云数据库 GaussDB
8.102
特性描述

文档版本 01

发布日期 2024-04-30

华为云计算技术有限公司

商标声明


和其他华为商标均为华为技术有限公司的商标。
本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为云计算技术有限公司

地址： 贵州省贵安新区黔中大道交兴功路华为云数据中心 邮编：550029

网址： https://www.huaweicloud.com/

目 录

1 分布式版

1.1 面向应用开发的基本功能

1.1.1 支持标准 SQL

可获得性

本特性自V300R002C00版本开始引入。

特性简介

SQL是用于访问和处理数据库的标准计算机语言。SQL标准的定义分成核心特性以及可选特性，绝大部分的数据库都没有100%支撑SQL标准。

GaussDB数据库支持SQL:2011大部分的核心特性，同时还支持部分的可选特性，为使用者提供统一的SQL界面。

客户价值

标准SQL的引入为所有的数据库厂商提供统一的SQL界面，减少使用者的学习成本和应用程序的迁移代价。

特性描述

具体的特性列表请参见《开发者指南》中“SQL参考 > SQL语法”章节。

特性增强

V300R002C00引入分区表特性，扩展建表SQL语法。

V300R002C00引入外部表特性，扩展建表SQL语法。

503.1.0引入DATABASE LINK特性，仅在ORA兼容模式下可用，扩展标准SQL语法。

特性约束

无。

依赖关系

无。

1.1.2 支持标准开发接口

可获得性

本特性自V300R002C00版本开始引入。

本特性自GaussDB 503.0.0版本开始引入ECPG。

特性简介

支持ODBC 3.5及JDBC 4.0标准接口。

支持ECPG、ECPG用来处理嵌入式SQL-C程序。

客户价值

提供业界标准的ODBC及JDBC接口，保证用户业务快速迁移至GaussDB。

提供ECPG常用标准接口，保证用户嵌入式SQL-C业务快速迁移至GaussDB。

特性描述

目前支持标准的ODBC 3.5及JDBC 4.0接口，其中ODBC支持SUSE、Win32、Win64平台，JDBC无平台差异。

ECPG具体的特性列表请参见《开发者指南》中“应用程序开发教程 > > 基于ecpg开发”章节。

特性增强

增加JDBC负载均衡和JDBC对接第三方日志框架功能。JDBC负载均衡特性，可均衡CN上的业务压力；JDBC对接第三方日志框架功能可满足用户对日志管控的需求。

特性约束

无。

依赖关系

无。

1.1.3 事务支持

可获得性

本特性自V300R200C00版本开始支持。

特性简介

事务支持指的就是系统提供事务的能力，GaussDB支持全局事务的ACID，保证Shared-Nothing架构下全局事务的原子性、一致性、隔离性和持久性；通过两阶段提

交协议来保证事务在各个节点上状态的一致性，避免不同节点出现提交和回滚不一致的现象。

客户价值

事务支持及数据一致性保证用户数据能够被准确的计算并正确的存储。

特性描述

分布式事务支持即支持分布式系统下全局事务的ACID。

- A：Atomicity原子性

  整个事务中的所有操作，要么全部完成，要么全部不完成，不可能停滞在中间某个环节。

- C：Consistency一致性

  事务必须始终保持与系统处于一致的状态，不管在任何给定的时间并发事务的数量多少。

- I： Isolation隔离性

  隔离状态执行事务，使它们好像是系统在给定时间内执行的唯一操作。如果有两个事务运行在相同的时间内，执行相同的功能，事务的隔离性将确保每一事务在系统中认为只有该事务在使用系统。

- D：Durability持久性

  在事务完成以后，该事务对数据库所作的更改便持久的保存在数据库之中，并不会被回滚。

全局事务管理器GTM统一分配事务id，依据二阶段提交协议进行事务的提交，保证全局事务的一致性。事务提交先发送预提交命令给各节点，各节点先预提交，预提交都成功后再最后正式提交。

支持事务的默认隔离级别是读已提交。保证不会读到脏数据。

事务分为单语句事务和事务块，相关基础接口如下：

- Start transaction：事务开启。
- Commit：事务提交。
- Rollback：事务回滚。

另有Set transaction可设置隔离级别、读写模式或可推迟模式。详细语法请参见《开发者指南》中“SQL参考 > SQL语法 > SET TRANSACTION”章节。

