

# Alibaba Cloud

## AnalyticDB for PostgreSQL Product Introduction

Document Version: 20200818

## 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 because of product version upgrade, adjustment, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and an updated version of this document will be released through Alibaba Cloud-authorized channels from time to time. You should 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 this document based on the "status quo", "being defective", and "existing functions" of its products and services. 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 take legal responsibility for any errors or lost profits incurred by any organization, company, or individual arising from download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, take responsibility for any indirect, consequential, punitive, contingent, 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 contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, 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 this document 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 directly contact Alibaba Cloud for any errors of this document.

# Document conventions

Style	Description	Example
 <b>Danger</b>	A danger notice indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	 <b>Danger:</b> Resetting will result in the loss of user configuration data.
 <b>Warning</b>	A warning notice indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	 <b>Warning:</b> Restarting will cause business interruption. About 10 minutes are required to restart an instance.
 <b>Notice</b>	A caution notice indicates warning information, supplementary instructions, and other content that the user must understand.	 <b>Notice:</b> If the weight is set to 0, the server no longer receives new requests.
 <b>Note</b>	A note indicates supplemental instructions, best practices, tips, and other content.	 <b>Note:</b> You can use Ctrl + A to select all files.
>	Closing angle brackets are used to indicate a multi-level menu cascade.	Click <b>Settings&gt; Network&gt; Set network type</b> .
<b>Bold</b>	Bold formatting is used for buttons, menus, page names, and other UI elements.	Click <b>OK</b> .
<b>Courier font</b>	Courier font is used for commands	Run the <code>cd /d C:/window</code> command to enter the Windows system folder.
<i>Italic</i>	Italic formatting is used for parameters and variables.	<code>bae log list --instanceid</code> <i>Instance_ID</i>
[ ] or [a b]	This format is used for an optional value, where only one item can be selected.	<code>ipconfig [-all -t]</code>
{ } or {a b}	This format is used for a required value, where only one item can be selected.	<code>switch {active stand}</code>

# Table of Contents

1.Overview	05
2.Scenarios and benefits	07
3.Terms	09
4.Features and limits	10
5.Precautions	11

# 1. Overview

AnalyticDB for PostgreSQL, formerly known as HybridDB for PostgreSQL, is a massively parallel processing (MPP) data warehousing service designed to analyze large volumes of data online.

AnalyticDB for PostgreSQL is developed based on the open source Greenplum Database project and enhanced with in-depth extensions by Alibaba Cloud. It is compatible with ANSI SQL 2003 syntax and the PostgreSQL and Oracle database ecosystems. It also supports row-oriented storage and column-oriented storage. AnalyticDB for PostgreSQL processes petabytes of data online at a high performance and supports highly concurrent online queries. This makes it a competitive data warehousing solution in various industries.

## Features

- Adaptable to variable workloads with no optimization required.

AnalyticDB for PostgreSQL is fully compatible with SQL 2003 syntax and partially compatible with Oracle syntax. It also supports PL/SQL stored procedures. It offers new-generation query optimizers to relieve you from the need to optimize complex SQL statements.

- Analyzes petabytes of data within seconds.

AnalyticDB for PostgreSQL uses an MPP scale-out architecture to respond to queries for petabytes of data within seconds. It performs 10 times better than conventional database engines by using vectorized computing and intelligent columnstore indexing.

- Provides high availability and always-on connectivity.

AnalyticDB for PostgreSQL supports distributed transactions, ACID (short for atomicity, consistency, isolation, durability), redundancy for all nodes and data, plus automatic monitoring and failover.

