

数据湖计算 DLC

产品简介



腾讯云

【 版权声明 】

©2013–2024 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分內容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

【 商标声明 】

及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

【 服务声明 】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

【 联系我们 】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或 95716。

文档目录

产品简介

产品概述

产品优势

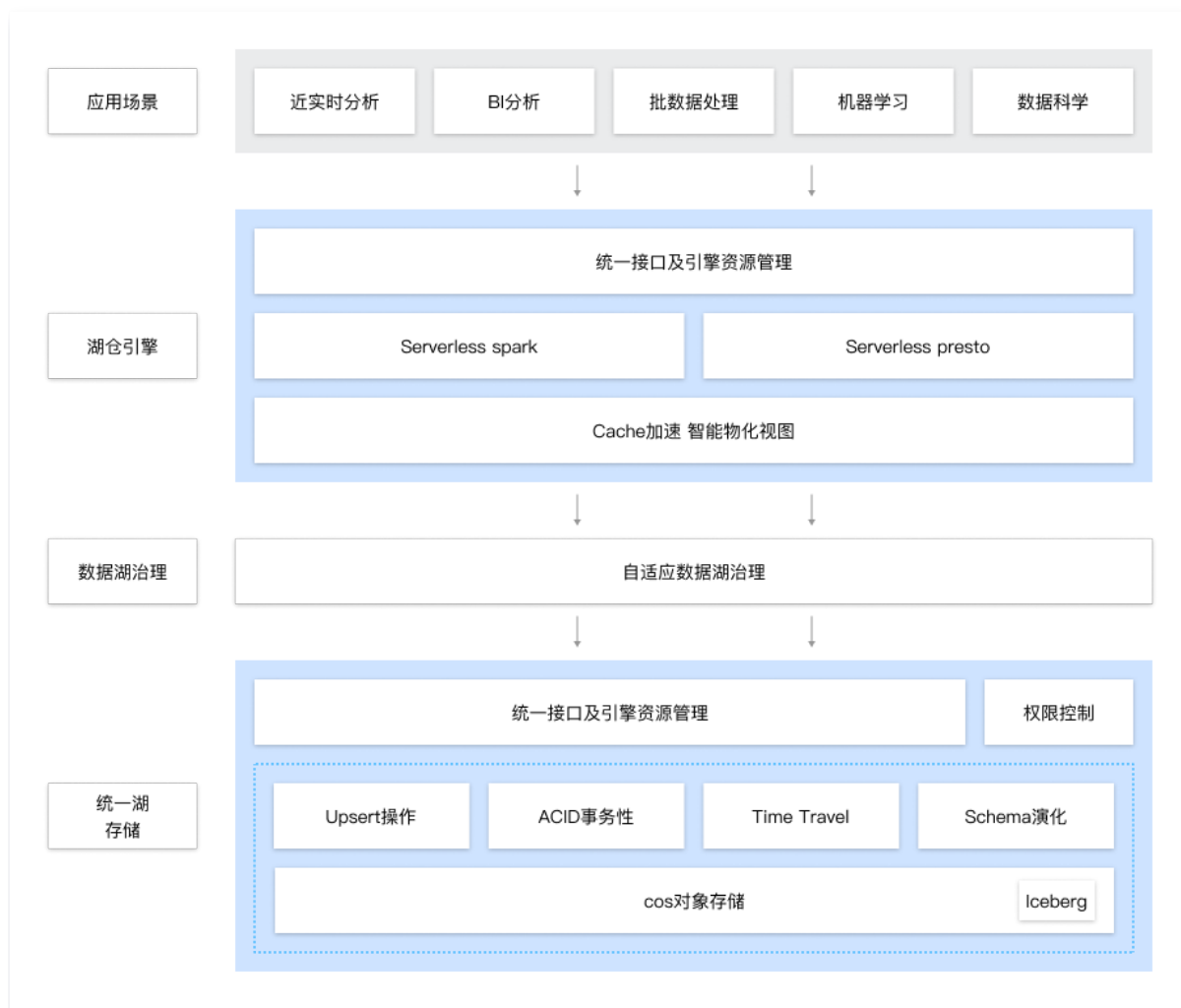
应用场景

产品简介

产品概述

最近更新时间：2023-09-20 16:35:31

数据湖计算DLC (Data Lake Compute, DLC) 提供了敏捷高效的数据湖分析与计算服务。服务采用无服务器架构 (Serverless)，开箱即用。使用标准 SQL 语法即可完成数据处理、多源数据联合计算等数据工作，有效降低用户数据分析服务搭建成本及使用成本，提高企业数据敏捷度。



