



DAMS

中国数据智能管理峰会

DATA & AI MANAGEMENT SUMMIT

京东EB级全域大数据平台建设和治理之路

演讲人：包勇军

Agenda

01 数据的价值

02 行业大数据平台现状

03 京东全域大数据平台的建设历程

04 未来的发展方向





01

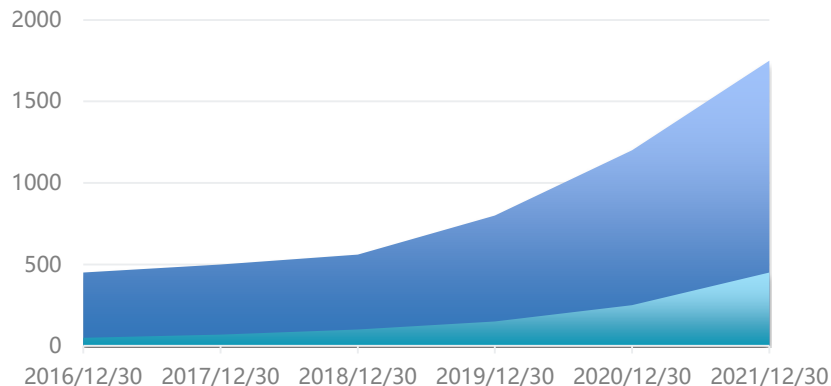
数据的价值

The world's most valuable resource is no longer oil, but data



The Economist (经济学人期刊) , 2017年, 5月版

数据洞察助力业务增长 (\$billion)



■ 上市公司 ■ 初创公司

Forrester报告显示，数据洞察驱动业务发展，领先者相比落后者：

1.78倍业务增长能力、**2.28倍**客户响应速度、**2.40倍**维持竞争优势

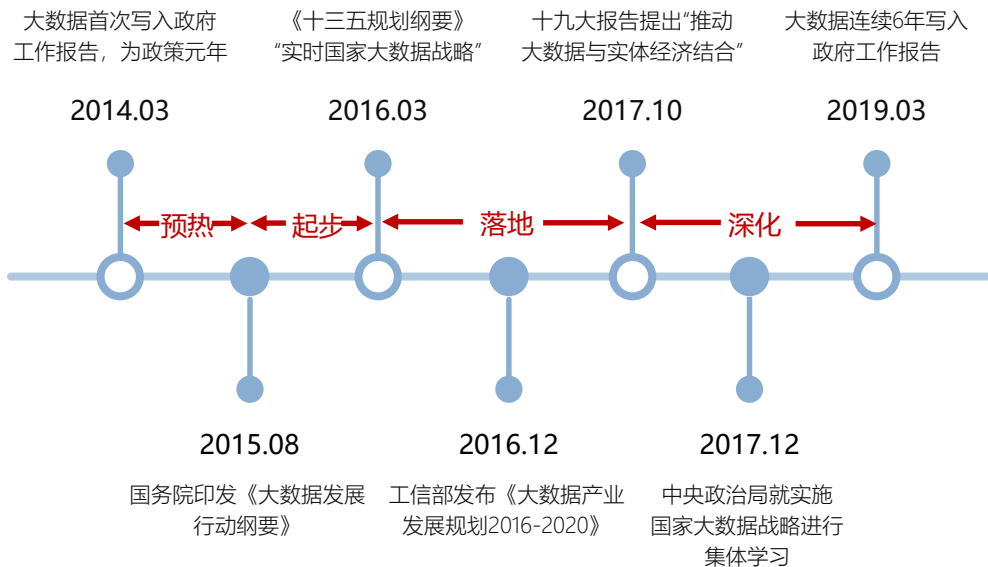
《数据中台2019年行业分析报告发布》

全球市值前十公司变化 科技公司成主流

1990	2000	2010	2020
日本电信电话公司	微软	中国石油	苹果
东京三菱银行	通用电气	埃克森美孚	沙特阿拉伯国家石油
日本兴业银行	NTT Docomo	微软	微软
三井住友银行	思科	工商银行	亚马逊
丰田汽车	沃尔玛	沃尔玛	谷歌
日本富士银行	英特尔	建设银行	阿里巴巴
日本第一劝业银行	日本电信电话公司	必和必拓	脸书
IBM	埃克森美孚	汇丰银行	腾讯
日本联合银行	朗讯	巴西国家石油	伯克希尔哈撒韦
埃克森美孚	德国电信	苹果	VISA