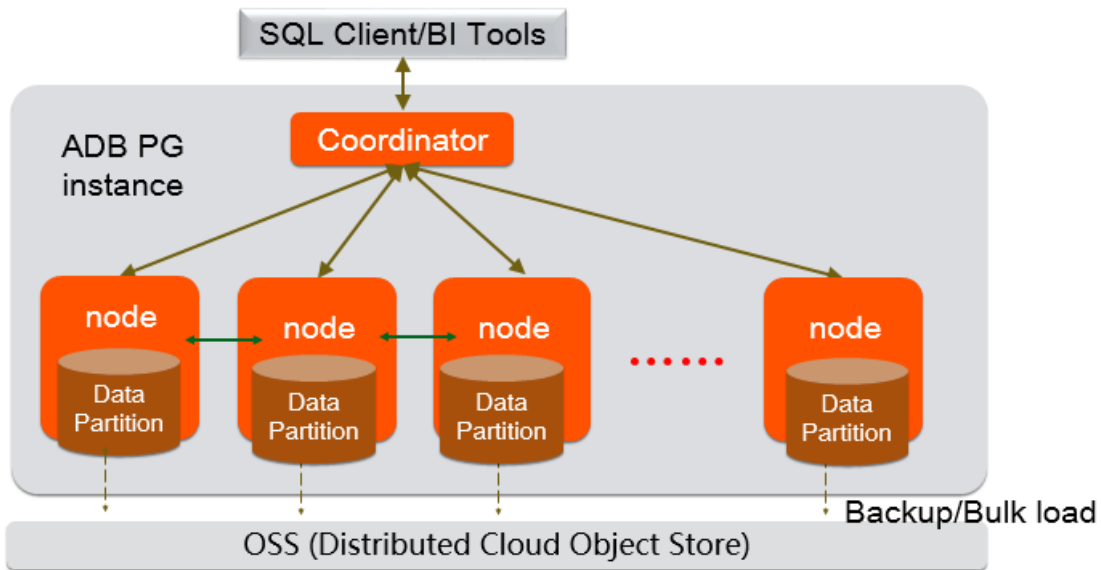
- Compatible with a wide variety of ecosystems.

AnalyticDB for PostgreSQL supports mainstream business intelligence (BI) and extract, transform, load (ETL) tools. For example, it is integrated with the PostGIS extension to analyze geographic data and with the MADlib library to provide more than 300 built-in machine learning algorithms.

- Enables data interconnection.

AnalyticDB for PostgreSQL can synchronize data with various data sources by using tools such as Alibaba Cloud Data Transmission Service (DTS) and Dataworks. It also supports highly parallel access to OSS and Data Lake Analytics.

## Architecture



AnalyticDB for PostgreSQL uses the MPP architecture, which enables an instance to be composed of multiple **compute nodes**. You can add compute nodes to an instance to linearly scale out its storage capacity and maintain a stable response time. Each instance is composed of one coordinator node and multiple compute nodes.

- **Coordinator node**
  - Receives query requests and determines distributed query plans.
- **Compute node**
  - Provides massively parallel processing.
  - Stores data in dual copies on each partition.
  - Automatically backs up data to OSS on a regular basis.