特性约束

无。

特性增强

无。

依赖关系

无。

1.1.4 函数及存储过程支持

可获得性

本特性自V300R002C00版本开始引入。

特性简介

函数和存储过程是数据库中的一种重要对象，主要功能是将用户特定功能的SQL语句集进行封装，并方便调用。

客户价值

1. 允许客户模块化程序设计，对SQL语句集进行封装，方便调用。
2. 存储过程会进行编译缓存，可以提升用户执行SQL语句集的速度。
3. 系统管理员通过对执行某一存储过程的权限进行限制，能够实现对相应的数据的访问权限的限制，避免了非授权用户对数据的访问，保证了数据的安全。

特性描述

GaussDB支持SQL标准中的函数及存储过程，其中存储过程兼容了部分主流数据库存储过程的语法，增强了存储过程的易用性。

特性增强

无。

特性约束

无。

依赖关系

无。

1.1.5 支持 SQL hint

可获得性

本特性自V300R002C00版本开始引入。

特性简介

支持SQL Hint影响执行计划生成。

客户价值

提升SQL查询性能。

特性描述

Plan Hint为用户提供了直接影响执行计划生成的手段，用户可以通过指定join顺序，join、stream、scan方法，指定结果行数，指定重分布过程中的倾斜信息等多个手段来进行执行计划的调优，以提升查询的性能。

特性增强

无。

特性约束

无。

依赖关系

无。

1.1.6 数据导入导出

可获得性

本特性自V300R002C00版本开始引入。

特性简介

GaussDB提供的数据导入和导出服务。

客户价值

适于大批量的数据导入和导出，比如数据搬迁等场景。

特性描述

提供了多种数据导入、导出功能：

1. 基于CN组件的导入、导出。
2. 基于GDS组件的导入、导出，详情请参见**高速并行数据加载**。

特性增强

无。

特性约束

导入导出的相关约束参见《管理员指南》中“导入数据”和“导出数据”章节。

依赖关系

无。

1.1.7 COPY 接口支持容错机制

可获得性

本特性自V300R002C00版本开始引入。

特性简介

支持将COPY过程中的部分错误导入到指定的错误表中，并且保持COPY过程不被中断。

客户价值

提升COPY功能的可用性和易用性，提升对于源数据格式异常等常见错误的容忍性和鲁棒性。

特性描述

GaussDB提供用户封装好的COPY错误表创建函数，并允许用户在使用COPY From指令时指定容错选项，使得COPY From语句在执行过程中部分解析、数据格式、字符集等相关的报错不会报错中断事务，而是被记录至错误表中，使得在COPY From的目标文件即使有少量数据错误也可以完成入库操作。用户随后可以在错误表中对相关的错误进行定位以及进一步排查。

特性增强

无。

特性约束

无。

依赖关系

无。

1.2 高性能

1.2.1 分布式 CBO 优化器

可获得性

本特性自V300R002C00版本开始引入。

特性简介

GaussDB优化器是基于代价的优化 (Cost-Based Optimization，简称CBO)。

客户价值

GaussDB CBO优化器能够在众多分布式计划中依据代价选出最高效的执行计划，最大限度的满足客户业务要求。

特性描述

在CBO优化器模型下，数据库根据表的元组数、字段宽度、NULL记录比率、distinct值、MCV值、HB值等表的特征值，以及一定的代价计算模型，计算出每一个执行步骤的不同执行方式的输出元组数和执行代价(cost)，进而选出整体执行代价最小/首元组返回代价最小的执行方式进行执行。

