Alibaba Cloud HybridDB for PostgreSQL

FAQ

Issue: 20190221

Legal disclaimer

Alibaba Cloud reminds you to carefully read and fully understand the terms and conditions of this legal disclaimer before you read or use this document. If you have read or used this document, it shall be deemed as your total acceptance of this legal disclaimer.

1. **You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.**
2. **No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company, or individual in any form or by any means without the prior written consent of Alibaba Cloud.**
3. **The content of this document may be changed due to product version upgrades , adjustments, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and the updated versions of this document will be occasionally released through Alibaba Cloud-authorized channels. You shall pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.**
4. **This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides the document in the context that Alibaba Cloud products and services are provided on an "as is", "with all faults " and "as available" basis. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity , applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not bear any liability for any errors or financial losses incurred by any organizations, companies, or individuals arising from their download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, bear responsibility for any indirect, consequential, exemplary, incidental, special, or punitive damages, including lost profits arising from the use**

or trust in this document, even if Alibaba Cloud has been notified of the possibility of such a loss.

5. By law, all the content of the Alibaba Cloud website, including but not limited to works, products, images, archives, information, materials, website architecture, website graphic layout, and webpage design, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of the Alibaba Cloud website, product programs, or content shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates).
6. Please contact Alibaba Cloud directly if you discover any errors in this document.

Generic conventions

Table -1: Style conventions

Style	Description	Example
	This warning information indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	 Danger: Resetting will result in the loss of user configuration data.
	This warning information indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	 Warning: Restarting will cause business interruption. About 10 minutes are required to restore business.
	This indicates warning information, supplementary instructions, and other content that the user must understand.	 Notice: Take the necessary precautions to save exported data containing sensitive information.
	This indicates supplemental instructions, best practices, tips, and other content that is good to know for the user.	 Note: You can use Ctrl + A to select all files.
>	Multi-level menu cascade.	Settings > Network > Set network type
Bold	It is used for buttons, menus , page names, and other UI elements.	Click OK.
`Courier font`	It is used for commands.	Run the `cd /d C:/windows` command to enter the Windows system folder.
Italics	It is used for parameters and variables.	`bae log list --instanceid` *Instance_ID*
[] or [a|b]	It indicates that it is a optional value, and only one item can be selected.	`ipconfig` *[-all|-t]*

Style	Description	Example
{} or {a|b}	It indicates that it is a required value, and only one item can be selected.	`swich {stand | slave}`

Contents

1 How to get the IP address of a local client

Symptom

Due to the complexity and diversity of the network, the user may not be able to correctly locate the IP address of the local client which should be add in the instance whitelist. This article describes how to view the IP of a local client.

Here is an example to denote the detailed steps of the operation.

Procedure

1. Add `0.0.0.0/0` to the whitelist of the HybridDB for PostgreSQL instance.
   a. Log on to the *HybridDB for PostgreSQL console*.
   b. Select the region where the target instance is located.
   c. Click the ID of the instance to go to the Basic Information page of the instance.
   d. In the left-side navigation pane, click Security Controls.
   e. In the Whitelist Settings page, click Modify under the default whitelist group to go to the Modify Group page.
   f. Delete the default address 127.0.0.1 from the whitelist and then add `0.0.0.0/0` to the whitelist.

      Note:
      The `0.0.0.0/0` address which allows any IP to access the database involves a high security risk, and this configuration should be deleted as soon as possible.

   g. Click OK to finish the operation.

