

# Alibaba Cloud AnalyticDB for PostgreSQL

常見問題

檔案版本：20190903

# 法律聲明

---

阿里雲提醒您在閱讀或使用本文檔之前仔細閱讀、充分理解本法律聲明各條款的內容。如果您閱讀或使用本文檔，您的閱讀或使用行為將被視為對本聲明全部內容的認可。

1. 您應當通過阿里雲網站或阿里雲提供的其他授權通道下載、擷取本文檔，且僅能用於自身的合法合規的商務活動。本文檔的內容視為阿里雲的保密資訊，您應當嚴格遵守保密義務；未經阿里雲事先書面同意，您不得向任何第三方披露本手冊內容或提供給任何第三方使用。
2. 未經阿里雲事先書面許可，任何單位、公司或個人不得擅自摘抄、翻譯、複製本文檔內容的部分或全部，不得以任何方式或途徑進行傳播和宣傳。
3. 由於產品版本升級、調整或其他原因，本文檔內容有可能變更。阿里雲保留在沒有任何通知或者提示下對本文檔的內容進行修改的權利，並在阿里雲授權通道中不時發布更新後的使用者文檔。您應當即時關注使用者文檔的版本變更並通過阿里雲授權渠道下載、擷取最新版的使用者文檔。
4. 本文檔僅作為使用者使用阿里雲產品及服務的參考性指引，阿里雲以產品及服務的”現狀“、“有缺陷”和“當前功能”的狀態提供本文檔。阿里雲在現有技術的基礎上盡最大努力提供相應的介紹及操作指引，但阿里雲在此明確聲明對本文檔內容的準確性、完整性、適用性、可靠性等不作任何明示或暗示的保證。任何單位、公司或個人因為下載、使用或信賴本文檔而發生任何差錯或經濟損失的，阿里雲不承擔任何法律責任。在任何情況下，阿里雲均不對任何間接性、後果性、懲戒性、偶然性、特殊性或刑罰性的損害，包括使用者使用或信賴本文檔而遭受的利潤損失，承擔責任（即使阿里雲已被告知該等損失的可能性）。
5. 阿里雲網站上所有內容，包括但不限於著作、產品、圖片、檔案、資訊、資料、網站架構、網站畫面的安排、網頁設計，均由阿里雲和/或其關係企業依法擁有其智慧財產權，包括但不限於商標權、專利權、著作權、商業秘密等。非經阿里雲和/或其關係企業書面同意，任何人不得擅自使用、修改、複製、公開傳播、改變、散布、發行或公開發表阿里雲網站、產品程式或內容。此外，未經阿里雲事先書面同意，任何人不得為了任何營銷、廣告、促銷或其他目的使用、公布或複製阿里雲的名稱（包括但不限於單獨為或以組合形式包含”阿里雲”、Aliyun”、“萬網”等阿里雲和/或其關係企業品牌，上述品牌的附屬標誌及圖案或任何類似公司名稱、商號、商標、產品或服務名稱、網域名稱、圖案標示、標誌、標識或通過特定描述使第三方能夠識別阿里雲和/或其關係企業）。
6. 如若發現本文檔存在任何錯誤，請與阿里雲取得直接聯絡。

## 通用約定

格式	說明	範例
	該類警示資訊將導致系統重大變更甚至故障，或者導致人身傷害等結果。	 <b>禁止：</b> 重設操作將丟失使用者配置資料。
	該類警示資訊可能導致系統重大變更甚至故障，或者導致人身傷害等結果。	 <b>警告：</b> 重啟操作將導致業務中斷，恢復業務所需時間約10分鐘。
	用於補充說明、最佳實務、竅門等，不是使用者必須瞭解的內容。	 <b>说明：</b> 您也可以通過按Ctrl + A選中全部檔案。
>	多級菜單遞進。	設定 > 網路 > 設定網路類型
粗體	表示按鍵、菜單、頁面名稱等UI元素。	單擊 確定。
<code>courier</code> 字型	命令。	執行 <code>cd / d C :/ windows</code> 命令，進入Windows系統檔案夾。
##	表示參數、變數。	<code>bae log list --instanceid Instance_ID</code>
[ ]或者[a b]	表示可選項，至多選擇一個。	<code>ipconfig [-all -t]</code>
{ }或者{a b}	表示必選項，至多選擇一個。	<code>swich {stand   slave}</code>