特性增强

无。

特性约束

无。

依赖关系

无。

1.2.2 全并行分布式执行

可获得性

本特性自V300R002C00版本开始引入。

特性简介

GaussDB全并行分布式执行是基于MPP(Massively Parallel Processing)架构的分布式并行执行架构。

客户价值

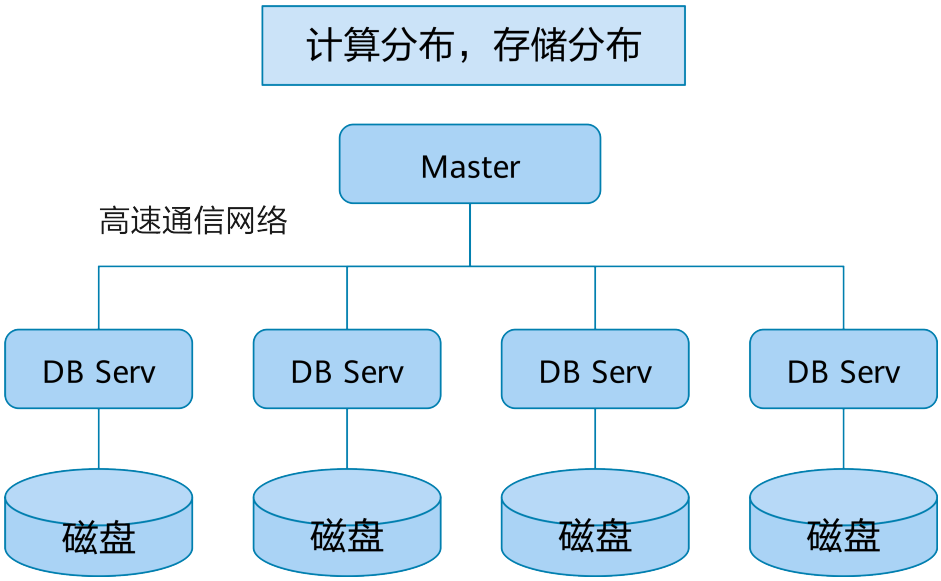
GaussDB 全并行分布式执行能够利用MPP架构的优势，充分利用集群硬件资源，提高数据库并发能力、查询性能，并可实现横向扩展。

特性描述

GaussDB采用MPP (Massively Parallel Processing) 架构，该架构的特点是：任务并行执行，数据分布式存储（本地化），分布式计算，私有资源（CPU、内存、磁盘、网络等），可横向扩展，Shared-Nothing。因此GaussDB天然具备大规模并行数据处理的能力。

在MPP架构的基础上，GaussDB又增加了Streaming流式计算框架，增强了所有计算节点之间的数据交换能力。目前数据库几乎所有的操作都可以实现全并行分布式执行，例如数据扫描、表连接、数据聚合等。且GaussDB支持最大256个DN分片的横向扩展能力，其计算能力相比于传统数据库有着巨大的优势。

图 1-1 MPP 架构图



特性增强

无。

特性约束

无。

依赖关系

无。

1.2.3 高性能事务处理

可获得性

本特性自V300R002C00版本开始支持分布式事务高性能处理能力。

特性简介

高性能事务处理指的是系统在保证事务ACID的前提下，最大限度处理并发业务的能力。其中包括单节点下事务性能优化，跨节点分布式事务的性能优化。

客户价值

在提供ACID的基本属性条件下，提供高吞吐高扩展的服务，满足客户对业务性能及事务隔离性的要求。

特性描述

在保证事务ACID的前提条件下，优化单节点事务、多节点事务的处理流程，减轻中心事务管理节点的压力，以支持高吞吐性及高扩展性。在GTM-Lite模式下，中心事务管

理节点GTM的压力进一步减轻，事务处理流程得到优化，消除GTM可支持的最大并发上限，事务处理性能及扩展性得到提升。

特性约束

无。

特性增强

V500R001C00版本提供GTM-Lite模式。

依赖关系

无。

1.2.4 高速并行数据加载

可获得性

本特性自V300R002C00版本开始引入。

特性简介

数据并行导入（加载）的核心思想是充分利用所有节点的计算能力和I/O能力以达到最大的导入速度。GaussDB的数据并行导入实现了对指定格式（支持CSV/TEXT格式）的外部数据高速、并行入库。

客户价值

高速、并行入库是和传统的使用INSERT语句逐条插入的方式相比较，入库性能得到提升。原理是，并行导入过程中：

- CN只负责任务的规划及下发，把数据导入的工作交给了DN，释放了CN的资源，使其有能力处理外部请求。
- 各个DN都参与数据导入的工作，充分利用各个设备的计算能力及网络带宽，提高数据导入的整体性能。

