ApsaraDB HybridDB for PostgreSQL

Product Introduction

Product Introduction

What is ApsaraDB HybridDB for PostgreSQL

ApsaraDB HybridDB for PostgreSQL is a distributed cloud database that is composed of multiple groups to provide online MPP (Massively Parallel Processing) data warehousing service. HybridDB for PostgreSQL is developed based on the Greenplum Open Source Database program and is enhanced with some in-depth extensions by Alibaba Cloud. This service includes the following features:

Compatible with Greenplum, allowing you to directly use all the tools that support Greenplum.

Supports features including OSS storage, JSON data type, and HyperLogLog approximating analysis.

Complies with SQL 2008 standard query syntax and OLAP aggregate functions, providing a flexible hybrid analysis capability.

Supports hybrid storage mode with data stored in both rows and columns, enhancing analytics performance.

Uses exclusive MetaScan and SortKey technologies to optimize your experience.

Supports Data compression techniques to reduce storage costs.

Provides online expansion and performance monitoring services to free you from complicated MPP cluster O&M and enable DBAs, developers, and data analysts to focus on improving enterprise productivity through SQL and creating core value.

Features and limits

This document lists the basic features and limits of ApsaraDB HybridDB for PostgreSQL.

Features

Covers the core functions of Greenplum Database. For details, see Summary of Greenplum Features.

Supports MetaScan and SortKey.

Supports the ORCA optimizer.

Supports distributed stored procedures in PL/PGSQL and PL/JAVA.

Supports multiple extensions, such as PostGIS, MADlib, fuzzystrmatch, orafunc, pgcrypto, and intarray. (You can use the CREATE EXTENSION command to create them.)

Supports using OSS_EXT extension to read data from or write data to Alibaba Cloud OSS (Object Storage Service), and supports the gzip compression to reduce the external table storage cost.

Supports the JSON data type and the HyperLogLog type. (You can use the CREATE EXTENSION command to create them.)

Limits

For limits of the core functions, see Summary of Greenplum Features.

Permission limits: The initial user of HybridDB for PostgreSQL (the root user) has the permission for creating databases (CREATEDB) and users (CREATEROLE), but does not have the super user permission (SUPERUSER). That is,

The root user cannot perform operations requiring the super user permission. For example, the root user cannot run file functions such as pg_ls_dir.

The root user has the permission to view and modify the data of all other non-super users, and terminate (Kill) the connections of other non-super users.

Not support the PL/R and PL/Java extensions.

Supports creating the PL/Python extensions, but does not support creating functions by using PL/Python.

Not support the gpfdist tool.

Not support MapReduce interfaces, gphdfs storage interfaces and local external tables.

Not support automatic backup and recovery. HybridDB for PostgreSQL keeps two copies of data, and you can back up data by using the pg_dump tool.

Instance types

HybridDB for PostgreSQL instances are categorized into the following type families:

High-performance: The name of this type family starts with "gpdb.group.**segsdx** ". This type family features a better I/O capability that secures higher performance.

High-capacity: The name of this type family starts with "gpdb.group.**seghdx** ". This type family features a larger and more affordable space to meet higher storage demands.

When selecting a type family, you only need to take the storage space and computing capability into consideration.

Besides, HybridDB for PostgreSQL supports OSS-based external table extensions and data compression in external storage through gzip. You can store data that is not required for real-time computation in an external storage to further reduce storage costs.

Instance types

The **high-performance** instances offer the following types:

Instance type	CPU	Memory	Disk
gpdb.group.segsdx1	1 Core	8 GB	80 GB SSD
gpdb.group.segsdx2	2 Cores	16 GB	160 GB SSD
gpdb.group.segsdx1	16 Cores	128 GB	1.28 TB SSD

6			

The **high-capacity** instances offer the following types:

Instance type	CPU	Memory	Disk
gpdb.group.seghdx4	4 Cores	32 GB	2 TB HDD
gpdb.group.seghdx36	36 Cores	288 GB	18 TB HDD

Advantages

Easy to Use

Compatible with NoSQL. You can directly manipulate JSON and XML data using SQL.

Accurate geolocation is achieved by the PostGIS extension that supports spatial data types and related spatial functions.

Support multiple extensions, which satisfy the needs of various application scenarios.

Supports unlimited storage expansion with OSS cloud storage.

Seamlessly integrates with other Alibaba Cloud mainstream products to easily implement complex business applications.

Outstanding Performance

Supports hybrid row-column storage. In OLAP analysis, the performance of column storage is up to 100 times that of row storage.

Supports high-performance parallel import of OSS data to avoid single-channel import performance bottlenecks.

With a cost-based SQL query optimizer, multi-table join query performance is several times

higher than MySQL.

You can expand the calculation unit, CPU, memory, and storage space in the same ratio as required to improve OLAP performance.

Safety and Stability

Supports distributed ACID data consistency, cross-node transaction consistency, dual-node data redundancy, and SLA-guaranteed 99.9% availability.

Implements unattended failover and disaster recovery with the active-standby architecture.

Provides secure infrastructure with distributed deployment and protection of computing units, servers, and cabinets.

Supports SQL auditing, field-level access isolation, and IP address whitelisting for protection against DDoS attacks.

Concepts

The following table lists the basic concepts involved in HybridDB for PostgreSQL.

Term	Description

Group	The operation unit in HybridDB for PostgreSQL.
	    - A HybridDB for PostgreSQL instance is composed of multiple groups.
	    - Increasing the number of groups can improve the linear performance.

Group type	The package of computing resources available.
	    - Each group type includes CPU, I/O, memory, and disk resources. Different group types have different performance.
	    - Resources in one group are allocated

	to the same physical host.

Number of groups	The number of purchased groups.
	   - The minimum unit is two.
	   - Each group type has different group
	     number limit.

MPP	Massively Parallel Processing, a distributed Shared Nothing computing architecture. MPP improves the performance by parallel computing of multiple shared-nothing nodes (known as groups in HybridDB for PostgreSQL).

Release notes

End of beta period: December 1, 2016

Product updates

Updates of zones and HybridDB for PostgreSQL console

High-performance SSD specification is available in China North 2 (Beijing), China East 1 (Hangzhou), and China East 2 (Shanghai).

High-capacity HDD specification is available in China East 1 (Hangzhou).

The HybridDB for PostgreSQL console supports "resource monitoring".

The HybridDB for PostgreSQL console supports "SQL auditing". This feature is offered for free before June 1, 2017.

Kernel features updates

JSON (including attribute operations and function-based indexing) is supported. For details, see Operations of JSON data.

HyperLogLog is supported. For details, see Use HyperLogLog.

OSS_EXT supports writing operations. For details, see Import data to OSS in parallel.

Bug fixes

The delete operation bug in the community edition is fixed.

A HyperLogLog bug that may cause system restart is fixed.

Public beta launched: July 11, 2016

Product performance

Developed based on open-source GreenPlum Database 4.3.

Supports the ORCA optimizer.

Supports extension-based management and the following extensions:

OSS_EXT: reads files stored on OSS (Open Storage Service).

PL/pgSQL: supports creating functions using PL/pgSQL language. This extension is nested in the default database.

orafunc: compatible with some Oracle functions.

PostGIS: supports geographic information data.

MADlib: a machine learning function library.

fuzzystrmatch: fuzzy matching of strings.

Available region: China North 2 (Beijing).