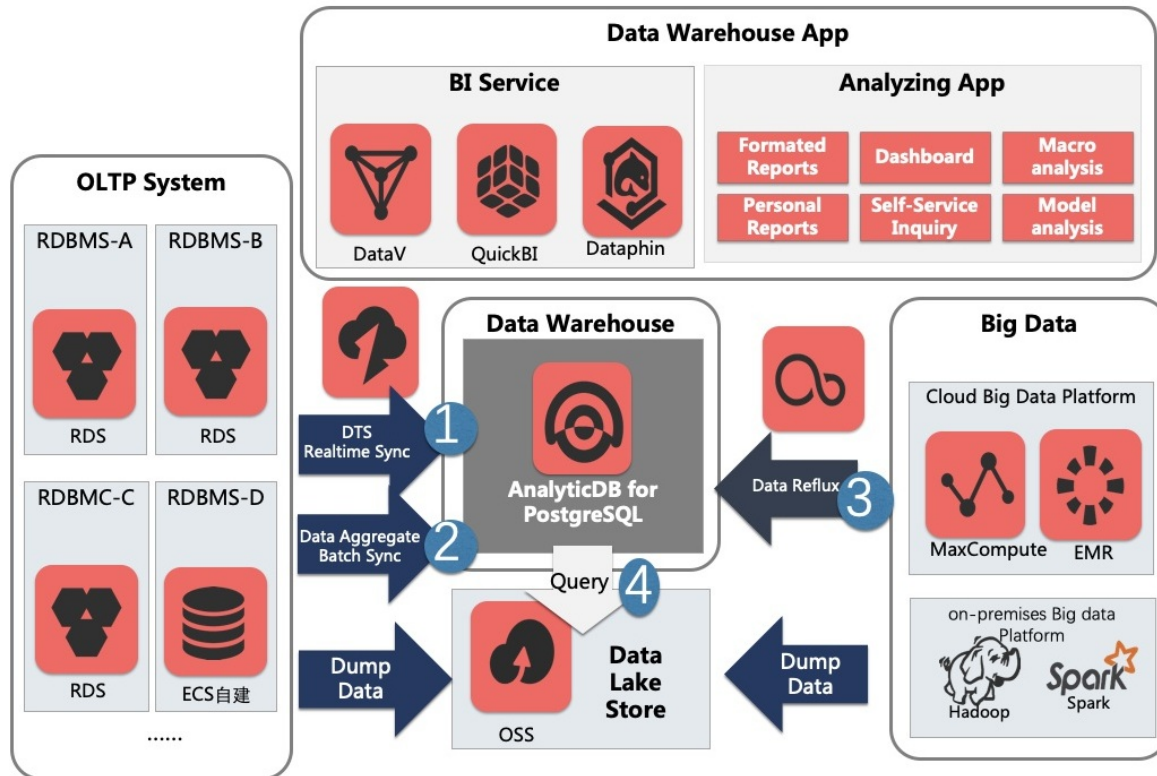
## More

- If you require technical assistance, [submit a ticket](#).
- For information about the open source Greenplum Database project, visit <http://www.greenplum.org>.

**Note** Since August 23, 2019, the basic building block of an AnalyticDB for PostgreSQL instance has been changed from compute group to compute node. A compute group contains multiple partitions, whereas a compute node corresponds to an MPP partition. This simplifies the instance type definition and complies with cluster database naming conventions. For details about the mappings between compute nodes and compute groups, see [Mappings between compute node types and compute group types](#).

## 2.Scenarios and benefits

### Scenarios



- **Data warehousing service**

Data Transmission Service (DTS) can synchronize data in real time in production system databases such as ApsaraDB RDS for MySQL, ApsaraDB RDS for PostgreSQL, and ApsaraDB RDS for PolarDB and traditional databases such as Oracle and SQL Server. Data can also be batch synchronized to AnalyticDB for PostgreSQL through the data integration service (DataX). AnalyticDB for PostgreSQL supports complex Extract, Transform, and Load (ETL) operations on large amounts of data. These tasks can also be scheduled by Dataworks. AnalyticDB for PostgreSQL also provides high-performance online analysis capabilities and can use Quick BI, DataV, Tableau, and FineReport for report presentation and ad hoc query.

- **Big data analytics platform**

You can import huge amounts of data stored in MaxCompute, Hadoop, and Spark to AnalyticDB for PostgreSQL through DataX or OSS for high-performance analysis, processing, and exploration.

- **Data lake analytics**

AnalyticDB for PostgreSQL can use an external table mechanism to access the huge amounts of data stored in OSS in parallel and build an Alibaba Cloud data lake analytics platform.

### Benefits

In OLAP business scenarios, AnalyticDB for PostgreSQL has the following benefits:

- ETL for offline data processing

AnalyticDB for PostgreSQL provides the following benefits that make it ideal to optimize complex SQL queries and aggregate and analyze huge amounts of data:

- Supports standard SQL, OLAP window functions, and stored procedures.
- Uses an ORCA-based distributed SQL optimizer to make complex queries without the need for tuning.
- Uses the MPP architecture capable of processing petabytes of data in seconds.
- Provides column store-based high-performance scanning of large tables and high compression ratio.

- Online high-performance query

AnalyticDB for PostgreSQL provides the following benefits for real-time exploration, warehousing, and updating of data:

- Allows you to write and update high-throughput data through INSERT, UPDATE, and DELETE operations.
- Supports a wide variety of row store indexes such as B-tree and bitmap indexes and can respond to queries in milliseconds.
- Supports distributed transactions, standard database isolation levels, and HTAP.

- Multi-model data analysis

AnalyticDB for PostgreSQL provides the following benefits for processing of a variety of unstructured data sources:

- Supports the PostGIS extension for geographic data analysis and processing.
- Takes advantage of the MADlib extension, a library of in-database machine learning algorithms, to implement an AI-native database.
- Provides high-performance retrieval and analysis of unstructured data such as images, speech, and text through vector retrieval.
- Supports formats such as JSON and can process and analyze semi-structured data such as logs.