特性描述

以Hash分布策略为例介绍GaussDB的数据导入过程。数据并行导入的流程请参见图1-2。

图 1-2 数据并行导入

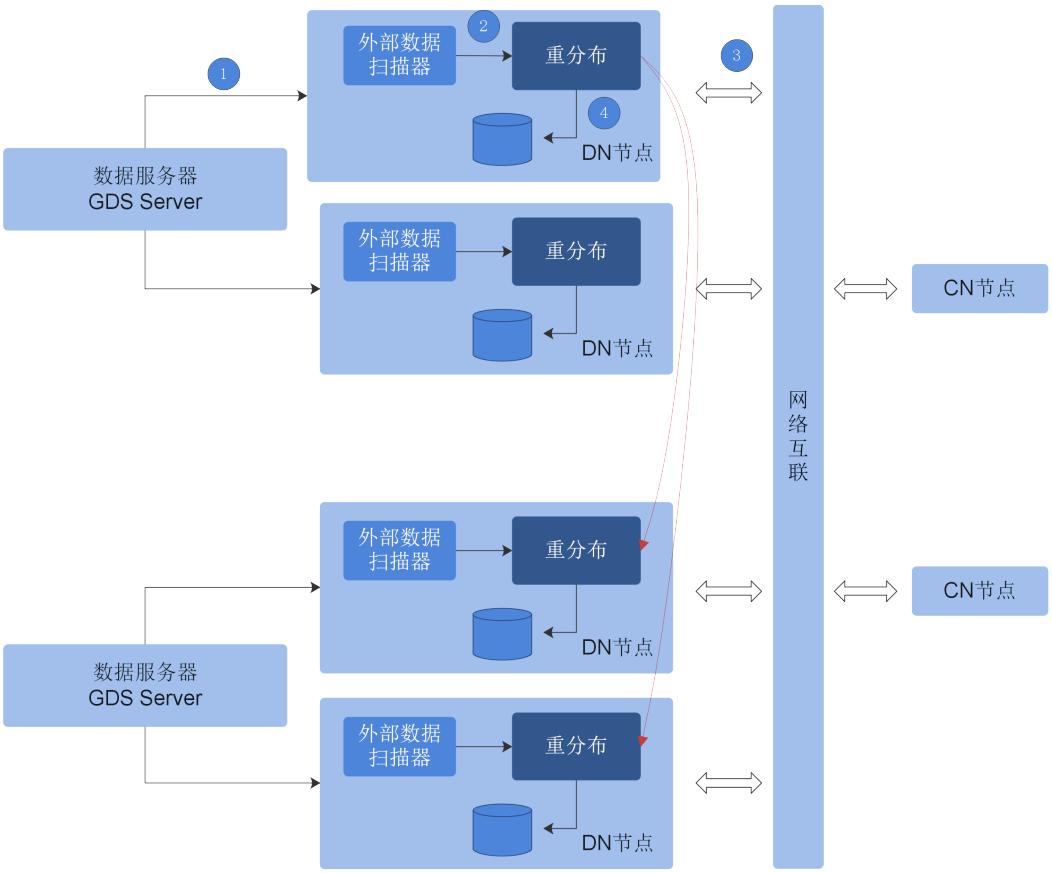


表 1-1 流程说明

流程	说明
创建Hash分布策略的表	业务应用在CREATE TABLE时预先设定Hash分布策略（指定表的某个属性作为分布字段）。
设定分区策略	应用程序在CREATE TABLE时还可以预先设定分区（指定表的一个属性作为分区字段），每个数据节点内部的每个Hash的数据都将按照设定的分区规则做相同的分区处理。
1	启动数据导入后，GDS将指定的数据文件分割成固定大小的数据块。
2	每个数据节点并行需从GDS下载这些数据块。
3 4	各个数据节点并行需处理数据块，从中解析出一条数据元组，每一个元组根据分布列计算出来的Hash值判断存储的物理位置： • 如果Hash在其他网络节点，则需要通过网络重分布到目标数据节点。 • 如果Hash在本地节点，则存储在本地数据节点。

流程	说明
数据写入分区	数据到达Hash所在的节点后还将根据分区逻辑写入对应的分区数据文件。 在数据写入分区表时，GaussDB还提供了Exchange（交换分区）的技术来提升写入性能。
GDS：全称Gauss Data Service，GDS服务用来管理数据源，可以在数据服务器上部署多个GDS服务来提升数据加载的性能。	

特性增强

V300R002C00版本支持非法字符容错。

特性约束

导入导出的相关约束参见《管理员指南》中“导入数据”和“导出数据”章节，建议用户使用个人账户进行GDS操作，不要使用root用户。

依赖关系

无。

1.2.5 支持 SMP 并行技术

可获得性

本特性自V300R002C00版本开始引入。

特性简介

GaussDB的SMP并行技术是一种利用计算机多核CPU架构来实现多线程并行计算，以充分利用CPU资源来提高查询性能的技术。

客户价值

SMP并行技术充分利用了系统多核的能力，来提高重查询的性能。

特性描述

SMP并行技术通过多线程多子任务并行执行的机制实现系统计算资源的充分高效使用。显然SMP多线程轻量执行的模式无疑能够解决MPP架构部署上的不足。

1. 首先，SMP并行执行是在线程级别上来完成任务的并行执行，理论上是可以使并行执行的子任务数达到物理服务器核数的上限。
2. 其次，SMP并行线程是在同一个进程内，可以直接通过内存进行数据交换，不需要占用网络连接与带宽。降低了限制MPP系统性能提升的网络因素的影响（本版本DN内线程间的数据交换还依赖于网络连接，后续版本再做优化）。
3. 最后，由于并行子任务启动后不需要附带其他后台工作线程，这样可以增加系统计算资源的有效利用率。

