

Cloud Data Warehouse for PostgreSQL

Product Introduction

Product Documentation



Copyright Notice

©2013-2024 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

Trademark Notice



All trademarks associated with Tencent Cloud and its services are owned by Tencent Cloud Computing (Beijing) Company Limited and its affiliated companies. Trademarks of third parties referred to in this document are owned by their respective proprietors.

Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud's products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.

Contents

Product Introduction

Overview

Use Cases

Node Specification

Cluster Architecture

Product Introduction

Overview

Last updated : 2024-02-02 15:15:26

Cloud Data Warehouse for PostgreSQL (CDWPG, formerly Snova Data Warehouse) offers a simple, fast, and cost-effective petabyte-level data warehouse solution in the cloud. It is compatible with the open-source data warehouse Greenplum and based on the massively parallel processing (MPP) architecture. It allows you to leverage diversified open-source PostgreSQL tools for ad-hoc query and analysis, ETL processing, and visual exploration of massive amounts of data in the cloud. With seamless cloud data integration, you can also easily analyze petabytes of data on data engines such as COS, TencentDB, and Elasticsearch Service.

Benefits

Elastic scaling

CDWPG delivers convenient and elastic scalability, allowing you to scale hundreds of nodes and adjust their configurations through simple operations through the console or TencentCloud API. You can choose to scale the compute unit, CPU, memory, and storage proportionally to better match your business needs and progress.

Ease of use

Cluster management, monitoring, and maintenance can be done in the console, freeing you from the cumbersome Ops of the underlying infrastructure. CDWPG fully conforms to the ANSI 2008 standard for SQL, so you can build an enterprise-grade data warehouse with standard SQL. You can also directly query COS data without preloading needed.

Seamless integration

CDWPG supports unlimited expansion of the storage capacity on top of COS. It can work with a wide variety of tools and solutions to enable high-speed data import from multiple sources (such as traditional relational databases, CKafka, and Oceanus) for aggregation and analysis in the cloud.

Best-in-class performance

CDWPG boasts a distributed MPP framework to linearly scale the storage and computing capabilities. With hybrid row/column-oriented storage, it offers an optimal storage solution for your business to query data several times faster than traditional data warehouses thanks to its deeply optimized query engine.

High security and reliability

CDWPG has two nodes in a redundant way to implement failover and disaster recovery imperceptible to users. It adopts distributed deployment to protect compute units, servers, and cabinets to secure critical data infrastructure. Furthermore, your clusters are independently deployed in isolated VPCs for more secure data access.

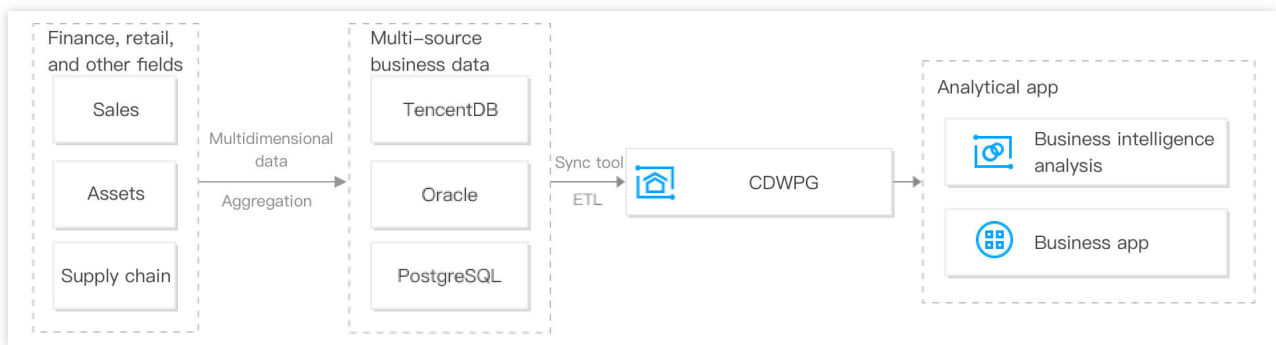
Use Cases

Last updated : 2024-02-02 15:15:26

Operational Analysis and Decision Making

In the fields of finance and retail, business data of the sales, assets, and supply chain need to be aggregated and analyzed for you to stay on top of operations and make accurate and efficient decisions.

