# Data Lake Compute

# Product Introduction

# Product Documentation

# Contents

# Product Introduction

# Overview

Last updated：2022-08-16 09:41:58

Data Lake Compute provides agile and efficient data lake analysis and computing services. With its serverless design, it frees you from the concerns over underlying architecture and compute resource maintenance. Its standard SQL syntax enables federated analysis and computing of data stored with COS and other cloud data services. In addition, it eliminates your need to set up traditional data layers, greatly accelerating the preparations for massive data analysis while enhancing your data agility.

Data Lake Compute is a SaaS-like, cloud-native data lake solution that is agile and smart with the following features:

- It provides stable, secure, and high-performance compute resources based on Tencent Cloud's big data kernel enhancement capabilities.
- It implements data processing and multi-source federated data computing by using the standard SQL syntax to greatly reduce your costs to set up and use data analysis services while improving your data agility.
- It offers more economical cost models and accurate cost control capabilities through storage-computing separation and elastic scaling within minutes.
- It delivers SaaS-like out-of-the-box services that eliminate your concerns over the underlying architecture and compute resource maintenance, thereby making it easier for you to get started with data computing.

# Features

### 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.

### Enterprise-grade agile data platform

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.

## Multi-source 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.

## 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.

# Product Architecture
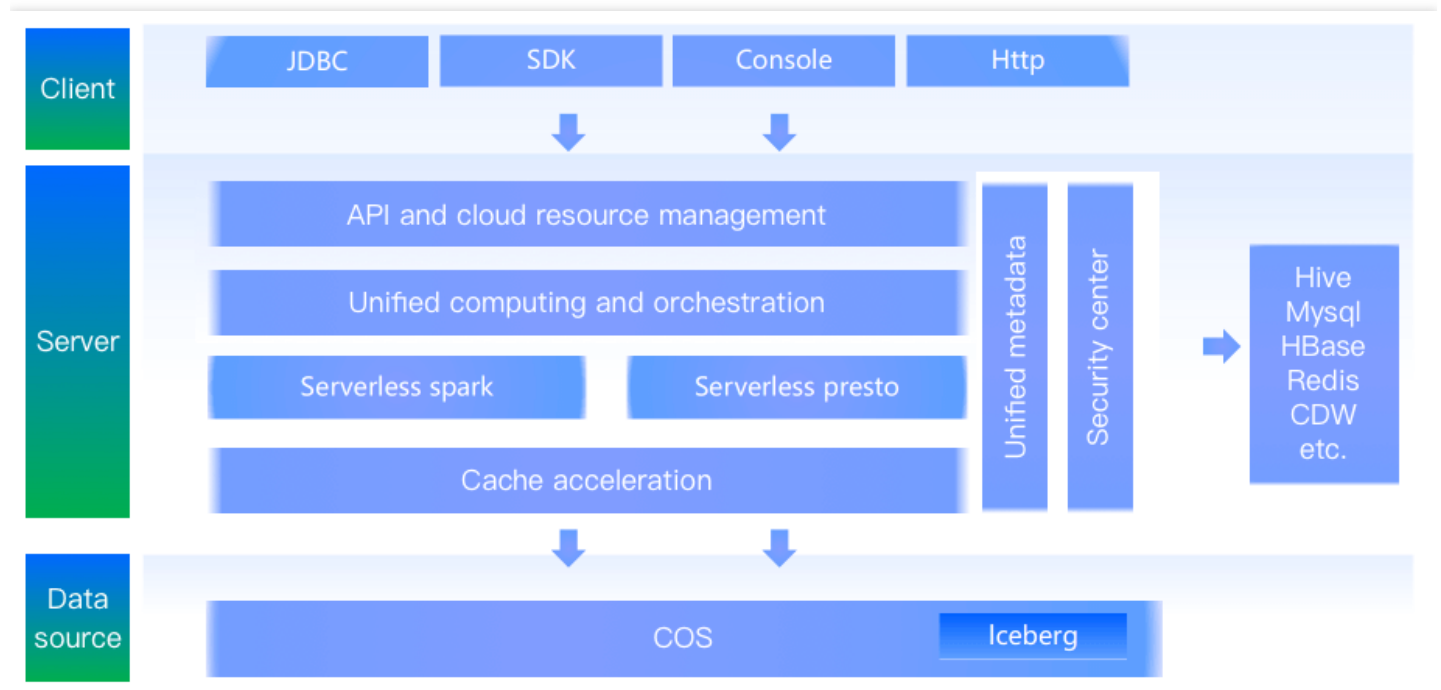


- Provides JDBC and SDK development modes and a visually interactive console for various use cases.
- Supports multiple data sources and unifies metadata management and column-level data security management for data consistency and security.