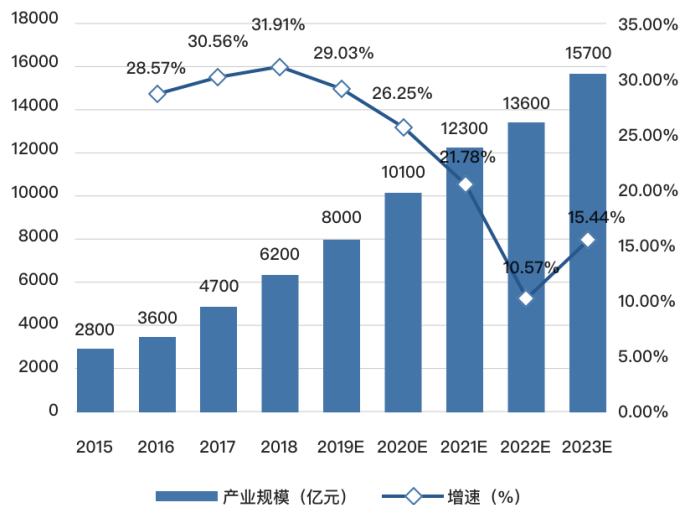
金融投资 通讯及硬件 互联网软件 石油 大众消费

国家“大数据”政策相继出台



大数据市场规模呈稳步上升态势，从19年8000亿元增长至23年的1.57万亿

2015-2023年中国大数据市场产值预测（亿元）



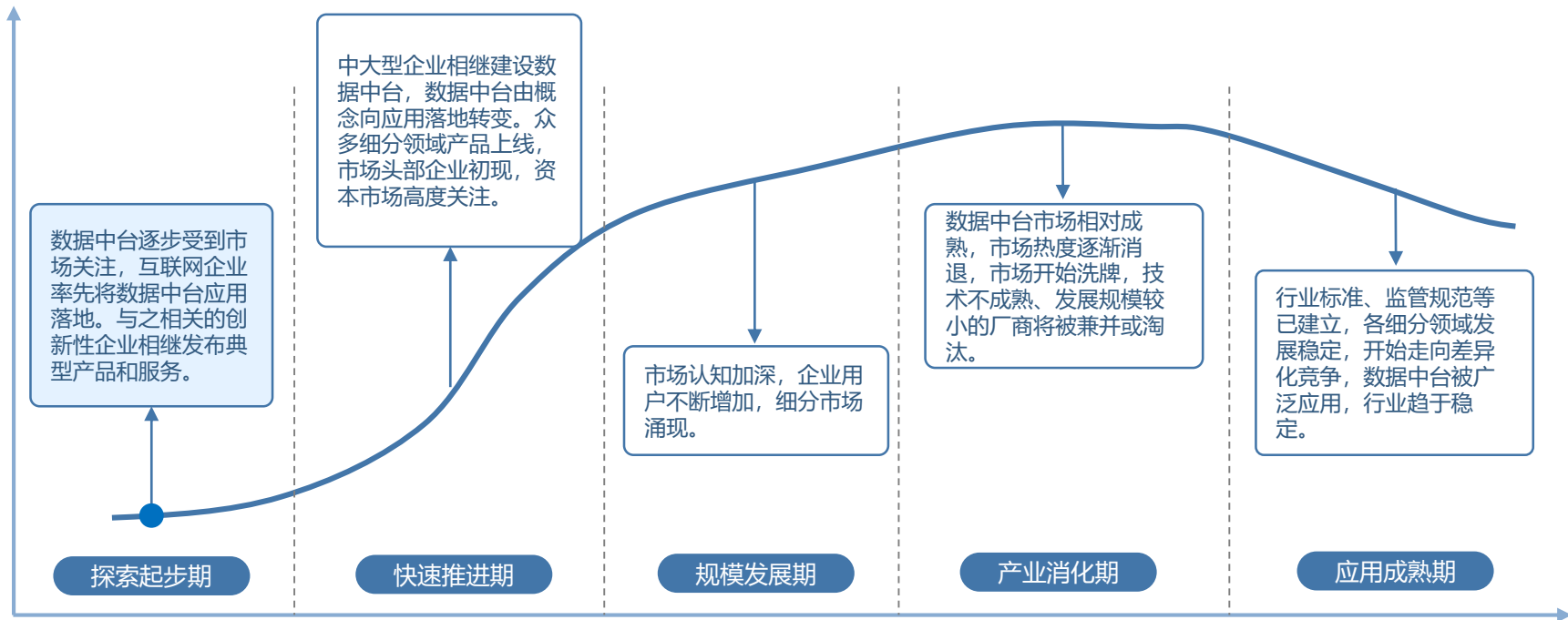
来源于：公开资料整理



02

行业大数据平台现状

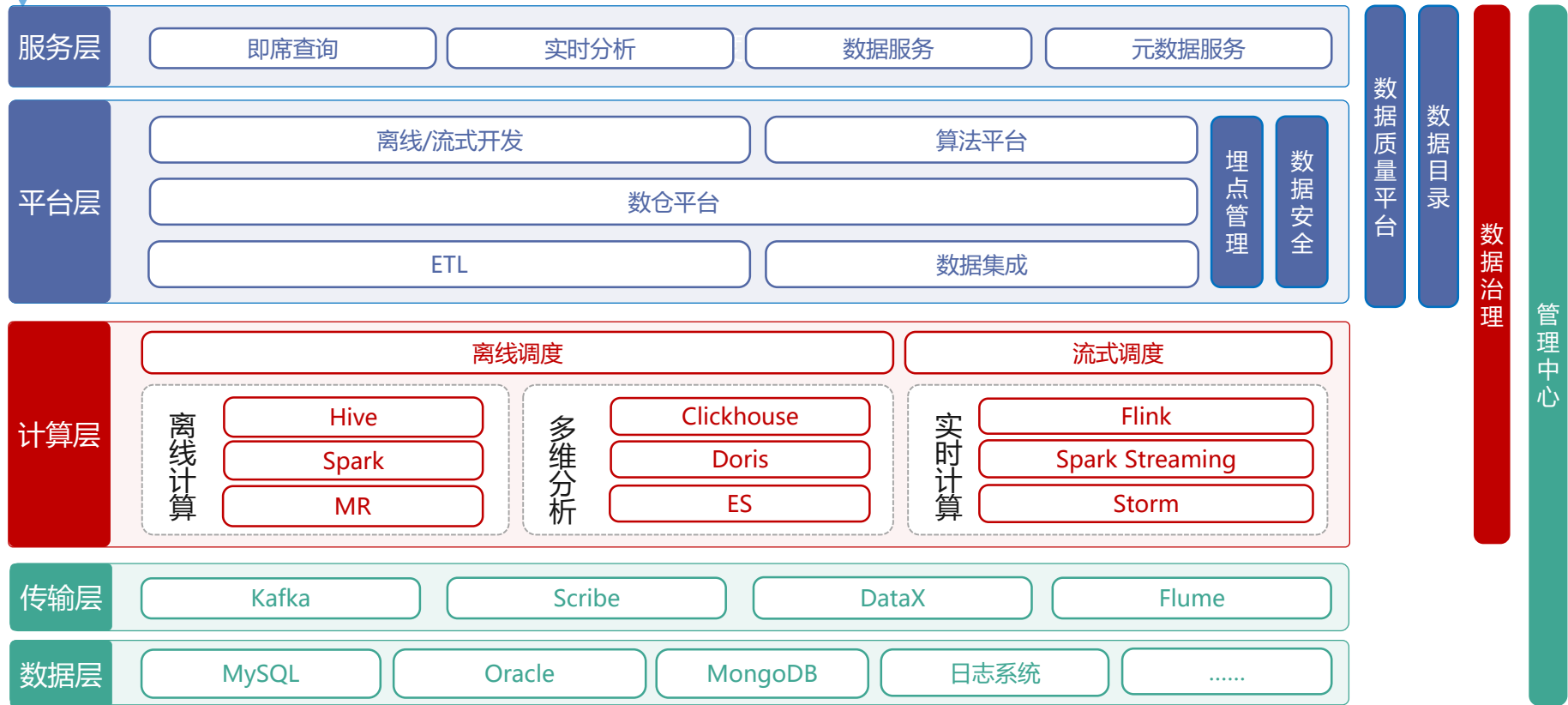
数据平台发展阶段



来源于：公开资料整理

行业大数据平台架构