用户无需进行传统的数据分层建模，大幅缩减了海量数据分析的准备时间，并可更进一步联合数据开发治理平台 WeData、数据集成 (DataInLong) 等腾讯大数据生态产品，快速构建企业级云原生实时湖计算平台。

- **统一数据开发:** 对接腾讯云数据开发治理平台 WeData，实现数据湖的统一集成、开发、治理及应用。
- **高性能湖仓分析引擎:** 同时支持 Spark 和 Presto 两种引擎，统一多引擎 SQL 语法，查询缓存加速，更加敏捷、高效完成数据查询分析。
- **统一元数据服务:** 提供元数据统一管理、完善的权限体系，满足多租户使用场景。
- **自适应数据治理:** 具备湖格式智能治理能力，高效处理流式写入下的小文件合并，历史版本的过期快照删除。
- **统一湖存储:** 基于 COS 及 Iceberg 增强的数据存储，保障数据 ACID 事务性，支持物化视图缓存加速层。

主要功能

数据探索： SaaS 化开箱即用的数据湖分析

使用标准 SQL 即可完成对数据湖的轻松查询，兼容 SparkSQL，无需理解不同数据设施的数据构造，帮助客户实现由数据库场景无缝升级到大数据场景。同时支持对多源异构数据进行联合查询分析，包括MySQL、EMR Hive(COS)、EMR Hive(HDFS)等。

数据作业： 极致弹性、高效降本 spark批处理

面向大数据 + AI 场景，基于原生 Spark 的批处理、流计算能力支持用户通过数据任务进行复杂的数据处理、ETL 等操作。支持机器学习及AI场景常用依赖包管理，快速构建 AI 场景的大数据基座，同时具备完善的数据访问策略管理功能，支持配置数据访问策略，保障数据安全。

数据管理： 易用、全面的数据湖综合治理能力

提供数据湖统一元数据管理视图，可创建、编辑数据湖整体数据目录，及新建、查询、删除数据库表及数据视图，消除数据孤岛。同时支持备湖格式智能数据治理，用户无需关注复杂的数据湖格式治理与优化，DLC 将智能处理各种因频繁碎片写入而产生的海量小文件以及孤儿快照，综合提升数据湖查询性能。

数据引擎： 超大规模计算扩展，弹性降本增效

提供灵活、高效的 Spark 和 Presto 云原生计算引擎弹性管理，支持多种扩缩容规则配置，显著降低数据湖查询分析综合成本，充分贴合业务真实使用曲线。DLC 以低成本、高弹性的云原生数据湖解决方案，助力企业建立统一数据资产，最大化发挥性能优势，赋能业务应用敏捷创新。

产品优势

最近更新时间：2023-09-20 16:35:31

敏捷易用

- 数据湖计算 DLC 提供 SaaS 化体验，无需选配、安装、调优，开箱即用。
- 用户无需关注复杂的数据湖底层运维及调优，使用标准 SQL 语法即可轻松开始数据分析。

高效降本

- 数据湖计算 DLC 采用存算分离的海量大数据分析架构，大数据组件容器化实现快速灵活部署，基于云原生对象存储的方式实现无限拓展、弹性伸缩。
- 支持按量计费，降低用户查询分析的成本，如采用数据分区或列式压缩格式，可进一步节约成本。

湖仓一体

- 面向跨湖仓一体架构的统一 SQL 分析及批作业处理，通过一套统一的数据湖架构完整支持企业级 BI、机器学习及数据科学场景。
- 灵活搭建 EB 级湖仓存储，支持超大规模机器学习及近实时数仓分析。

极致性能

- 采用无服务器（Serverless）架构，用户无需关注底层运维，计算资源即用即毁，系统根据算力需求提供秒级伸缩和动态扩容能力。
- 提供高性能数据引擎，高效建模提升查询效率，零成本缓存加速方案带来极致性能体验。
- 产品实现：交互式查询、批量查询、智能分析等多种场景全覆盖。

