

中国大数据分析平台 行业研究报告

©2022.12 iResearch Inc.



行业界定：大数据分析平台逐渐由产品态转向集成态，行业边界模糊。在技术架构上，主要包含数据采集与存储、计算、分析与决策三个层级。在 OLAP 之上融合了深度学习等技术，在提升数据分析深度和广度的同时，也极大增加了数据服务在业务侧的低门槛和友好性，满足用户运用数据分析驱动业务发展的需求。



市场情况：尽管行业边界泛化，市场参与者众多，但按照部署模式、架构分类及能力补给，可分为以下五类：1) 以云上数据湖方案为主的公有云厂商；2) 以本地化大数据分析平台为主的传统软件服务商；3) 提供轻量化数仓架构的数据库/数仓厂商；4) 为数据应用层提供服务能力的软件供应商；5) 提升数据应用能力的人工智能厂商。行业市场整体呈现竞合状态。



架构选型：搭建平台前用户首先需要明确自身的数据体量和业务场景需求。在明确大数据分析平台需要具备的基本功能后，再决定平台搭建过程中使用的大数据处理框架和工具。在分层架构中，数据分析层的组件选型和整体搭建十分关键，尤其是存储引擎的选型直接决定了离线、在线、实时三大场景的支撑和算力效率的高低。



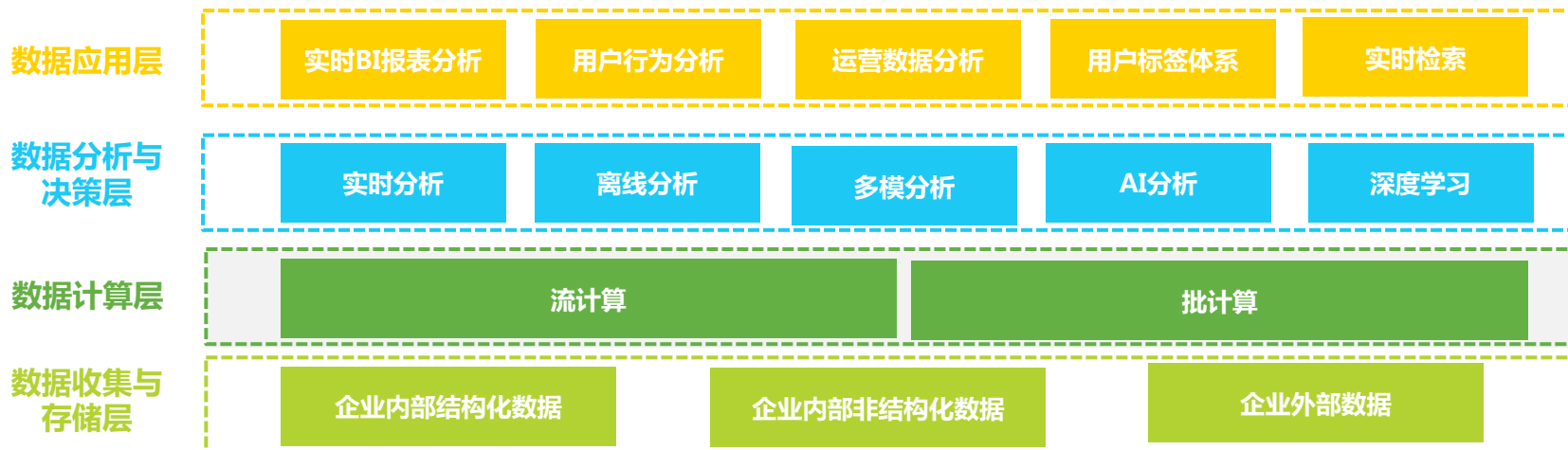
趋势洞察：传统架构下的湖仓分体引发数据孤岛，造成实施、运维和成本问题。湖仓一体架构在数据和查询层面形成一体化架构，突破实时性和并发度、集群规模受限、非结构化数据无法整合、建模路径冗长、数据一致性弱等瓶颈。同时，平台融合 AI 自主学习和自适应能力，增强用数人员的分析和决策能力。

大数据分析平台行业概述	1
大数据分析平台市场分析	2
大数据分析平台构建建议	3
行业应用与典型案例实践	4
大数据分析行业投资分析	5

驱动业务的全场景数据分析平台，提供实时、多维的数据分析和智能决策

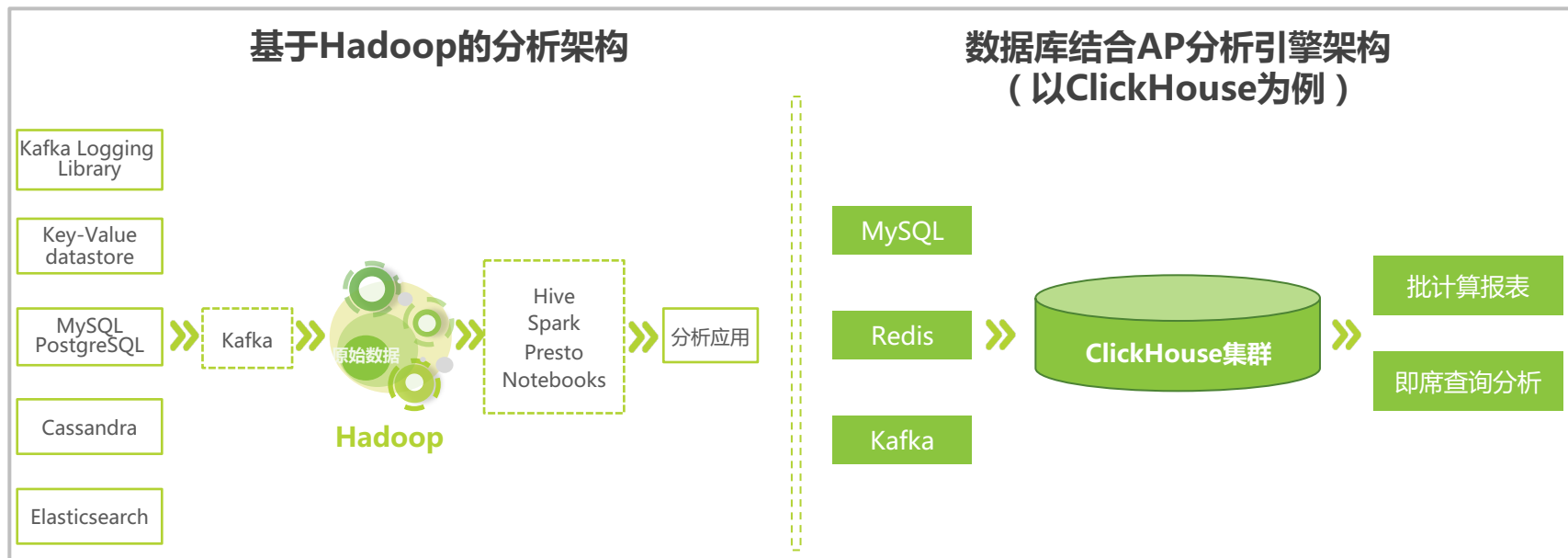
大数据分析平台，是企业用户在大数据环境下用于分析与决策的平台。按技术架构划分，主要包含数据收集与存储、数据计算、数据分析与决策三个层级。从服务边界来看，大数据分析平台概念小于数据中台，强调平台的数据分析与决策能力，弱化了数据本身的规划、治理与服务；在 OLAP 之上，又融合了深度学习等技术，在提升数据分析深度和广度的同时，也极大增加了数据服务在业务侧的低门槛和友好性。企业通过构建大数据分析平台，聚拢各业务系统数据，打通全渠道组织各业务维度，用数据分析驱动业务，满足企业级宽表实时分析、实时 BI 报表分析、用户行为分析、自助分析、AI 智能分析等全方位需求。

大数据分析平台技术框架及核心组件



技术沿革（一）

平台技术架构持续更新迭代，由离线处理向实时分析演进



架构剖析

• 基于 Hadoop 分析架构的流程原理：

各类结构化数据通过采集管道进入 Kafka，Spark 实时消费 Kafka 的数据，写入集群内的 HDFS，RDS 数据库中的数据通过 Spark 每天一次全量扫表同步至 HDFS。HDFS 存储汇总用户数据，对数据库数据定期执行 snapshot。

• 基于 Hadoop 分析架构的优缺点：

优点：借助 Hadoop 集群的高并发能力，实现百 TB 到 PB 级数据的离线计算和处理，同时数据存储在 HDFS 上，存储成本低。

缺点：数据定期入库，数据计算的时效性通常是 T+1。

架构剖析

• 数据库结合 AP 分析引擎架构的流程原理：

将平台架构引入 TP 引擎结合 AP 引擎实现实时分析平台，各类结构化数据同步至分析引擎后可进行交互分析。

• 数据库结合 AP 分析引擎架构的优缺点：

优点：舍弃了传统离线大数据架构，实现实时批量计算，在 GB 到100TB 级别的计算有了很大提升，BI 人员无需等待 T+1的离线计算后得到最终结果，大幅提升数据资产的商业价值。

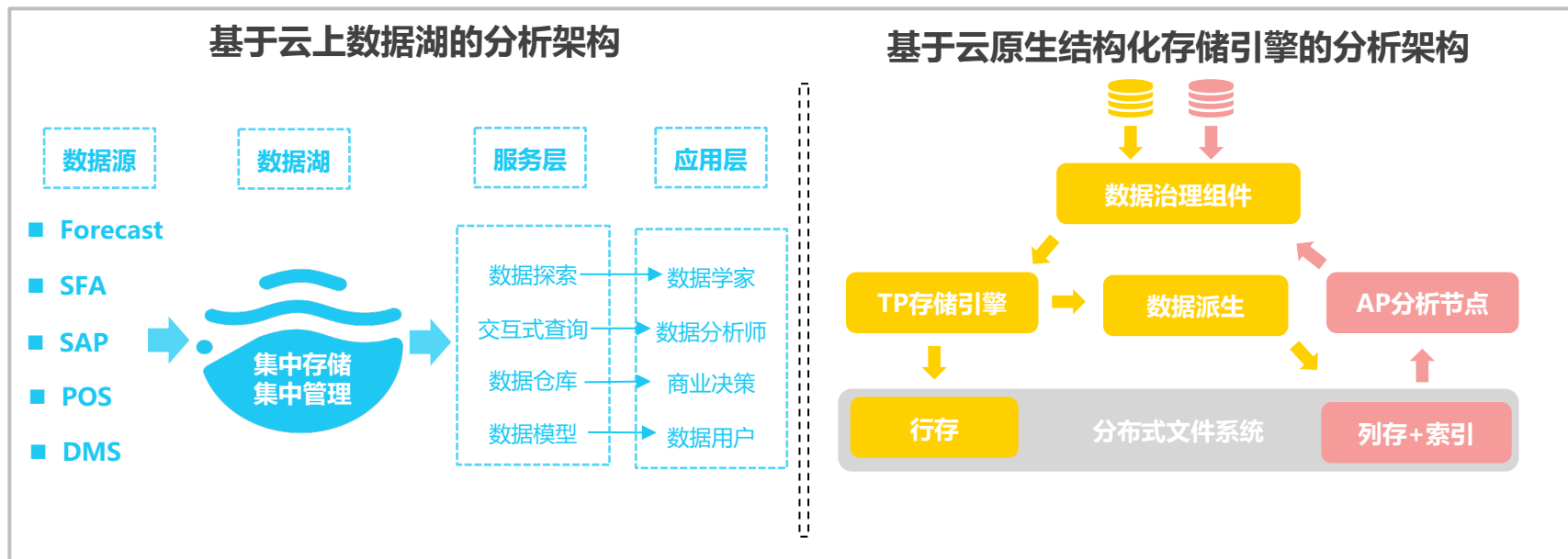
缺点：在处理百 PB 级以上数据时，ClickHouse 架构的扩展能力、复杂场景计算和存储成本相对 Hadoop 方案较弱。

来源：艾瑞咨询研究院根据公开资料整理及绘制。

来源：艾瑞咨询研究院根据公开资料整理及绘制。

技术沿革（二）

平台技术架构持续更新迭代，产品在云上落地和升级



架构剖析

基于云上数据湖的分析架构的流程原理：

可理解为借助云原生存储引擎，基于传统 Hadoop 方案的云上落地和升级，保留自建 HDFS 集群的分布式存储可靠性和高吞吐能力，借助数据湖降低传统方案的运维和存储成本。