特性增强

无。

特性约束

不满足条件的索引扫描不支持并行执行，具体情况如下：

1. 第一次升级到支持索引并行扫描版本，升级未提交期间，不支持索引并行扫描；
2. 不支持hash、psort索引类型；
3. 不支持bitmapscan；
4. STREAM不可用，QUERY_DOP等于1，基表为复制表。

依赖关系

依赖于全并行分布式执行框架。

1.2.6 Ustore 存储引擎

可获得性

本特性自503.1.0版本开始引入。

特性简介

In-place Update（原地更新）行存储引擎，简称Ustore。相比于Append Update（追加更新）行存储引擎，Ustore存储引擎可以提高数据页面内更新的HOT UPDATE的垃圾回收效率，有效降低多次更新元组后存储空间占用的问题。

Ustore存储引擎采用NUMA-aware的Undo子系统设计，使得Undo子系统可以在多核平台上有效扩展；同时采用多版本索引技术，解决索引清理问题，有效提升了存储空间的回收复用效率。

Ustore存储引擎结合Undo空间，可以实现更高效、更全面的闪回查询和回收站机制，能快速回退人为“误操作”，为GaussDB提供了更丰富的企业级功能。Ustore基于Undo回滚段技术、页面并行回放技术、多版本索引技术、xLog无锁落盘技术等实现了高可用高可靠的行存储引擎。

Ustore完全支持ACID特性：

- 原子性（Atomicity）：原子事务是一系列不可分割的数据库操作。在事务完成（分别提交或中止）之后，这些操作要么全部发生，要么全部不发生。
- 一致性（Consistency）：事务结束后，数据库处于一致状态，保留数据完整性。
- 隔离性（Isolation）：事务之间不能相互干扰。Ustore支持读已提交隔离级别，事务只能读到已提交的数据而不会读到未提交的数据，这是缺省值。
- 持久性（Durability）：即使发生崩溃和失败，成功完成（提交）的事务效果持久保存。

客户价值

- 针对OLTP场景，实现Inplace-update，利用Undo实现新旧版本分离存储；降低类似于AStore存储引擎由于频繁更新或闪回功能开启导致的数据页空间膨胀，以及相应引起的索引空间膨胀。

- 通过在DML操作过程中执行动态页面清理，去除VACUUM依赖，减少由于异步数据清理产生的大量读写IO。通过Undo子系统，实现事务级的空间管控，旧版本集中回收。

特性描述

Ustore的关键特性如下：

- In-place Update存储模式：Ustore存储引擎将最新版本的“有效数据”和历史版本的“垃圾数据”分离存储。将最新版本的“有效数据”存储在数据页面上，并单独开辟一段Undo空间，用于统一管理历史版本的“垃圾数据”，因此数据空间不会由于频繁更新而膨胀，“垃圾数据”集中回收效率更高。
- 回滚段设计：回滚段简称Undo，负责历史记录的插入、查询以及Undo空间的分配与释放等操作，北向对接Ustore，南向对接Buffer Pool。基于历史版本直接进行回收，实现了自治式的空间管理机制，减少了I/O时的性能抖动。同时实现了多个后台线程的并发访问，降低并发业务冲突竞争，从而提高性能。
- 基于页面的并行回放技术：Ustore利用多线程技术加速日志回放，Startup线程从磁盘中读取xLog日志，把组装好的xLog记录通过Dispatcher线程分配给多个回放线程进行回放。Dispatcher线程基于页面号进行xLog记录的分配，分配更加均匀，各个回放线程并行执行回放，提高了回放的速度。
- 闪回：数据库恢复技术的一环，能够使得DBA有选择性的高效撤销一个已提交事务的影响，将数据从人为的不正确的操作中进行恢复。在采用闪回技术之前，只能通过备份恢复、PITR等手段找回已提交的数据库修改，恢复时长需要数分钟甚至数小时。采用闪回技术后，通过闪回Drop和闪回Truncate恢复已提交的数据库Drop/Truncate的数据，只需要秒级，而且恢复时间和数据库大小无关。Ustore支持闪回表、闪回查询、闪回truncate、闪回drop，而且适用于分区表。
- UBtree：与原有的Btree索引相比，索引页面增加了事务信息，使得UBtree索引具备MVCC能力以及独立过期旧版本回收能力。In-place Update引擎支持UBtree 索引，UBtree也是In-place Update引擎的默认索引类型。支持并行创建索引、索引空间管理算法优化，索引空间进一步压缩。