安全加固

- 基于腾讯云成熟的 VPC 网络隔离技术，保证多租户隔离。
- 通过腾讯云品质的安全加固，实现高可靠的数据的安全。
- 细粒度的权限管控，操作更加安全。

数据科学

- 数据湖计算 DLC 快速支持各种机器学习能力，满足一站式智能分析场景。
- 产品提供可视化能力助力用户通过预测分析实现数据洞察。

生态融合

- 无缝融合腾讯云数据生态，可以直接读取云存储服务数据。
- 具备良好的跨平台兼容性，支持各类上层数据应用。

核心能力

统一SQL，高性能引擎：100%兼容spark、presto，覆盖交互式、离线分析场景。Alluxio local cache+亲和性调度方案。Sort Benchmark冠军性能增强

Iceberg独特治理能力：国内独家的Iceberg数据湖一体化治理能力，客户无需关注Iceberg小文件合并、孤儿快照删除等治理顽疾，完全透明使用数据湖

Serverless极致降本：通过计算引擎优化与弹性伸缩帮助客户降低数据分析成本，客户验证算力资源节约30%

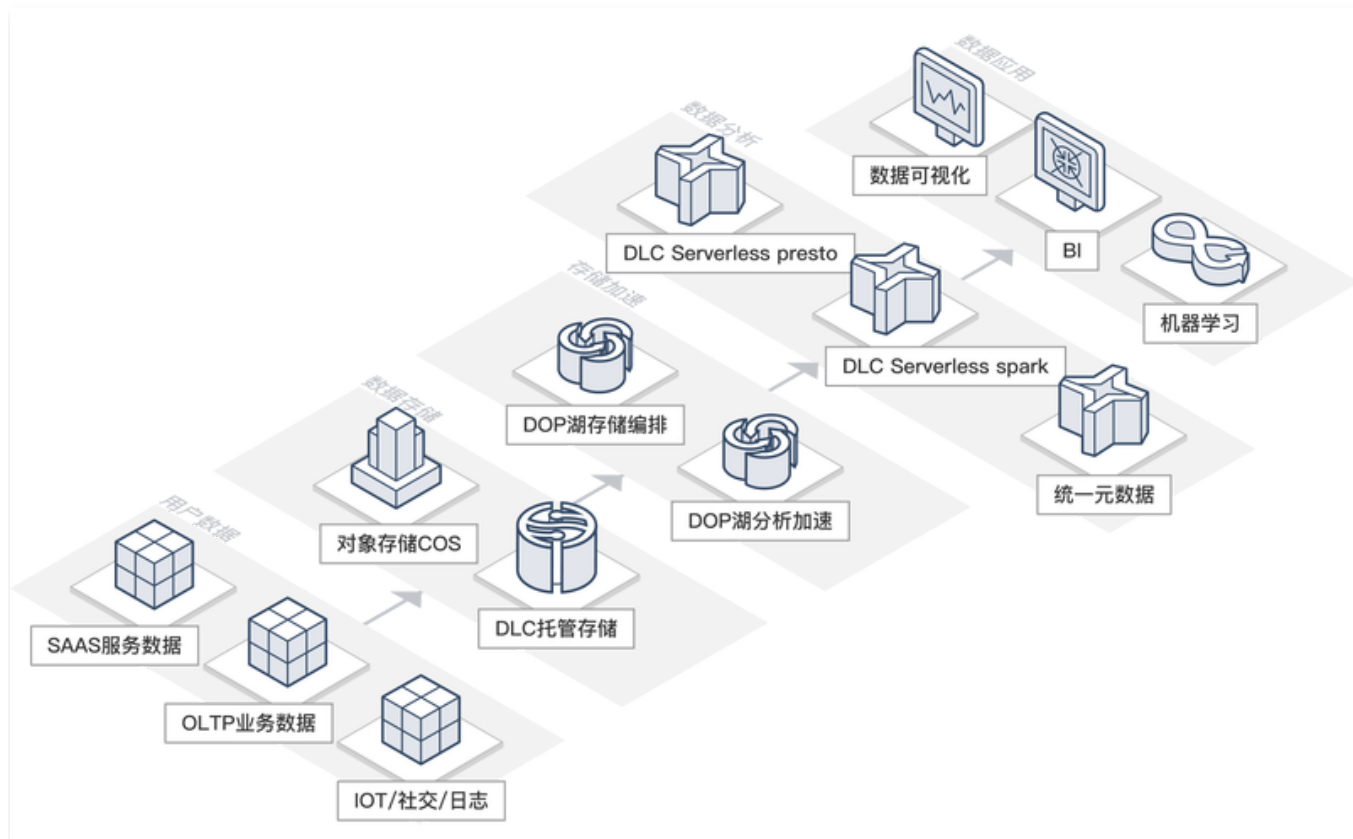
一站式数据湖管理，企业级0运维:统数据湖管理，快捷交付，0运维即可即用。成熟技术团队24小时在线服务

