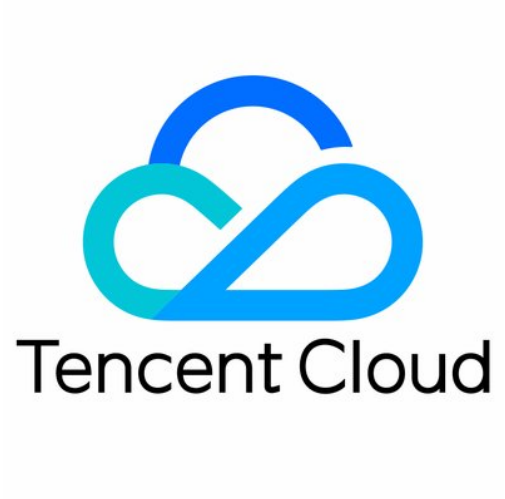
Data Lake Compute

Product Introduction

Product Documentation



Contents

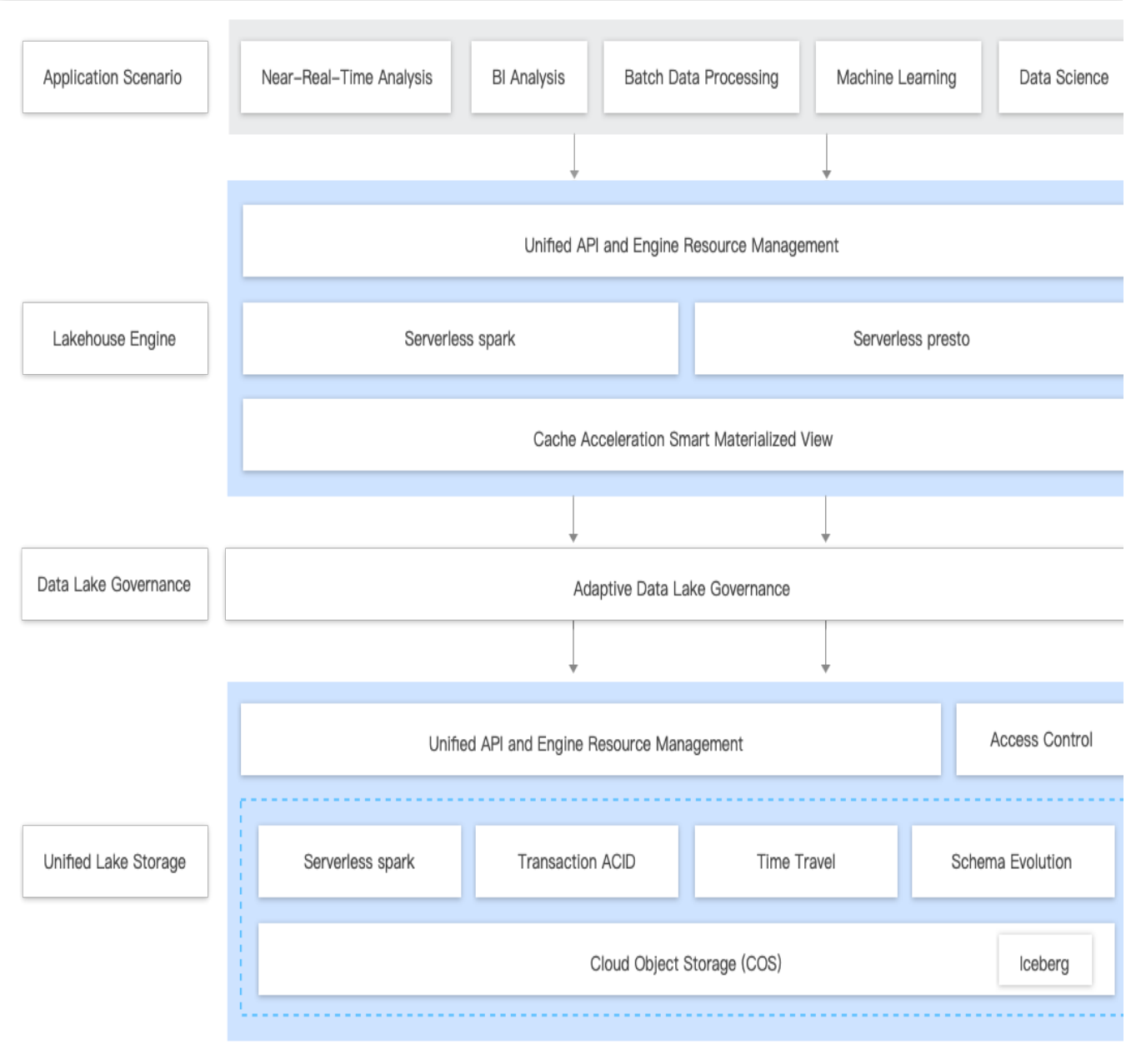
Product Introduction
- Overview
- Strengths
- Use Cases