基于云上数据湖的分析架构的优缺点：

优点：对大数据平台的使用者做了区分和定义，针对不同的使用场景，数据的使用方式，分析复杂度和时效性也会有不同。

缺点：数据湖方案本身并没有解决传统方案的所有痛点。

架构剖析

基于云原生结构化存储引擎的分析架构的流程原理：

将类似第二阶段和第三阶段的融合，在线库和分析库隔离，不依赖在线库数据；全量数据支持高效批量计算，分析结果集支持即席查询，支持实时写入实时流计算。

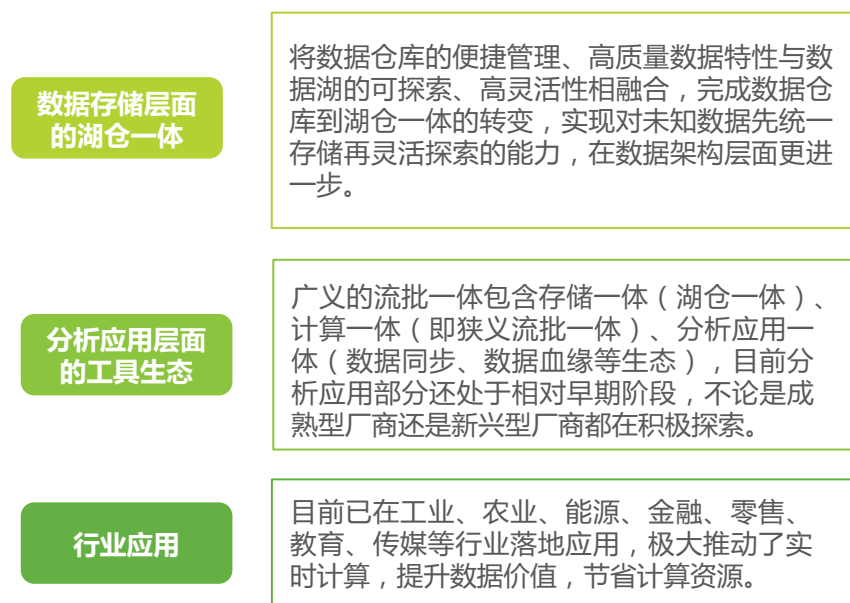
基于云原生结构化存储引擎的分析架构的优点：

优点：在具备宽表合并高吞吐低成本存储的同时，可以提供 TB 级别数据即席查询和分析的能力，无需过度依赖额外的计算引擎，实现高效实时分析能力。

流批一体：统一开发、统一计算、逻辑一致、降低成本

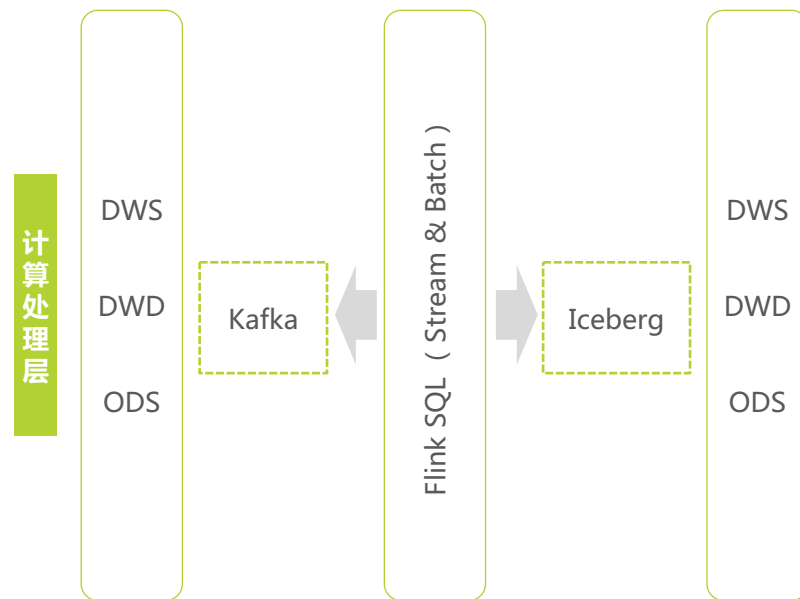
采用流计算+交互式分析双引擎架构，流计算负责基础数据，交互式分析引擎是中心，流计算引擎对数据进行实时 ETL 工作，与离线相比，降低了 ETL 过程的 latency。交互式分析引擎自带存储，通过计算存储的协同化，实现高写入 TPS、高查询 QPS 和低查询 latency，从而做到全链路的实时化和 SQL 化，实现用批的方式做到实时分析和按需分析，并能快速响应业务变化，两者配合实现1+1>2的效果。流批一体实现了建立一套统一的系统，由同一个开发团队开发，同时支持流式计算和批量计算，提供一致的编程环境，降低开发和运维成本，减少资源浪费，提高数据口径的一致性。

流批一体的技术趋势及行业应用



来源：艾瑞咨询研究院自主研究及绘制。

流批一体的技术框架



来源：艾瑞咨询研究院自主研究及绘制。

核心产品（一）

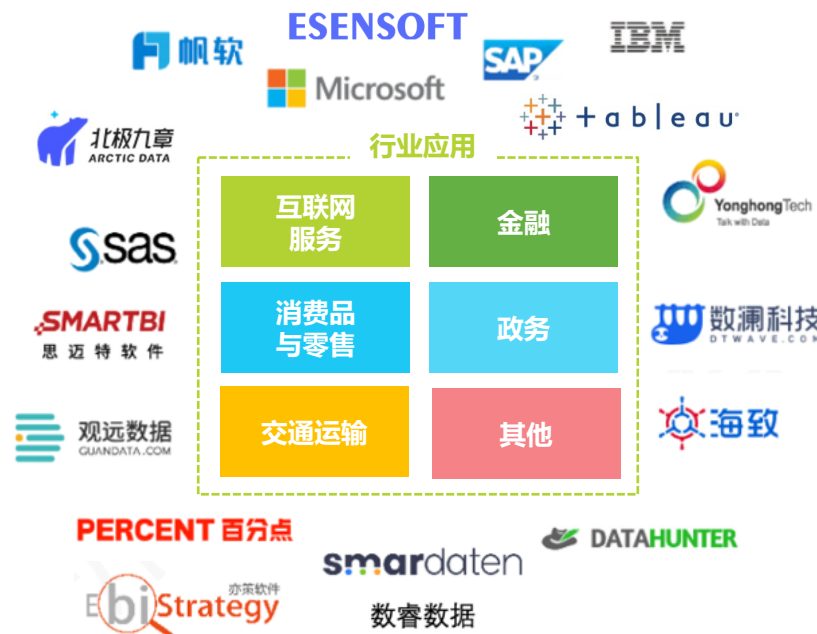
商业智能 BI：通过数据整合分析实现商业价值

商业智能（BI，Business Intelligence）是大数据分析最典型应用领域，是由数据库、数据仓库、数据湖、湖仓一体、ETL、OLAP、数据挖掘、机器学习和人工智能等技术组成的一套完整解决方案。随着大数据处理技术的发展，商业智能的洞察和分析能力进一步提升，数据分析和可视化的门槛不断降低，企业实现不同层级的拖拽式自助分析和多种类型的图表展示，并在统一平台进行整合和共享，获得不同层级的数据洞察，最终用于商业决策。机器学习和人工智能在商业智能中扮演越来越重要的角色。

BI 的技术发展趋势

BI SaaS化	云上落地是商业智能最大的技术发展趋势
一站式平台化	商业智能趋于集成数据仓库提供存储功能，集成 python 及 R 语言提供数据挖掘，延伸范围越来越广
BI 与新技术融合	商业智能与流程自动化 RPA 和人工智能等新技术深度融合
自助式 BI	业务人员可独立分析数据，生成定制报告，不再依赖 IT 部门或数据分析师
协作 BI	商业智能与社交媒体和网络技术等协作工具融合，允许共享报告，增加涉众和专家互动，提升商业决策质量

BI 的行业应用及典型企业



来源：艾瑞咨询研究院自主研究及绘制。

来源：艾瑞咨询研究院自主研究及绘制。

核心产品（二）

数据孪生与增强分析：释放数据潜力，加强数据价值转化

数据孪生：利用物理模型、传感器更新、运行历史数据，集成多学科、多物理量的仿真过程，在虚拟空间中完成映射，在信息化平台创建虚拟的“数字孪生体”，融合多源异构数据，打通企业数据孤岛，让数据在业务侧发挥更大价值。

增强分析：2017年由 Gartner 首次提出，并将其定义为下一代数据和分析范式，通过将机器学习和人工智能运用于现有的操作流程中，使数据管理和分析自动化，从而更有效地进行数据分析。增强分析使更多的用户获得更深入的数据洞察，减少了当前依赖 IT 处理所带来的效率问题和口径偏差。

数据孪生的技术趋势及产品应用

增强分析的技术趋势及产品应用

与5G万物互联场景强耦合

通过感知采集数据，对物理实体进行动态监测和即时描述，可视化展现，提供系统内部生成的信息或对象质效评价信息，用作系统故障排除工具，优化改进性能，揭示要素间的复杂关系，预测物理实体未来状态，辅助操作和决策。

纳入智慧城市的顶层框架

随着数字孪生技术日趋成熟，基础设施的运行状态和市政资源的调配情况都可由感知件传入数字孪生城市，帮助城市管理者分析决策。城市的长“生命周期”特性，让数字孪生发挥更加深远的价值。

与智能驾驶产业创新融合

在智能驾驶领域，数字孪生测试系统对整车测试场景虚拟重构，在虚拟场景中通过物理仿真工具对测试对象和流程做全程复现，评估分析测试结果，实现不同场景的测试。

增强数据准备

一是可视化交互，通过拖拉拽的方式实现可视化的数据配置、数据源的混合以及数据清洗工作，让数据准备更加快捷；二是算法辅助，利用 ML 和 AI 技术实现部分流程的自动化。

增强数据分析

一是自动洞察，代替一部分分析师的工作，从数据中发现潜在信息和价值；二是自动可视化，根据数据分析结果自动选择可视化的方式进行展示，与 NLQ、NLG 等技术配合，将大大加快整个分析流程。

增强机器学习

通过算法将特征工程、模型选择与参数优化，以及深度神经网络结构搜索等机器学习过程中的关键步骤自动化，帮助数据科学家更高效地得到满意的模型。

产品应用

主要应用于 BI 领域，作为 AI 深度分析模块集成 BI 产品，使数据洞察更为精准化、自动化、智能化。

以强大算力提升用户的数据洞察分析，推动行业整体发展

1) 放大数据分析价值：大数据分析平台基于集体智慧的分布式数据驱动决策，使用者的数据洞悉分析能力、用数效率和决策产出得到极大提升，数据资产的商业价值充分体现。2) 降低分析门槛：低代码和无代码的分析工具极大降低了业务人员使用数据分析的门槛，有利于企业形成数据驱动文化，提高数据驱动效能，让企业各个层级的人员都能够参与到数据分析当中。3) 降低分析成本：大数据分析平台缩短了从数据提取到离线分析，再到报告制作的周期，无需重复提数，边际成本趋近于零，显著降低了时间和人力成本。

大数据分析平台核心价值



1. 展示

关键数据指标以数据报表、DashBoard 等可视化方式呈现，便于企业整体把握业务核心数据及发展趋势。

2. 解读

通过简单分析和解读产品或用户行为的一些现象或数据变化，了解现象发生或数据波动的原因。

3. 归因

收集多方面的复杂原因，提前解决矛盾以避免问题恶化或提升数据指标，总结沉淀为对产品和用户的深刻洞察。

4. 预测

对产品或用户行为的未来趋势做出预判，指导经营活动，例如：预测 DAU、销量、订单量等。



1. 用户保留与发展

指标拆解分析，发现问题，寻找突破口，帮助企业制定精准方案获取新用户，并提升老用户的满意度。

2. 商业价值转化

生成数据报表和实时监控指标体系，拆解、挖掘、构建用户画像，进行日常运营决策，提升经营业绩。

3. 产品研发与升级

数据湖方案较其他大数据产品更强调“海量异构数据统一存储、多源数据统一管理、多计算引擎统一调用”的能力。

4. 产业链供应链协同

链接产业链全要素，实现供产销信息的实时同步，优化产能配置，提升企业的柔性生产能力，提升产业链效率。

5. 风险管理

通过分布式即席查询、数据可视化等手段，为企业提供风险识别并持续监控，主动探索风险应对策略并评估控制效果。

构建满足场景需求的大数据分析平台，选型评估十分关键

1) 架构可扩展性：企业构建大数据分析平台之初以小规模项目起步，待业务规模增加后再考虑复杂的解决方案。这时架构的可扩展性就显得十分重要，确保业务数据规模上升后架构横向扩展的能力是关键。2) 数据实时性：随着流批一体等技术的不断完善和推广，数据的实时处理、分析和输出展示都显得十分重要。3) 数据模式灵活性：在企业初创阶段，数据经创新探索后才能沉淀，灵活度要求较高，数据湖架构较为适用；企业成熟阶段数据规模和处理成本上升，平台成长性决定了业务发展的持续性。4) 数据可接入性：企业构建大数据分析平台时应提供给开发者丰富、开放及资料完整的应用程序 API 接口。

产品功能

系统能力

系统集成能力：

关系型数据、非关系型数据、爬虫数据，离线数据和实时数据及实时数据处理能力

系统开发能力：

对离线任务和实时任务的支持、开发调度配置等

数据资产管理能力：

数据标准管理、元数据管理、数据质量管理等

数据科学能力：

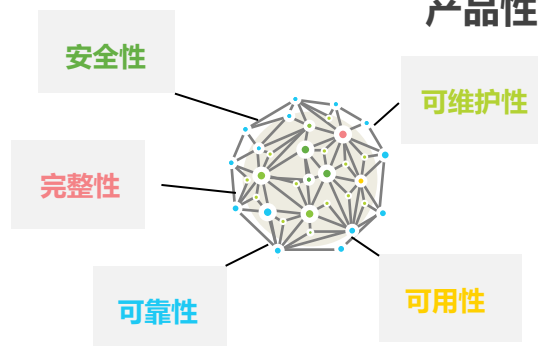
数据上传、预处理、特征工程、模型评估、模型发布等功能，主流算法框架和语言等

数据应用能力：

API 创建、生成、发布、执行、审批、鉴权、限流等功能，标签开发、标签分类、标签目录管理、业务确权等功能

数据能力

产品性能



重点考察

最大节点数量、吞吐能力、并发能力、运算速度、相应时间、安全性等

公司规模	厂商状况和业务规模影响其能否为企业提供长效服务
咨询能力	方案咨询和定制能力取决于垂直行业的knowhow积累
服务案例	厂商同行业客户服务经验、标杆案例和案例数量等
市场价格	项目价格包括产品费用、项目实施费用和维护服务费用，也是选择厂商的重要指标