- Supports Spark and Presto engines with unified SQL syntax.
- Caches and accelerates queries for more agile and efficient data query and analysis.
- Integrates COS and Iceberg data storage to ensure the ACID of data transactions.

# Strengths

Last updated：2022-08-16 09:41:58

## Agility and ease of use

- Data Lake Compute delivers an out-of-the-box SaaS experience that frees you from specification selection, installation, and fine-tuning.
- It supports the standard SQL syntax for you to seamlessly switch between multiple engines for easier data analysis.
- It supports Tencent's many types of cloud-based data infrastructures. This implements multi-source federated analysis of heterogeneous data through a unified data view with no need to load additional data.

## Cost reduction

- 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 and elastic scaling on top of cloud object storage.
- It is billed in a pay-as-you-go manner to reduce the costs of data query and analysis.
- It adopts partitions or columns for compressed data storage to further lower costs.

## 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：2022-08-16 09:41:58

## 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.
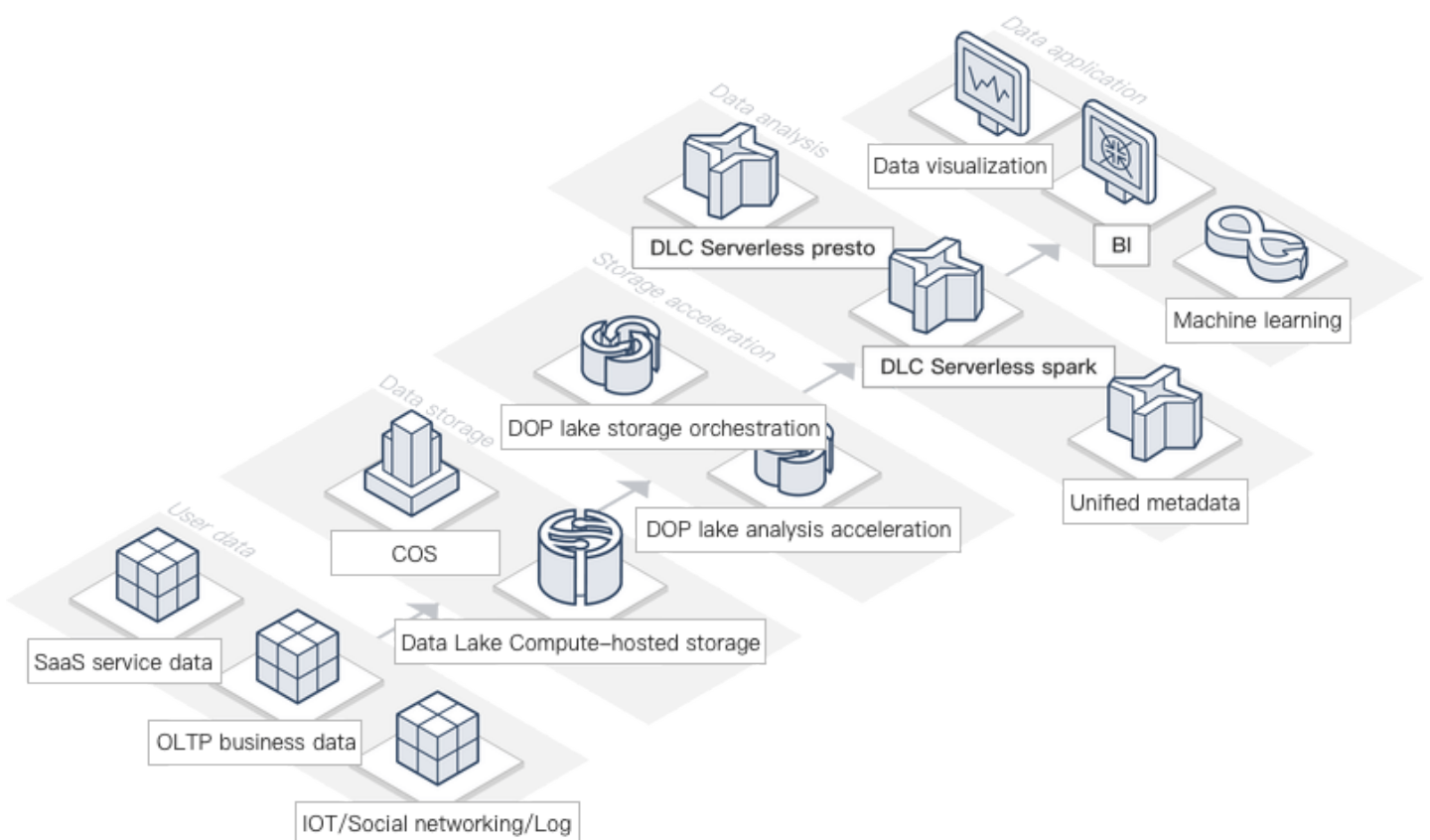