特性增强

无。

特性约束

Ustore设计几乎能够覆盖SQL和未来特性集，支持大多数的SQL标准，也支持常见的数据库特性。下面介绍Ustore的各种约束。

Ustore不支持以下特性：

- 不支持串行化隔离级别。
- 对于支持row movement的分区表，不支持并发更新或删除同一行操作。
- 不支持的DDL功能：在线vacuum full/cluster、在线alter table（除新增字段、重命名等无需全量重写数据的操作外）、table sampling。
- 不支持BRIN索引。
- 不支持批量访存接口。不支持rowid语义。
- 不支持单事务块或语句中既包含Astore表又包含Ustore表。
- 数据表与回滚段要同为页式。

依赖关系

无。

1.2.7 自适应压缩

可获得性

本特性自V300R002C00版本开始引入。

特性简介

数据压缩是当前数据库采用的主要技术。数据类型不同，适用于它的压缩算法不同。对于相同类型的数据，其数据特征不同，采用不同的压缩算法达到的效果也不相同。自适应压缩正是从数据类型和数据特征出发，采用相应的压缩算法，实现了良好的压缩比、快速的入库性能以及良好的查询性能。

客户价值

数据入库和频繁的海量数据查询是用户的主要应用场景。 在数据入库场景中，自适应压缩可以大幅度地减少数据量，成倍提高IO操作效率，将数据簇集存储，从而获得快速的入库性能。当用户进行数据查询时，少量的IO操作和快速的数据解压可以加快数据获取的速率，从而在更短的时间内得到查询结果。

特性描述

目前，数据库已实现了RLE、DELTA、BYTEPACK/BITPACK、LZ4、ZLIB、LOCAL DICTIONARY等多种压缩算法。数据库支持的数据类型与压缩算法的映射关系如下表所示。

-	RLE	DELTA	BITPACK/ BYTEPACK	LZ4	ZLIB	LOCAL DICTIONARY
Smallint/Int/Bigint/Oid Decimal/Real/Double Money/Time/Date/ Timestamp	√	√	√	√	√	-
Tinterval/Interval/Time with time zone/	-	-	-	-	√	-
Numeric/Char/Varchar/ Text/Nvarchar2 以及其他支持数据类型	√	√	√	√	√	√

特性增强

支持对压缩算法进行不同压缩水平的调整。

特性约束

仅支持列存。

依赖关系

开源压缩软件LZ4/ZLIB。

1.2.8 分区

可获得性

本特性自V300R002C00版本开始引入。

特性简介

在GaussDB分布式系统中，数据分区是在一个节点内部按照用户指定的策略对数据做进一步的水平分表，将表按照指定范围划分为多个数据互不重叠的部分。

客户价值

对于大多数用户使用场景，分区表和普通表相比具有以下优点：

- 改善查询性能：对分区对象的查询可以仅搜索自己关心的分区，提高检索效率。
- 增强可用性：如果分区表的某个分区出现故障，表在其他分区的数据仍然可用。

特性描述

目前GaussDB数据库支持的分区表为范围分区表、列表分区表和哈希分区表：

- 范围分区表：将数据基于范围映射到每一个分区，这个范围是由创建分区表时指定的分区键决定的。这种分区方式是最为常用的。

  范围分区功能，即根据表的一列或者多列，将要插入表的记录分为若干个范围（这些范围在不同的分区里没有重叠），然后为每个范围创建一个分区，用来存储相应的数据。用户在CREATE TABLE时增加PARTITION参数，即表示针对此表应用数据分区功能。
- 列表分区表：将数据中包含的键值分别存储在不同的分区中，依次将数据映射到每一个分区，分区中包含的键值由创建分区表时指定。
- 哈希分区表：将数据根据内部哈希算法依次映射到每一个分区中，包含的分区个数由创建分区表时指定。

