

大数据平台的演进之路

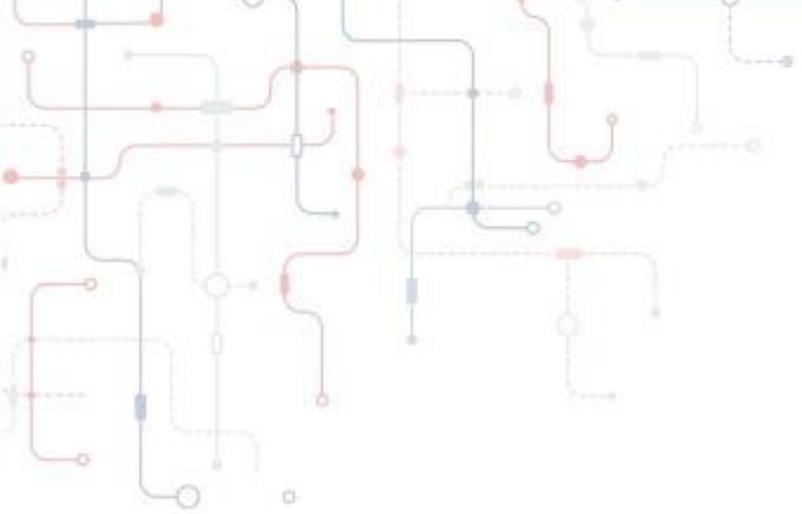


袋鼠云大数据产品总监-申杭



15658160712





目录

| **大数据平台的发展历程**

