

HybridDB for PostgreSQL

产品简介

产品简介

产品概述

云数据库 HybridDB for PostgreSQL (ApsaraDB HybridDB for PostgreSQL) 是一种在线分布式云数据库，由多个计算组组成，可提供大规模并行处理（MPP）数据仓库服务。HybridDB for PostgreSQL 基于 Greenplum Database 开源数据库项目开发，由阿里云深度扩展后，具备如下特性：

兼容 Greenplum，用户可以直接使用所有支持 Greenplum 的工具。

独有的MetaScan和SortKey技术，大幅提升了性能。

支持 OSS 存储、JSON 数据类型、HyperLogLog 预估分析等功能特性。

通过符合 SQL 2008 标准查询语法及 OLAP 分析聚合函数，提供灵活的混合分析能力。

支持行存储和列存储混合模式，分析性能优越。

支持数据压缩技术，存储成本低廉。

提供在线扩容、性能监测等服务，用户无需再进行复杂的大规模 MPP 集群的运维管理工作，使 DBA、开发人员及数据分析师只需专注于如何通过 SQL 提高企业的生产力、创造核心价值。

功能与限制

本文介绍介绍了云数据库 HybridDB for PostgreSQL 的基本功能和功能限制。

基本功能

具有 Geenplum Database 的核心功能，详情参见文档说明。

支持独有的MetaScan和SortKey技术。

支持 ORCA 优化器。

支持通过过程语言 PL/pgSQL 和 PL/Java 进行分布式存储过程编写。

支持 PostGIS、MADlib、fuzzystrmatch、orafunc、pgcrypto、intarray 等插件（需使用 CREATE EXTENSION 命令创建）。

支持使用 OSS_EXT 插件，从阿里云对象存储 OSS（Object Storage Service）读取数据或写入数据到 OSS，并支持gzip压缩，以大幅节省外部表存储成本。

支持 JSON 数据类型和 HyperLogLog 类型（需使用 CREATE EXTENSION 命令创建）。

功能限制

核心功能上的限制参见文档说明。

权限限制：HybridDB for PostgreSQL 的初始用户（称为“根用户”）有创建数据库（CREATEDB）、创建用户（CREATEROLE）的权限，但没有超级用户（SUPERUSER）权限，因此无法执行要求超级用户权限的操作。例如，执行 pg_ls_dir 等文件函数。但根用户有权限查看和修改所有其他非超级用户的数据，终止（Kill）其他非超级用户的连接等。

不支持 PL/R 和 PL/Java 插件。

支持 PL/Python 插件创建，但不支持使用 PL/Python 语言创建函数。

不支持 gpfdist 工具。

不支持 MapReduce 接口、gphdfs 存储接口以及本地外部表。

暂不支持自动备份和恢复功能。HybridDB 会保存两份数据，用户也可以使用 pg_dump 工具自行备份。

实例规格

目前，HybridDB for PostgreSQL 支持如下两种实例规格：

高性能：规格名称以 gpdb.group.segsdx 开始，特点是可以提供更好的 I/O 能力，带来更高的性能。

高容量：规格名称以 gpdb.group.seghdx 开始，特点是可以提供更大、更实惠的空间，满足更高的存储需求。

因此，在选择实例规格时，建议您根据需要的存储空间和计算能力进行选择。

同时，HybridDB for PostgreSQL 支持基于 OSS 的外部表扩展，并可通过 gzip 实现外部存储上的数据压缩，不需要参与实时计算的数据可以存放到外部存储以进一步节省存储成本。

规格详情

高性能实例的规格信息如下表所示：

【高性能】规格	CPU	内存	存储空间
gpdb.group.segsdx1	1 Core	8 GB	80 GB SSD
gpdb.group.segsdx2	2 Cores	16 GB	160 GB SSD
gpdb.group.segsdx16	16 Cores	128 GB	1.28 TB SSD