用户可以在实际使用中根据需要调整建表时的分区键，使每次查询结果尽可能存储在相同或者最少的分区内（称为“分区剪枝”），通过获取连续I/O大幅度提升查询性能。

实际业务中，时间经常被作为查询对象的过滤条件。因此，用户可考虑选择时间列为分区键，键值范围可根据总数据量、一次查询数据量调整。

特性增强

支持范围分区表，列表分区表的增加，删除、切割、合并、清空、交换功能。

支持哈希分区表清空、交换功能。

特性约束

无。

依赖关系

无。

1.2.9 高级分析函数支持

可获得性

本特性自V300R002C00版本开始引入。

特性简介

无。

客户价值

提供窗口函数来进行数据高级分析处理。窗口函数将一个表中的数据进行预先分组，每一行属于一个特定的组，然后在这个组上进行一系列的关联分析计算。这样可以挖掘出每一个元组在这个集合里的一些属性和与其他元组的关联信息。

特性描述

简单举例说明窗口分析功能：

分析某一部门内每个人的薪水和部门平均薪水的对比。

```
SELECT depname, empno, salary, avg(salary) OVER (PARTITION BY depname) FROM empsalary;
depname | empno | salary | avg
-----------+-------+--------+-----------------------
develop | 11 | 5200 | 5020.0000000000000000
develop | 7 | 4200 | 5020.0000000000000000
develop | 9 | 4500 | 5020.0000000000000000
develop | 8 | 6000 | 5020.0000000000000000
develop | 10 | 5200 | 5020.0000000000000000
personnel | 5 | 3500 | 3700.0000000000000000
personnel | 2 | 3900 | 3700.0000000000000000
sales | 3 | 4800 | 4866.6666666666666667
sales | 1 | 5000 | 4866.6666666666666667
sales | 4 | 4800 | 4866.6666666666666667
(10 rows)
```

可以看到，通过这个avg(salary) OVER (PARTITION BY depname)分析函数，每一个人的薪水和部门的平均薪水很容易计算出来。

目前，系统支持row_number(), rank(), dense_rank(), percent_rank(), cume_dist(), ntile(),lag(), lead(),first_value(), last_value(), nth_value()分析函数。具体的函数用法和语句请参见《开发者指南》中“SQL参考 > 函数和操作符 > 窗口函数”章节。

特性增强

无。

特性约束

无。

依赖关系

无。

1.2.10 数据倾斜优化技术

可获得性

本特性自V300R002C00版本开始引入。

特性简介

数据倾斜问题是分布式架构的重要难题，特别是在运行时产生的数据倾斜。GaussDB针对数据倾斜问题给出了完整的解决方案，包括存储倾斜和计算倾斜两大问题。

客户价值

解决了分布式下的数据倾斜问题，提高了集群的横向扩展能力。

特性描述

存储倾斜和计算倾斜的优化如下：

- 针对存储层的优化，GaussDB提供了丰富的视图用于查看数据存储的倾斜情况。
- 针对计算倾斜，GaussDB提出了RLBT（ Runtime Load Balance Technology ），利用统计信息或者hint的方式来识别可能出现的倾斜值，然后对倾斜部分数据和非倾斜部分数据分别进行处理。例如在join时，对非倾斜数据按照hash进行重新分布，对于倾斜数据按照round robin进行重新分布。

特性增强

无。

特性约束

无。

依赖关系

依赖全并行分布式执行。

1.2.11 SQL by pass

可获得性

本特性自V300R002C00版本开始引入。

特性简介

通过对TP场景典型查询的定制化执行方案来提高查询性能。

客户价值

提升TP类查询的性能。

特性描述

在典型的OLTP场景中，简单查询占了很大一部分比例。这种查询的特征是只涉及单表和简单表达式的查询，因此为了加速这类查询，提出了SQL-BY-PASS框架，在parse层对这类查询做简单的模式判别后，进入到特殊的执行路径里，跳过经典的执行器执行框架，包括算子的初始化与执行、表达式与投影等经典框架，重写一套简洁的执行路径，并且直接调用存储接口，这样可以大大加速简单查询的执行速度。