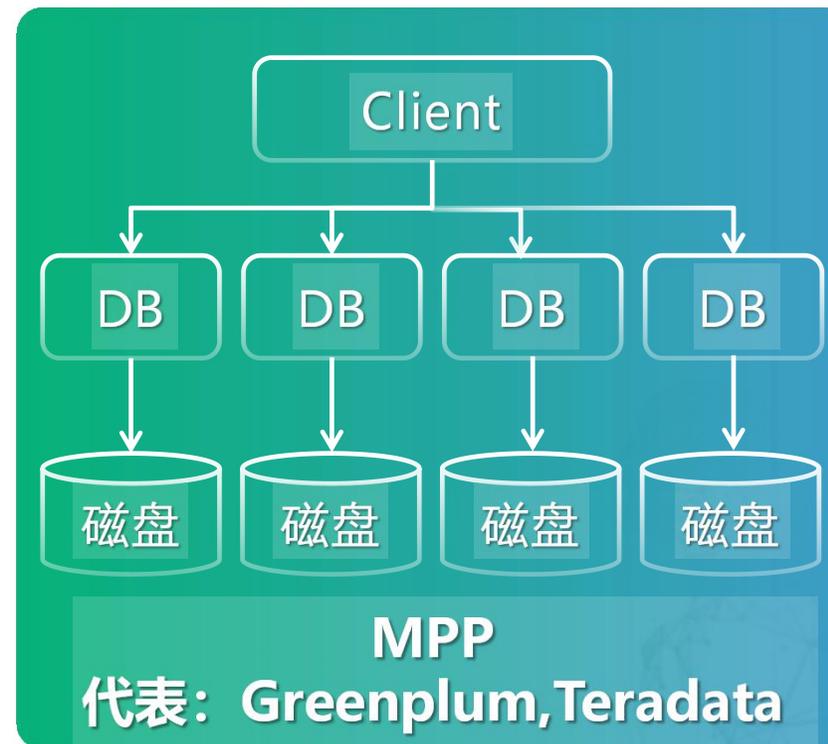
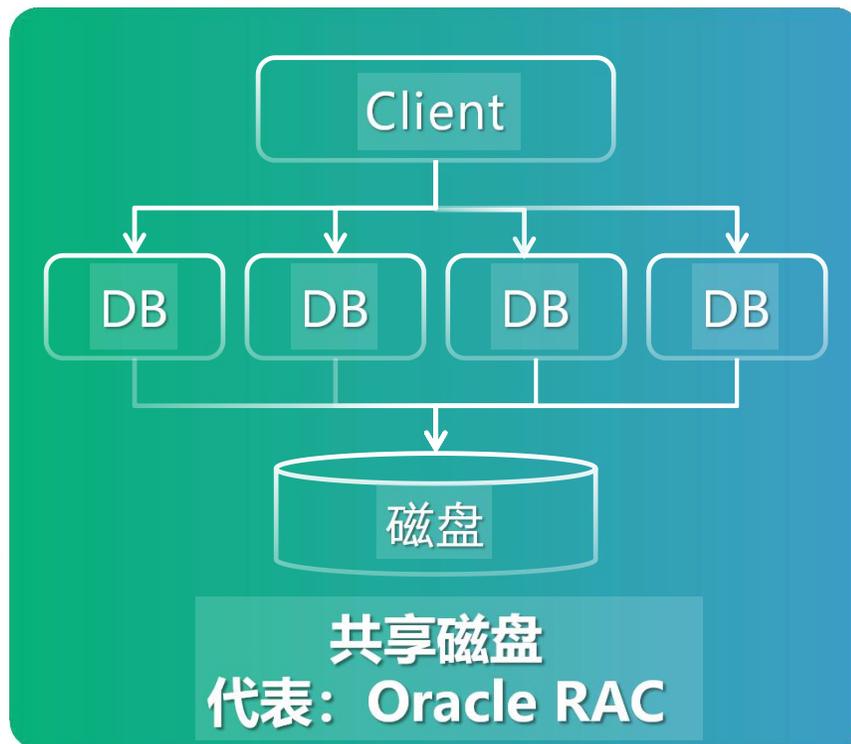
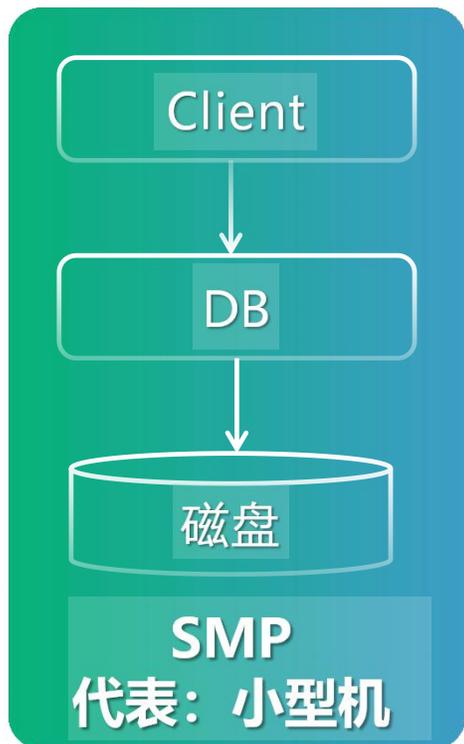
| **数据中台介绍**

