ApsaraDB HybridDB for PostgreSQL

Product Introduction

Product Introduction

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. It has 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.

Supports data compression technologies to cut down 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.

This section lists the basic features and functional limitations of ApsaraDB HybridDB for PostgreSQL.

Basic features

Covers the core functions of Greenplum Database. For details, see Pivotal Greenplum Database Documentation.

Supports ORCA optimizer.

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

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

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

Supports JSON data type and HyperLogLog type. (Use CREATE EXTENSION command to create them.)

Functional limitations

For limitations of the core functions, see Pivotal Greenplum Database Documentation.

Permission limitations: The initial user of HybridDB for PostgreSQL (known as the "root user") has permissions for creating databases (CREATEDB) and users (CREATEROLE), but has no superuser permissions (SUPERUSER). That is the root user is unable to perform operations requiring the superuser permissions. For example, to perform file functions of pg_ls_dir among others. But, the root user has the permissions to view and modify the data of all other non-superusers, and terminate (Kill) the connection of other non-superusers.

PL/R and PL/Java extensions are not supported.

Supports creating PL/Python extension, but does not support creating functions using PL/Python.

The gpfdist tool is not supported.

MapReduce interfaces, gphdfs storage interface and local external tables are not supported.

Automatic backup and recovery functions are not supported for the time being. HybridDB for PostgreSQL keeps two copies of data, and you can also back up data using the pg_dump tool.

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

Term	Description

Group	The operation unit in HybridDB for PostgreSQL. - One HybridDB for PostgreSQL instance is composed of multiple groups. - Increasing the number of groups can result in improving the linear performance.
Group type	The available pack of computing resources. - Each group type includes CPU, IO, 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 2. - Each group type has different group number limit.
MPP	Massively Parallel Processing, a distributed Shared Nothing computing architecture. MPP improves the performance through parallel computing by multiple share-nothing nodes (known as groups in HybridDB for PostgreSQL).

Officially commercialized on 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 fixing

The delete operation bug in the community edition is fixed.

A HyperLogLog bug that may cause system reboot is fixed.

Public beta launched on July 11, 2016

Product performance

Developed based on open-source GreenPlum Database 4.3.

Supports 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).