



企业一站式数据开发与治理平台解决方案

2024-01-07



01

背景与挑战

数据治理存在的背景以及挑战



PART

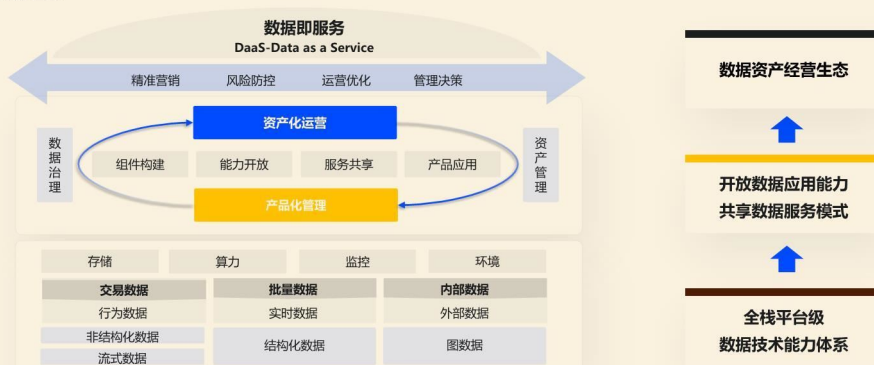
企业普遍存在数据现状问题

随着数据时代的到来，以Google、Facebook、Amazon为代表的企业已经完成了从IT巨头到DT巨头的转变。这些公司借助其在大数据、云计算、人工智能的技术发展优势，快速实现业务数据化、数据资产化和企业经营数据化，加速商业价值的转化，在引领技术风向的同时获取了巨大的商业成功。



数据服务体系的数据驱动模式

依托大数据技术，以数据服务、数据产品的体系化建设为枢纽，推进多元化数据应用建设，健全数据治理体系，加速数据资产化管理，激发数据动能。



企业在实行数据战略时需要关注什么

专注与数据，将数据视为战略资产

- 你有什么数据？
- 这些数据在哪里？
- 如何治理和保护它们？
- 如何访问它们？
- 如何发挥他们的价值？



完善的数据应用能力对企业未来至关重要

数据在不断的增长



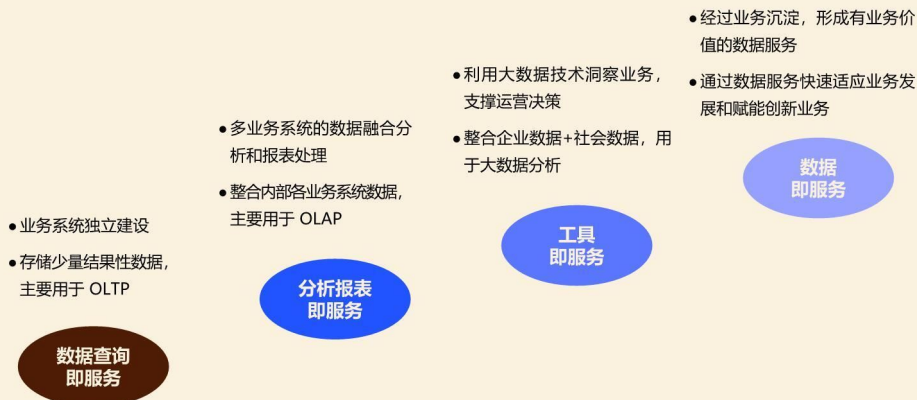
数据来源：Statista, 2020年11月

但还未被充分利用

未来两年企业数据将以 **42.2%** 的速率保持高速增长，但与此同时，调研结果显示，企业运营中的数据只有 **56%** 能够及时捕获，而这其中，仅有 **57%** 的数据得到了利用，**43%** 的采集数据并未被激活。也就是说，仅有32%的企业数据价值能够被激活。

数据来源：IDC和希捷科技, 2020年

企业在数据应用能力的建设路径



未来的企业数据中台需要什么

- 1** 支持数据的一站式开发治理和服务, 提供智能化的数据任务调度与监控
- 2** 弹性的, 可插拔的计算、查询引擎及数据治理工具组件
- 3** 开源兼容, 能够支持各种开源大数据组件集成, 满足客户不同场景需求
- 4** 支持混合云, 与云原生技术深度融合, 优化底层存储和计算引擎架构
- 5** 支持流批一体、湖仓一体的现代化数据架构设计

02

数据中台方案

数据中台方案及内容

PART

数据中台解决方案

数据开发与治理平台

从数据汇聚到数据服务的一站式大数据智能研发与治理平台，支持实时/离线计算引擎，集成了数据集成、数据开发、生产运维、实时分析、资产管理、数据质量、数据安全、数据共享等核心数据工艺，承上启下，让数据从采集到展现、从分析到驱动应用实现一站式平台服务

大数据引擎套件

提供可插拔的数据引擎服务，可根据业务需求在数据处理时编排引擎处理模式，并依托云原生化的特性进一步提升引擎的调度和运行效率。



数据应用与专家服务

提供在数据中台建设中企业所需要的各类专家服务和数据应用定制化服务，帮助企业打通数据业务闭环，实现数据价值赋能数据运营。

大数据运维管控平台

提供跨大数据生态的智能化大数据挖掘群运维管理能力，提升企业对大数据平台的运维效率，降低运维成本。

大数据引擎套件

- 提供覆盖全业务场景的计算引擎包含了离线计算、流式计算、数据查询、数据分析、机器学习等各个领域。
- 引擎模块以容器化封装杜绝与平台层的依赖，可根据业务场景需求提供热插拔式的调用为业务场景提供最佳的服务。
- 基于容器的租户隔离能力，大幅提升资源隔离和管控能力
- 提供弹性调度能力有效提高集群的利用率，并可实现多集群多云的调度编排能力