来源：艾瑞咨询研究院自主研究及绘制。

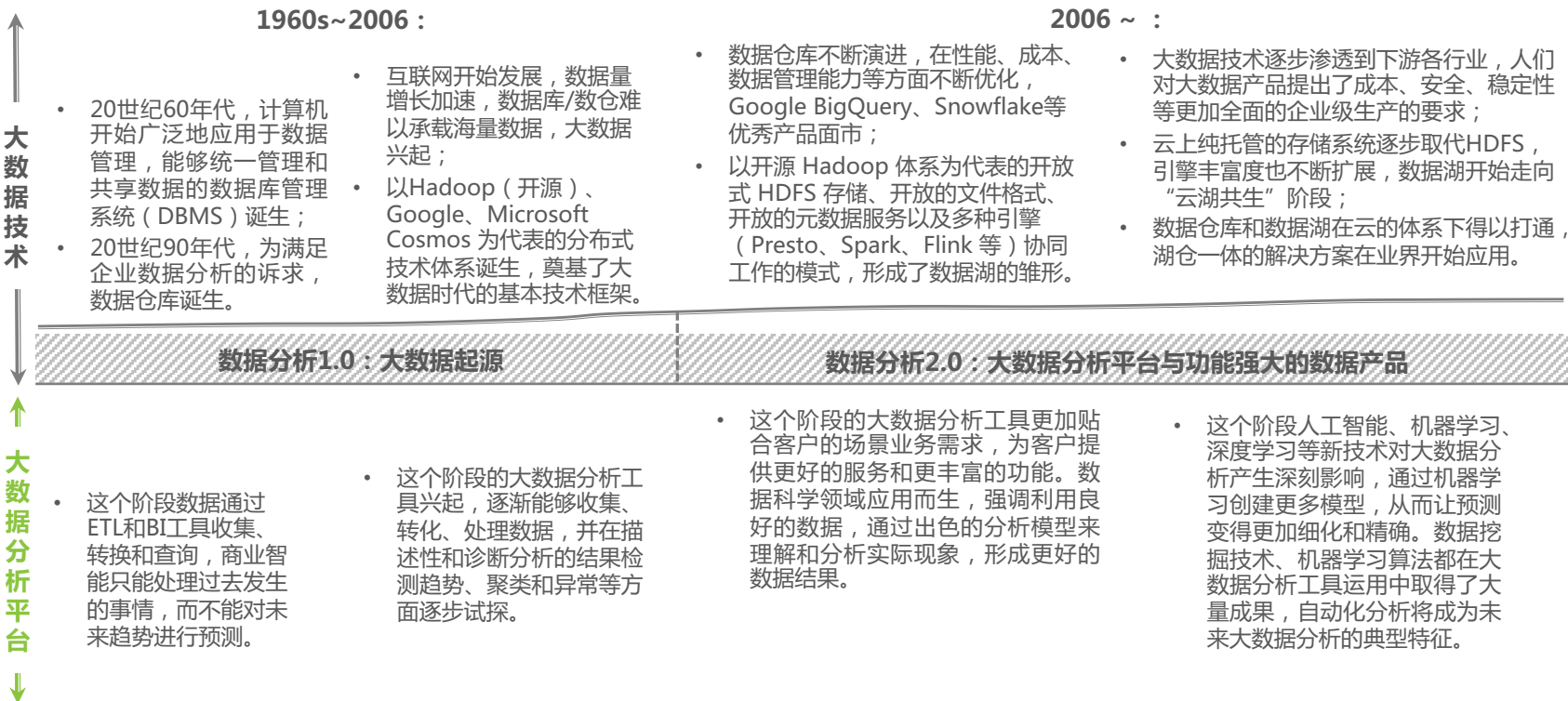
来源：艾瑞咨询研究院自主研究及绘制。

大数据分析平台行业概述	1
大数据分析平台市场分析	2
大数据分析平台构建建议	3
行业应用与典型案例实践	4
大数据分析行业投资分析	5

大数据分析平台随技术革新和市场需求不断演进

随着增强分析演进到了人工智能时代，未来将成为大数据分析平台的核心特性。湖仓一体技术不断完善，未来也将成为大数据分析的基础底座。流批一体技术让有界数据和无界数据实现了打通，让数据分析获得了更好的一致性。中国大数据分析平台正在逐年发展及突破，公有云厂商及其他行业厂商纷纷在做尝试。我国大数据分析市场整体处于高速发展阶段，未来发展空间广阔。

中国大数据分析平台行业发展历程



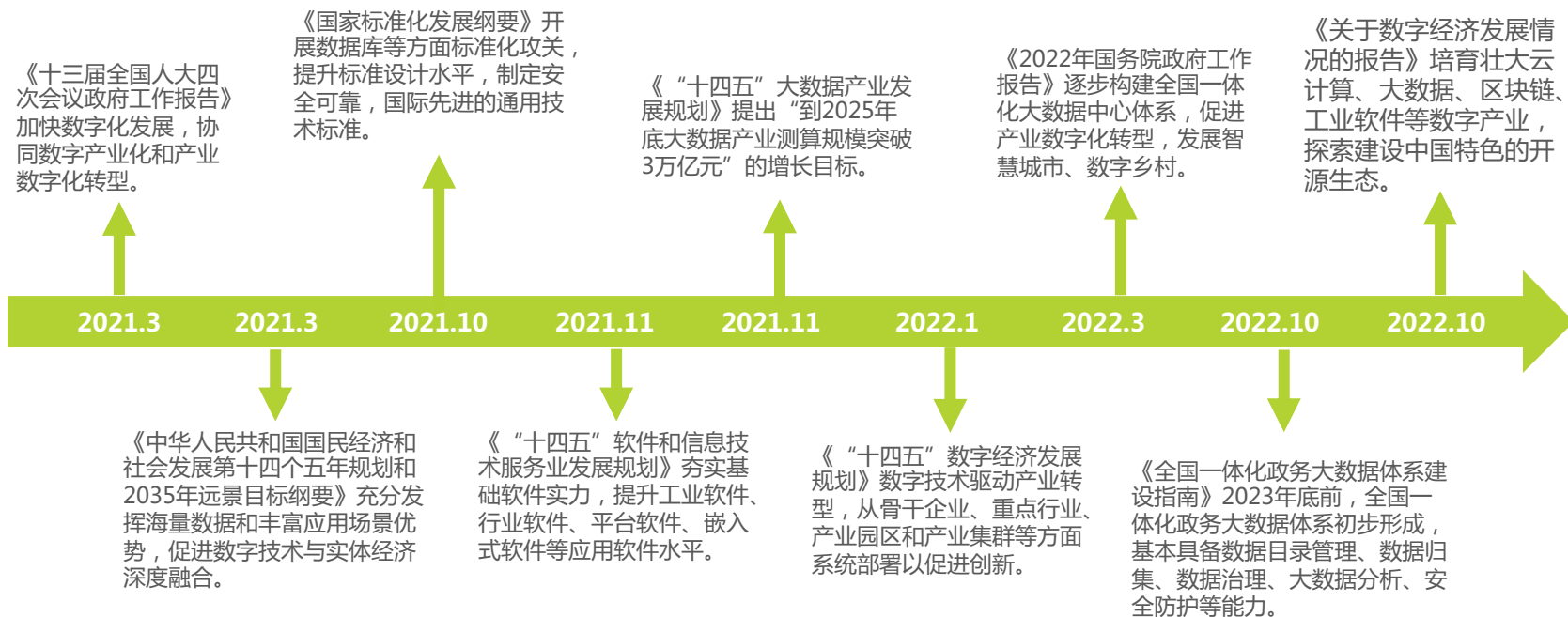
来源：艾瑞咨询研究院自主研究及绘制。

驱动因素：政策面

政府不断出台大数据相关政策，促进产业发展

随着5G、云计算、人工智能等新一代信息技术发展，信息技术与传统产业加速融合，数字经济蓬勃发展。大数据产业作为战略性新兴产业，是激活数据要素潜能的关键支撑，是加快经济社会发展质量变革、效率变革、动力变革的重要引擎。2021年11月印发的《“十四五”大数据产业发展规划》在“十三五”规划的产业规模1万亿元目标基础上，提出“到2025年底大数据产业测算规模突破3万亿元”的增长目标。在2022年10月的《关于数字经济发展情况的报告》中提出，2023年底前，全国一体化政务大数据体系初步形成，基本具备数据目录管理、数据归集、数据治理、大数据分析、安全防护等能力。近年来，大数据政策已布局政务、金融、工业、制造、制造、交通、能源、医疗等多个领域。

2021-2022年中国大数据重点法律法规和产业政策脉络



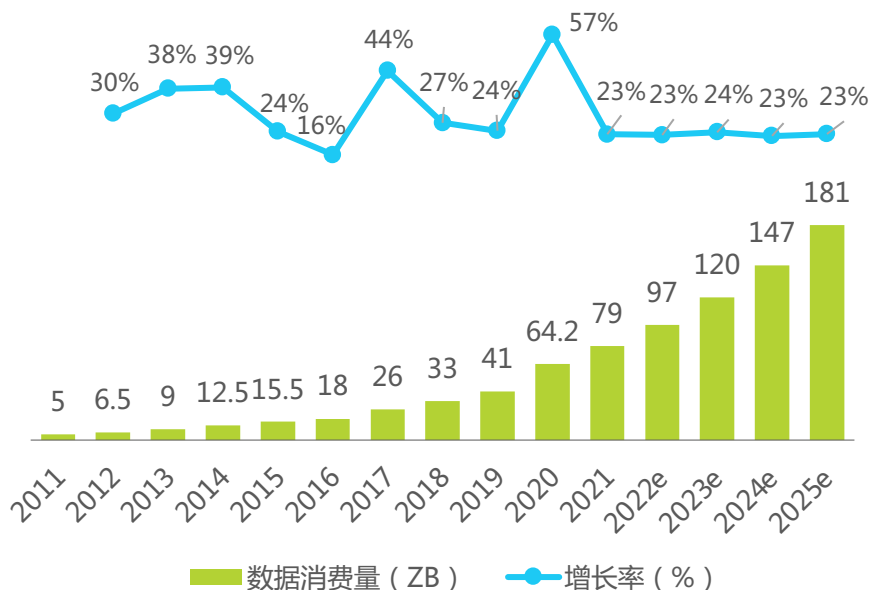
来源：公开资料，艾瑞咨询研究院整理及绘制。

驱动因素：宏观面

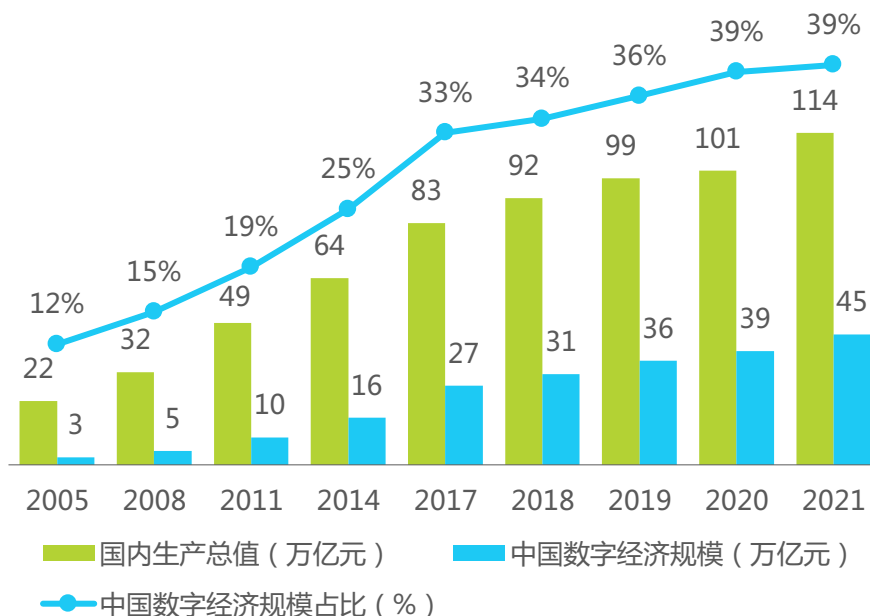
全球数据消费量和国内数字经济规模稳步增长

近年来大数据发展浪潮席卷全球，全球数据消费量稳步增长。随着数据资产的不断积累，用数场景和需求趋于复杂。据 IDC 预测，2021年全球大数据和分析支出达2157亿美元，比2020年增长10.1%，且未来五年全球大数据和分析支出还将继续增长，2021-2025年预测期内复合年增长率（CAGR）预计为12.8%。放眼国内，数字经济发展热潮兴起，数字化转型需求放量。大数据与云计算、人工智能、区块链等新一代信息技术加速融合创新，驱动我国大数据产业生态日渐完善，应用领域不断丰富。大数据在关键技术领域加快自主创新尤其在基础软件领域，各类自主研发的大数据平台产品百花齐放，合力向数据共享、流通、开放的方向持续演进。

2011-2025年全球数据消费量及增长率



2005-2021年中国数字经济规模及全国占比



来源：Statista 2022，艾瑞咨询研究院整理及绘制。

来源：中国电信招股说明书，艾瑞咨询研究院整理及绘制。

驱动因素：微观面

受业务需求驱动，打通各业务维度，提供分析和决策洞察

大数据创造价值的关键在于大数据应用，随着大数据技术的发展，大数据应用已经渗透至各行各业的企业当中。在企业的业务系统中，除了与业务直接相关的数据存储在数据库外，还有海量的系统监控数据和业务日志产生。随着企业数据资产的日积月累，能够全面覆盖日常经营、输出分析结果的指标体系成为企业的必需品。除了面对错综复杂的业务数据，集团性企业还需聚拢各业务系统的数据。因此，企业迫切希望能够打通全渠道来组织各业务维度，让业务数据能够更持久的存储，并提供实时/离线分析，帮助企业高层进一步了解企业的宏观运营面、基本面、财务面，帮助业务部门及时锁定潜在问题，提供精细化运营。由此，大数据分析平台应用而生。

大数据分析平台主要解决的问题

丰富的数据源支持

- 面对丰富的数据源，大数据分析平台提供统一的数据接入，便于后续体系化分析
- 数据接入主要包括文件日志、数据库日志、关系型数据库和应用程序等的接入
- 相关组件起到了上游数据源与分析平台存储接口的结构作用

数据格式延迟绑定

- TP（事务型）数据库和数据仓库常采用写入型 schema，即基于业务需求预先定义 schema，适合变化少的固定业务，不适合分析型业务
- 大数据分析平台的分析系统多采用读取型 schema，数据在分析时才会根据数据类型进行相应的处理

TP 和 AP 隔离

- TP 型业务适合行存储，而 AP 型业务适合列存储，分析业务的大规模全量扫描会影响在线业务的 SLA
- 大数据分析平台的典型处理方式是复制存储，面相多维分析需求，重塑数据分布、格式、索引，优化系统的分析性能

存算弹性扩缩容

- 业务数据随着各个业务系统运转而日积月累，企业普遍面临系统架构改造优化和数据迁移
- 大数据分析平台的搭建会对应考虑数据的分层存储和存储引擎的选择，保证存算能力可按需扩缩容，并提供存算资源的编排优化

来源：中国电信招股说明书，艾瑞咨询研究院整理及绘制。

产业链上中下游厂商与中游大数据分析厂商产品存在交叉

产业图谱



注释：以上厂商与行业为不完全列举，排名不分先后。

来源：艾瑞咨询研究院自主研究及绘制。

主要包括一体化解决方案、产品服务和行业应用

中国大数据分析平台的商业模式大致可分为三类：1) 解决方案，为企业搭建大数据分析系统，按照构建和部署大数据系统的费用+每年的维护/升级服务费用进行收取。2) 产品服务，产品化服务模式包括情报挖掘、舆情分析、销售追踪、精准营销、个性化推荐、可视化、网站/APP 分析工具等，订阅式的按需购买，按年/月收费，持续更新。3) 行业应用，大数据与传统行业碰撞形成的商业模式，利用大数据获得行业洞察，实现更多的收益。没有直接的变现，通过大数据技术深层挖掘价值，节约成本，优化行业模式。