京东零售数据算法通道委员会



01

数据规模增加：服务器增长的边际效益越来越低

02

如何对不断累积的海量数据去芜存菁，形成正向循环

03

实时数据的需求越来越多，开发门槛高、周期长、落地难

04

缺少“中台”统筹导致烟囱式开发，带来业务困扰和资源浪费

05

指数级数据增长下的永无止境的高时效性需求，从分钟->秒->毫秒

06

传统的规则化大数据分析无法满足更精准的需求

07

疫情、新基建激发大数据建设诉求，政企数字化转型迫在眉睫



03

京东全域大数据平台的建设历程

京东大数据平台总体规模

集群规模

服务器规模**数万**

计算能力

日计算任务**数百万**级别

存储能力

总存储**EB+级**

日增几十**PB**

峰值能力

每秒处理消息数**数十亿级**

处理延时**秒级**

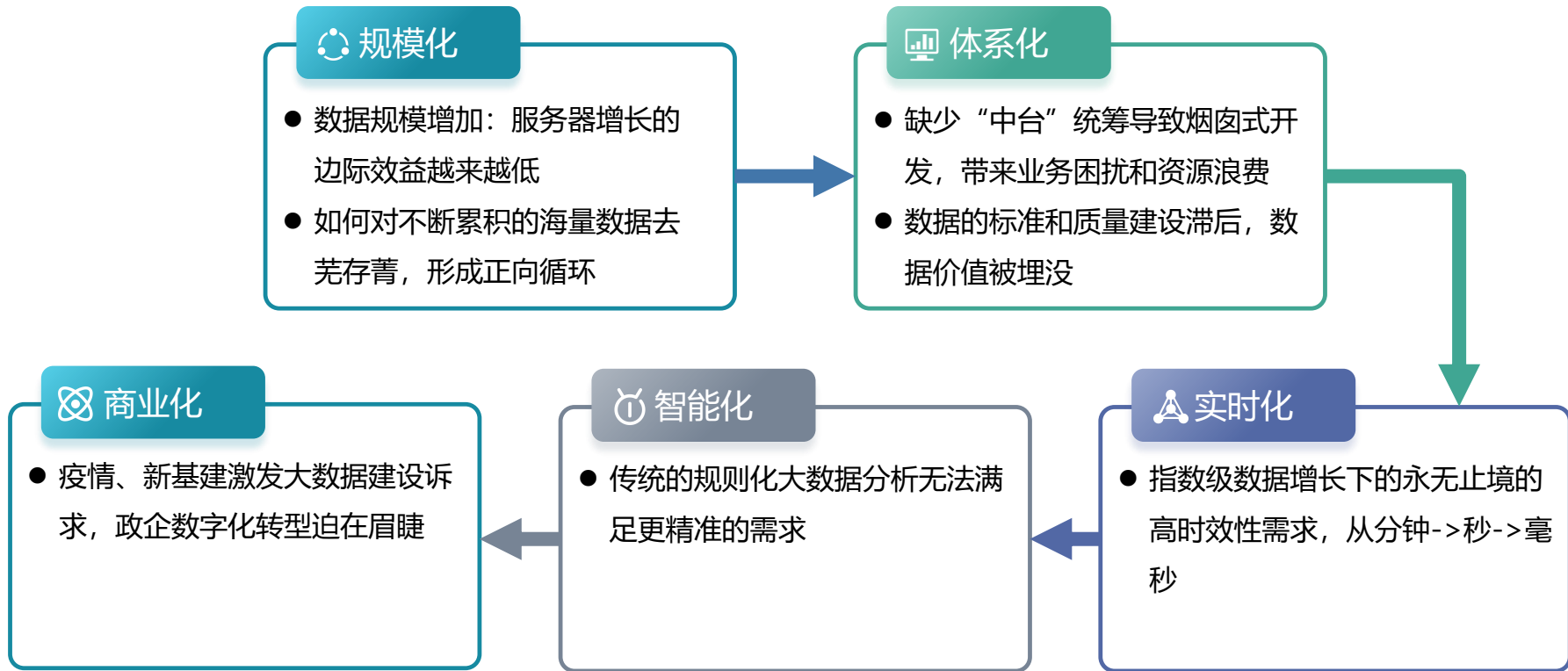


京东旅行

京东国际

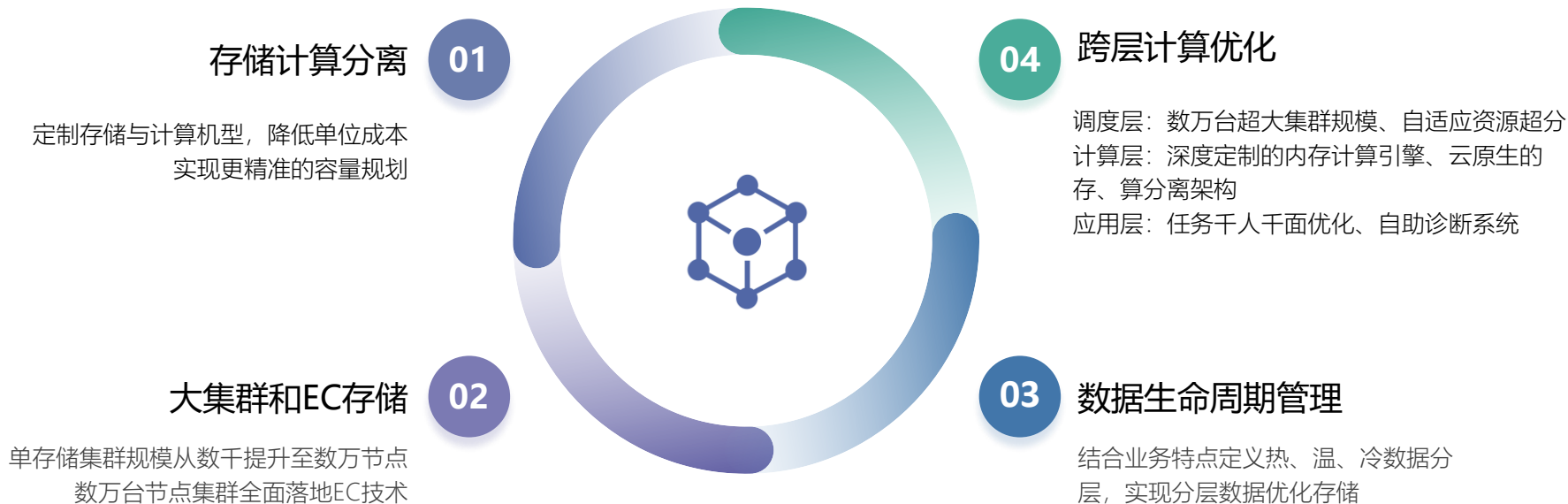
京东之家

海量的存储计算、分析的需求、多业态的业务场景



无论是**堆机器**、**堆集群**、**堆人力**，无法在数据的快速膨胀、业务的高速增长和平台的稳定、易用、高效上取得比较好的平衡





自动化运维运营，通过系统化来解决平台规模化可靠性易用性，达到低碳降本和业务提效，支持业务快速布局和起量