Product Introduction

Overview

Last updated：2025-01-03 15:27:27

Data Lake Compute (DLC) offers agile and efficient data lake analytics and computation services. With its serverless architecture, it is ready to use out of the box. By utilizing standard SQL syntax, it can accomplish data processing and multi-source data joint computation, effectively reducing the cost of setting up and using data analysis services, and enhancing the agility of enterprise data.



Users are not required to perform traditional data layer modeling, significantly reducing the preparation time for massive data analysis. Furthermore, it can be combined with Tencent's big data ecosystem products such as WeData and DataInLong to swiftly construct an enterprise-level, cloud-native, real-time lake computation platform.

Unified Data Development: Integration with Tencent Cloud's WeData platform for unified integration, development, governance, and application of data lakes.

High-Performance Lakehouse Analysis Engine: Supports both Spark and Presto engines, unifies multi-engine SQL syntax, and accelerates query caching, thereby completing data query analysis more swiftly and efficiently.

Unified Metadata Service: Offers unified metadata management and a comprehensive permission system to meet multi-tenant usage scenarios.

Adaptive Data Governance: Equipped with intelligent lake format governance capabilities, efficiently handling small file merging under stream writing and deletion of expired snapshots from historical versions.

Unified Lake Storage: Data storage enhanced by COS and Iceberg, ensuring ACID transactionality of data and supporting a materialized view cache acceleration layer.

Main Feature

Data Exploration: Ready-to-use SaaS-based data lake analysis.

Standard SQL can be used to easily query data lakes, compatible with SparkSQL, eliminating the need to understand the data structure of different data facilities, and assisting customers in seamlessly upgrading from database scenarios to big data scenarios. It also supports joint query analysis of heterogeneous data from multiple sources, including MySQL, EMR Hive(COS), EMR Hive(HDFS), and more.

Data Jobs: Ultimate elasticity and cost-effective Spark batch processing.

Aimed at big data + AI scenarios, it leverages the batch processing and stream computing capabilities of native Spark to support users in performing complex data processing and ETL operations through data tasks. It supports the management of commonly used dependency packages in machine learning and AI scenarios, swiftly constructing a big data foundation for AI scenarios. Additionally, it possesses a comprehensive data access policy management function, supporting the configuration of data access policies to ensure data security.

Data Management: User-friendly and comprehensive capabilities for holistic governance of data lakes.

Provides a unified metadata management view for data lakes, enabling the creation and editing of the overall data directory of the data lake, as well as the creation, querying, and deletion of database tables and data views, thereby eliminating data silos. It also supports intelligent data governance for backup lake formats. Users need not concern themselves with complex data lake format governance and optimization. DLC will intelligently handle a large number of small files and orphan snapshots generated by frequent fragmented writing, thereby comprehensively enhancing the performance of data lake queries.

Data Engine: Massive scale computation expansion, elastic cost reduction and efficiency enhancement.

Offers flexible and efficient elastic management of Spark and Presto cloud-native computing engines, supports various scaling rules configurations, significantly reduces the comprehensive cost of data lake query analysis, and closely aligns with the actual business usage curve. As a low-cost, highly elastic cloud-native data lake solution, Data Lake Compute empowers businesses to establish unified data assets, maximize performance advantages, and enable agile innovation in business applications.

Strengths

Last updated：2024-09-18 18:01:03

Agility and ease of use

DLC offers a SaaS-based experience that is ready to use without the need for additional selection, installation, or optimization.
Users can easily start data analysis using standard SQL syntax without worrying about complex underlying Ops or performance tuning of data lake.