商业模式

市场发展趋势

解决方案



中国政府用户和一些重点行业（如金融、电信等）更多选择整体解决方案的本地部署及私有云/行业云模式，体现在大数据硬件采购占比显著高于国外。
行业特点：数据安全要求高，强政策导向，预算充裕。

产品服务



数量庞大的中小企业更倾向于订阅式的产品服务接入大数据分析领域，开箱即用，按需付费，极大降低了技术门槛和资金投入门槛。
行业特点：企业模式灵活，创新性强，对大数据新兴技术敏感度高，但资本开支有限。

行业应用



传统行业结合大数据技术派生的新商业模式，深化行业洞察，例如：智慧医疗系统（大数据+医疗）、工业4.0（大数据+制造业）。虽无直接变现，但推广复制潜力大，通过深挖垂直行业，补充前两种商业模式。

来源：艾瑞咨询研究院自主研究及绘制。

国产化

国产化大数据分析产品蓬勃发展，对比国外产品的精细化，国产化产品呈现出功能一体化，集成数据整合、数据加工、数据治理、数据分析、数据可视化于一身。

SaaS化

随着公有云技术不断演进，订阅模式下的 SaaS 产品表现出云端部署、按需调用、持续更新、省去本地运维的诸多优点，让国内企业数字化转型轻装上阵。

平民化

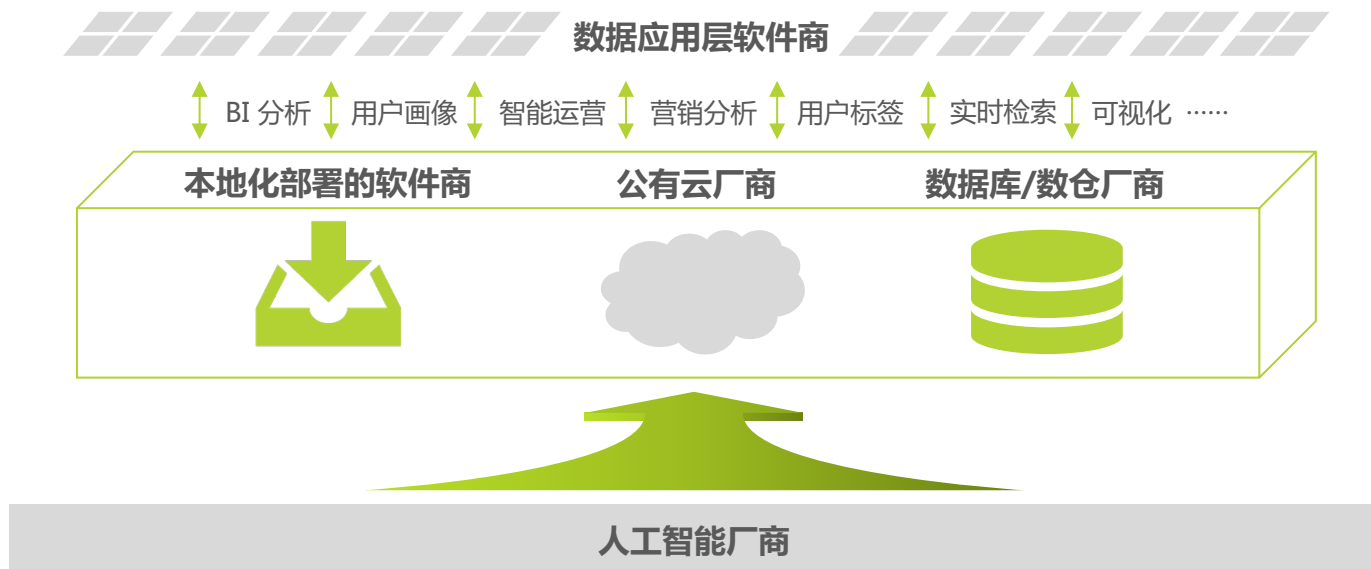
企业部署大数据分析平台的技术门槛、资金投入门槛和使用门槛不断被降低，低代码和无代码模式的数据分析应用越来越偏向于业务侧主导，数据技术更好的服务于企业业务本身。

来源：艾瑞咨询研究院自主研究及绘制。

行业边界泛化，市场参与者众多，服务类型多样

大数据分析平台逐渐从产品态向集成态发展，行业市场的参与者众多，服务类型丰富多样，行业边界趋于模糊，但也可大致分为以下几类：1) 公有云厂商借助云原生能力自然演进存算分离架构，提供方便接入各类数据、降低存储和运维成本的数据湖方案。2) 与云厂商以 PaaS 形式提供服务不同，传统软件商提供以本地化部署为主的一体化大数据分析平台解决方案。3) 国产数据库/数据仓库厂商融合创新技术，自主研发存储和分析性能优异的产品和架构方案。4) 为大数据分析平台的应用层提供 BI 分析、用户画像、智能运营、可视化发布等能力的软件服务商，与前述市场参与者逐步建立起合作生态。此外，人工智能厂商提供的 AI 能力让数据应用进一步延伸，让数据接入、清洗、存储、分析、训练到可视化输出的过程更加自动化，加强了场景需求与数据分析的自适应能力。

大数据分析平台行业的市场格局



来源：艾瑞咨询研究院自主研究及绘制。

国内外大数据分析平台产业探索方向及落地方式有所不同

国内外大数据分析平台的目标客户都主要锁定在特定行业、具有商业前景的企业，致力于为其提供成长初期缺乏的资源，实现商业价值快速增长。虽然核心诉求一致，但受限于体制、经济与文化等差异，产品的探索方向及落地形式不尽相同。国外产品更注重创客文化及高技术投资回报，倾向于以股票收获溢价作为主要的盈利方式，通过技术积累与项目展示收获口碑；国内大数据分析平台紧密围绕政策导向和产业价值定位制定预期发展目标，通过打通产学研加速资源交换与聚焦，为企业获得收益，不断积累资源与品牌影响力形成雪球效应。

国内外大数据分析平台产品对比

国内产品

数加平台

数据深度整合、计算、挖掘，将计算的结果通过可视化工具进行个性化的数据分析和展现。优点是功能完善，提供 SQL 查询。

数说立方

优点是体验和功能良好，集数据处理、特征工程、建模、文本挖掘为一体的机器学习平台，支持 SaaS，私有化部署，有权限管理。

神测分析

产品支持私有部署、任意维度的交叉分析，并帮助客户搭建专属的数据仓库。优点是专注用户行为数据分析，提供 SQL 查询。

国外产品

Tableau

产品界面采取拖拽式，操作性强；数据兼容性强，适用于多种数据文件与数据库。优点是产品功能完善和良好的图形展现与客户感知。

QlikView

优点是视图种类丰富，界面简洁，互动性强；可通过各类可视化效果，将 Qlik 扩展到任何应用程序中，支持使用标准的和最新的网络 API。

Spotfire

优点是交互界面形象易懂，对业务人员操作友好，便于进行复杂的数据分析，无需建数据仓库就可直接从多个异构数据源提取数据进行分析。

在数据兼容性、性能、开放性和行业经验方面仍存痛点

大数据分析平台的应用痛点反馈

数据兼容性

现阶段采购大数据分析平台的企业 IT 基础往往并非完全空白，企业内部积累了多类数据库和多种数据文件。因此，企业希望大数据分析平台能尽可能多的兼容多种格式的数据、多种平台和操作系统。

性能提升

企业在实际使用时发现大数据分析平台在处理结构化数据和GB级数据时差异不大，但处理非结构数据和TB/PB级数据时性能下降较大。对于数据波幅较大，且要求实时分析的企业，这是最大的痛点。

开放性

许多企业都希望拥有一定的自主权，不与供应商完全绑定。由于业务模式的变化，会经常有二次开发的需求。所以，企业会希望大数据分析平台足够开放，且提供足够多的接口和可视化的开发工具。

全链路能力

企业在使用时发现，现阶段国内可以提供全链路服务的供应商较少，因此只能采购多家供应商来满足自身从采集治理到分析可视化的需求。尤其是技术水平较弱的企业更希望厂商提供全面的服务。

行业理解

此类痛点集中于制造业和航空业等专业性较强的行业，他们具有大数据分析平台的需求，但与此类供应商接触时发现其不理解行业业务，因此主要考虑自己进行开发，或者寻求原来的硬件合作商帮助。

功能丰富性

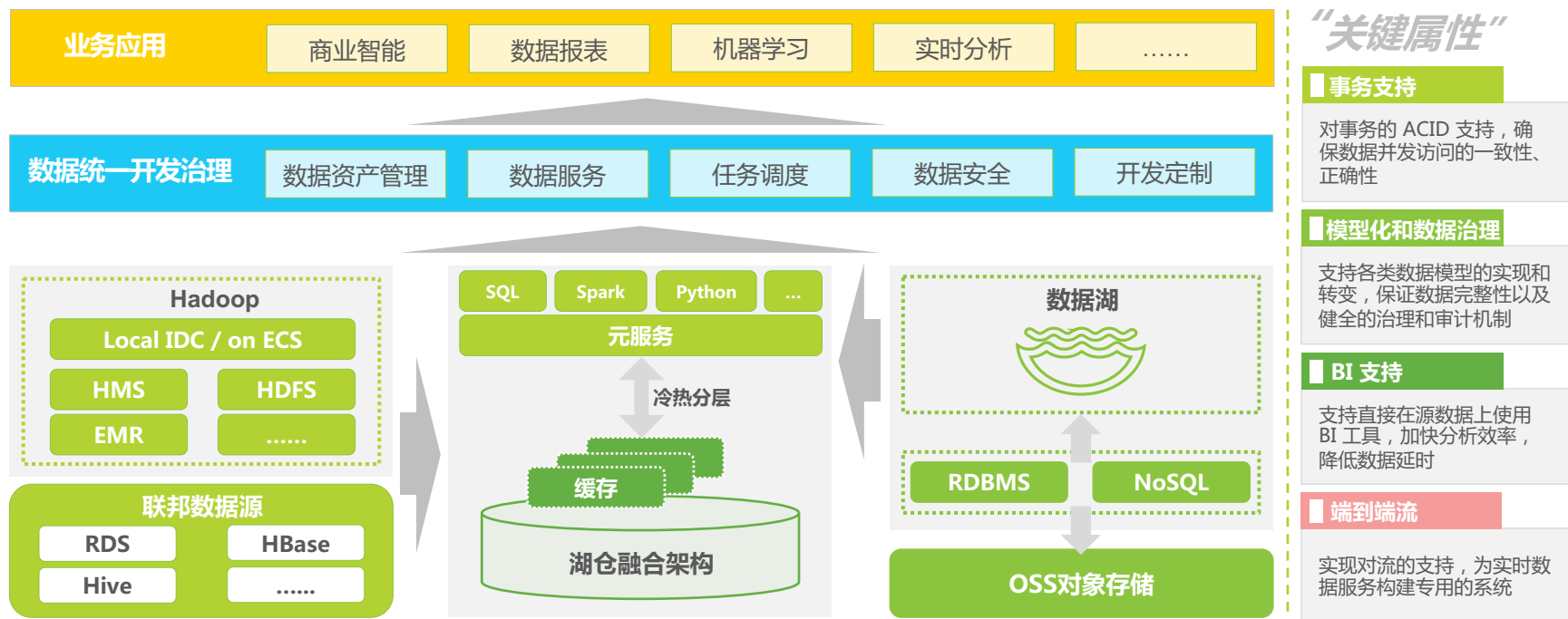
企业希望大数据分析平台除了基本的分析能力和常规的图表展示外，还可以增加更丰富的功能，例如多维交叉分析，自定义分析字段，3D大屏展示，可视化图表嵌入等。

趋势一：架构演进

湖仓融合，发挥海量、多模、实时的数据处理能力

湖仓融合作为开放式的数据架构和管理模式，将数仓建于数据湖之上，融合二者优势，优化企业的基础技术栈。湖仓融合打通底层异构数据源/平台，支持多种数据类型并存，实现数据共享。数据入湖后可直接加工处理，避免数据多份冗余和流动导致的算力、网络及成本开销。相比传统数仓和数据湖方案，湖仓一体架构增强了实时业务处理以及非结构化数据的治理能力，优势突出体现于：1) 完善的数据管理能力；2) 丰富的计算引擎支持；3) 更高的数据实时性；4) 更高的开放性。此外，数据安全、访问控制以及数据探索等企业级系统不可或缺的功能都可在湖仓融合架构中部署、测试和管理。

湖仓融合技术理念



来源：艾瑞咨询研究院自主研究及绘制。

趋势二：AI 融合

围绕 AI 核心能力，增强人员分析和决策能力

大数据分析随着人工智能的发展而不断演进，多层次、多维度的提升了数据使用人员的分析与决策能力。自新冠疫情爆发以来，企业业务环境发生了巨大变化，但从未削弱 AI 和机器学习发挥的重要作用。随着业务决策变得更具关联性、情境性和连续性，企业通过 AI 工程编排和优化系统，来适应、抵御或吸收各种干扰因素，提升自适应的 AI 能力，以迅速适应场景变化，提供更加快速灵活的决策。自然语言处理（NLP）增强了计算机系统对自然语言的准确识别、分析和处理，让搜索式分析成为全新的可视化交互方式，系统智能将自然语言结构的问题转化为 SQL 语句进行查询，易用性和自服务程度更高，对业务人员的使用门槛更低。

大数据分析 & AI 技术的协同创新

数据民主化

企业所有成员，尤其是非专业技术成员，都能轻松应用数据资源，开展应用分析，做出业务决策，推动更好的客户体验



自适应的 AI 能力

- 通过 AI 工程编排和优化系统，来适应、抵御或吸收干扰因素，加强 AI 自适应管理能力，用以敏捷适应业务场景的变化，提供更加快速灵活的决策
- 通过提升 AI 信任、风险和安全管理，进而提升 AI 在业务目标实现和企业内外部数据管理中的效果，即自适应的 AI 风险管理

数据编制

通过对现有的、可发现的和可推断的元数据资产进行持续分析，来支持数据系统跨平台的设计、部署和使用，从而实现灵活的数据交付

基于图形技术的场景分析

业务场景和分析需求的数据存储于图形中，基于相似性、制约因素、路径等识别和创建进一步的场景，利用数据点之间的关系及数据点本身实现深入分析。基于图形技术的分析和 AI 模型将进一步取代建立在传统数据基础上的分析模型