## 目錄

---

法律聲明.....	I
通用約定.....	I
1 如何查看錶和資料庫的資料量大小? .....	1
2 目前 HybridDB for PostgreSQL 支援哪些ETL工具? 如何從其他資料來源定期同步資料到 HybridDB for PostgreSQL? .....	2
3 如何擷取更多技術資訊.....	3
4 如何選擇執行個體規格? .....	4
5 磁碟鎖定之後如何清理資料? .....	6
6 如何使用表分區? .....	7
7 如何使用 HybridDB for PostgreSQL 列存和壓縮功能? .....	8
8 執行個體“節點”規格與原“計算群組”規格的對應關係? .....	9

# 1 如何查看錶和資料庫的資料量大小?

假設表的模式為< schemaname >, 表名為< tablename >。

- 執行以下命令，查詢一張表的總大小（單位為MB，包含表的索引和資料）：

```
select pg_size_pr etty ( pg_total_r elation_si ze ('< schemaname >.< tablename >'));
```

- 執行以下命令，查詢表的資料大小（單位MB，不包括索引）：

```
select pg_size_pr etty ( pg_relatio n_size ('< schemaname >.< tablename >'));
```

- 執行以下命令，查詢分區表所有分區的總大小（單位MB，包含表的索引和資料）：

```
select schemaname , tablename , round ( sum ( pg_total_r elation_si ze ( schemaname || '.' || partitiont ablename )) / 1024 / 1024 ) " MB " from pg_partiti ons where schemaname = '< schemaname >' and tablename = '< tablename >' group by 1 , 2 ;
```

- 執行以下命令，查詢一個Schema下面的所有表的總大小（單位MB，包括索引和資料）：

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

- 執行以下命令，查詢每個資料庫的大小（單位MB）：

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

## 2 目前 HybridDB for PostgreSQL 支援哪些 ETL 工具？如何從其他資料來源定期同步資料到 HybridDB for PostgreSQL？

---

目前支援的 ETL 工具如下：

- [阿里雲的 Data Integration 服務\(Data Integration\)](#)：阿里雲提供的 ETL 工具。在 Data Integration 服務中，將 HybridDB for PostgreSQL 配置為一個 PostgreSQL 資料庫，即可實現其他資料來源（RDS、MaxCompute、TableStore 等）到 HybridDB for PostgreSQL 的資料同步。
  - 您可以直接從其他資料來源讀取資料，寫入到 HybridDB for PostgreSQL 中。
  - 如果資料量較大，需要並發匯入，則建議您先通過 Data Integration 服務把資料從其他資料來源匯入到 OSS，再通過 OSS 外部表格匯入 HybridDB for PostgreSQL。
- [Pentaho Kettle Data Integration 軟體](#)：開源的 ETL 工具。
  - 支援將資料先通過 Kettle 匯入到本地磁碟，再通過 COPY 或 OSS 匯入到 HybridDB for PostgreSQL。
  - 也支援將 OSS 儲存掛載為本地虛擬磁碟，通過 Kettle 匯入到此磁碟，最後通過 HybridDB for PostgreSQL 的 OSS 外部表格匯入到 HybridDB for PostgreSQL 中。
- [Informatica 軟體](#)：商業化的 ETL 工具。
- [彩虹橋](#)：阿里雲雲市場中的 ETL 工具，供應商業化服務。
- [dbsync](#)：阿里雲提供的開來源資料庫同步工具。
  - 支援從 MySQL、PostgreSQL 並發同步資料到 HybridDB for PostgreSQL。
  - 支援簡單的資料轉換。
  - 支援通過解析 Binlog，准即時地從 MySQL 同步資料到 HybridDB for PostgreSQL。
- 其他支援 Greenplum 的 ETL 工具。

## 3 如何擷取更多技術資訊

---

1. 關注云棲社區中的 - Greenplum Database 技術圈：<https://yq.aliyun.com/groups/13>
2. 通過工單系統向我們諮詢。
3. Greenplum Database 開源社區官方資料：<http://www.greenplum.org>

## 4 如何選擇執行個體規格？

---

在雲資料庫 HybridDB for PostgreSQL 中建立執行個體和升級執行個體規格時，您都需要選擇[計算群組規格](#)和[計算群組數量](#)。目前，HybridDB for PostgreSQL 支援多種計算群組規格，詳情參見[規格總覽](#)。每種計算群組規格可以選擇不同的計算群組數量，具體可參見[規格組合速查表](#)。