Cost efficiency

DLC leverages a storage-compute separation architecture for massive big data analysis. Its containerized big data components enable rapid and flexible deployment, while cloud-native COS allows for unlimited scalability and auto scaling.
DLC supports pay-as-you-go billing, reducing the cost of query and analysis. Additionally, using data partitioning or columnar compression formats can further optimize cost savings.

Unified lakehouse architecture

DLC enables unified SQL analysis and batch processing of jobs for cross-lakehouse architecture, fully supporting enterprise-grade BI, machine learning, and data science scenes within a single data lake architecture.
It allows for the flexible construction of EB-level lakehouse storage, supporting large-scale machine learning and near-real-time data warehouse analysis.

Superior performance

Data Lake Compute is serverless, so you don't need to worry about the underlying Ops. The system terminates compute resources after use and scales instantly and dynamically as computing power requirements change.
It comes with high-performance data engines and efficient models to boost the query efficiency. As a cache acceleration solution with zero costs and superior performance, it covers interactive query, batch query, smart analysis, and much more use cases.

Security enhancement

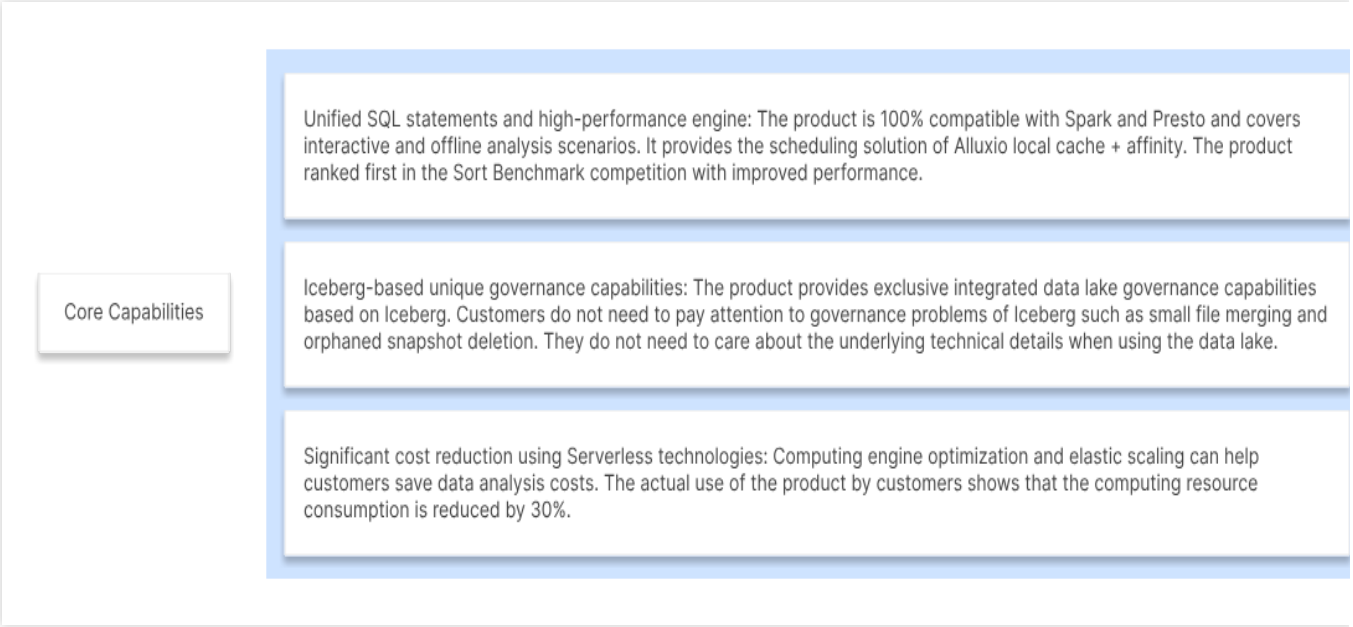
Data Lake Compute adopts Tencent Cloud's mature VPC network isolation technology to ensure that tenants are isolated at the network level.
It further achieves high data reliability and security thanks to Tencent Cloud's superior security enhancement.
It enables fine-grained permission control to make operations more secure.

