数据加速器 GooseFS

产品简介

腾讯云

版权所有：腾讯云计算（北京）有限责任公司

【版权声明】

©2013-2025 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内容。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】

腾讯云

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：2024-10-31 17:11:42

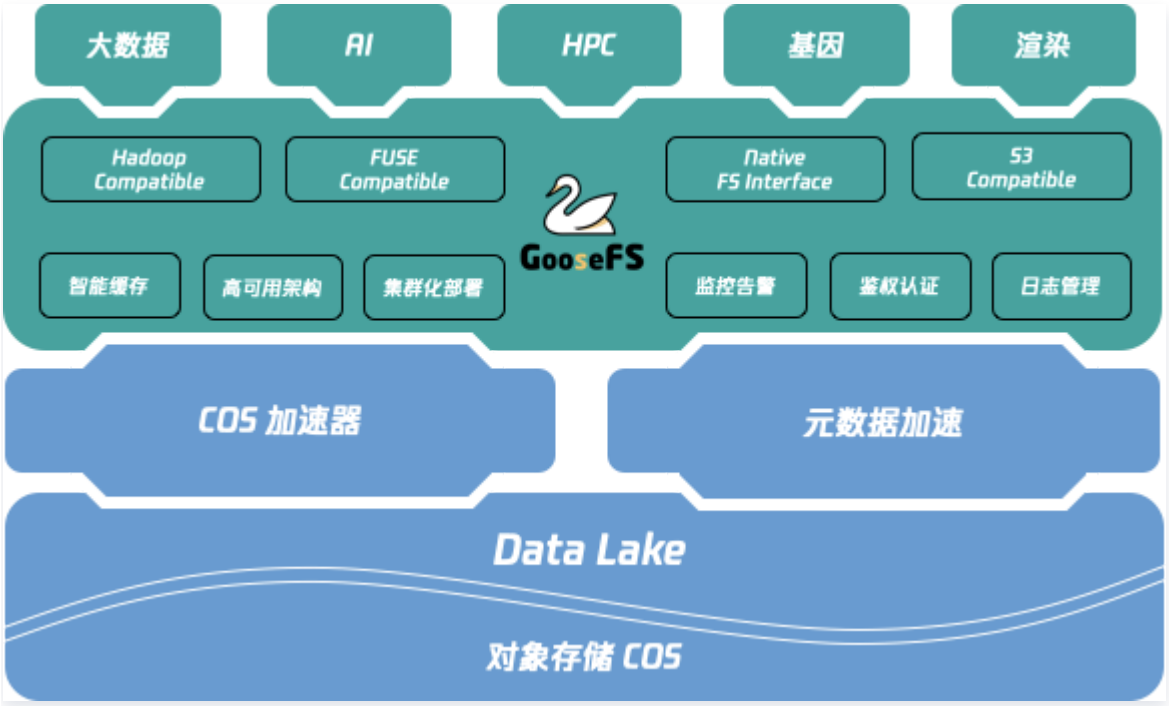
数据加速器（Data Accelerator Goose FileSystem，GooseFS），是由腾讯云推出的高可靠、高可用、弹性的数据加速服务。依靠对象存储（Cloud Object Storage，COS）作为数据湖存储底座的成本优势，为数据湖生态中的计算应用提供统一的数据湖入口，加速海量数据分析、机器学习、人工智能等业务访问存储的性能；采用了分布式集群架构，具备弹性、高可靠、高可用等特性，为上层计算应用提供统一的命名空间和访问协议，方便用户在不同的存储系统管理和流转数据。

产品功能

数据加速器有两个子产品：GooseFS 和 GooseFSx。

数据加速器 GooseFS

GooseFS 旨在提供一站式的缓存解决方案，在利用数据本地性和高速缓存，统一存储访问语义等方面具有天然的优势；GooseFS 在腾讯云数据湖生态中扮演着“上承计算，下启存储”的核心角色，如下图所示。



GooseFS 提供了以下功能：

1. 缓存加速和数据本地化（Locality）：GooseFS 可以与计算节点混合部署提高数据本地性，利用高速缓存功能解决存储性能问题，提高写入对象存储 COS 的带宽。
2. 融合存储语义：GooseFS 可以支持 HDFS 、POSIX 等协议，支持对接腾讯云 COS、云 HDFS 等多种云存储，适用于多种生态和应用场景。
3. 统一的腾讯云相关生态服务：包括日志、鉴权、监控，实现了与 COS 操作统一。

4. 提供 Namespace 管理能力：针对存储在腾讯云 COS、云 HDFS 等多种业务数据，提供不同的读写缓存策略和生命周期管理能力。
5. 感知 Table 元数据功能：对于大数据场景下数据 Table，提供 GooseFS Catalog 用于感知元数据 Table，提供 Table 级别的 Cache 预热。