2. Connect to the HybridDB for PostgreSQL instance using a local client. See *Connect to a HybridDB for PostgreSQL database* to download and install the psql client. Issue the following statement to connect to the database:

```
psql -h yourgpdbaddress.gpdb.rds.aliyuncs.com -p 3432 -d postgres -U
 gpdbaccount
```

Parameter descriptions are as follows:

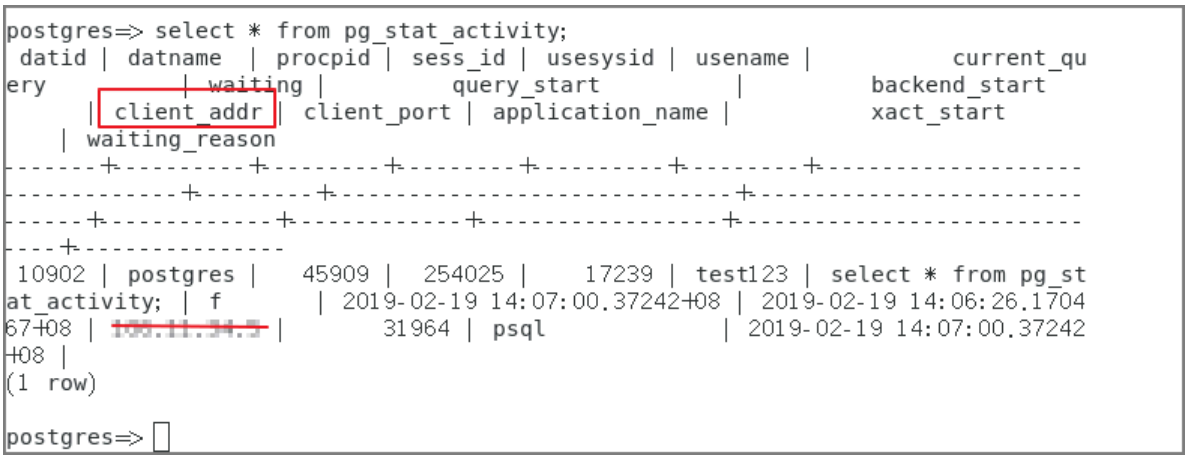
- -h: specifies the host address.
- -p: specifies the port number.
- -d: specifies the database (the default database is postgres).
- -U: specifies the connected user.

You can view more parameters by performing `psql-- help`. And in the psql prompt, you can view more supported psql commands by performing `\?`.

3. Run the following statement in the SQL command line window in the database to query the IP address of the local client.

```
select * from pg_stat_activity;
```

The value in the `CLIENT_ADDR` of the query result is the IP address of the local client.



4. Remove `0.0.0.0/0` from the whitelist, and add the IP address in the previous step to access the database with a client.

2 How to view the data size of a table or a database?

Suppose that the table schema is `<schemaname>` and the table name is `<tablename>`.

- Run the following command to query the total size of a table (unit: MB, including the table index and data):

```
select pg_size_pretty(pg_total_relation_size('<schemaname>.<
tablename>'));
```

- Run the following command to query the data size of a table (unit: MB, excluding the index):

```
select pg_size_pretty(pg_relation_size('<schemaname>.<tablename>'));
```

- Run the following command to query the total size of all the partitions in a partition table (unit: MB, including the table index and data):

```
select schemaname,tablename,round(sum(pg_total_relation_size(
schemaname || '.' || partitiontablename))/1024/1024) "MB" from
 pg_partitions where schemaname='<schemaname>' and tablename='<
tablename>' group by 1,2;
```

- Run the following command to query the total size of all the tables under a schema (unit: MB, including the index and data):

```
select schemaname ,round(sum(pg_total_relation_size(schemaname
||'.'||tablename))/1024/1024) "Size_MB" from pg_tables where
schemaname='<schemaname>' group by 1;
```

- Run the following command to query the size of each database (unit: MB):

```
select datname,pg_size_pretty(pg_database_size(datname)) from
pg_database;
```

3 What ETL tools does HybridDB for PostgreSQL currently support? How to synchronize data from other data sources to HybridDB for PostgreSQL on a regular basis?