Data portfolio

Data Lake Compute quickly supports a wide variety of machine learning capabilities to accommodate use cases of one-stop smart data analysis.
It offers visualization capabilities to help you gain data insights through predictive analysis.

Ecosystem integration

Data Lake Compute is seamlessly integrated into Tencent Cloud's data ecosystem for direct access to data stored in COS.
It is compatible with numerous platforms and supports a diversity of upper-layer data applications.

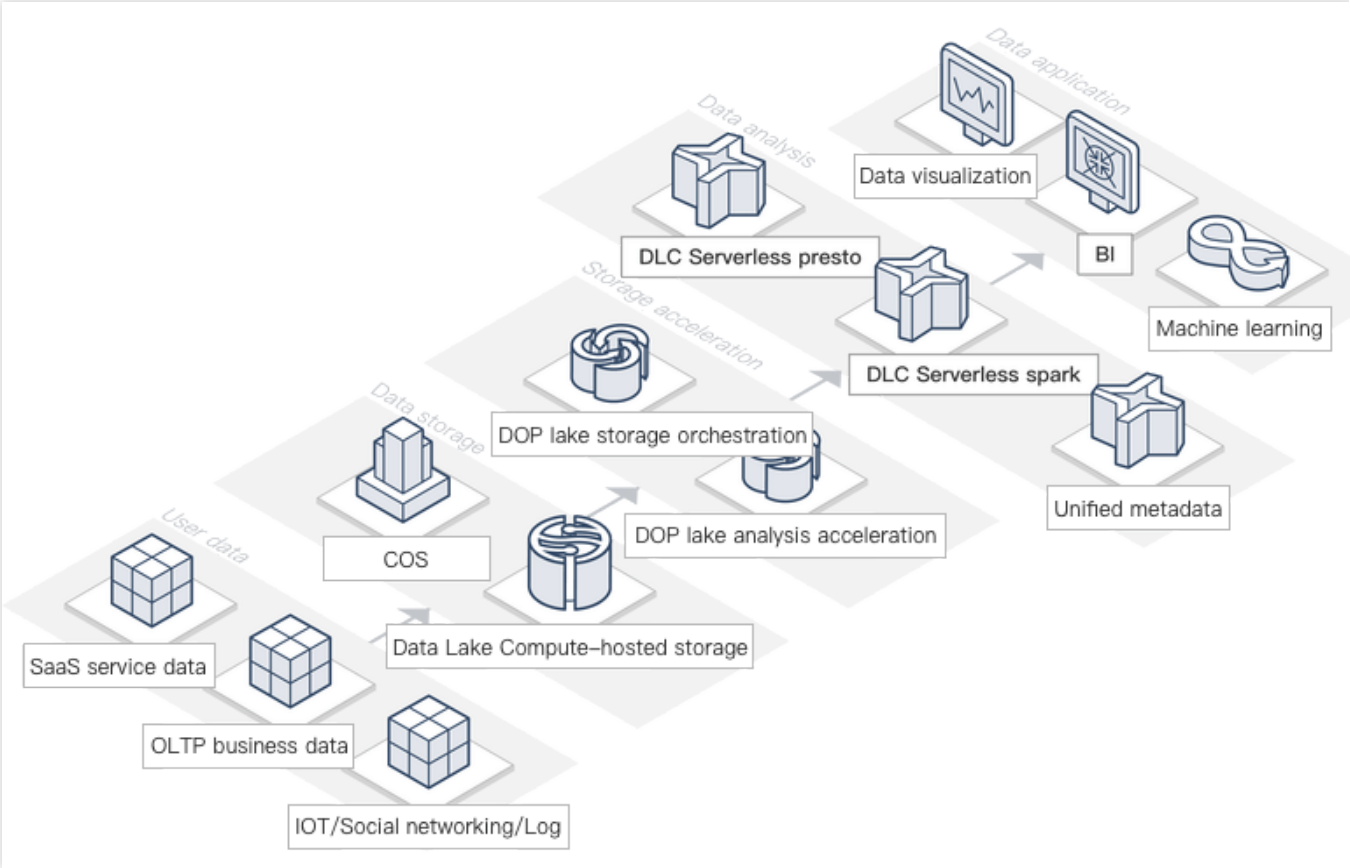


Use Cases

Last updated：2025-01-03 15:27:27

Agile and Real-Time Data Lake Analysis

Data Lake Compute leverages a big data analysis architecture with separated storage and computing. It enables fast and flexible deployments based on big data component containerization and implements unlimited expansions on top of cloud object storage. Its advanced cloud-native elastic model fits virtually any type of business to reduce your costs. As a cost-effective and highly elastic cloud data lake solution, it helps you unify data assets and maximize performance for agile and innovative business applications.