| **数栈产品介绍**

| **数据中台案例**



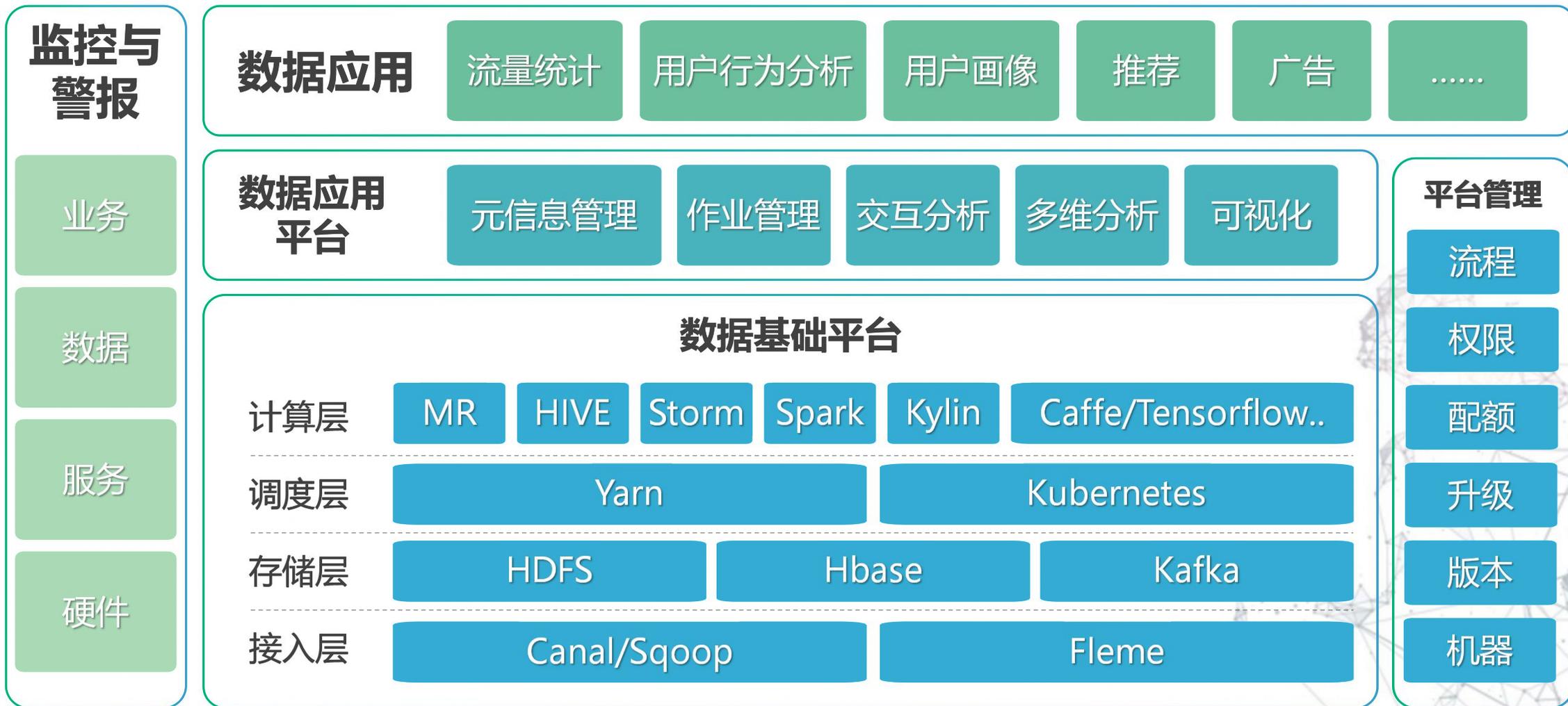
传统数据仓库计算引擎



面临的问题

- 在数据量爆发性增长的情况下，计算能力面临瓶颈
- 主要面向离线计算处理，缺少对实时计算/机器学习/即系查询等场景的支持
- 数据管理/数据治理能力弱

新时代的计算平台架构



数据计算的频率



OLTP/OLAP
混用



离线



准实时



实时

特点:

- ◎ 数据实时性越来越高
- ◎ 容错性高
- ◎ 处理量越来越多

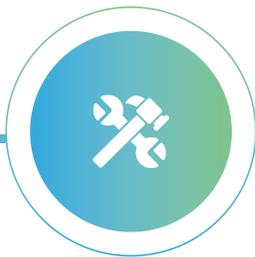
大数据价值

- ◎ 辅助决策->数据业务化
- ◎ 成本中心->利润中心
- ◎ 价值模糊->可量化



用户画像

- 人群透视
- 人群圈选
- 人群放大



精准营销

- 广告
-



搜索

- 搜索优化
-

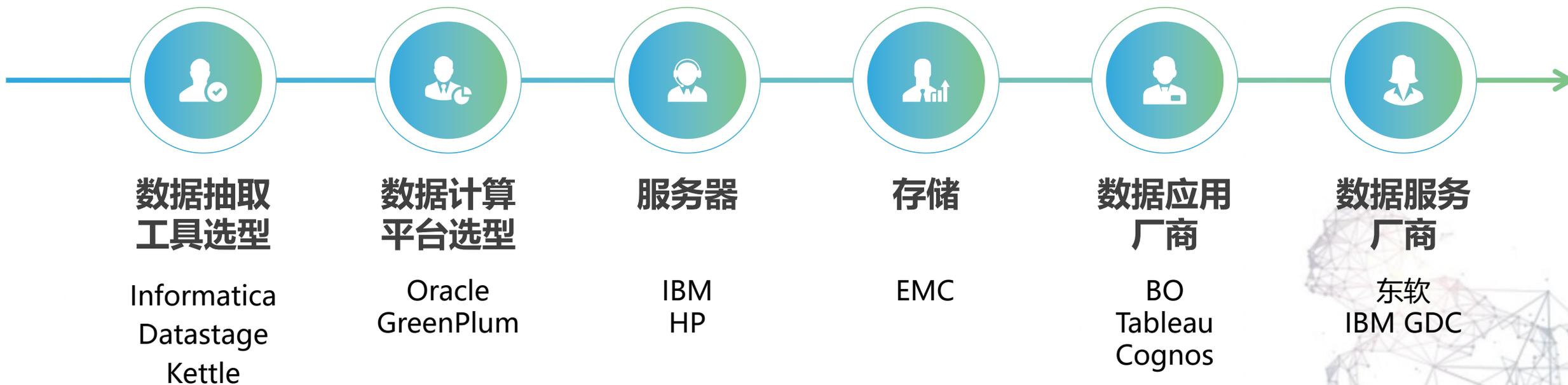


AI

- 个性化推荐
- 图像识别
- 风控
- 自然语言处理

2018 AI先行者大会

大数据平台建设步骤



缺点

- ◎ 产品分散, 厂商多, 周期长
- ◎ IOE体系, 产品费用高
- ◎ 运维管理成本高

数据中台定位

做为企业全域数据的共享能力中心，旨在提供数据采集、数据模型、数据计算和萃取、数据治理、数据资产、数据服务等全链路的一站式产品+技术+方法论服务，构建面向业务应用的数据智能平台



战略决心

- ◎ 决策管理层“大中台、小前台”战略创新的决心
- ◎ 驱动组织和流程进化，构建全域打通的数据中台
- ◎ 协调业务部门和技术部门，中台战略的**落地重于设计**