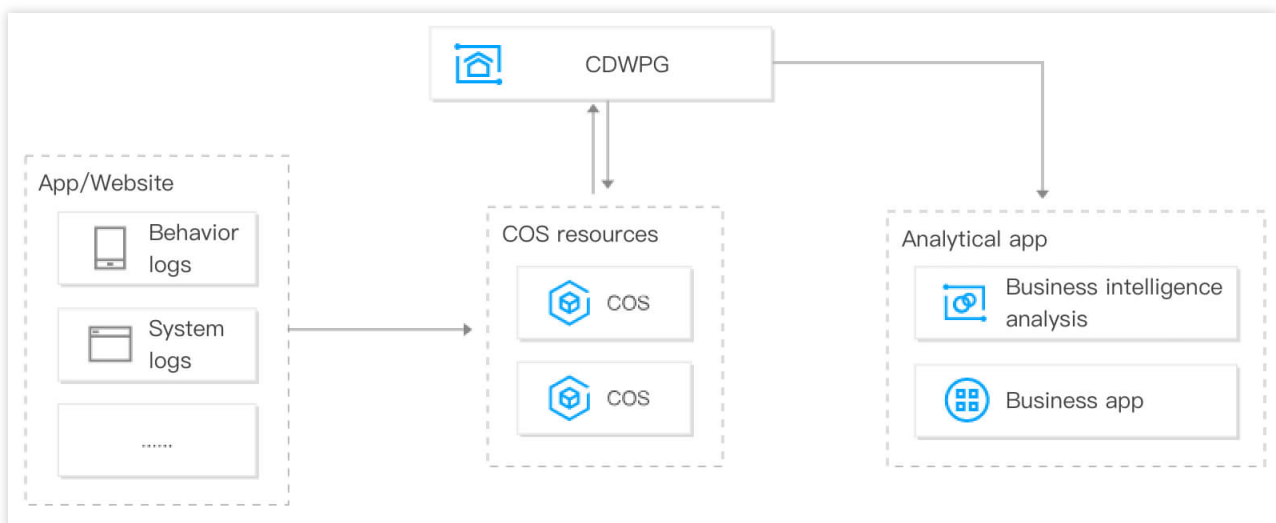
CDWPG allows you to import data from TencentDB, Oracle, and PostgreSQL through sync or ETL tools. It enables the analysis of heterogeneous data from multiple sources to assist with business decision making.



Massive Log Analysis

In internet finance, gaming, O2O, and other industries, CDWPG can analyze user behaviors, system logs, orders, and other structured or semi-structured data at the petabyte level economically and efficiently.

You can store massive amounts of data in COS without importing it into CDWPG in advance and then quickly analyze and apply such data by just writing SQL statements.



Real-Time Insights into User Behaviors

In the fields of internet and gaming, CDWPG analyzes user behaviors in real time to optimize operations strategies and enhance the efficiency of resource operations.

Oceanus can also help to filter, aggregate, and analyze users' real-time operation data, and then sync the results to CDWPG instantly. With CDWPG's powerful SQL analysis and efficient query capabilities, you can stay up to date with the user behavior trends and adjust operations strategies timely.



Node Specification

Last updated : 2024-02-02 15:15:26

CDWPG supports the following node specifications:

Storage-elastic: The underlying storage device is enhanced cloud disk with elastic scalability.

We recommend you select node specifications based on the storage and computing power required by your business.

Meanwhile, CDWPG supports COS-based external table extensions and achieves data compression on external storage via gzip. In this case, data that is not involved in real-time computing can be stored externally to further reduce the storage costs.