搜索式分析

使用机器学习和自然语言处理（NLP）来自动化和处理数据，系统准确识别、分析和处理自然语言，智能将自然语言结构的问题转化为 SQL 语句进行查询，极大降低业务人员的数据分析门槛

趋势三：场景多元

数据分析场景呈现多元化趋势，不断向多维纵深发展

随着数字化转型的不断深入，企业的精细化运营更加聚焦于通过数据分析来提升效率和优化生产。因此用户对于取数的时效、维度、深度以及交互的方式方法（移动、托拉拽、操作的友好体验）要求越来越高，分析场景和分析需求不断深化，使得我们对大数据分析平台所提供的算力、计算和存储组件、资源调度等要求趋于复杂。企业的分析场景和需求呈现多元化发展，既要满足前端业务人员的实时分析，提供运营人员实时查询当前经营数据，又要向中台建设人员提供统一存算平台，以满足高并发 SQL 复杂查询访问。

大数据分析场景的多元化发展

企业应用

数据分析场景多元化

既要满足前端经营数据的实时查询和分析，又要满足后端高并发复杂查询

情景驱动，知识图谱利用率提升

引入客户行为数据，实时追踪客户监控变化

业务侧主导数据分析应用

进行客户分级，搭建差异化分析指标体系

决策驱动数据分析

实时进行 MRR 多维度分析

市场发展

国产化产品的高速发展

支持采集终端用户行为、服务器日志、业务数据和第三方等多方数据源

SaaS 化产品的平民化

利用深度学习和语义分析模型构建推荐引擎

数据生态系统复合化程度加强

为用户生成个性化推荐

数据资源化竞争加剧

分析用户转化情况和渠道转化情况，形成及时反馈

架构需求

统一多模型大数据分析架构

运用 SDK 对网页、App、小程序等多端进行数据采集，并整合外部数据

分布式系统架构广泛应用

同时支持可视化界面和 SQL 创建用户标签

云原生大数据平台架构

统计分析用户行为特征和标签构成特征

实时计算和运营分析

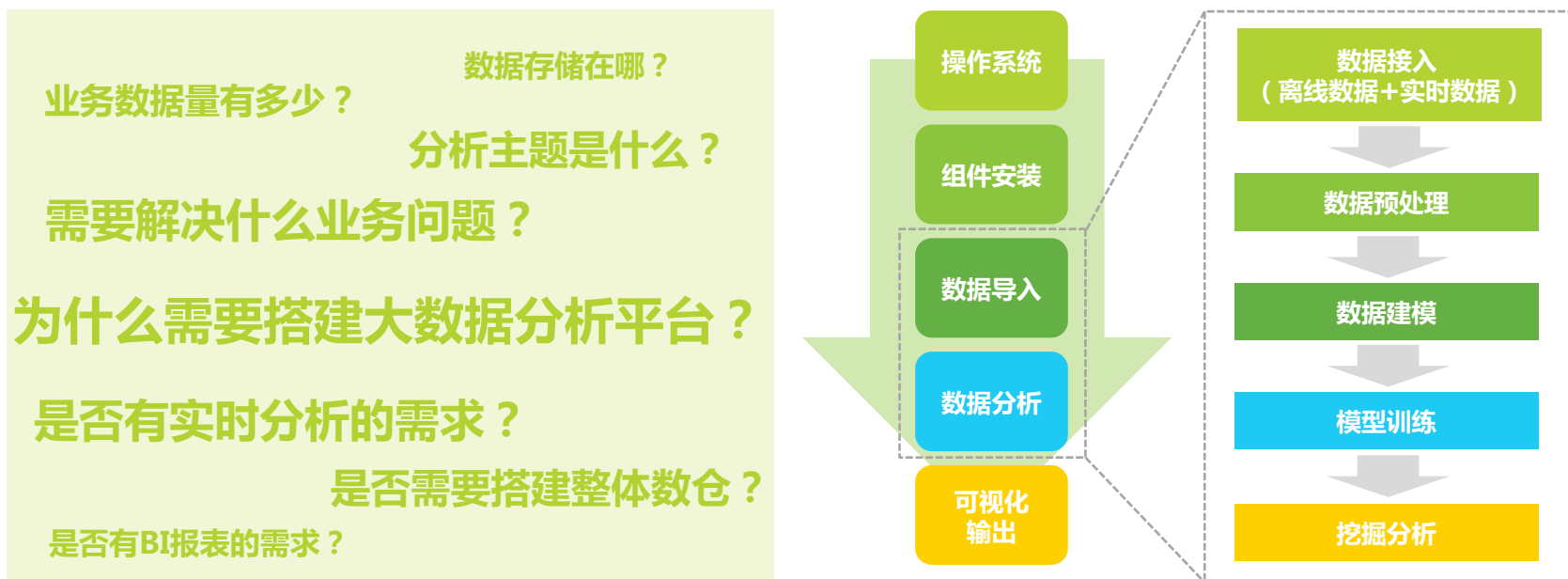
支持标签数据导出，进行二次加工和其他应用

大数据分析平台行业概述	1
大数据分析平台市场分析	2
大数据分析平台构建建议	3
行业应用与典型案例实践	4
大数据分析行业投资分析	5

明确业务场景需求，基于数据体量选定平台框架和功能组件

在搭建大数据分析平台前，用户首先要明确自身的数据体量和业务场景需求，希望通过大数据分析平台得到哪些信息，需要接入哪些数据，进行哪些主题分析，最终实现哪些功能。在明确大数据分析平台需要具备的基本功能后，再决定平台搭建过程中使用的大数据处理框架和工具，并将其有机结合以完成海量数据的挖掘和分析。在构建大数据分析平台时，首先要建设企业的基础数据中心，构建统一的数据存储体系，统一数据建模。其次，集中组建数据处理中心，下沉数据处理能力，并通过统一的数据管理监控体系，保障平台系统的稳定运行。最后，构建数据应用中心，统一输出数据服务，满足业务需求，体现数据价值。

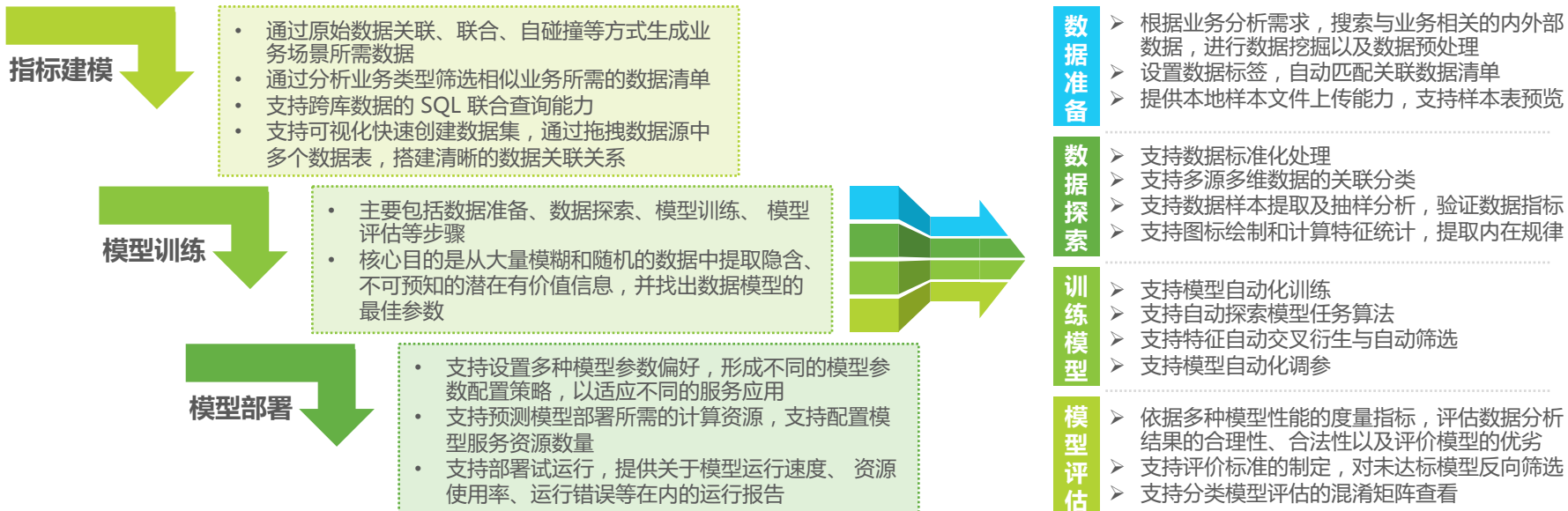
搭建大数据分析平台的整体思路



基于场景需求，选定分析指标，通过模型训练构建分析能力

大数据分析平台的建设核心是分析能力的建设。不论用户采用何种部署方式，数据分析能力的建设都万变不离其宗。首先，根据业务场景需求选定指标进行建模，重点建设数据构造、合并和统计处理的运算能力。接着进行模型训练，从大量有噪声的、不完全的、模糊和随机的数据中挖掘多源多维数据间的关联性。通过多维分析数据，加深对数据的理解，提取可能对业务结果相关的影响因子，探索数据的内在规律特征，并寻找模型最佳参数，支撑分析模型对业务的定量和定性分析。在完成指标建模、模型训练后，对满足业务分析需求的模型进行部署调试，形成可被调用的服务能力，为其它业务系统、模型提供数据分析能力。此外，大数据分析平台应具备基础框架功能，支持多厂家、多技术类型模型导入，提供对应功能和工作流程设计，保障分析能力实施落地。

大数据分析能力建设



来源：艾瑞咨询研究院自主研究及绘制。

依据行业特征、数据体量以及场景需要自由选择部署方式

大数据分析平台的部署方式主要分为本地化部署和云上部署。本地化部署根据用户数据体量的大小，又可分为基于 Hadoop 生态的平台搭建和“数据库+AP 分析引擎”的数仓方案。本地化部署的优点是自主可控和灵活度高，缺点是性能调优和运维复杂，自身技术能力要求和综合成本高。云上部署以公有云厂商提供的“低成本存储+弹性存算引擎”的数据湖方案为主，在保留 HDFS 集群分布式存储可靠性和高吞吐能力的前提下，提供一站式云上 PaaS 能力，实现各类数据快速便捷入湖，用户无需考虑兼容、安全、性能调优以及运维。尽管国内主流数据湖方案的底层存储系统仍以 Hadoop 的分布式架构构建为主，但架构上层拥有的读写优化、内存加速、数据融合等特性是云原生数据湖融合第三方开源组件的价值体现，是本地化部署的 Hadoop 方案所不具备的。

本地化部署 VS 云上部署

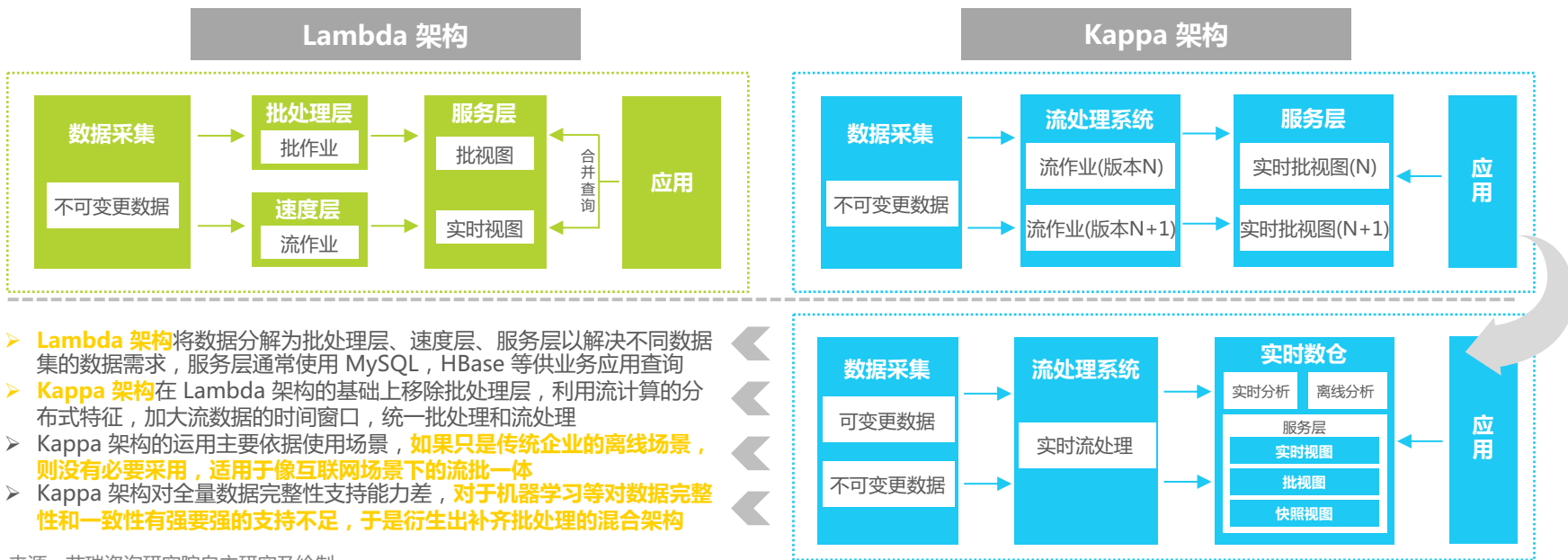
<h4>基于 Hadoop 的大数据方案</h4> <p>用户画像</p> <ul style="list-style-type: none">用户数据总量达到百 TB 乃至 PB 量级，每年新增数据量超过 100 TB 以上，适合构建 Hadoop 平台 <p>搭建方式</p> <ul style="list-style-type: none">业务场景以低时效要求的离线场景为主，业务侧对分析结果敏感度低，可尝试开源自建；否则，建议成熟稳定的商业版	 <p>本地化部署</p>	 <p>云上部署</p>	<h4>云上数据湖方案</h4> <p>用户画像</p> <ul style="list-style-type: none">公有云用户（80%来自互联网行业，20%为非互联网行业），业务数据已上云，使用云上 PaaS 平台在开发难度、使用体验、运维成本等方面达到最优传统行业用户的混合云部署（如政府、金融、医药公司等），虽本地建有大数据平台，但希望通过云上平台提供新算力资源，进行例如仿真计算在内的二次计算 <p>方案亮点</p> <ul style="list-style-type: none">基于 Hadoop 的分布式架构构建底层存储系统，利用云原生数据湖的架构优势，一方面实现多源异构数据的快速便捷入湖，降低存储成本；另一方面融合第三方开源产品插件，提供读写优化、内存加速、数据融合等性能提升；同时免去了用户性能调优、兼容、安全及运维等方面的烦恼
<h4>“数据库+AP 分析引擎”的数仓方案</h4> <p>用户画像</p> <ul style="list-style-type: none">用户的数据总量在 GB 到 100 TB 级别，日增数据在几十 GB 至百 GB，建议采用数仓方案 <p>方案亮点</p> <ul style="list-style-type: none">在有限数据体量下，借助存储引擎自身的存储格式和计算下推，支持实时批量计算，实时展现分析结果，性能亮眼			