资产盘点

电商全业务域数据体系
统一数据标准和口径



资产治理

资产治理产品化
提供自动化工具



资产分析

端到端全链路资产分析诊断
资产ROI评估

✓ 年节约机器和人力成本数十亿元

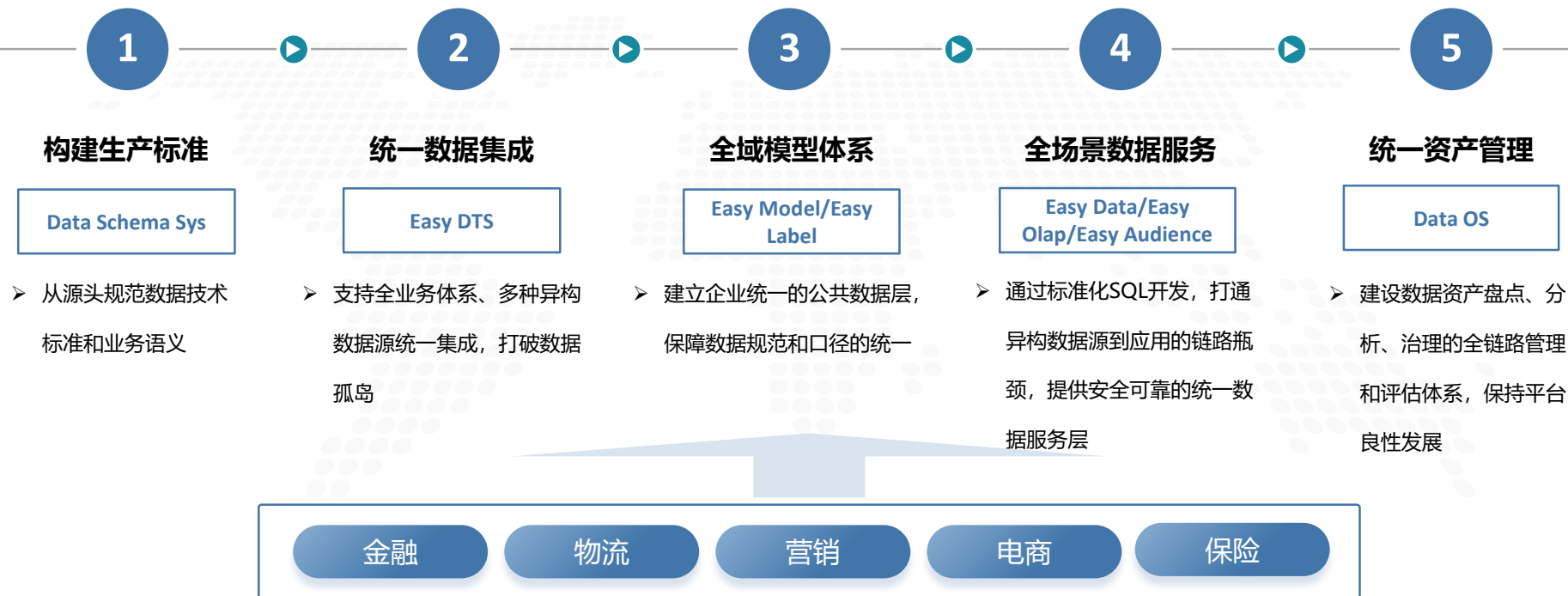
体系化：业务快速并行发展带来的挑战

京东零售数据算法通道委员会

野蛮生长，管理成本急剧增加



- ✓ **数据烟囱：**重复建设，缺乏连接、无法形成合力
- ✓ **数据腐化：**随着时间流逝，数据价值流失
- ✓ **缺乏治理：**高数据价值甄别难