应用场景

最近更新时间：2023-09-20 16:35:32

敏捷实时的数据湖分析

DLC 采用存算分离的海量大数据分析架构，基于大数据组件容器化实现快速灵活部署，基于云原生对象存储的方式实现无限拓展，结合 DLC 先进的云原生弹性模型，充分贴合业务真实使用曲线，真正为您节省成本。DLC 以低成本、高弹性的云原生数据湖解决方案，助力企业建立统一数据资产，最大化发挥性能优势，赋能业务应用敏捷创新。



典型场景

企业日志批量查询

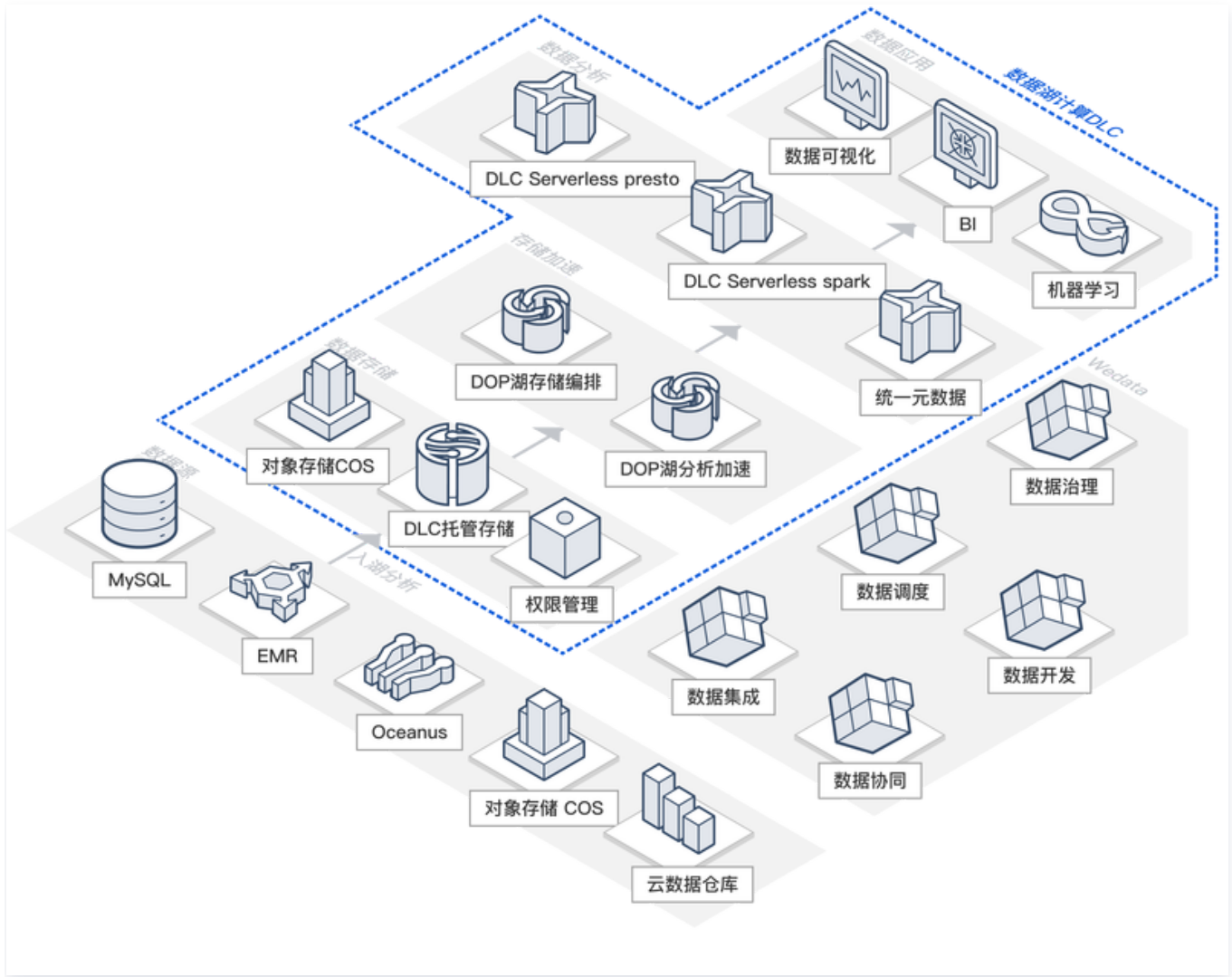
企业日志数据通常以 json、文本文件等格式进行存储，用户可以将日志数据存储到 COS 中，直接使用标准 SQL 快捷地对 COS 中的海量数据进行批量分析，并将查询结果快速生成数据报表，轻松实现数据可视化，大幅提升工作效率。DLC 只需要几步简单的配置，就可以将云上日志服务的数据导入到 DLC 中，快速上手加速分析。