## 3. Terms

The following table lists the basic concepts of AnalyticDB for PostgreSQL.

Term	Description
Massively Parallel Processing (MPP)	A distributed shared-nothing computing architecture. MPP uses a large number of nodes that do not share resources with each other to perform parallel computing and improve performance. A node is known as a data partition in AnalyticDB for PostgreSQL.
compute node	The unit for allocating cluster resources in AnalyticDB for PostgreSQL. An instance is composed of multiple compute nodes. When the number of compute nodes increases, the storage capacity scales out but the query response time does not change. A compute node is a unit of computing resources that includes fixed CPU cores, memory, and storage. Each compute node type contains a data partition in the MPP architecture.
number of compute nodes	The number of compute nodes purchased for a cluster instance. A single instance supports up to 4,096 nodes. The storage space and compute resources of cluster instances increase linearly as the number of compute nodes increases.
data partition	A unit for computing and storing data in the MPP architecture. Table data is stored in different data partitions by partition key.

## 4. Features and limits

This topic lists the basic features and limits of AnalyticDB for PostgreSQL (formerly known as HybridDB for PostgreSQL).

### Features

- Provides the key features of Greenplum Database. For more information, see [Summary of Greenplum Features](#).
- Supports exclusive MetaScan and SortKey technologies. For more information, see [Use sort keys in column-oriented tables](#).
- Supports the next-generation ORCA optimizer based on Cascade Framework.
- Supports distributed stored procedures in PL/pgSQL and PL/Java. For details about PL/Java stored procedures, see [Use PL/Java UDF](#).
- Supports vector retrieval of unstructured data.
- Supports multiple extensions such as PostGIS, MADlib, fuzziystrmatch, orafunc, pgcrypto, intarray, and Roaring Bitmap. You can use the CREATE EXTENSION command to create them. For more information, see [Manage extensions](#).
- Allows you to use the OSS\_EXT extension to read data from or write data to Alibaba Cloud Object Storage Service (OSS), and provides gzip compression to reduce the external table storage cost. For more information, see [Import or export OSS data by using OSS external tables](#).
- Supports the JSON data type. For more information, see [Manage JSON and JSON-B data](#).

 **Note** The JSONB data type is not supported.

- Supports the HypeLogLog data type. For more information, see [Use HyperLogLog](#)
- AnalyticDB for PostgreSQL automatically performs a full backup every weekend. You can submit a ticket to modify the backup time and period. You can also use the pg\_dump tool to perform backups on your own.

### Limits

- For details about limits of the core features, see [Summary of Greenplum Features](#).
- Permission limits: The root user of AnalyticDB for PostgreSQL has permissions to create databases (CREATEDB) and roles (CREATEROLE), but does not have superuser permissions (SUPERUSER) and cannot perform operations that require such permissions. For example, the root user cannot run file functions such as pg\_ls\_dir. However, the root user has permissions to view and modify data of other non-superusers and terminate their connections.
- PL/R extensions are not supported.
- PL/Python extensions can be created but functions cannot be created with PL/Python.
- The gpfdist tool is not supported. Note: Nodes can load data in parallel based on OSS external tables. For more information, see [Import or export OSS data by using OSS external tables](#)
- MapReduce APIs, gphdfs storage APIs, and local external tables are not supported.

## 5. Precautions

This topic describes precautions you must consider before you use AnalyticDB for PostgreSQL.

You do not need to conduct basic database O&M operations such as ensuring high availability and installing security patches for purchased AnalyticDB for PostgreSQL instances. However, you must pay attention to the following situations:

- **Instance upgrade:** Instances are read-only while they are being upgraded. Transient disconnections may occur after the upgrade is completed for up to 30 seconds. Make sure that your application is configured with automatic reconnection policies by using mechanisms such as the connection pool.
- **Switchover:** The coordinator and compute nodes each adopts a primary/secondary high availability architecture. When an error occurs in the primary node, the service is automatically switched to the secondary node within 30 seconds. A transient disconnection of up to 30 seconds may occur during the switchover. Make sure that your application is configured with automatic reconnection policies by using mechanisms such as the connection pool.