人才投入

- ◎ 降低大数据的技术门槛
- ◎ 在线化知识沉淀和传承
- ◎ “方法论+工具保障+合作厂商”降低对高端人才的依赖

工具保障

- ◎ 分布式计算&存储&调度架构
- ◎ 企业级数据仓库 + 实时流计算
- ◎ 覆盖大数据全生命周期的一站式可视化平台



大数据平台理念对比

计算平台多样性

离线计算/流计算/图计算/机器学习/分析引擎，支持更多的应用场景

1



4



2



5



开发方式变革

纯web化开发方式，一站式数据处理流程

数据治理

脏数据管理，元数据管理，数据质量，数据资产

方法论的产品化

数仓建设规范，标准，产品化

3



6



数据应用

BI报表/大屏/营销引擎/DMP

数据中台价值



全面梳理数据资产
构建全域数据中心



落地战略和组织保障
培养业务创新的土壤



数据资产管理在线化
形成数据人才成长的摇篮



通过业务的不断滋养
打造业务和数据的闭环

数据中台实施路径



◎ 部署计算平台, 开发套件, 数据治理, 数据应用引擎等产品



◎ 在平台的基础上构建数据仓库, 数据治理, 数据应用等服务

◎ 持续迭代, 推进

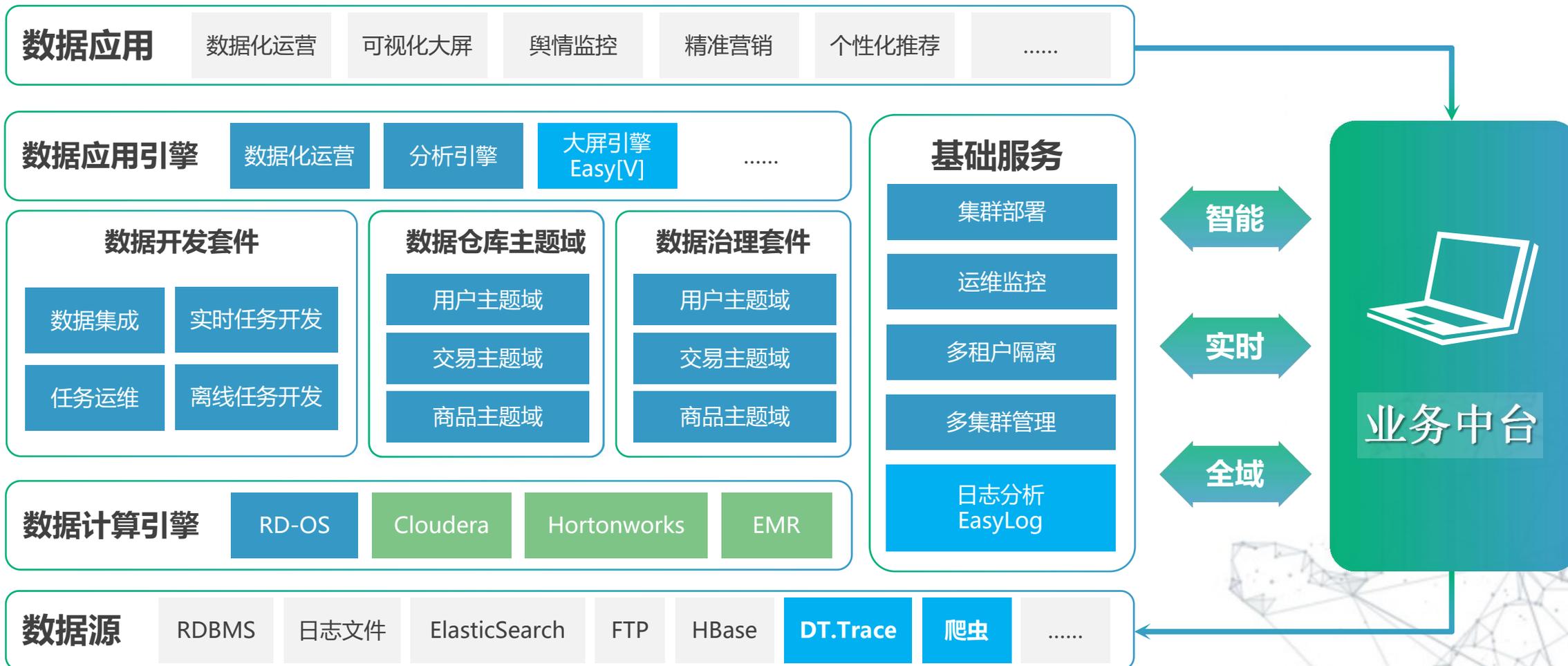


◎ 高层重视, 自上而下推动

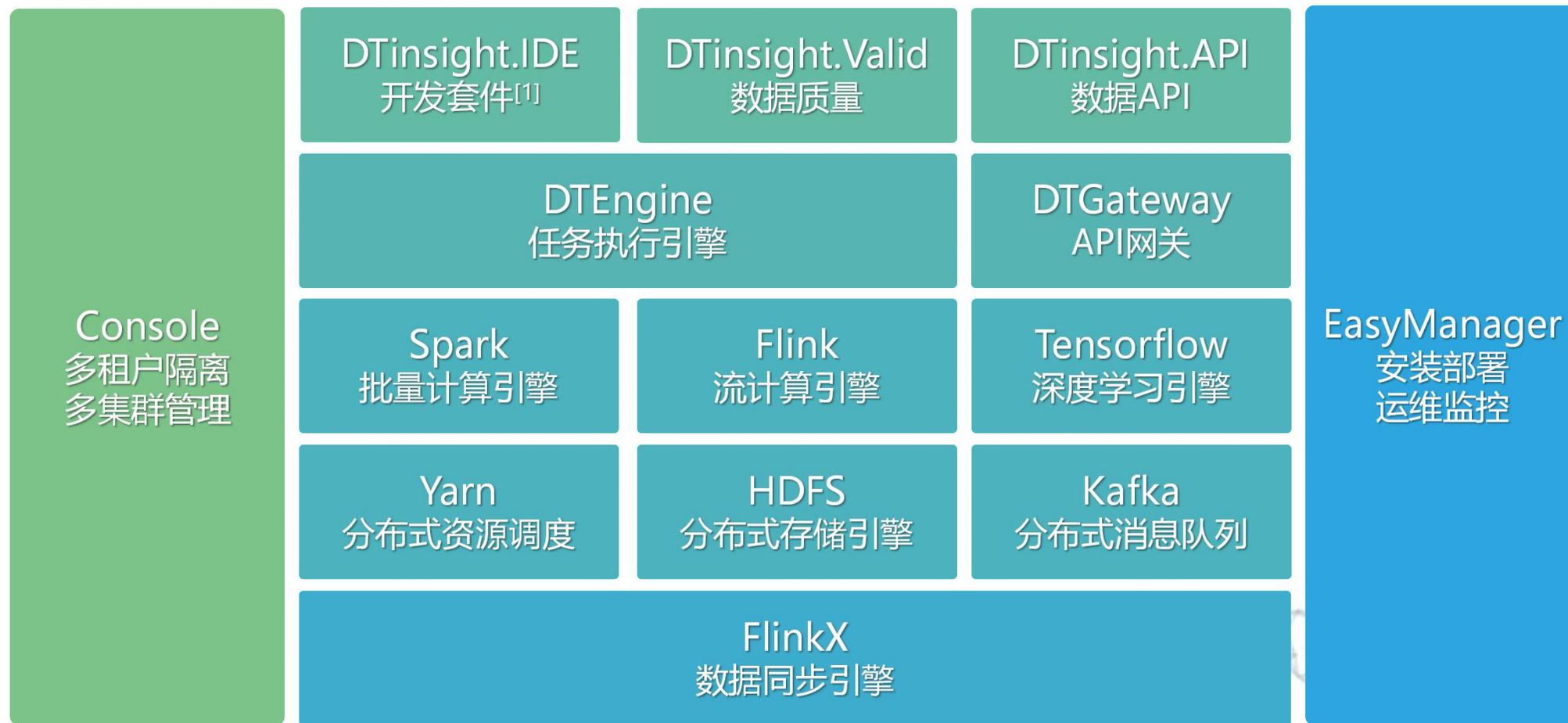
◎ 组织变更, 提升中台部门的战略地位

◎ 数据中台人员配备(数据开发/算法开发/分析师/前后端研发/业务人员)