来源：艾瑞咨询研究院自主研究及绘制。

从离线、在线及实时场景出发，按需选择和组合分析架构

广义而言，大数据分析平台不再局限于产品态，更趋近于包含数据采集层、存储层、调度层、计算层、交互分析层、数据服务层等的集成态。如果从技术架构的角度进行抽象，大数据分析平台的架构都可归属于 Lambda 或 Kappa 架构。若从场景角度进一步抽象，又可拆分为离线、在线以及实时分析架构。在自下而上分层的集成态中，三种分析架构的差异主要源于数据分析层中存算引擎的选用，以满足各自的分析场景。从技术角度而言，数据分析层的部署最为复杂，但也最富于创新，既有云原生数据湖的存算分离与弹性扩缩容，也有本地化部署下，基于 Docker 技术的平台解耦，解决物理服务器资源供给弹性不足的问题，满足存算能力的横向扩展。在落地实施时，用户的分析场景又趋于融合，既有 HTAP 数仓方案的融合框架缩影，也有融合 AP 和 TP 场景的海量大数据分析平台，用户皆可按需选取。

Lambda 架构 VS Kappa 架构



- **Lambda 架构**将数据分解为批处理层、速度层、服务层以解决不同数据集的数据需求，服务层通常使用 MySQL, HBase 等供业务应用查询
- **Kappa 架构**在 Lambda 架构的基础上移除批处理层，利用流计算的分布式特征，加大流数据的时间窗口，统一批处理和流处理
- Kappa 架构的运用主要依据使用场景，**如果只是传统企业的离线场景，则没有必要采用，适用于像互联网场景下的流批一体**
- Kappa 架构对全量数据完整性支持能力差，**对于机器学习等对数据完整性和一致性有强要强的支持不足，于是衍生出补齐批处理的混合架构**

来源：艾瑞咨询研究院自主研究及绘制。

采取自建方式部署，应着重关注分析层存算引擎的组合搭建

对于采取本地化部署的用户，按照数据量级大致可分两类：其一，年新增数据量在100TB 级的传统行业头部用户，由于本地数据体量大，上云的带宽成本高，同时还可能受到合规监管的约束，多采用开源自建或采购商业版 Hadoop 生态。其二，年新增数据量在 TB 级别以下的中小企业，选择抛开架构繁复的 Hadoop生态，自建“数据库+AP 分析引擎”的数仓方案。自建大数据分析平台必然涉及组件选择，尤其是数据分析层的组件集成，直接影响场景支撑和效率提升。而在数据分析层的存算引擎中，存储引擎的选择显得尤为重要。不难发现抛开计算引擎本身的性能，数据的宽表合并、CRUD、批量计算、实时流计算、即席查询等，都依赖大数据分析平台自身的存储引擎。

大数据分析平台的组件选择



来源：艾瑞咨询研究院自主研究及绘制。

打破传统架构下的技术异构，统一数据能力提升业务价值

传统 Hadoop 架构和以 MPP 为主的数仓架构都无法真正适应云平台。Hadoop 将存储和计算部署在同一物理集群以拉近与数据的距离，仅在同一集群下实现了存算分离，而 MPP 数据库本身存算耦合。传统架构下的湖仓分体引发数据孤岛的原因有三：第一，异构技术架构；第二，集群规模受限；第三，集群高并发受限。数据孤岛进而造成实施、运维和成本的问题。湖仓一体技术呼之欲出——在数据和查询层面形成一体化架构，解决实时性和并发度、集群规模受限、非结构化数据无法整合、建模路径冗长、数据一致性弱、性能瓶颈等问题，降低数据管理门槛和运维成本。从架构演进方向来看，国内以基于 Hadoop 的改造方案为主，从事务特性出发进行优化，如 Hudi 和 Iceberg 等，基于 HDFS 或 S3 实现支持事务的存储层，其他与 Hadoop 区别不大。另一方面，以 Snowflake 为代表，基于多云的数仓架构方案在存算分离等方面的特性更具前瞻性，值得持续关注。

架构演进方向及业务价值

	基于 Hadoop 的改造方案 (以 Hudi 为例)	基于云原生数仓架构的方案 (以 Snowflake 为例)
改造背景	Apache Hudi 是由 Uber 工程师为满足内部数据分析需求而设计的，Hudi 的数据写入不绑定 Spark，也可以使用 Hudi 自带写入工具	Snowflake 的湖仓一体方案基于 Snowflake 的云原生数据仓库，Snowflake 实现了事务的支持，仅支持对象存储
实现原理	使用 SparkSQL/Flink 作为 SQL 计算引擎，Spark Streaming/Flink 作为流处理引擎，存储使用 HDFS/S3 对象存储，Update/Delete 的事务实现由 Hudi 实现	使用 Snowflake 作为 SQL 计算引擎，Spark Streaming/Flink 作为流处理引擎，S3 对象存储，Update/Delete 事务由 Snowflake 原生提供支持

业务价值

实时 T+0

全量数据 T+0 的流处理和实时按需查询，满足事前预测、事中判断和事后分析

一份数据

所有用户（BI 用户、数据科学家等）可以共享同一份数据，避免数据孤岛

超高并发

支持数十万用户使用复杂分析查询，并发访问同一份数据

数据一致

通过支持完善的事务机制，保障不同用户同时查询和更新同份数据时的一致性

云原生

适合云环境，自由增减计算和存储资源，按用量计费，节约成本

多类型数据

支持关系表、文本、图像、视频等结构化数据和非结构化数据存储

大数据分析平台行业概述	1
大数据分析平台市场分析	2
大数据分析平台构建建议	3
行业应用与典型案例实践	4
大数据分析行业投资分析	5

行业聚焦-总览

指导政策覆盖政务、金融、零售、医疗、交通和教育等领域

中国大数据分析平台各赛道政策方向

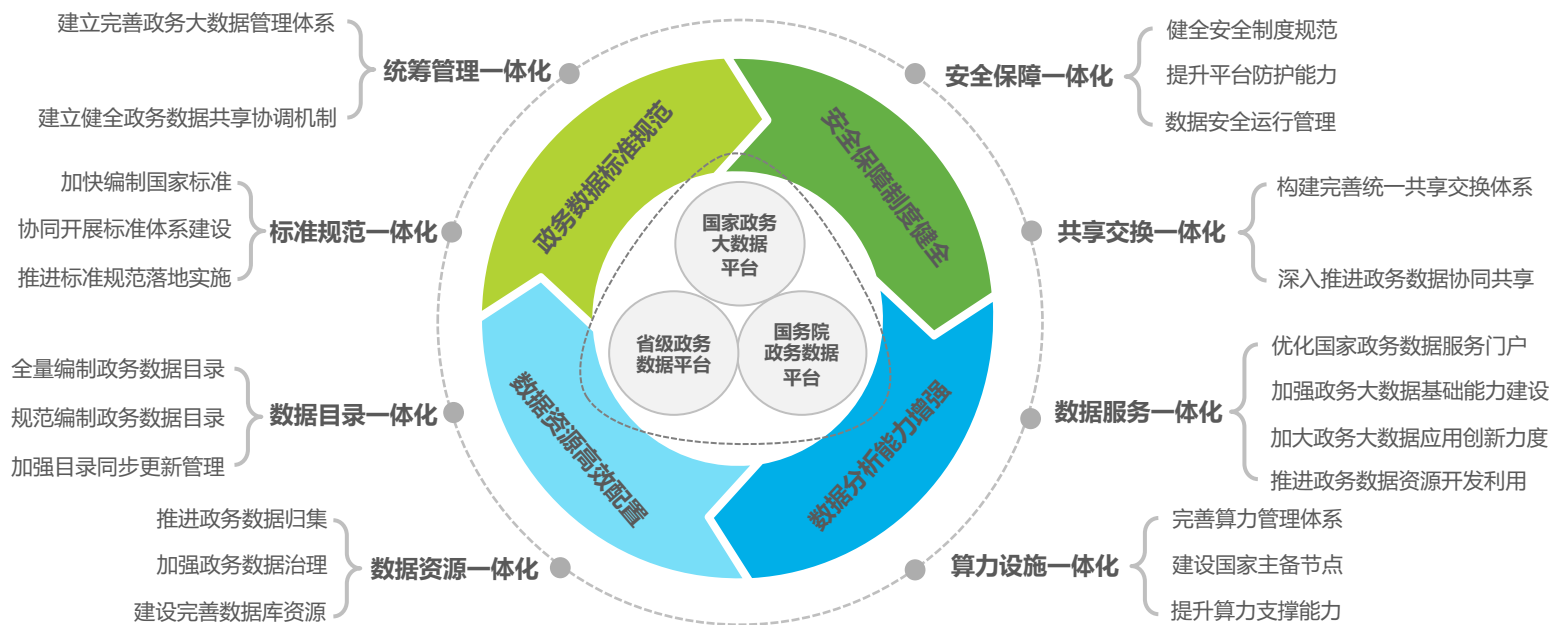
子赛道	时间	政策发布机关	政策名称
政务	2022.10	国务院办公厅	《全国一体化政务大数据体系建设指南》
金融	2021.12	中国人民银行	《金融大数据平台总体技术要求》
医疗	2018.7	国家卫生健康委员会	《国家健康医疗大数据标准、安全和服务管理办法（试行）》
	2016.6	国家卫生计生委规划与信息司	《关于促进和规范健康医疗大数据应用发展的指导意见》
交通	2018.3	交通运输部办公厅、国家旅游局办公室	《关于加快推进交通旅游服务大数据应用试点工作的通知》
	2016.8	交通运输部	《关于推进交通运输行业数据资源开放共享的实施意见》
教育	2022.10	中共中央办公厅、国务院办公厅	《关于新时代进一步加强科学技术普及工作的意见》
	2022.9	民政部办公厅	《民政部关于落实国务院加强数字政府建设的指导意见》
水利	2017.5	水利部	《关于推进水利大数据发展的指导意见》
农业	2016.10	农业部	《农业部关于推进农业农村大数据发展的实施意见》
能源	2016.7	中国煤炭工业协会、中国煤炭运销协会	《推进煤炭大数据发展的指导意见》

来源：中国政务网国务院政策文件库，艾瑞咨询研究院根据公开资料自主研究及绘制。

构建全国一体化政务大数据体系，加快政府服务型建设

2022年10月国务院发布《全国一体化政务大数据体系建设指南》，就整合构建全国一体化政务大数据体系作出部署，提出加强数据汇聚融合、共享利用，促进数据高效流通使用，充分释放政务数据资源价值，提高政府管理水平和服务效能。全国一体化政务大数据体系聚焦“惠民”、“善政”、“兴业”和“城市管理”，依托城市大数据相关平台和应用支撑，汇聚城市管理各相关部门业务数据，集中展示和分析。基于大数据推进政府改革，转变执政理念，创新治理方式，由电子政务向智慧政务升级，提高政府工作效能。通过大数据推动产业结构优化升级，催生基于大数据、网络经济的新兴产业，促进投资，拉动经济增长。综合提升城市运行管理、政务服务、城市综合管理决策和产业转型升级等方面能力。

全国一体化政务大数据体系构建



来源：艾瑞咨询研究院根据公开资料自主研究及绘制。

基于大数据分析的精准营销、实施管控、分析决策

相比其他行业，金融大数据分析起步早、技术高、发展快，国家对金融行业大数据发展整体秉持鼓励和支持的基调。一方面，出于对金融行业系统性和非系统性风险双重考虑，国家对金融大数据分析出台了细致的监管政策，涉及数据收集、数据治理标准、大数据软件应用等多个方面。另一方面，金融作为数据密集型行业，数据体量大，数据逻辑性强，对实时性、安全性和稳定性的要求高，同时结构化数据占比高，在分析工具成熟度方面具有明显优势。最后，金融行业的数据分析应用场景广泛，包括精准营销、风险控制、客户关系管理、反欺诈检测、反洗钱检测、决策支持、股票预测、宏观经济分析

与预测等多个方面。大数据分析在金融行领域蕴含了巨大的潜力和挑战。

金融行业的大数据分析需求和场景应用



来源：艾瑞咨询研究院根据公开资料自主研究及绘制。

线上线下场景融合，智能设备终端接入，数据分析日趋复杂

新零售，即依托互联网、大数据、人工智能等手段，升级改造商品环节，深度融合线上服务、线下体验以及现代物流，重塑“人-货-场”业态结构与生态圈的零售新模式。新零售商在不同的业务场景下，布局各类智能终端设备，进行数据采集、算法运行和数据交互等操作。由于不同环节接入的设备终端采集的数据种类、敏感程度、商业价值不同，因而数据收集、处理场景和共享链条更加多样和复杂。随着新零售从消费终端获取的数据量和类别“爆发式”增长，对大数据分析的维度和深度提出了新的要求，包括从“人脸验证”“人脸识别”到“人脸分析”的创新应用、“不杀熟”和“不强制”的算法模型和个性化推荐，以及数据交互的具体场景和风险识别等。