版权所有：腾讯云计算（北京）有限责任公司

产品优势

最近更新时间：2023-09-20 20:30:51

数据加速器（Data Accelerator Goose FileSystem，GooseFS）在数据湖场景中具有如下几点明显的优势：

数据 I/O 性能

GooseFS 部署提供近计算端的分布式共享缓存，上层计算应用可以透明地、高效地从远端存储将需要频繁访问的热数据缓存到近计算端，加速数据 I/O 性能。GooseFS 提供了元数据缓存功能，可以加速大数据场景下查询文件数据以及列出文件列表等元数据操作的性能。配合大数据存储桶使用，还可进一步加速重命名文件的操作性能。此外，业务可以按需选择 MEM、SSD、NVME 以及 HDD 盘等不同的存储介质，平衡业务成本和数据访问性能。

存储一体化

GooseFS 提供了统一的命名空间，不仅支持了对象存储 COS 存储语义，也支持 HDFS、K8S CSI 以及 FUSE 等语义，为上层业务提供了一体化的融合存储方案，简化业务侧运维配置。存储一体化能够打通不同数据底座的壁垒，方便上层应用管理和流转数据，提升数据利用的效率。

生态亲和性

GooseFS 全兼容腾讯云大数据平台框架，也支持客户侧自定义的本地部署，具备优秀的生态亲和性。业务侧不仅可以在腾讯云弹性 MapReduce 产品中使用 GooseFS 加速大数据业务，也可以便捷地将 GooseFS 本地化部署在公有云 CVM 或者自建 IDC 内。此外，GooseFS 支持透明加速能力，对于已经使用腾讯云 COSN 和 CHDFS 的用户，只需做简单的配置修改，即可实现不修改任何业务代码和访问路径的前提下，自动使用 GooseFS 加速 COSN 和 CHDFS 的业务访问。

版权所有：腾讯云计算（北京）有限责任公司

应用场景

最近更新时间：2023-09-20 20:30:51

开源生态数据湖

客户基于开源 Hadoop 生态构建大数据处理与分析，会面临计算资源与存储资源扩容速度不匹配、存储系统需对接多数据源的问题。

推荐产品

推荐数据加速器 GooseFS。

主要能力

- 计算存储分离
  通过计算与存储分离，实现计算资源弹性伸缩，满足客户对计算资源的灵活调度。
- 多数据源支持
  可对接多种数据源，允许存储任意规模的结构化、半结构化、非结构化数据。
- 高性能业务架构
  通过数据加速器（Data Accelerator Goose FileSystem，GooseFS）、元数据加速器、AZ 加速器等多级加速服务，提升计算业务访问性能。

交互式查询数据湖

客户在对象存储（Cloud Object Storage，COS）中存储了多种数据源数据，包括实时计算数据，需要对其中的数据进行 OLAP 分析并进行数据可视化展示。

主要能力

- 多数据源支持
  可对接多种数据源，允许存储任意规模的结构化、半结构化、非结构化数据。
- 性能加速
  通过数据加速器、元数据加速器、AZ 加速器等多级加速服务，实现超越本地 HDFS 的性能。

机器学习数据湖

在经典机器学习场景中，训练数据量大，同时要求很大的内网带宽。

主要能力

- 超大带宽
  可以提供超大的内网带宽，满足机器学习场景大带宽需求。
- 多数据源支持
  可对接多种数据源，允许存储任意规模的结构化、半结构化、非结构化数据。

版权所有：腾讯云计算（北京）有限责任公司

- 性能加速
  通过数据加速器、元数据加速器、AZ 加速器等多级加速服务，实现超越本地 HDFS 的性能。

云原生数据湖

通过容器服务，结合 Flink、TensorFlow 等开源应用，搭建云原生数据 ETL 集群和分析集群，实现计算资源的弹性化；通过数据加速器、元数据加速器、AZ 加速器等多级加速服务，提升计算业务访问性能；通过对象存储服务作为数据湖存储底座，实现海量异构数据的低成本存储。

主要能力

- 计算存储分离
  通过计算与存储分离，实现计算资源弹性伸缩，满足客户对计算资源的灵活调度。
- 高性能业务架构
  通过数据加速器、元数据加速器、AZ 加速器等多级加速服务，提升计算业务访问性能。
- 丰富生态支持
  可存储 Parquet、ORC 多种格式数据源，支持 Spark、Presto、Flink 等多种大数据插件。

版权所有：腾讯云计算（北京）有限责任公司