您在選擇計算群組規格和計算群組數量的時候，需要考慮下面兩個因素：

- 所需要的儲存空間
- 所需要的計算能力

### 儲存空間

每種計算群組規格和計算群組數量組合對應的儲存空間 = 計算群組規格的儲存空間 \* 計算群組數量

在選擇需要的儲存空間時，需要注意以下幾點：

- 選擇的儲存空間要略大於實際評估結果。因為在資料處理的過程中，會產生一些記錄檔和臨時檔案。
- 合理選取分布鍵，盡量避免資料扭曲。分布鍵選擇應該盡量平均，否則容易造成資料扭曲。資料扭曲會使某個計算群組儲存空間耗盡，而其他計算群組儲存空間空閒，儲存資源使用率較低。

### 計算能力

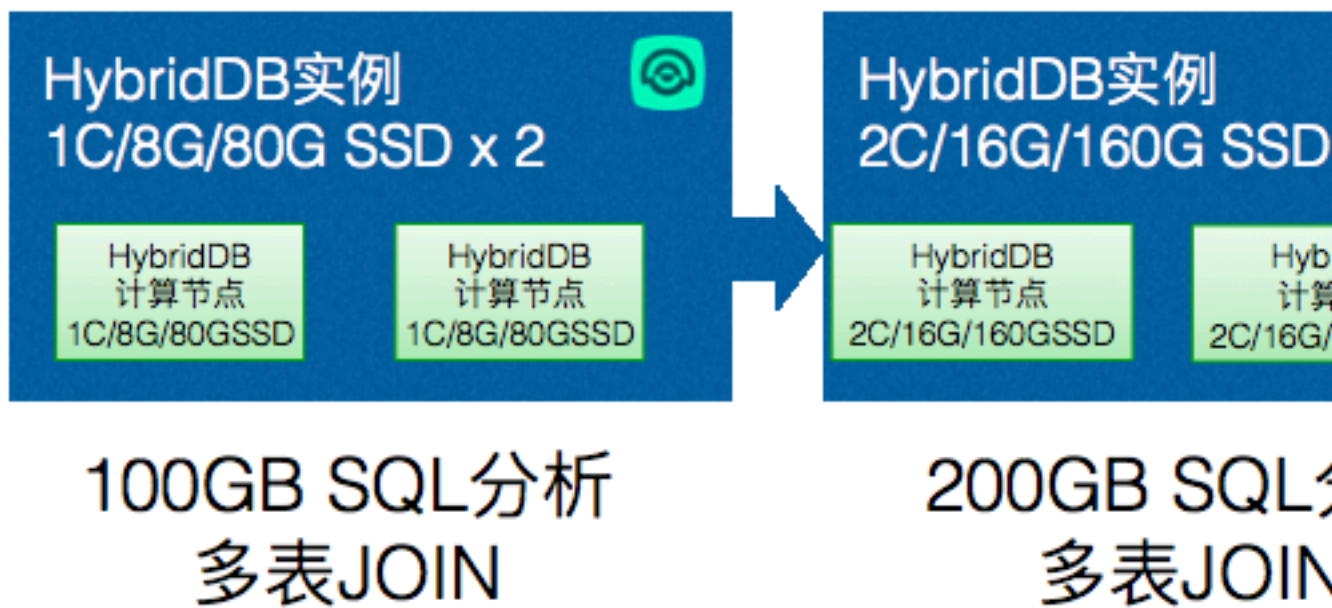
每種計算群組規格和計算群組數量組合對應相應計算能力。計算能力取決於其計算群組規格類型，CPU個數，記憶體大小和計算群組數量。

目前，HybridDB for PostgreSQL 提供兩類規格：

- 高效能，規格名稱以gpdb.group.segsdx開始，提供更好的 I/O 能力，帶來更高的效能。
- 高容量，規格名稱以gpdb.group.seghdx開始，提供更大、更實惠的空間，滿足更高的儲存需要。

除了計算群組規格類型之外，其他三個因素的增加也可以線性地增加效能。以高效能規格為例，下面三種情況，SQL 執行時間基本相近：





另外，您還要考慮每種計算群組規格和計算群組數量組合對應的價格，詳見 [產品定價資訊](#)。

## 5 磁碟鎖定之後如何清理資料？

### 磁碟滿判定條件

任何一個計算群組磁碟空間達到限制或者主節點（Master）磁碟空間達到限制，就會鎖定整個執行個體。

### 檢查執行個體是否鎖定

直接連接資料庫，執行以下命令，檢查執行個體是否被鎖定：

```
show rds_force_ trans_ro_n on_sup ;
```

根據返回結果，如果 `rds_force_ trans_ro_n on_sup` 的值為 `on`，表示執行個體已被鎖定，資料庫已處於唯讀狀態。

### 如何清理資料

當執行個體因磁碟滿被鎖定時，HybridDB for PostgreSQL 仍然支援對資料表進行truncate/drop/grant 操作。清理資料後，如果磁碟空間降到閾值以下，過大約5分鐘執行個體會自動解鎖。