服务价值

- 成本优化：存算分离云原生数据湖架构，按使用量付费，精准控制成本投入。
- 强大易用：统一 SQL 语法，上手使用简单，更快的查询速度。

助力企业敏捷搭建数据中台

数据湖计算 DLC 本身作为新型数据架构，本身提供轻量敏捷易用和低成本闭环大数据分析能力，用户可以借助数据湖提供的统一元数据管理视图，消除数据孤岛。同时也可以组合云上丰富大数据产品的优势，满足各类数据实时、离线分析场景，全方面解决企业各类问题。通过数据的便捷快速流动，可以有机组合不同云上产品的能力和优势，使得 DLC 得以作为企业最佳数据中台和数据启动场所。



典型场景

统一元数据视图

用户在云上可能存在多套元数据视图，例如 EMR、DLC 和其他各类数据源产品，DLC 内置企业级统一元数据视图，可以集中管理和使用不同数据源的元数据，敏捷快速地构建企业元数据中心，无缝在不同产品和版本之间切换。典型代表同一份元数据，可以在 DLC 和 EMR 不同产品之间便捷切换使用。

一份数据敏捷泛场景分析

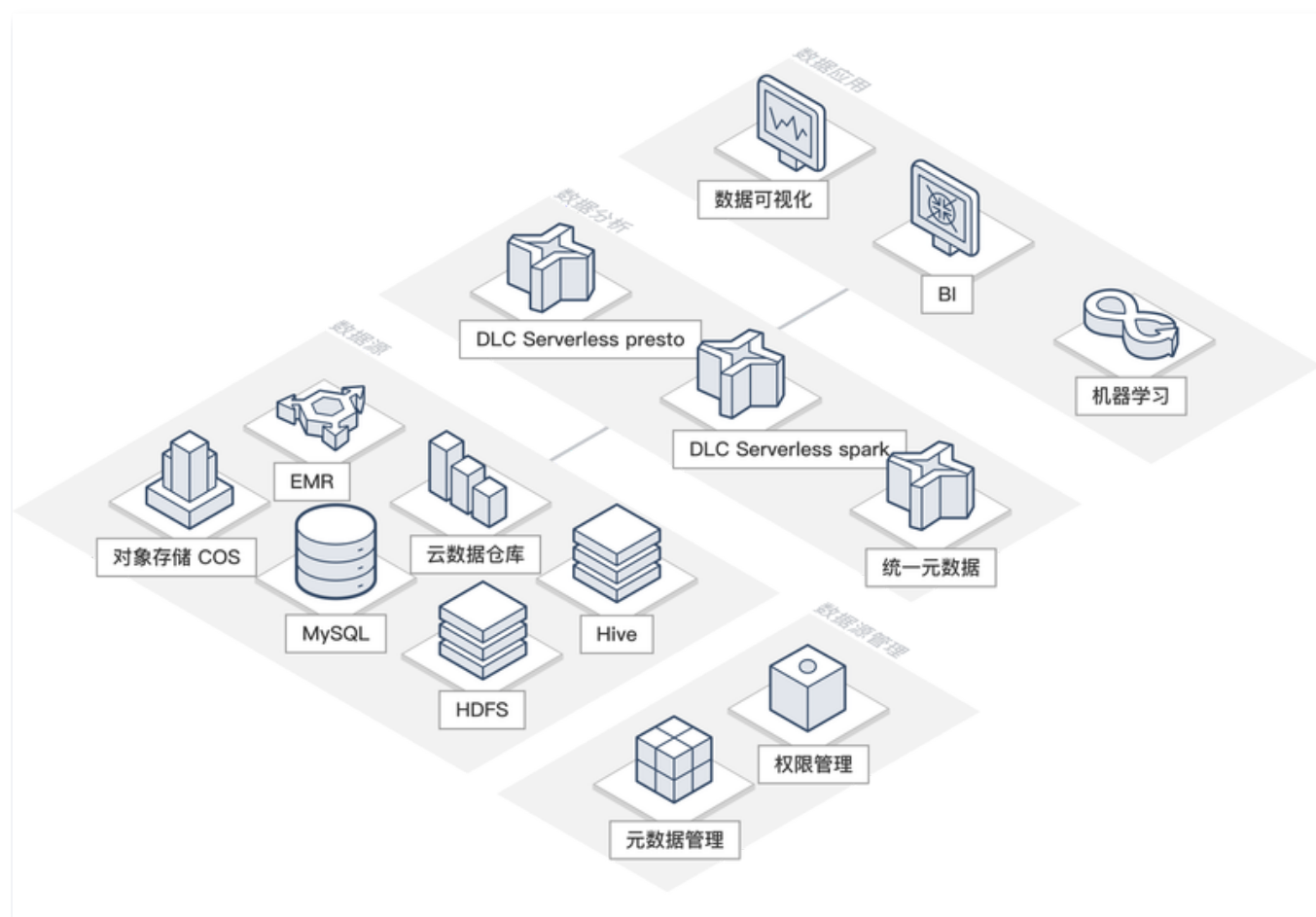