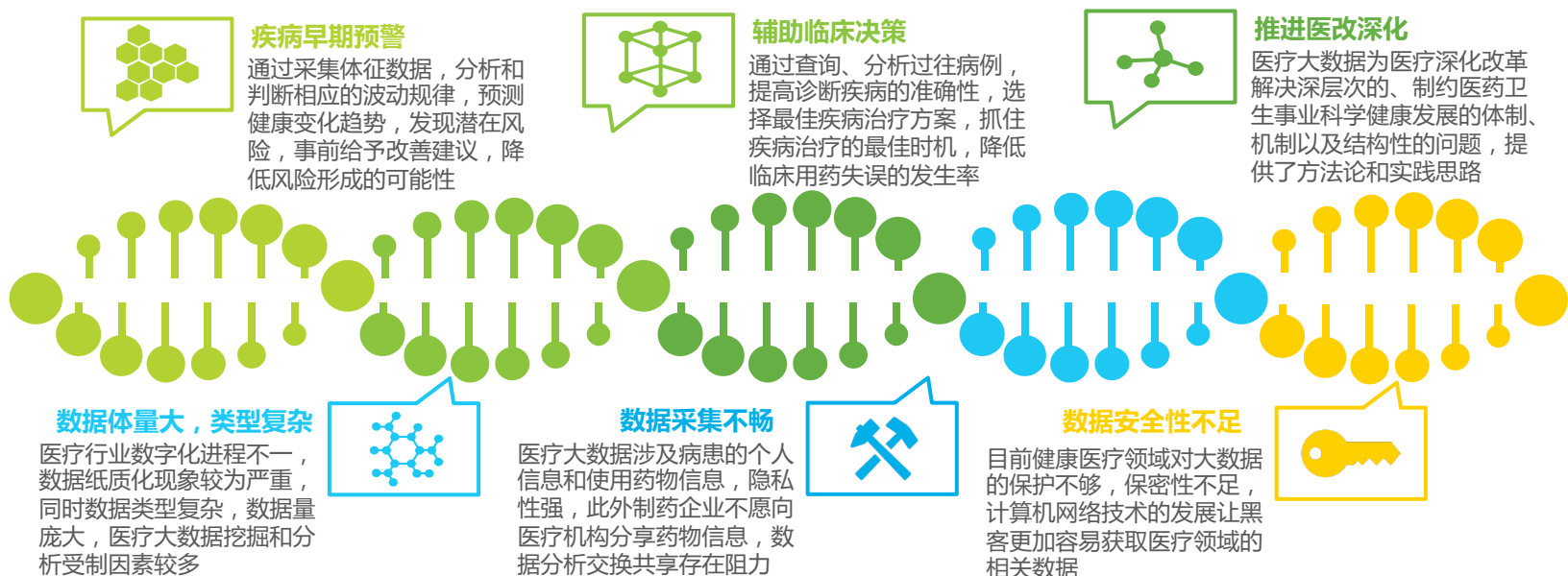
新零售模式下的算法模型和个性化推荐



数据体量庞大，类型繁杂，与健康 and 生命安全息息相关

与其他行业的数据相比，医疗行业的数据更为重要、复杂。不仅与健康 and 生命安全息息相关，而且数据结构和类型也更加庞杂和繁琐。医疗大数据的来源包括：1) 医疗数据资源，如电子病历数据、临床检验数据、医学影像数据、医患行为数据等；2) 行业数据资源，如商业健康险、医保、新药研发、医药销售等；3) 学科相关数据资源，如生命科学、环境科学等；4) 产生于互联网的关于疾病、健康或寻医的话题、搜寻内容、购药行为以及网站访问记录等。医疗大数据呈现数据规模大、数据结构多样、增长速度快、数据价值高等显著特征。大数据分析在健康医疗领域的广泛应用，能够大幅提高对患者治疗的安全系数，为患者制定更有针对性的治疗方案，并有效降低医疗成本，意义重大。

医疗行业大数据分析的作用和难点



来源：艾瑞咨询研究院根据公开资料自主研究及绘制。

交通大数据独有地域和时域动态演化特征，实时性显著

智慧交通将大数据技术、传感技术、信号技术等与城市交通系统相结合，建立智能化网络交通体系，进而对交通系统覆盖区域实现全过程、实时性监督，确保城市交通系统的正确运行。大数据分析平台依托云存储和分布式数据处理技术，实现信息资源的存储与分配，通过各个路口安装的监控设备，对交通资源进行整合，在大数据技术的分化处理工作模式下，对交通环境进行立体化分析，以此提供更加全面的数据服务和交通体系网络。交通行业的大数据分析平台以人机交互系统、监控系统、信号传输系统、警报系统、信息传输系统、导航系统等为主，结合智能城市发展的需要，将大数据的数据采集、智能服务、安全指令等应用于智能城市建设，充分实现大数据时代智能城市交通系统的完善建设。

智慧交通大数据分析的特点和应用



时空移动性

交通事件具有**地域和时域特征**。交通大数据在时间和空间维度具有动态演化特征



交通规划

发现交通片区趋势和模式，支撑长期决策和城市规划，**分析交通流量，交通信号灯**



多维结构特征

交通大数据的**多维结构特征**明显，如时间信息、空间信息、天气、驾驶员信息等



交通预测

使用数据挖掘和机器学习算法，对可能的交通事件做出预测，**交通拥堵和事故预测**



社会关联性

交通大数据存在于信息空间和物理世界，**二者与人类社会有机连接和互动**



时效性明显

交通管理和规划决策更注重分析近期甚至实时数据，数据挖掘和分析的实时性高



交通监督

实时自适应交通控制系统、车牌车辆识别、检测交通事故、分析道路危险状况



交通指引

从历史轨迹数据集中挖掘有规律的线路，提供乘车共享建议，**路径推荐、导航系统**



数据采集是教育大数据挖掘和分析的重点突破方向

教育大数据产生于教育实践活动，既包括校园环境下的教学活动、管理活动、科研活动以及校园生活，也包括家庭、社区、博物馆、图书馆等非正式环境下的学习活动；既包括线上教育活动，也包括线下的教学活动。教育大数据挖掘和分析的难点主要在于数据采集，是由于教育大数据的数据源主要来自“人”和“物”两个方面。其中，“人”包括学生、教师、管理者和家长，因此数据采集需要综合应用多种技术，每种技术采集的数据范围和重点都有所不同。大数据分析在教育领域的应用主要包括：1) 教育数据挖掘。对学习行为和学习过程进行量化、分析和建模，探索和改进教学顺序和内容等领域模型，促进有效学习。2) 学习分析技术。对学习及其学习环境的数据测量、收集和分析，理解和优化学习过程和学习环境，构建出学习者特征并推送个性化资源。

教育大数据分析的数据采集和应用



教育大数据采集

01 平台采集技术

- 日志搜索分析
- 在线学习与管理平台
- 移动 APP
- 网络爬虫采集

02 图像识别技术

- 学生考试成绩数据
- 学习记录和练习数据
- 课程笔记数据

03 视频录制技术

- 学习过程情感数据
- 校园安全数据
- 课堂教学数据

04 物联感知技术

- 学生学习行为数据
- 学生体质数据
- 学生生活数据
- 设备状态数据

优化教学策略

利用数据分析结果，对知识掌握程度及潜在水平形成认识，优化教学方法、合理安排教学活动与顺序

自适应教学

通过大数据分析形成可视化报告，能够将学习者隐性的思维过程、能力素养等难以观测的维度显性化

个性化教学

对学习者的学习过程与行为数据进行挖掘、聚类和分析，有利于教师开展真正意义上的个性化教学

教学干预预警

通过数据分析知晓学习交互发生过程，洞察学习偏好和路径等规律，及时提供恰当干预措施

基于自然语言搜索的智能分析工具，深掘业务数据价值

北极九章于2018年成立，立足2B市场，对标美国BI服务商 ThoughtSpot。产品聚焦企业数据的搜索式分析引擎，界面与一般搜索引擎无异，亮点是实现中文自然问句与数据库交互，搜索引擎智能识别和分析常规业务问题，对数据库深度搜索后自动生成数据分析，形成数据洞察赋能业务决策。搜索式的分析引擎获取数据更加敏捷灵活，数据分析的使用门槛和成本更低，一线业务人员获取数据分析能力更加轻松。产品生态方面，数字化时代下数据量陡增，对数据挖掘和分析型 BI 提出更高要求，搜索式分析引擎是传统BI功能和性能的重要补充。

北极九章增强分析引擎技术架构

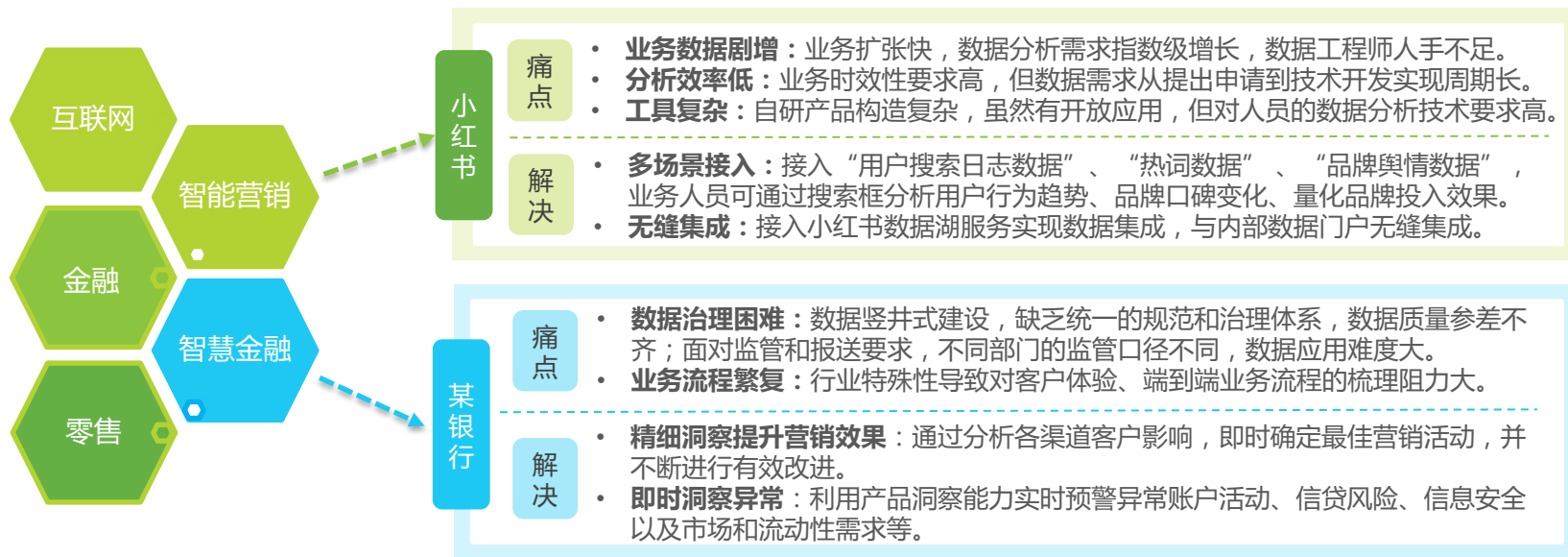


来源：北极九章，艾瑞咨询研究院整理及绘制。

产品面向业务人员，与传统BI形成合作生态

1) 产品定位方面：相比传统 BI 的操作复杂、学习成本高，搜索式数据分析产品直接面向非数据分析专业的业务人员，交互方式简单，显著降低使用门槛，让数据真正创造业务价值。两类产品从底层思路和设计逻辑有本质差异，非直接竞争关系。2) 产品技术方面：北极九章通过建立索引来获取多源异构数据，在涉及一对多查询时（如不同品牌间，或同一品牌在不同时间段对比），相对传统 BI，在响应速度方面具备相对优势。3) 产品体验方面：产品界面与搜索引擎无异，自然语言输入便可获取数据洞察，使用零门槛；产品基于云原生架构，开箱即用，大大降低部署成本。

北极九章产品服务领域与应用案例



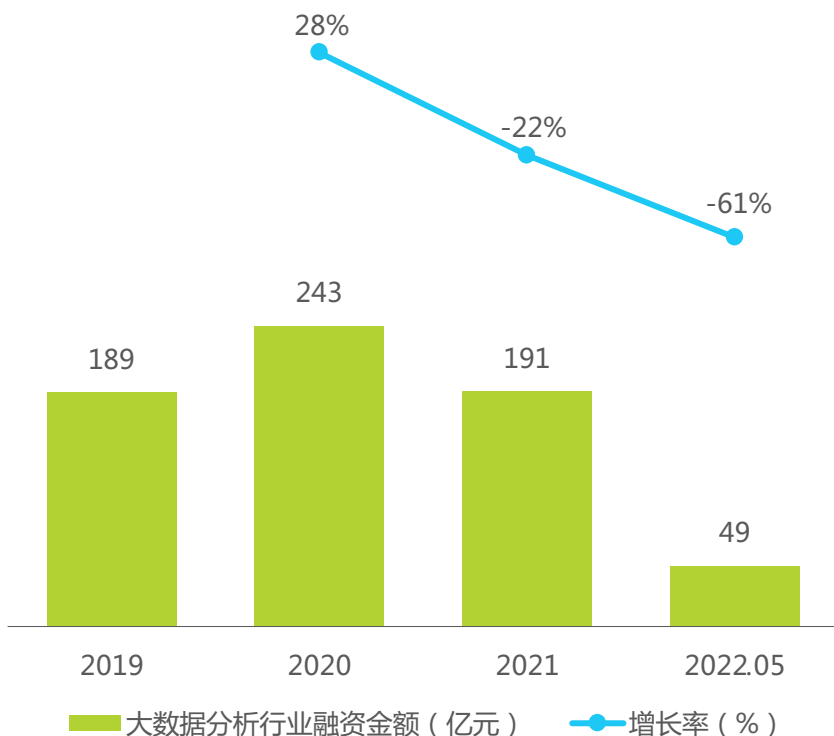
大数据分析平台行业概述	1
大数据分析平台市场分析	2
大数据分析平台构建建议	3
行业应用与典型案例实践	4
大数据分析行业投资分析	5