Typical use cases

Batch log query

Unlike the typical practice of storing enterprise log data as JSON and text files, you can store data in COS and then use standard SQL statements through Data Lake Compute to batch query and analyze massive amounts of data, with

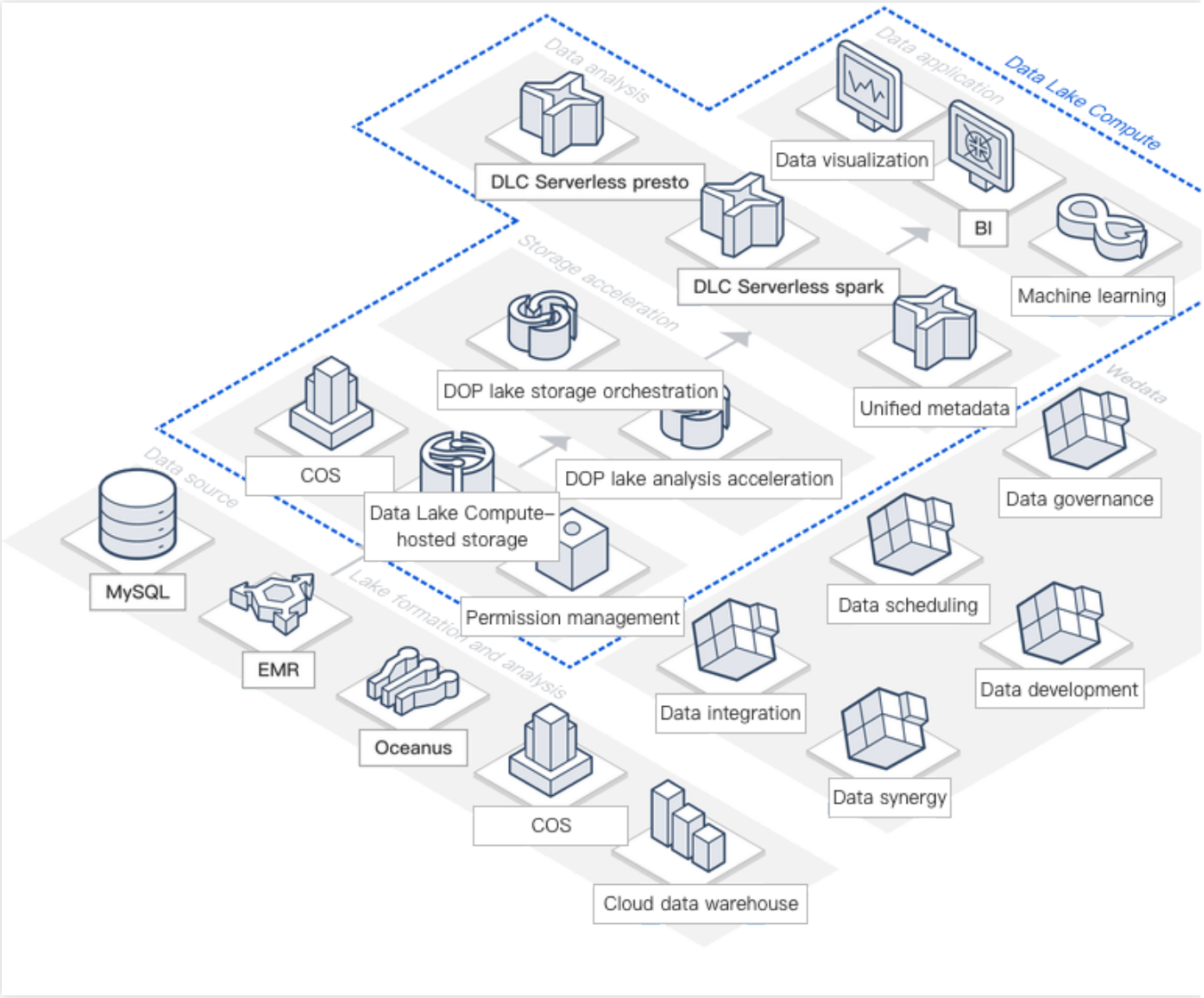
data reports generated quickly. In this way, Data Lake Compute visualizes your data and boosts your productivity. With a few simple configurations, you can also import cloud-based log service data into a data lake for agile analysis.