在大数据生态中，Presto 擅长交互式分析，Spark 擅长 ETL 任务，通过 DLC 提供的统一语法和轻量化集群能力，可以做到同一份数据无缝在不同引擎间切换，满足各类场景。也可以通过结合 WeData 将数据导入导出到其它数十种数据产品和数据源，如 EMR、CDW、ES、数据库、日志服务等，通过数据的便捷流动，充分发挥不同产品的优势。

服务价值

- 开箱即用：无需多余运维，节省运维成本。
- 元数据管理：支持多种数据源，统一元数据管理，打破数据孤岛。
- 场景全覆盖：数据分析、应用场景覆盖全面，数据集成、协同、调度、开发、治理全覆盖。

敏捷数据湖联邦分析

DLC 帮助客户实现由数据库场景无缝升级到大数据场景，支持对多源异构数据进行联合查询分析，包括对象存储、云数据库、其他数据服务等。用户通过统一的数据视图，使用标准的 SQL 即可快速实现多源数据联邦分析，打通数据孤岛，发挥数据价值。



典型场景

跨业务数据联合查询

企业内部不同部门和业务线通常基于业务体系采用不同的数据架构，业务数据存放在不同的存储系统中，存在数据割裂的情况：如交易型数据存储到关系型数据库、活跃数据存储到 Redis、历史记录存储到对象存储。DLC 帮助用户实现异构数据打通，用户可以跨多个数据源进行联合分析，助力用户跨业务数据分析更加迅捷。