投资整体分析

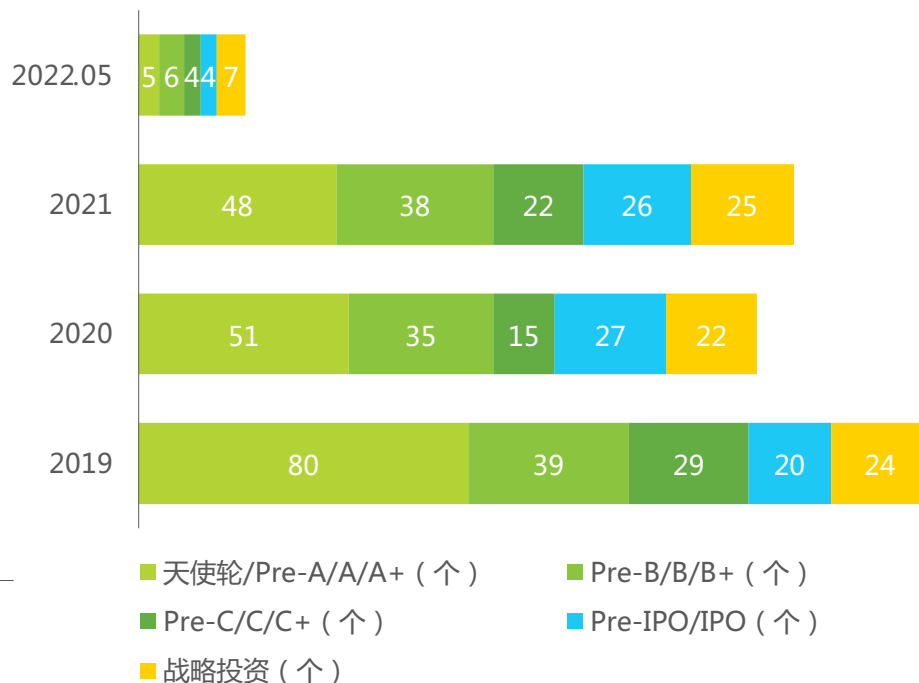
受疫情影响，大数据分析行业融资金额及融资频次缩水明显

2019年-2022年5月，中国大数据分析行业的项目（公司）共计328个，融资事件达534次，项目平均融资事件数为1.63次，有162个项目（公司）（占比49.39%）获得B轮及以上融资。244个项目（公司）已披露累计获得近672.22亿元融资金额。

2019-2022年5月中国大数据分析行业融资金额及增长率



2019-2022年5月中国大数据分析行业各轮次融资事件



来源：IT桔子，艾瑞咨询研究院整理及绘制。

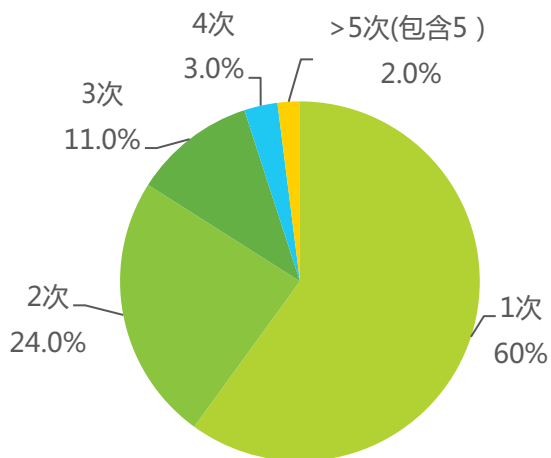
来源：IT桔子，艾瑞咨询研究院整理及绘制。

投资轮次分析

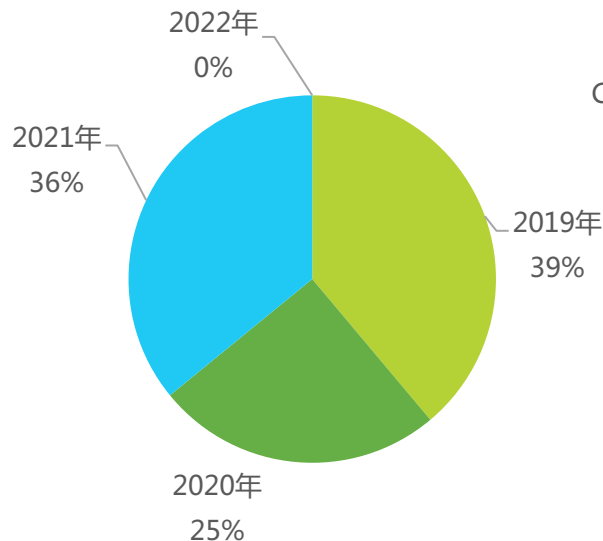
超六成项目难获得后续融资，融资活动集中于头部项目，市场风险偏好下行，资本更青睐产品集成度高的中后期企业

2019-2022年5月，中国大数据分析行业的项目仅有131（39%）个项目获得两轮及以上融资，197（61%）个项目仅获得一轮融资。三年内仅获得一轮融资的197个项目中，从起投年份看，64.4%起投年份为2019或2020年；从当前轮次看，56.93%处于融资早期阶段（种子/天使轮、A轮、B轮），说明中国大数据分析行业的融资事件相对集中于少部分头部项目，超六成早期项目难以获得后续融资。早期项目多为初创或中小企业，技术实力和研发能力弱，资金投入大，在长期没有外部资金注入的情况下，经营和管理缺乏创新动力，将会被市场淘汰。

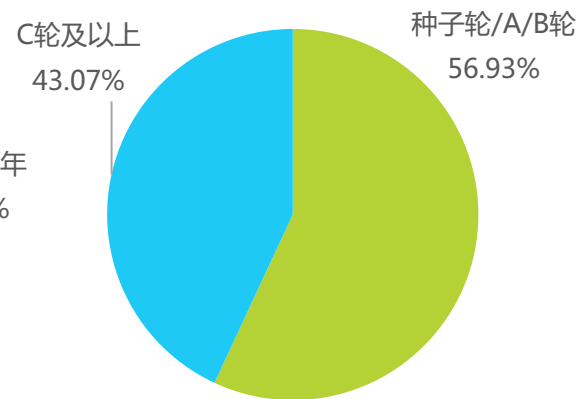
项目融资次数占比



项目起投年份占比



项目当年投资轮次占比



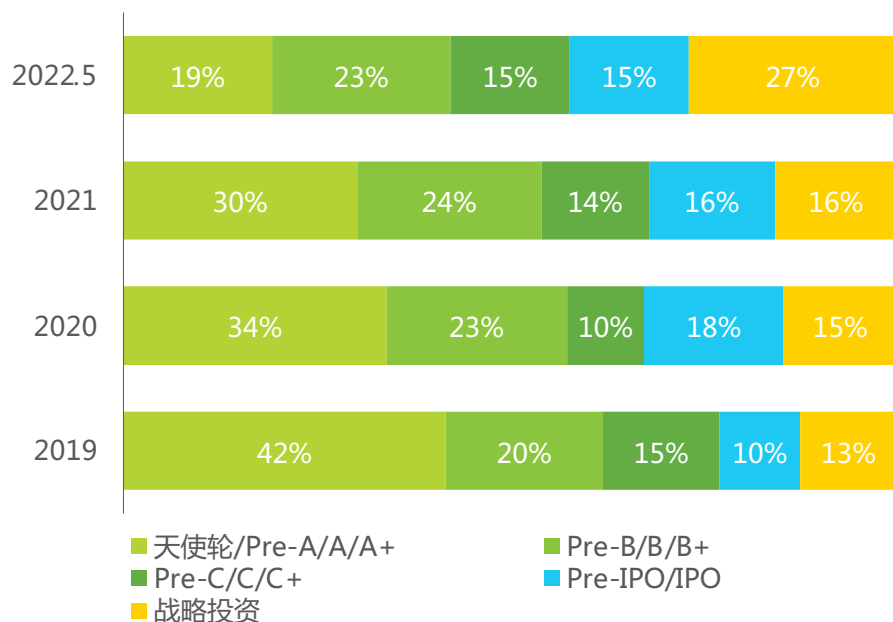
来源：IT桔子，艾瑞咨询研究院整理及绘制。

投资周期分析

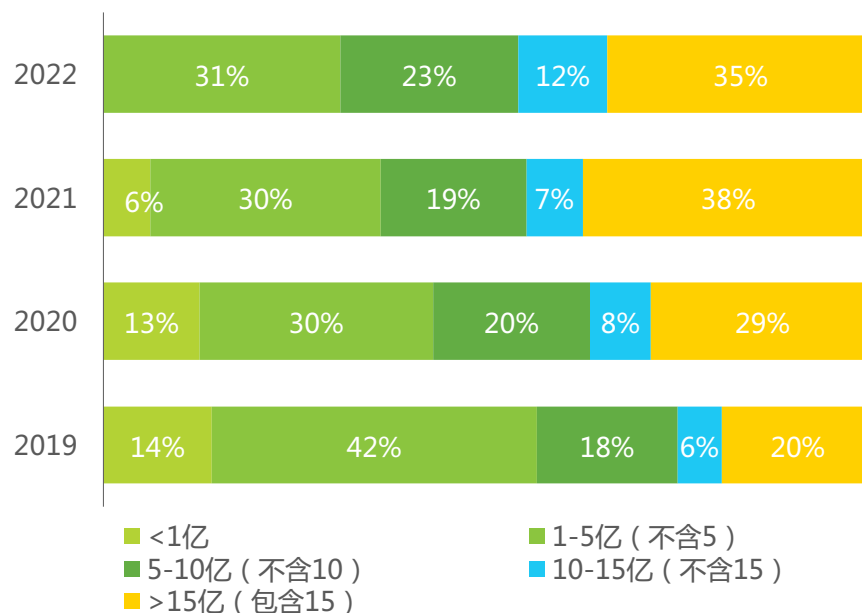
投资机构向头部聚拢，融资轮次偏向后期，资本与产业两端的马太效应凸显

2019-2022年5月，资本市场最初对大数据分析行业的早期项目有明显偏好，伴随大数据企业生命周期演进，资本更多投资于成长期和成熟期企业。高估值融资事件比重逐年提升，反映出大数据分析行业的马太效应显著，受资本青睐的头部企业持续获投，融资轮次多分布于中后期，资源高度集中。头部企业凭借关键技术和商业模式优势迅速占领市场，借助资本的力量发展壮大成为行业独角兽。对于初创企业而言，只有通过长期的技术积累，在垂直领域深挖，构建起技术壁垒，同时在商业模式上做出一定创新，才有可能在激烈的行业竞争中生存下来。

2019-2022年5月项目融资次数占比趋势



2019-2022年5月融资后项目估值



来源：IT桔子，艾瑞咨询研究院整理及绘制。

来源：IT桔子，艾瑞咨询研究院整理及绘制。

受疫情影响，市场风险偏好下行，投资风险需要重点关注

在国家宏观政策驱动下，数字经济转型仍是目前较为热门的投资赛道，但受疫情影响整体市场风险偏好下行，大数据分析行业的投融资活动需重点关注项目的投资风险。从产业链各环节投资价值上看，中游数据服务层的成长期和成熟期项目普遍已进去中后期轮次，价格偏高，综合性价比较低，新兴型项目难获得后续融资，退出风险较高，整体产业价值较小；建议重点关注中下游服务+应用的创新融合层，聚焦深刻理解应用行业、有效解决行业难点、重点应用2-3个关键行业的目标项目。

投资风险分析

市场竞争超预期风险

市场竞争超预期可能导致投资项目的市场份额、盈利能力和估值情况低于预期。

技术研发风险

大数据分析行业受技术驱动的影响显著，存在投资项目的技术研发不及预期风险。

数据合规使用风险

大数据分析行业涉及数据采集分析，若监管政策变化，员工及合作方、客户不当操作，可能产生数据合规风险，对投资项目的业绩和估值产生不利影响。

估值模式风险

大数据分析行业的研发和经营普遍具有较强的创新性，投资项目的业务模式和技术细分领域与市场可比公司可能存在差异，存在可比估值误差的风险。

业务拓展慢于预期

厂商业务面向大量垂直行业及政府客户，新业务拓展和应用落地存在低于预期风险。

客户流失或付费能力降低

随着行业竞争程度日益剧烈，存在客户流失风险，从而对公司规模增长和业务拓展形成负面影响。

艾瑞新经济产业研究解决方案



行业咨询

- 市场进入 为企业提供市场进入机会扫描，可行性分析及路径规划
- 竞争策略 为企业提供竞争策略制定，帮助企业构建长期竞争壁垒



投资研究

- IPO行业顾问 为企业提供上市招股书编撰及相关工作流程中的行业顾问服务
- 募 投 为企业提供融资、上市中的募投报告撰写及咨询服务
- 商业尽职调查 为投资机构提供拟投标的所在行业的基本面研究、标的项目的机会收益风险等方面的深度调查
- 投后战略咨询 为投资机构提供投后项目的跟踪评估，包括盈利能力、风险情况、行业竞对表现、未来战略等方向。协助投资机构为投后项目公司的长期经营增长提供咨询服务

关于艾瑞


艾瑞咨询是中国新经济与产业数字化洞察研究咨询服务领域的领导品牌，为客户提供专业的行业分析、数据洞察、市场研究、战略咨询及数字化解决方案，助力客户提升认知水平、盈利能力和综合竞争力。

自2002年成立至今，累计发布超过3000份行业研究报告，在互联网、新经济领域的研究覆盖能力处于行业领先水平。

如今，艾瑞咨询一直致力于通过科技与数据手段，并结合外部数据、客户反馈数据、内部运营数据等全域数据的收集与分析，提升客户的商业决策效率。并通过系统的数字产业、产业数据化研究及全面的供应商选择，帮助客户制定数字化战略以及落地数字化解决方案，提升客户运营效率。

未来，艾瑞咨询将持续深耕商业决策服务领域，致力于成为解决商业决策问题的顶级服务机构。

联系我们 Contact Us

 400 - 026 - 2099

 ask@iresearch.com.cn



企 业 微 信



微 信 公 众 号

法律声明

版权声明

本报告为艾瑞咨询制作，其版权归属艾瑞咨询，没有经过艾瑞咨询的书面许可，任何组织和个人不得以任何形式复制、传播或输出中华人民共和国境外。任何未经授权使用本报告的相关商业行为都将违反《中华人民共和国著作权法》和其他法律法规以及有关国际公约的规定。

免责条款

本报告中行业数据及相关市场预测主要为公司研究员采用桌面研究、行业访谈、市场调查及其他研究方法，部分文字和数据采集于公开信息，并且结合艾瑞监测产品数据，通过艾瑞统计预测模型估算获得；企业数据主要为访谈获得，艾瑞咨询对该等信息的准确性、完整性或可靠性作尽最大努力的追求，但不作任何保证。在任何情况下，本报告中的信息或所表述的观点均不构成任何建议。

本报告中发布的调研数据采用样本调研方法，其数据结果受到样本的影响。由于调研方法及样本的限制，调查资料收集范围的限制，该数据仅代表调研时间和人群的基本状况，仅服务于当前的调研目的，为市场和客户提供基本参考。受研究方法和数据获取资源的限制，本报告只提供给用户作为市场参考资料，本公司对该报告的数据和观点不承担法律责任。

合作说明

该报告由北极九章和艾瑞共同发起，旨在体现行业发展状况，供各界参考。

为商业决策赋能

EMPOWER BUSINESS DECISIONS

iResearch

艾 瑞 咨 询