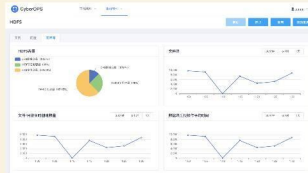
大数据运维管控平台

产品功能

- 向导式自动化平台安装、部署过程，支持集群中节点的自动化配置，覆盖 Hadoop生态及第三方各类大数据组件(Greenplum、Doris等)安装、升级、维护
- 支持对各类组件管理和集群各节点的可视化、图形化、指标化监控
- 支持统一的告警服务，提供消息、邮件及界面可视化等多种方式展示
- 支持基于整体平台或独立组件的版本升级，并保障数据的持续可用

客户价值

简化大数据组件安装、维护、监控预警、日志管理等服务，提升运维人员对大数据集群各服务的管理和维护的效率50%以上



产品优势

依赖关系自动化

一个配置能够自动的分发到所有的关联组件节点上

组件参数智能化

根据当前的硬件条件自动设置组件的配置内存、CPU、硬盘IO参数

运维监控可视化

快速找出性能瓶颈、错误日志等，给出优化建议

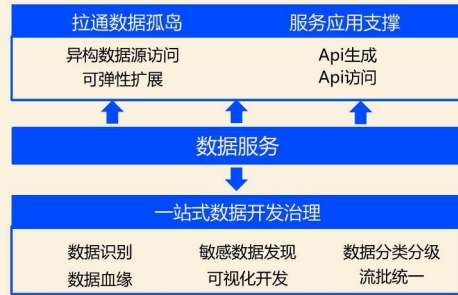
一站式数据开发与治理平台

产品功能

- 提供全链路的数据管理功能，从产生、流转、加工、使用、到销毁
- 多种粒度的元数据可视化管理分析功能，实现数据全生命周期的查看与溯源
- 提供基于组织架构的多层级资源配额配置的项目空间管理
- 支持各种主流关系型数据库和大数据平台，支持各种类型数据文件

核心模块

- 数据集成
- 元数据系统
- 统一调度系统
- 数据服务
- 数据治理套件



产品优势

智能调度	流批一体可视化	使用门槛无代码化
调度引擎实时获取服务器资源信息，通过智能调度算法动态分配任务至资源充分的服务器	可视化的流批一体开发模式，统一离线和实时数据处理	屏蔽底层数据源细节，丰富的开发组件，拖、拉、拽式的开发模式，降低使用门槛

数据中台数据分层

面向业务	ADS层	Application Data Store, 数据仓库应用数据集市层。完全贴合前端业务应用面向应用提供数据服务 关键词: 高度汇总, 应用指标
	DWS层	Data Warehouse Subject, 数据仓库主题数据层。按分析对象对实体进行数据整合, 轻度汇总 面向应用提供数据服务 关键词: 轻度汇总, 公共数据模型
面向数据	DWD层	Data Warehouse Detail, 数据仓库明细层数据, 面向数据进行数据域分类, 整合、清洗、形成一套标准化数据模型 关键词: 数据清洗, 数据模型
	ODS层	Operational Data Store, 操作性型数据, 指结构与源系统保持一致的全量数据对STG层数据进行数据全/增量合并以及数据清洗和标准化动作 关键词: 数据清洗, 数据模型
	STG层	STAGE, 结构与源系统保持一致的增量数据。汇聚业务系统源头数据是ETL加工过程的缓冲区 关键词: 数据清洗, 数据模型

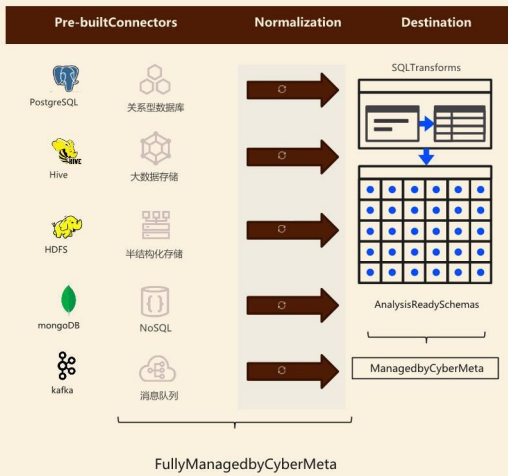
数据中台整体架构



平台优势



数据中台：数据集成



- ✓ 多数据源支持：支持各种主流关系型数据库和大数据平台，支持各种类型数据文件
目前支持数据源：MySQL、Oracle、SQLServer、Hive、FTP、Minio、Hbase、HDFS...
- ✓ 集成引擎采取异步并行处理的技术，实现流程中的每个组件多线程并行高效处理
- ✓ 单点容错机制，保证整体作业的高可用
- ✓ 基于日志的FlinkCDC实时数据采集方案

数据中台：数据开发



- ✓ 支持通过脚本、Java程序、API、SQL等多种方式进行数据查询分析
- ✓ 支持数据的多源、多类型的导入与导出
- ✓ 提供可视化的SQL执行IDE，支持不同类型的数据查询及结果输出
- ✓ 流批一体开发模式，统一离线和实时数据处理
- ✓ 调度引擎实时获取服务器的资源信息，基于智能调度算法动态分配任务至资源充分的服务器

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：

<https://d.book118.com/596100011035010142>