- ✓ **来源繁杂：**结构化、非结构化数据、多种异构数据源
- ✓ **业务复杂：**需求扩张速度与应接能力矛盾



01

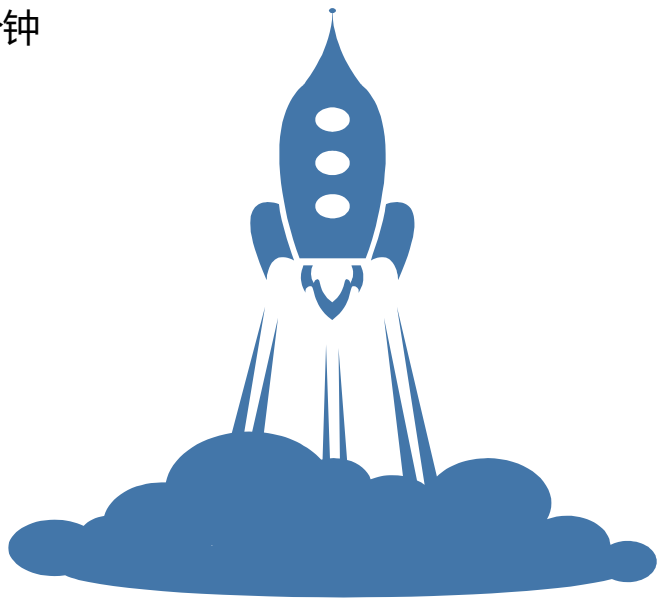
基于海量数据的秒级决策，高时效性需求：天->分钟
->秒->毫秒

02

大促洪峰压力

03

实时计算技术开发推广



实时化：解决方案 Easy Realtime实时计算平台

京东零售数据算法通道委员会

低延迟

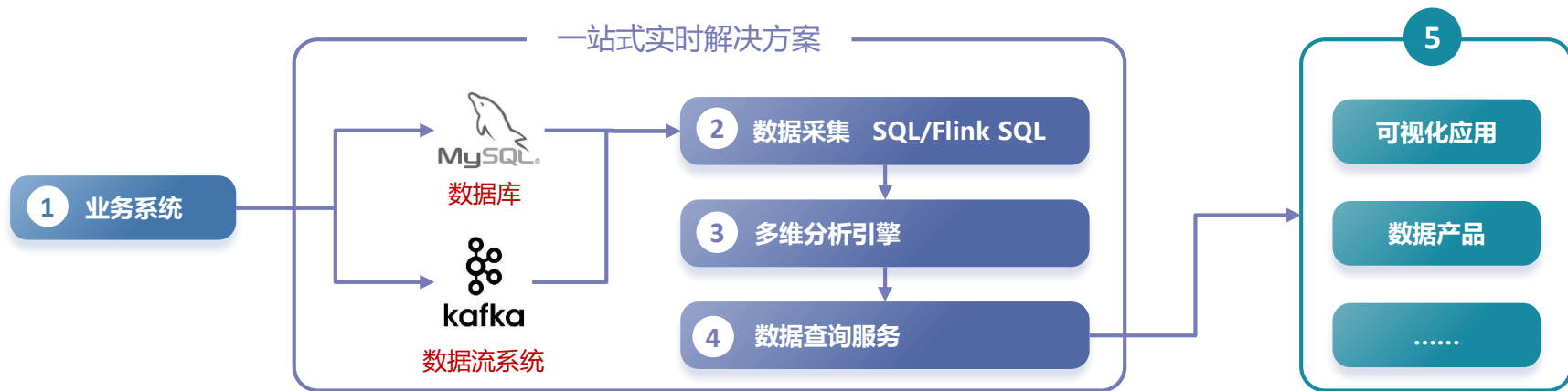
- 全方位深度定制的实时计算引擎
- 全链路的实时采集、传输、计算分析一体化计算框架