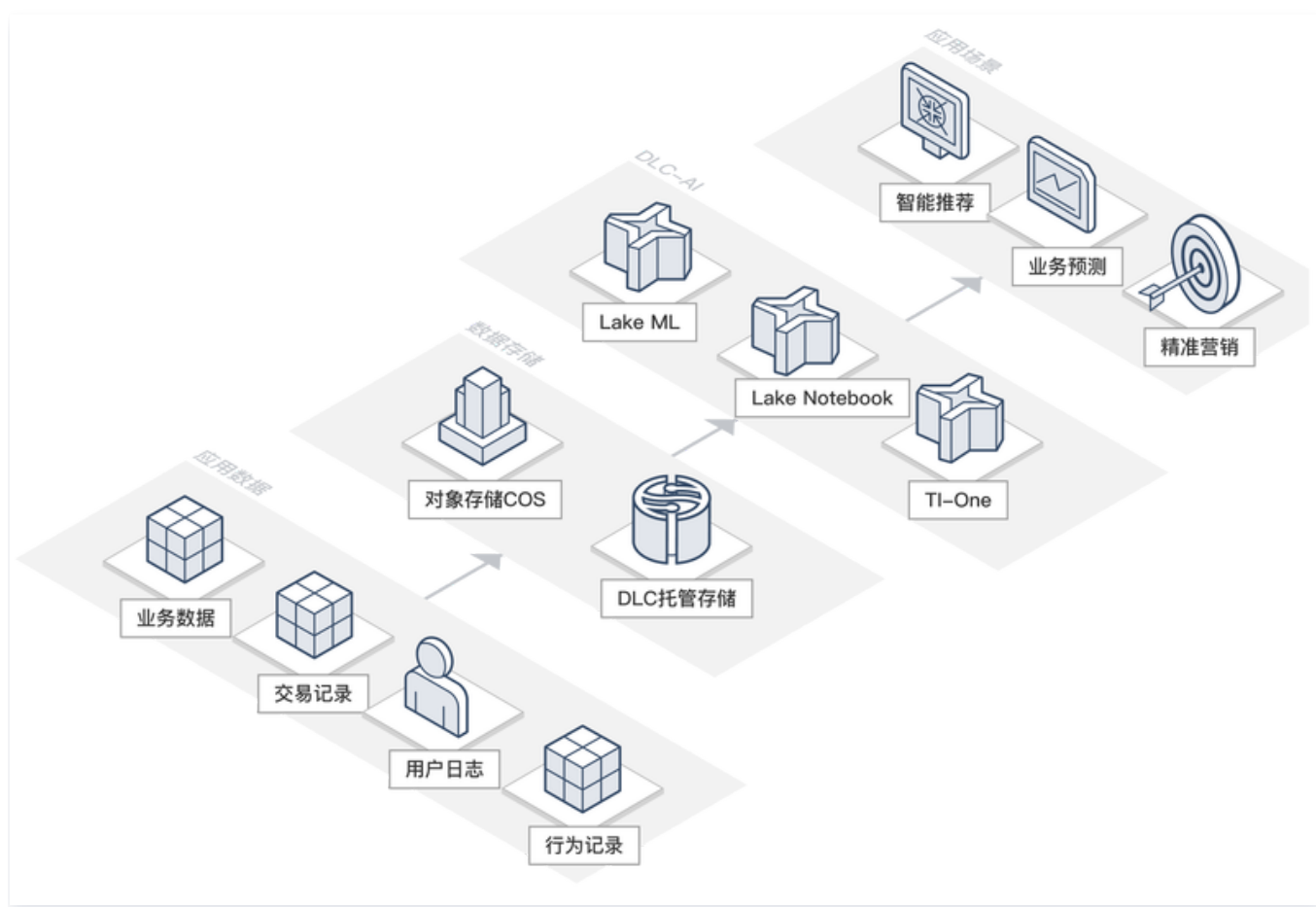
服务价值

- 开箱即用：无需构建数据传输管道，无需多余运维，节省运维成本。
- 安全高效：统一权限管理体系，数据权限精确到列，极致查询速度。

- 迅捷使用：无需适配编程语言，轻松实现跨业务分析。

丰富多元的数据湖科学分析

数据湖是 AI 场景的大数据基座，在经典机器学习场景和深度学习场景下服务用户，DLC 提供结合各种AI能力及平台，快速支持各种机器学习能力，在多种智能数据湖分析场景下，提供综合性解决方案。DLC 将多个行业数据免费开放给用户，无需数据获取、清洗即可直接进入数据分析阶段。产品提供强大的 BI 能力助力用户通过预测分析快速实现数据洞察。



典型场景

数据科学赋能业务增长

DLC 提供原生机器学习能力，融合成熟的机器学习平台，为用户提供完善的智慧分析解决方案，解决用户实际的业务问题，如智能推荐、召回策略等，助力企业实现业务增长。在机器学习场景下，用户面临数据量大，模型训练慢，算法效果差的问题。DLC 提供的机器学习算法模型开箱即用，用户可以快速基于数据构建机器学习模型，轻松预测业务成果。DLC 为用户提供 BI 能力，助力企业实现高效商业智能分析，提供企业运营效率。

服务价值

- 易用性强：无缝对接腾讯云机器学习平台，提供多种模型及 API 方便使用。
- 数据规范：对数据进行统一管理、治理，为数据科学提供标准化数据。