Service benefits

Cost-effective: Data Lake Compute is pay-as-you-go, allowing you to precisely control costs through its cloud-native data lake architecture with separated storage and computing.
Easy-to-use: You can easily get started with Data Lake Compute for faster queries through the unified SQL syntax.

Agile Setup of a Data Middleend

Data Lake Compute is a new data architecture with closed-loop big data analysis that is lightweight, agile, easy-to-use, and cost-effective. It has a unified metadata management view that allows you to break through data silos. It combines the strengths of many cloud-based big data services to accommodate real-time and offline data analysis scenarios and comprehensively solve a wide range of data problems. Moreover, with convenient and swift data flows, it features many of the capabilities and advantages of different cloud services, making it an ideal option for enterprises setting up a data middleend.



Typical use cases

Unified metadata view

Data Lake Compute enables you to unify all of your different metadata views such as EMR and other data sources into one. In this way, you can manage and use metadata from different sources in a centralized manner, build your metadata center with agility, and switch between products and versions seamlessly. Specifically, you can easily reuse the same metadata across products like Data Lake Compute and EMR.

Agile and versatile data analysis

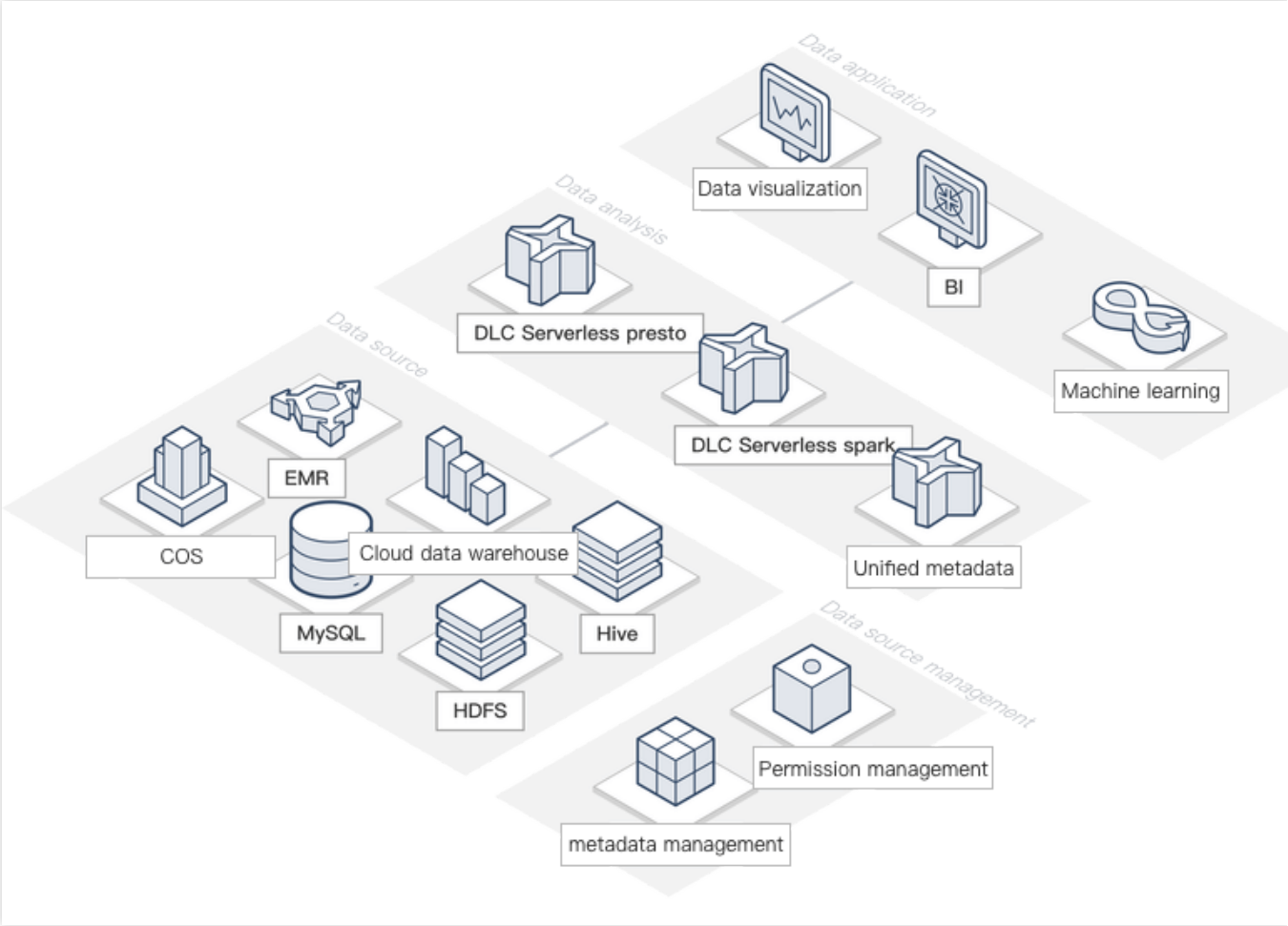
In the big data ecosystem, Presto excels in performing interactive analysis while Spark does well in ETL tasks. Data Lake Compute provides unified syntax and lightweight clustering capabilities, so the same data can go seamlessly between engines in different scenarios. It also works with WeData so data can be imported from and exported to dozens of products and data sources, such as EMR, CDW, ES, TencentDB, and CLS. This makes the most out of the strengths of each product through convenient data flows.