高可用

- 云原生资源调度、多租户资源隔离、自适应的跨机房容灾
- 全链路的实时监控自愈，实现分钟级故障恢复

易用性

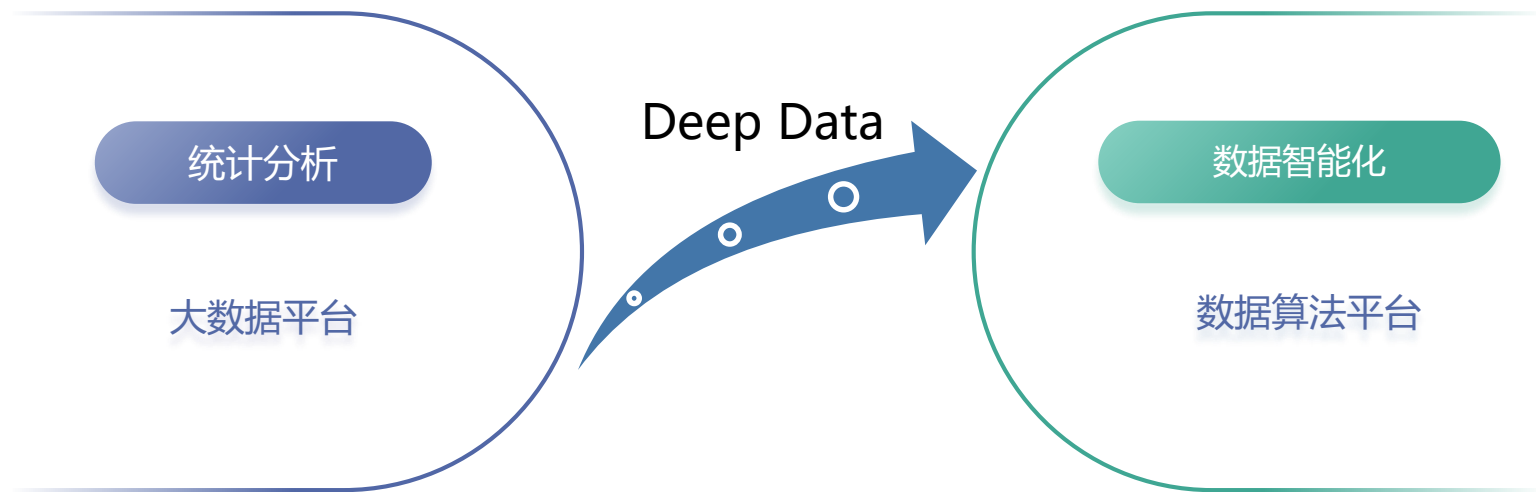
- “零代码”开发
- 一站式融合数据服务平台



智能化：对数据的深度理解是业务再增长的动力源

京东零售数据算法通道委员会

打造以高效精准业务决策为目标的数据智能化服务





挑战

01

大数据下的超大规模机器学习算力挑战

- 高性能的算法平台九数研发，提供了一套成熟的工业级解决方案

02

数据安全背景下的跨业务实体数据跨融合分析

- 研发联邦学习数据交换平台，为京东生态合作提供数据交换避难所

03

数据多模态的大量涌现，组织越来越复杂，应用越来越灵活

- 打造伽利略图计算框架，支持异构多数据类型的组织分析，可支撑数十亿节点、数百亿边的数据规模

智能化：九数(9N)商业分析和业务智能化平台

京东零售数据算法通道委员会

京东零售

京东健康

京东数科

京东物流

全渠道

.....