高容量型实例的规格信息如下表所示：

【高容量】规格	CPU	内存	存储空间
gpdb.group.seghdx4	4 Cores	32 GB	2 TB HDD
gpdb.group.seghdx36	36 Cores	288 GB	18 TB HDD

价格信息详情，参见 <https://www.aliyun.com/price/product#/gpdb/detail>。

产品优势

HybridDB for PostgreSQL 作为一个 HTAP 关系型数据库，具有如下优点。

简单易用

兼容NoSQL，用户可直接用SQL操作JSON和XML类型的数据。

高度兼容MySQL 5.6的语法和函数，并支持常用的Oracle函数。

通过 PostGIS 插件，支持空间数据类型和相关的空间函数，实现精确的地理定位。

支持多种插件，用户可以轻松地通过插件满足各种应用场景的需求。

支持OSS云存储扩展，可实现存储空间无限扩展。

与阿里云的其他主流产品无缝集成，用户可轻松实现复杂的业务应用。

性能卓越

支持行列混合存储。在OLAP分析时，列存储的性能和行存储的性能相比，最多可提高100倍的性能。

支持高性能OSS并行数据导入，避免单通道导入的性能瓶颈。

采用基于成本的SQL查询优化器，多表关联查询性能比MySQL高数倍。

用户可按需进行计算单元、CPU、内存、存储空间的等比扩展，从而提高OLAP性能。

安全稳定

支持分布式ACID数据一致性，实现跨节点事务一致，所有数据通过双节点同步冗余，SLA保障99.9%可用性。

采用主备架构，实现用户无感的故障转移和容灾备份。

分布式部署，计算单元、服务器、机柜三重防护，提高重要数据基础设施保障。

支持SQL审计，身份隔离深入到字段级别；支持IP白名单防DDoS攻击。

名词解释

下表列出了 HybridDB for PostgreSQL 所涉及到的基本概念：

名词	解释
计算组	HybridDB for PostgreSQL 中的运行单元，一个 HybridDB for PostgreSQL 实例由多个计算组组成，计算组数量的增加，可以线性提升性能。
计算组规格	用户可购买的计算资源单位，包括 CPU、I/O、内存和磁盘。不同计算组规格的性能不同，每个计算组内的资源将分配在同一台物理主机中。
计算组个数	所购买的计算组数量，最小单位为 2 个，不同计算组规格对应不同的可购买数量。
MPP	Massively Parallel Processing，一种分布式 Shared Nothing 计算架构，通过多个无共享的节点（HybridDB for PostgreSQL 中称为计算组）同时并行计算以提升性能。

发布历史

2016 年 12 月 01 日，正式商业化

产品更新

可用区及控制台更新

- 开通高性能 SSD 规格：华北 2（北京）、华东 1（杭州）、华东 2（上海）。
- 开通大容量 HDD 规格：华东 1（杭州）。
- 控制台支持“资源监控”功能。
- 控制台支持“SQL审计”功能，此功能在 2017 年 06 月 01 日前可免费使用。

内核功能更新

- 支持 JSON（包括属性操作和基于函数的索引），详情参见 JSON 数据类型操作。
- 支持 HyperLogLog，详情参见 HyperLogLog 使用。
- OSS_EXT 支持写入操作，详情参见 OSS 高速并行导入。

Bug 修复

修复社区版本 delete 操作的 bug。

修复 HyperLogLog 中一个可能导致系统重启的 bug。

2016 年 07 月 11 日，公测发布

产品性能

基于 Geenplum 开源项目 4.3 版本研发。

支持 ORCA 优化器。

支持插件化管理，支持如下插件：

- OSS_EXT：读取存放在 OSS（Open Storage Service）上的文件。
- PL/pgSQL：支持使用 PL/pgSQL 语言创建函数，默认数据库中自带此插件。
- orafunc：兼容 Oracle 的部分函数。
- PostGIS：支持地理信息数据。
- MADlib：机器学习方面的函数库。
- fuzzystmatch：字符串模糊匹配。

可售卖区域：华北 2（北京）。