**Typical use cases**

**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.
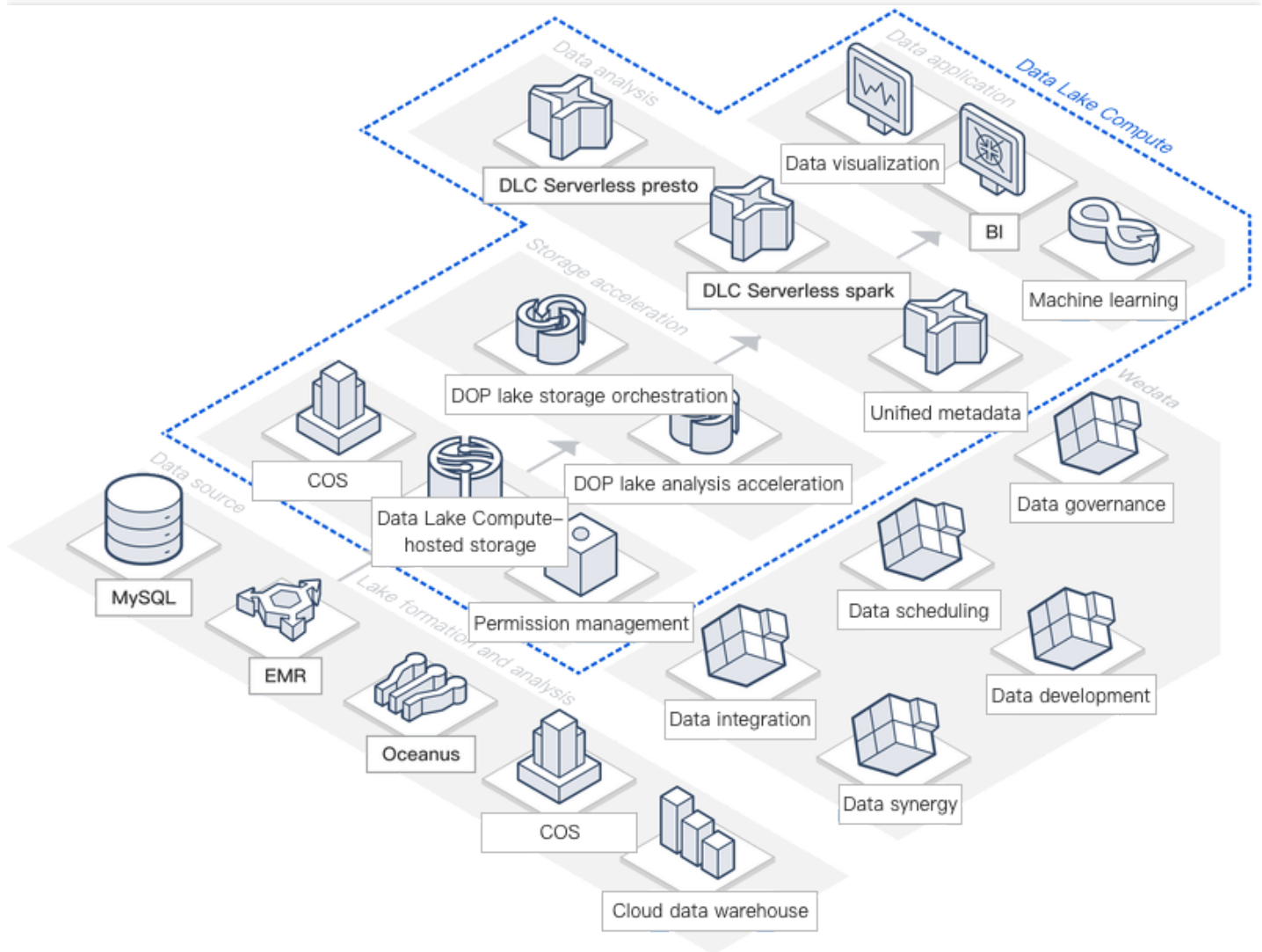


## Typical use cases

### 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.
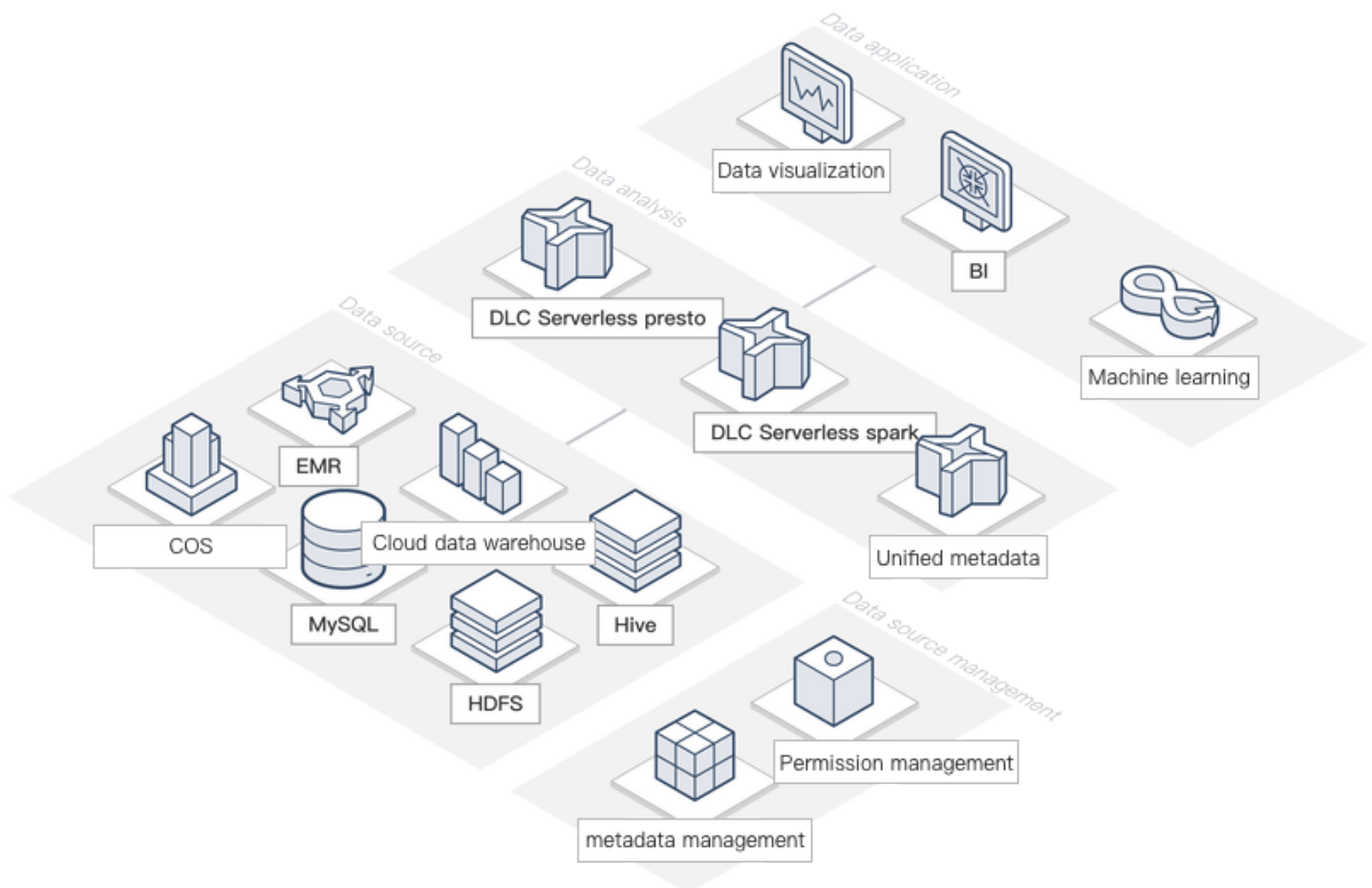
## Typical use cases

### 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.
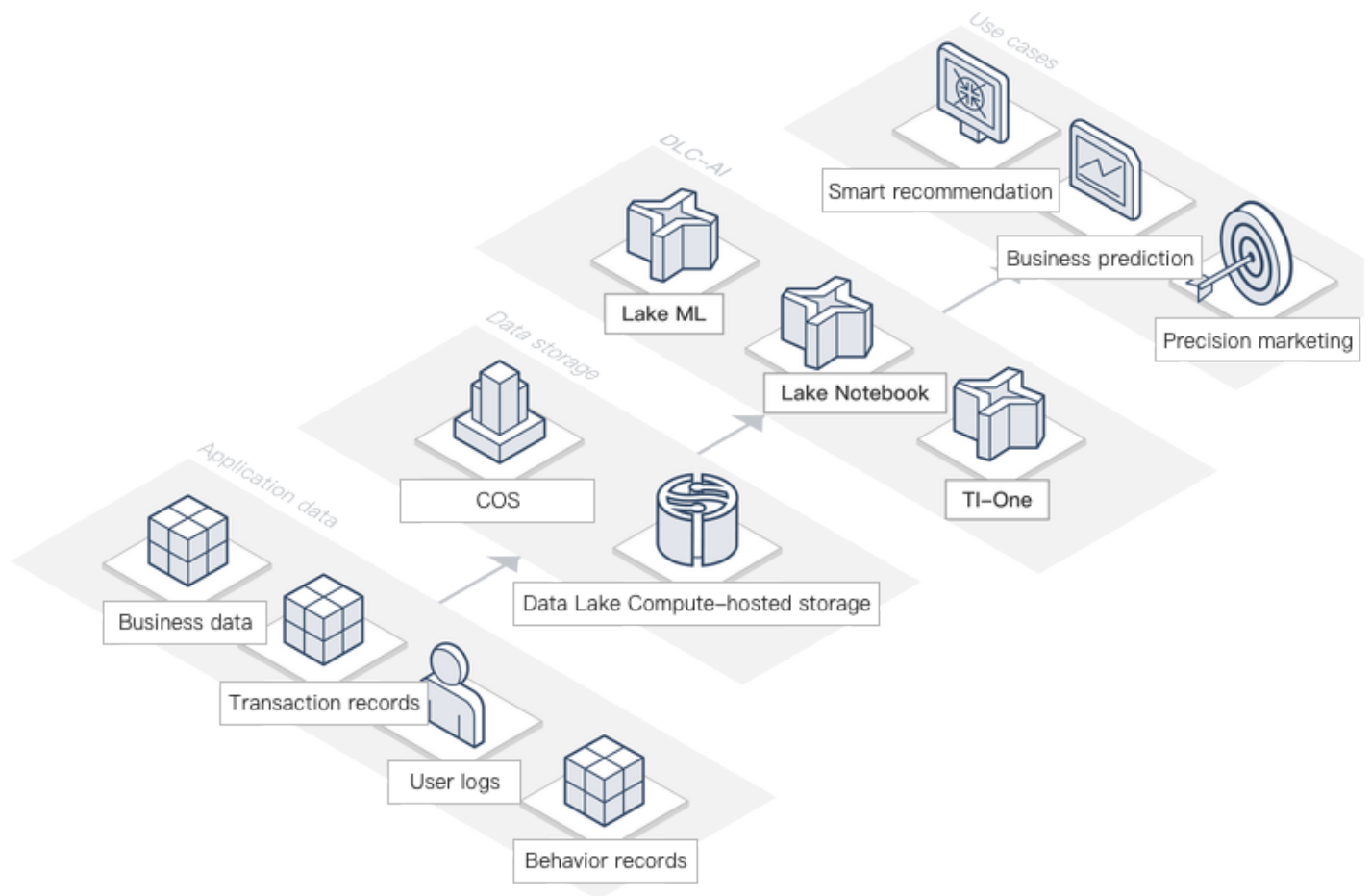


## Typical use cases

**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.