List of Storage-Elastic Models

Node specification

Specification Type	CPU (Cores)	Memory (GB)
4C16G	4	16
8C32G	8	32
16C64G	16	64

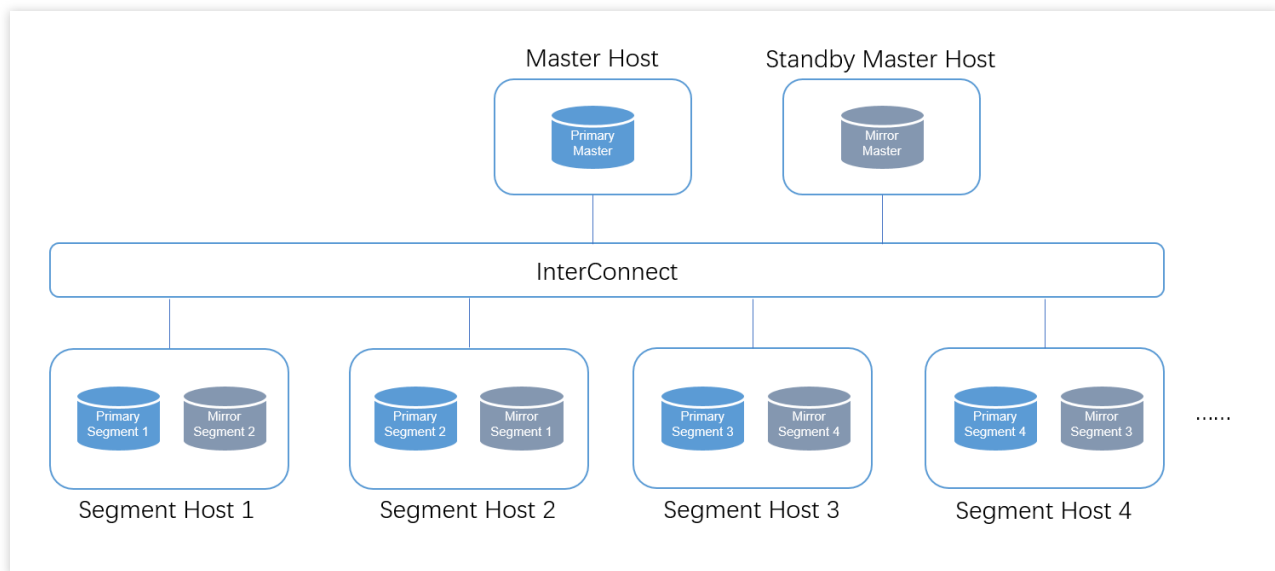
Node storage

Disk Type	Disk Capacity
Enhanced SSD	200 - 6000 GB
Premium Cloud Storage	200 - 6000 GB

Cluster Architecture

Last updated : 2024-02-02 15:15:26

CDWPG adopts the shared-nothing (SN) massively parallel processing (MPP) architecture as shown below, where the master and standby master nodes are deployed on two servers.



Master Node: A master node stores only data dictionaries but not business data. It is responsible for generating and distributing SQL execution plans to each segment node, interacting with clients, and verifying permissions.

Segment Node: A segment node stores business data and executes SQL statements distributed by the master node.

To ensure that each segment service is at the same performance level, each segment node server has the same resource configuration, and no model change will be made during scaling.

To ensure a high cluster availability, each segment node houses a primary segment of the current node and a mirror segment of another node that works as a standby node. When a segment node becomes unavailable, its mirror node will take over. When the master node becomes unavailable, the standby master node will take over to ensure the cluster availability. Changes will be automatically synced when the master node is restarted.

All the business data of the cluster is stored in the databases of all segment nodes, and each data table is sharded into each segment node. During analysis, all segment nodes work simultaneously to compute their part of data, greatly improving the efficiency.

CDWPG supports data backup and restoration. After data backup is enabled, each segment node will create a dump file containing data rebuild commands, and the master node will create several dumps files containing system targets, metadata files, DDL statements, and other information. During data restoration, all segments restore data from local backup files at the same time.