注意：

執行個體鎖定後，不支援 delete 操作。因為 delete 會寫 xlog，使空間佔用增大。

另外，您可以執行以下語句來查詢表的大小：

```
select pg_size_pr etty ( pg_total_r elation_si ze (' test '));
```

## 6 如何使用表分區？

對於資料庫中的事實表以及一些比較大的表，建議您使用表分區。使用表資料分割函數，方便您定期進行資料刪除（通過 `alter table drop partition` 命令可刪除整個分區的資料）和匯入（使用交換分區的方式，即 `alter table exchange partition` 命令可以加入新資料分區）。

HybridDB for PostgreSQL 支援 Range Partition（定界分割）、List Partition（列表分區）和 Composite Partition（多級分區）。但是，Range Partition 只支援利用數值或時間類型的欄位來分區。

以下是一個使用 Range Partition 的表的例子。

```
CREATE TABLE LINEITEM (
  L_ORDERKEY BIGINT NOT NULL ,
  L_PARTKEY BIGINT NOT NULL ,
  L_SUPPKEY BIGINT NOT NULL ,
  L_LINENUMB ER INTEGER ,
  L_QUANTITY FLOAT8 ,
  L_EXTENDED PRICE 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 , COMPRESSION = zlib , COMPRESSIONLEVEL = 5 , BLOCKSIZE = 1048576 , OIDS = false )
DISTRIBUTE BY ( l_orderkey )
PARTITION BY RANGE ( L_SHIPDATE ) ( START ( date ' 1992 - 01 - 01 ' ) INCLUSIVE END ( date ' 2000 - 01 - 01 ' ) EXCLUSIVE
EVERY ( INTERVAL ' 1 month ' ) );
```

## 7 如何使用 HybridDB for PostgreSQL 列存和壓縮功能?

在 HybridDB for PostgreSQL 中建立表時，預設使用行式儲存，並且不啟用壓縮。如需使用列存和壓縮功能，您必須在建表時指定列存和壓縮選項。例如，您可以在建表語句中加入以下子句，來啟用列存和壓縮功能。

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

一般情況下，推薦您使用列存和壓縮功能，尤其是在包含較多複雜查詢，或需要降低儲存成本的情境。



注意:

目前，HybridDB for PostgreSQL 僅支援 zlib 和 RLE\_TYPE 壓縮演算法；如果指定了 quicklz 演算法，會自動轉換為 zlib。

## 8 執行個體“節點”規格與原“計算群組”規格的對應關係？

AnalyticDB for PostgreSQL 在 2019年8月23日，執行個體規格定義變更，叢集執行個體由“計算群組”組成的規格定義改為由“節點”組成。本文主要用來描述“節點”和“計算群組”的對應關係。

引入節點定義取代計算群組定義，主要為了簡化 MPP 叢集的規格選型，遵循標準叢集資料庫的定義模式。一個節點只包含一個MPP資料分區，相對可以簡化MPP叢集的選型和使用，而一個計算群組根據其規格可能會包含多個MPP資料分區。

計算群組規格（不再使用）	節點	MPP資料分區數	資源合計	隨用隨付價格
1x2C SSD	2個 1核 SSD節點	2	2核/16GB記憶體/160GB使用者儲存/320GB總儲存	1.76元/小時
1x16C SSD	16個 1核 SSD節點	16	16核/128GB記憶體/1280GB使用者儲存/2560GB總儲存	14.08元/小時
4x4C SSD	4個 4核 SSD節點	4	16核/128GB記憶體/1280GB使用者儲存/2560GB總儲存	14.08元/小時
2x2C HDD	2個 2核 HDD節點	2	4核/32GB記憶體/2TB使用者儲存/4TB總儲存	4.60元/小時
4x4C HDD	4個 4核 HDD節點	4	16核/96GB記憶體/8TB使用者儲存/16TB總儲存	18.40元/小時

按如上對應關係，節點 和 計算群組 在相同資源和分區數下，效能和容量保持一致，價格保持一致。