ApsaraDB HybridDB for PostgreSQL

Product Introduction

Product Introduction

Product overview

ApsaraDB for HybridDB, an online MPP (Massively Parallel Processing) service, is a distributed cloud database composed of multiple groups and provides MPP data warehousing service. ApsaraDB for HybridDB is developed based on the Greenplum Open Source Database program and incorporates some in-depth extensions by Alibaba Cloud, which enhances it with the following features:

Compatible with Greenplum and you can 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, ApsaraDB HybridDB offers a flexible hybrid analysis capability.

Supports hybrid storage mode with data stored in both rows and columns, enhancing the 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 how to improve enterprise productivity through SQL and create core value.

Features and limitations

This article introduces the basic features and functional limitations of ApsaraDB for HybridDB.

Basic features

Include the core functionality of Greenplum Database. For details, see Pivotal Greenplum Database Documentation.

Support ORCA optimizer.

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

Support PostGIS, MADlib, fuzzystrmatch, orafunc, pgcrypto and intarray extensions (they should be created using the CREATE EXTENSION command).

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

Support JSON data type and HyperLogLog type (they should be created using the CREATE EXTENSION command).

Functional limitations

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

Permission limitations: The initial user of HybridDB (known as the “root user”) has permissions for creating databases (CREATEDB) and users (CREATEROLE), but has no superuser permission (SUPERUSER), hence is unable to execute operations requiring the superuser permission, for example, to execute file functions of pg_ls_dir among others. The root user has the permission to view and modify the data of all other non-superusers, and kill the connection of other non-superusers.

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

HybridDB supports the creation of 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 saves two copies of data, and you can also back up data using the pg_dump tool.

Concept explanations

The table below lists the basic concepts involved in this product:

Term	Description
Group	The operation unit in HybridDB. One HybridDB instance is composed of multiple groups. Increasing the number of groups can result in a linear performance increase.
Group type	The unit of computing resources purchasable by the user. It includes CPU, IO, memory and disk resources. Different group types have different performance. Resources of the same group type are allocated in the same physical host.
Number of groups	The number of purchased groups, minimum two. Different group types correspond to different number purchasable.
MPP	Massively Parallel Processing, a distributed Shared Nothing computing architecture. It improves the performance through parallel computing by multiple share-nothing nodes (known as groups in HybridDB).

ApsaraDB for HybridDB is developed based on the Greenplum Open Source Database program and incorporates some in-depth extensions by Alibaba Cloud. It is an online MPP (Massively Parallel Processing) distributed cloud database composed of multiple groups and provides MPP data warehousing service.

Release notes

Officially commercialized on December 1, 2016

Product updates

Updates of zones and HybridDB 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 console supports “resource monitoring”.

The HybridDB 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 https://help.aliyun.com/document_detail/48147.html.

HyperLogLog is supported. For details, see https://help.aliyun.com/knowledge_detail/44456.html.

OSS_EXT supports writing operations. For details, see https://help.aliyun.com/document_detail/35457.html.

Bug fixing

The delete operation bug in the community edition is fixed.

A HyperLogLog bug that may cause system reboot is fixed.

Others

HybridDB provides the “separated storage and computing” feature that is integrated and developed based on Greenplum Database and Apache HAWQ open-source programs. The feature is currently open to invited users for testing. For details, see https://help.aliyun.com/document_detail/48643.html.

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: the machine learning function library.

fuzzystrmatch: fuzzy matching of strings.

Available region: China North 2 (Beijing).