数栈：企业级一站式数据中台PaaS



数栈产品架构



开源生态兼容

- 兼容CDH、HDP、EMR等第三方开源分布式计算引擎

可视化操作界面

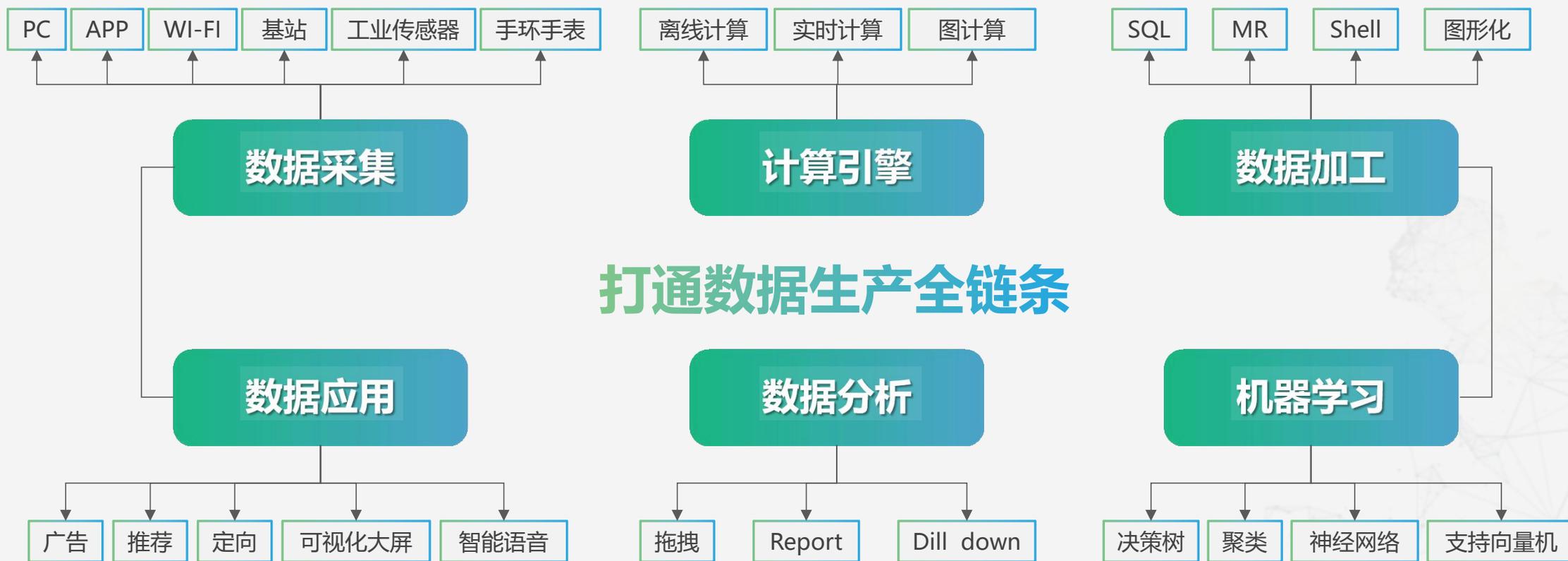
- 屏蔽底层复杂的分布式计算引擎，基于Web的开发平台，提高开发效率

全生命周期覆盖

- 全链路覆盖数据价值释放过程，支持实时、离线和智能计算



一站式大数据平台



案例1：国酒茅台数据中台

数据打通业务

数据智能应用

全局监控

电商大屏

云商大屏

销售大屏

业务异常监控分析

数据化运营

电商决策门户

营销决策门户

服务评价分析

ID识别与画像

价格监测舆情

B2C经营门户

分销经营门户

组织效能分析

产品画像

品牌卫士舆情

官商经营门户

流量分析

电商销量预测

经销商画像

市场洞察舆情

云商经营门户

物流仓储分析

渠道销量预测

经销商参谋

渠道管家舆情

产品顾问舆情

数据植入业务

反黄牛风控

商品推荐

商圈选址

广告营销

业务回流数据

数据智能基础



数据采集

数据埋点、解析入库

官方商城
茅台云商



数据架构设计

数据架构和模型设计

总线矩阵
规范定义



数据研发与运维

编码研发测试与运维

数据基础层
数据公共层



数据连接与萃取

ID识别与连接

用户ID识别
用户标签

基础设施

数据库（关系型、分析型）

数据/算法 计算引擎与存储

中间件（负载均衡等）

2018 AI先行者大会



案例1：国酒茅台数据中台



案例2：中金易云数据中台



云计算：大数据计算能力 × 行业专家：行业深度理解 × 合作伙伴：全面丰富的图书数据

2018 AI先行者大会

关注袋鼠云技术团队
回复“09”获取课件