Service benefits

Out-of-the-box service: Unnecessary Ops tasks and costs are saved.
Metadata management: Multiple data sources are supported to unify metadata management and break through data silos.
Full coverage: Data Lake Compute comprehensively covers data analysis and application scenarios, specifically, data integration, synergy, scheduling, development, and governance.

Agile and Federated Data Lake Analysis

Data Lake Compute helps you seamlessly transition from database to big data scenarios, where you can query and analyze multi-source heterogeneous data in the cloud from object storage, database, and other services. Its unified data view and standard SQL capabilities speed up federated data query and analysis, breaking down data silos while fully tapping into the value of data.



Typical use cases

Cross-business federated data query

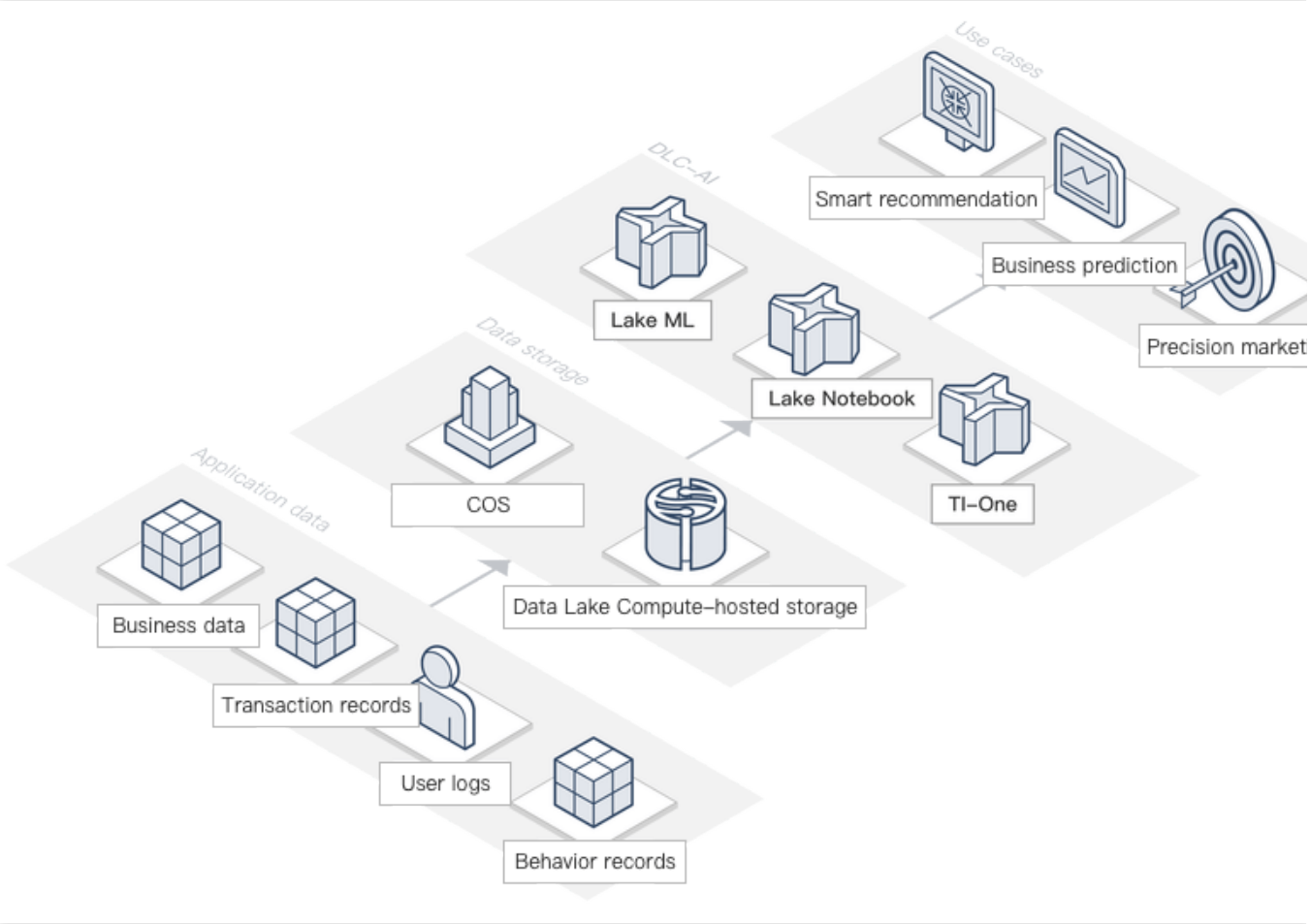
Enterprise departments and business lines often use different data architectures for their specific business systems. This means business data is dispersed in different storage systems, for example, transaction data in relational databases, active data in Redis, and historical records in object storage. With Data Lake Compute, you can align and analyze heterogeneous data from multiple sources to utilize your cross-business data more quickly.

Service benefits

Out-of-the-box: There is no need to set up data transfer pipelines, so unnecessary Ops and costs are saved.
Secure and efficient: The permission management system is unified and refined to the column level, making queries super fast.
Easy-to-use: Cross-business analysis can be easily implemented without programming language adaptation.

Rich and Diversified Data Lake Portfolio

A data lake is the foundation for big data in AI scenarios, including machine learning and deep learning. Connected to a wealth of AI capabilities and platforms, Data Lake Compute readily supports a multitude of machine learning capabilities and delivers comprehensive solutions to various smart data lake analysis applications. It opens up multiple industry databases free of charge so that you can perform data analysis without data acquisition and cleansing. It also provides strong BI capabilities to help you gain data insights through predictive analysis.



Typical use cases

Business growth empowered by data

Data Lake Compute offers native machine learning capabilities based on a sophisticated machine learning platform to provide a complete smart analysis solution. It helps solve your real-world business issues, such as smart recommendation and recall policies, and empower your business growth. Machine learning scenarios are often susceptible to problems like large data volumes, slow model training, and poor algorithm results. With this solution, you can enjoy out-of-the-box machine learning algorithm models to create data-driven models and predict business outcomes. You can also use its BI capabilities for efficient business analysis and improved operational efficiency.

Service benefits

Ease of use: The service is seamlessly connected to Tencent Cloud's machine learning platform, giving you access to a wealth of models and APIs.

Data standardization: Unified data management and governance provide more standardized data for data operations.