The following ETL tools are supported:

- *Alibaba Cloud Data Integration*: an ETL tool provided by Alibaba Cloud. In Data Integration, you can configure HybridDB for PostgreSQL as a PostgreSQL database to synchronize data from other data sources (RDS, MaxCompute, and TableStore) to HybridDB for PostgreSQL.
  - You can directly read data from other data sources, and write data to HybridDB for PostgreSQL.
  - If the data size is big and concurrent imports are required, we recommend that you first import the data to OSS from other data sources by using Data Integratio n, and then import the data by using the OSS external table to HybridDB for PostgreSQL.
- *Pentaho Kettle*: an open-source ETL tool.
  - You can import data through Kettle to an ephemeral disk first, and then import the data to HybridDB for PostgreSQL by using COPY or OSS.
  - You can also mount an OSS instance as a virtual ephemeral disk, import data through Kettle to that disk, and then import data to HybridDB for PostgreSQL by using the OSS external table.
- *Informatica*: a commercial ETL tool.
- *Bifrost*: an ETL tool in Alibaba Cloud marketplace. It now provides commercial services.
- *dbsync*: an open-source database synchronization tool provided by Alibaba Cloud.
  - Supports synchronizing data concurrently from MySQL and PostgreSQL to HybridDB for PostgreSQL.
  - Supports simple data conversion.
  - Supports parsing binlog to synchronize data from MySQL to HybridDB for PostgreSQL in a quasi-real-time manner.

- **Other ETL tools supporting Greenplum.**

4 How to learn more technical information?

You can learn more technical information from the *open-source community of Greenplum Database*. In addition, you can contact us by using the ticket system.

5 How to select instance type?

When creating or upgrading an instance in HybridDB for PostgreSQL, you must specify the *group type and number of groups*.

HybridDB for PostgreSQL supports multiple group types. For more information, see *Instance types*. You can also specify multiple groups with the same type. We recommend that you consider the following factors when selecting your group type and quantity:

- The storage space required
- The computing capability required

Storage space

The storage space of a group type and group quantity combo = the storage space of the group type * the number of groups.

Note the following when determining the expected storage space:

- Make the storage space slightly larger than the actual space evaluated. Because data processing can generate some logs and temporary files.
- Select distribution keys appropriately and avoid data skews. Allocate the distributi on keys evenly as much as possible, otherwise data skews may occur. Data skews can exhaust the storage space in a computing group and leave the storage space of other groups idle, resulting in low usage of storage resource.

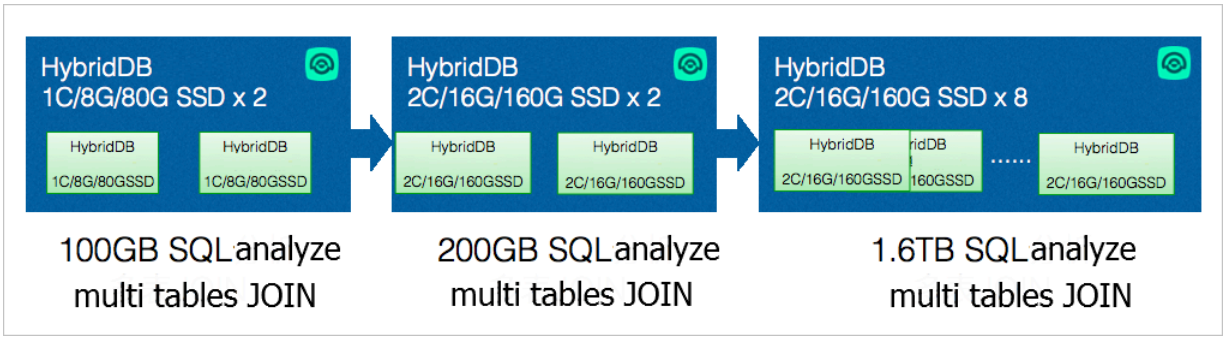
Computing capability

A group type and group quantity combo matches a specific computing capability. The computing capability is determined by the computing group type, number of CPUs, memory size, and the number of groups.