9N服务平台

精准匹配

时序预测

趋势分析

关联分析

9N算法诊断平台

仿真系统

效果诊断

9N核心算法引擎

9N-Deep
(模型训练引擎)

Galileo
(图计算引擎)

9N-FL
(联邦学习引擎)

9N-RL
(强化学习引擎)

9N-OL
(在线学习引擎)

9N-Cloud算法资源云化管理系统

一站式算法解决方案

● 赋能业务

- 服务京东全业务场景，助力业务高速发展

● 典型案例

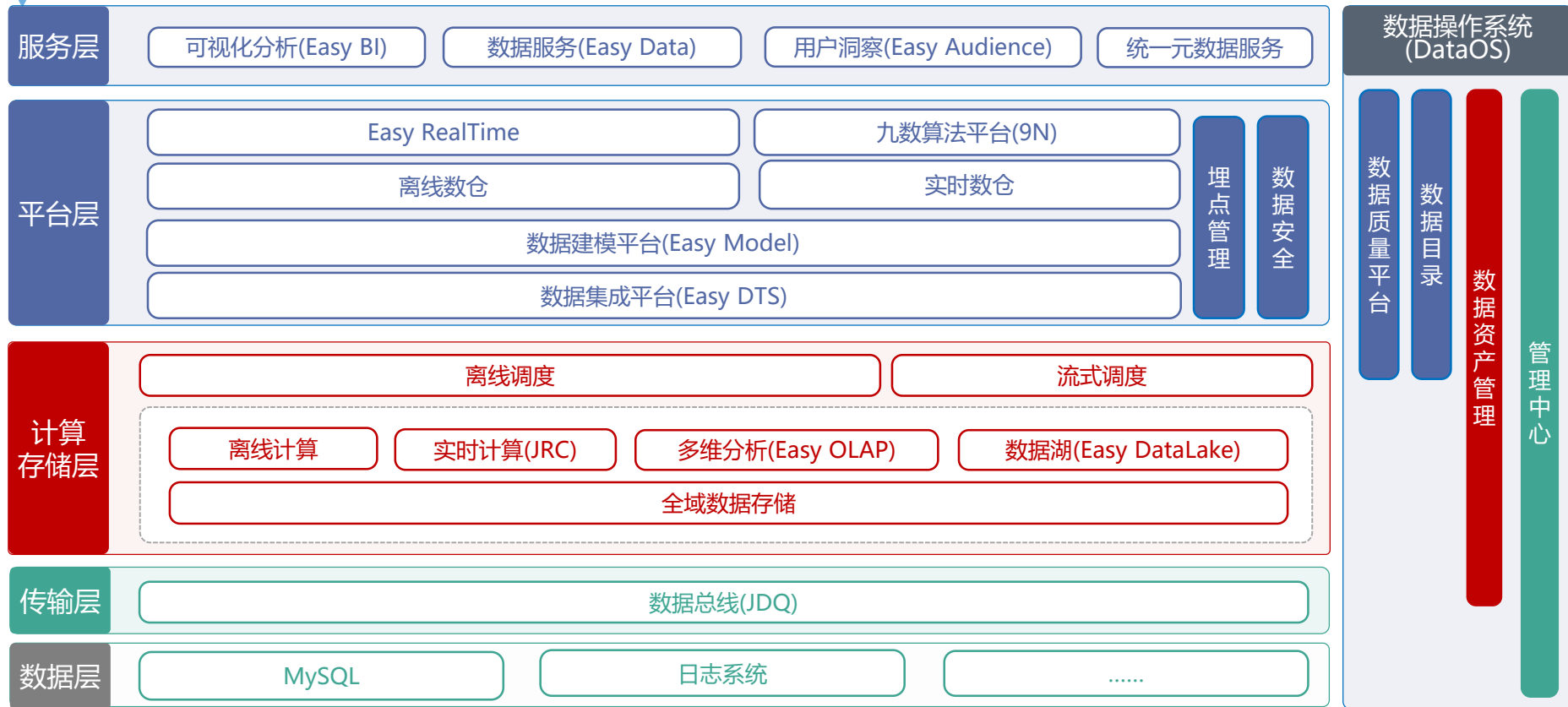
- 助力京东广告从规则化到数字化智能化的转型，加速无人投放广告高速发展，已经成为京东广告核心收入来源

● 开源回馈社区

- 9N-FL已开源
- Galileo图计算框架即将开源
- 9N算法平台整体开源计划中

京东全域大数据平台架构 Easy JData

京东零售数据算法通道委员会





04

未来发展方向



DAMS

中国数据智能管理峰会
DATA & AI MANAGEMENT SUMMIT

Q&A





DAMS

中国数据智能管理峰会

DATA & AI MANAGEMENT SUMMIT

THANK YOU!

