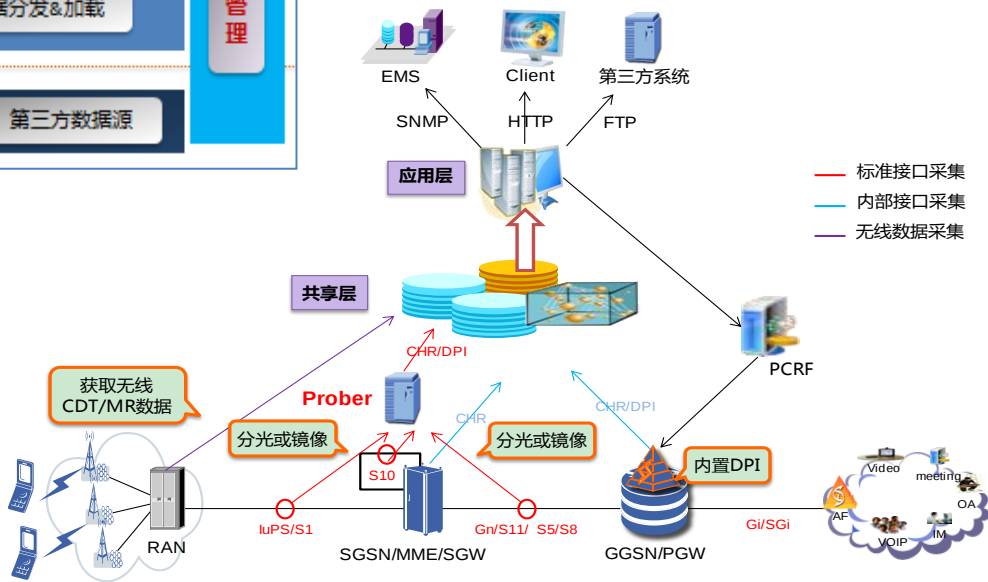
基于大数据平台的无线增值业务系统



无线增值业务系统是基于大数据技术支撑网络运维和运营的产品，立足于从客户的角度去感知和分析网络和业务信息

基于中兴的DAP大数据平台，无线增值业务系统具备强大的处理能力，提供灵活、开放融合的能力，可帮助运营商构建基于大数据技术的精细化运营平台。

- ✓ 灵活的秒级实时监控，可以实现网络质量、热点业务、VIP用户、热点区域和异常事件的全方位监控；
- ✓ 丰富的数据源支撑能力，既支持中兴网元内部接口采集数据，也支持Gn/IuPS/S1/S10/S11/S5/S8等多种标准接口的采集数据，还可通过接口适配运营商已有的数据源；



用户投诉处理系统

应用场景

- 海量话单数据的存储、分析，简化用户投诉根因分析的难度，并为网络优化提供依据；

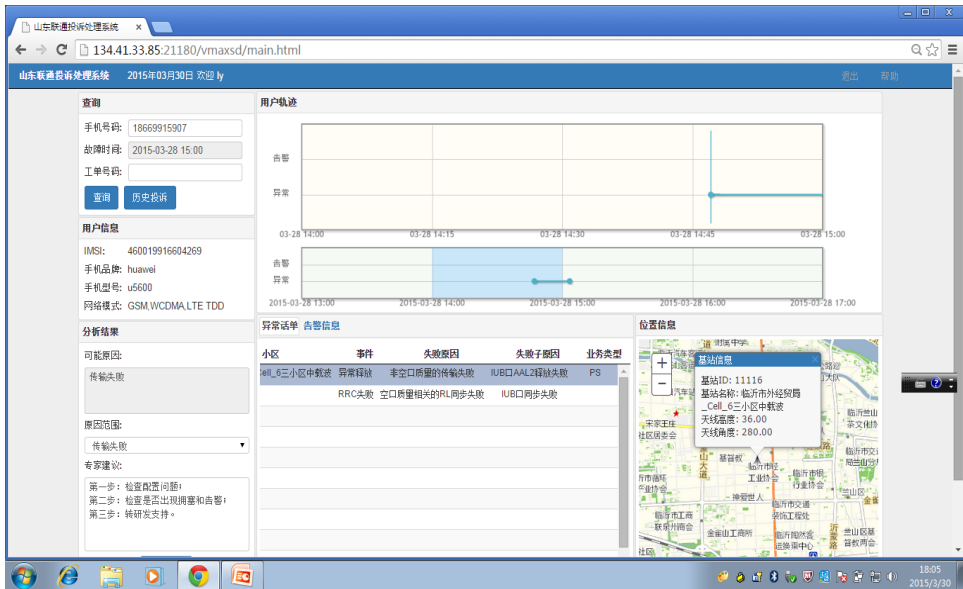
主要数据源

- RNC话单数据；
- 工程参数
- IMSI和MDN对应表

相关技术要求

- 海量话单数据的导入、安全存储；
- 海量数据的快速查询
- 系统的高可靠性

用户投诉处理



道路监控云应用系统，保障青奥安全

应用场景

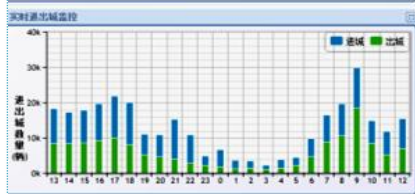
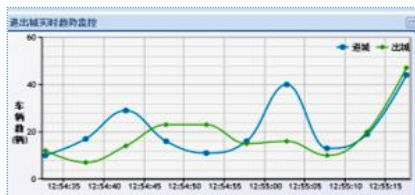
- 提供集中式海量数据存储管理和计算分析能力

主要数据源

- 各交通卡扣道路图像监控数据，包括车辆信息、卡口信息以及车辆监控图片、视频。

相关技术要求

- 接口要求：标准JDBC/SQL接口等
- 处理能力：支持百亿级记录准实时查询和处理分析



青奥安全保障

- 民航、高铁、宾馆人流统计和24h监控
- 进出城车辆24h监控，归属地统计
- 重点车辆24h监控、定位、轨迹查询
- 青奥基站人流量24h监控



大数据性能

■ 实时入库和查询快速响应能力：

对单个车牌1个月的轨迹确保在1秒内完成查询，6个月内的轨迹确保在5秒内完成查询，6-12个月内的轨迹则确保在10秒内完成查询

■ 数据分析处理能力：

支持基于监控数据开展多种应用，如过车轨迹查询、可疑套牌车辆分析、可疑高频度车辆分析等

■ 高性能的云存储平台承担对峰值流量超过10000条/秒，3年累计高达200亿条的总数据量的存储和处理

让我们在这里交流



ZTE_BigData : 发布中兴大数据原创思想

Thank you



未来，不等待