HybridDB for PostgreSQL has two group types available:

- High-performance group: the group type name starts with gpdb.group.segsdx. This type features a better I/O capability that secures higher performance.
- High-capacity group: the group type name starts with gpdb.group.seghdx. This type features a larger and more affordable space to meet higher storage demands.

Three other factors also contribute to linear increase of performance. Take a high -performance type for example. The SQL execution durations in the following use cases are close to each other:



In addition, you also need to take the price into account.

6 How to clear the data on a locked disk?

Analysis

When the disk space of any computing group reaches the limit or the disk space on a master node reaches the limit, the entire instance is locked. Connect to the database and run the following command to check whether your instance is locked:

```
show rds_force_trans_ro_non_sup;
```

If the `rds_force_trans_ro_non_sup` value is `on` in the response, it means that the instance is locked, and the database is read-only.

Fix

When an instance is locked because the disk is full, the truncate, drop, and grant operations are still supported on the data table. You can use these operations to clear the disk until the space falls within the threshold. After that, the instance is automatically unlocked in about five minutes.

 Note:
Deletion is not supported when the instance is locked. Because deletion requests writing data to the xlog, which can increase space usage.

You can also run the following statement to query the table size:

```
select pg_size_pretty(pg_total_relation_size('test'));
```

7 How do I modify parameters?

At present, HybridDB for PostgreSQL does not support global changes of parameters . But you can modify parameters in a connection (see Greenplum's parameter modification restrictions for details).

You can use the command alter role set = modify the parameters. This modification only applies to the specified user.

8 How to use table partition?

We recommend that you enable table partition for fact tables or large tables in the database. Table partition simplies the deleting data and importaning data operations on a regular basis. With table partition,

- You can use the `alter table drop partition` command to delete all the data in a partition.
- You can use partition exchange, that is, the `alter table exchange partition` command to add a new data partition.

HybridDB for PostgreSQL supports range partition, list partition, and composite partition. But range partition only supports partitioning by number or time-based fields.

An example table using range partitions is shown as follows.

```
CREATE TABLE LINEITEM (
L_ORDERKEY BIGINT NOT NULL,
L_PARTKEY BIGINT NOT NULL,
L_SUPPKEY BIGINT NOT NULL,
L_LINENUMBER INTEGER,
L_QUANTITY FLOAT8,
L_EXTENDEDPRICE FLOAT8,
L_DISCOUNT FLOAT8,
L_TAX FLOAT8,
L_RETURNFLAG CHAR(1),
L_LINESTATUS CHAR(1),
L_SHIPDATE DATE,
L_COMMITDATE DATE,
L_RECEIPTDATE DATE,
L_SHIPINSTRUCT CHAR(25),
L_SHIPMODE CHAR(10),
L_COMMENT VARCHAR(44)
) WITH (APPENDONLY=true, ORIENTATION=column, COMPRESSTYPE=zlib,
COMPRESSLEVEL=5, BLOCKSIZE=1048576, OIDS=false) DISTRIBUTED BY (
l_orderkey)
PARTITION BY RANGE (L_SHIPDATE) (START (date '1992-01-01') INCLUSIVE
END (date '2000-01-01') EXCLUSIVE EVERY (INTERVAL '1 month' ));
```

9 How to enable column-store and compression in HybridDB for PostgreSQL?

Row-store with no compression is used by default when you create a table in HybridDB for PostgreSQL. To enable column-store and compression, you must specify the column-store and compression options when creating the table. For example, you can add the following statement to the tabulation statements to enable the two features.

```
with (APPENDONLY=true, ORIENTATION=column, COMPRESSTYPE=zlib,
COMPRESSLEVEL=5, BLOCKSIZE=1048576, OIDS=false)
```

In general, we recommend that you enable column-store and compression, especially when you have a lot of complicated queries to the table, or when you want to reduce the storage cost.

 Note:

HybridDB for PostgreSQL currently only supports the zlib and RLE_TYPE compression algorithms. If you specify to use quicklz, HybridDB for PostgreSQL automatically converts the algorithm to zlib.