特性增强

无。

特性约束

无。

依赖关系

无。

1.2.12 支持 HyperLogLog

可获得性

本特性自V300R002C00版本开始引入。

特性简介

通过使用HyperLogLog相关函数，计算唯一值个数Count(Distinct)，提升性能。

客户价值

提升AP/TP类查询的性能。

特性描述

HLL（HyperLoglog）是统计数据集中唯一值个数的高效近似算法。它有着计算速度快，节省空间的特点，不需要直接存储集合本身，而是存储一种名为HLL的数据结构。每当有新数据加入进行统计时，只需要把数据经过哈希计算并插入到HLL中，最后根据HLL就可以得到结果。

HLL在计算速度和所占存储空间上都占优势。在时间复杂度上，Sort算法需要排序至少O(nlogn)的时间，虽说Hash算法和HLL一样扫描一次全表O(n)的时间就可以得出结果，但是存储空间上，Sort算法和Hash算法都需要先把原始数据存起来再进行统计，

会导致存储空间消耗巨大。而对HLL来说不需要存原始数据，只需要维护HLL数据结构，所以占用空间始终是1280字节常数级别。

GaussDB采用分布式HLL架构。DN承担计算HLL的任务然后在CN汇总，避免了CN计算瓶颈。

特性增强

无。

特性约束

无。

依赖关系

无。

1.2.13 NUMA 架构优化

可获得性

本特性自V500R001C00版本开始引入。

特性简介

NUMA架构优化，主要面向ARM处理器架构特点、ARMv8指令集等，进行相应的系统优化，涉及到从操作系统、软件架构、锁并发、日志、原子操作、Cache访问等一系列的多层次优化，从而大幅提升了GaussDB数据库在ARM平台上的处理性能。

客户价值

数据库的处理性能，如每分钟处理交易量（Transaction Per Minute），是数据库竞争力的关键性能指标，在同等硬件成本的条件下，数据库能提供的处理性能越高，那么就可以提供给用户更多的业务处理能力，从而降低客户的使用成本。

特性描述

- GaussDB根据ARM处理器的多核NUMA架构特点，进行了一系列的架构相关优化。一方面尽量减少跨核内存访问的时延问题，另一方面充分发挥ARM多核算力优势，所提供的关键技术包括重做日志批量插入、热点数据NUMA分布、CLog分区等，大幅提升TP系统的处理性能。
- GaussDB基于鲲鹏芯片所使用的ARMv8.1架构，利用LSE扩展指令集实现高效的原子操作，有效提升CPU利用率，从而提升多线程间同步性能、xLog写入性能等。
- GaussDB基于鲲鹏芯片提供的更宽的L3缓存cacheline，针对热点数据访问进行优化，有效提高缓存访问命中率，降低Cache缓存一致性维护开销，大幅提升系统整体的数据访问性能。
- 鲲鹏920，2P服务器（64cores*2，内存768 GB），网络10 GE，I/O为4块NVME PCIE SSD时，TPCC为1000warehouse，性能是150万 tpmC。

特性增强

- 支持重做日志批量插入，分区CLog，提升ARM平台下的数据库处理性能。
- 支持LSE扩展指令集的原子操作，提升多线程同步性能。

特性约束

无。

依赖关系

无。

1.2.14 物化视图

可获得性

本特性自V500R001C10版本开始引入。

特性简介

物化视图实际上就是一种特殊的物理表，物化视图是相对普通视图而言的。普通视图是虚拟表，应用的局限性较大，任何对视图的查询实际上都是转换为对SQL语句的查询，性能并没有实际上提高。而物化视图实际上就是存储SQL所执行语句的结果，起到缓存的效果。

客户价值

使用物化视图功能提升查询效率。

特性描述

支持全量物化视图和增量物化视图。全量物化视图只支持全量更新；增量物化视图同时还支持增量更新功能，用户可通过执行语句把新增数据刷新到物化视图中。

特性增强

无。

特性约束

全量物化视图支持的场景与CREATE TABLE AS语句基本一致，增量物化视图支持基表简单过滤查询和UNION ALL语句。

依赖关系

无。

1.2.15 分布式全局二级索引

可获得性

本特性自503.1.0版本开始引入。

特性简介

在指定的表上创建全局二级索引（Global Secondary Index，简称GSI）。

客户价值

全局二级索引允许用户定义与基表分布不一致的索引，从而实现非基表分布键点查/点查和范围查询性能提升，去除UNIQUE/PRIMARY KEY需包含基表分布键的约束。

特性描述

- 对于某一基表，定义按照用户指定方式分布的若干个GSI，GSI上除了索引键以及用户指定的CONTAINING列之外，同时记录数据元组对应的xc_node_hash（基表分布键的hash值）、tableoid（当基表为分区表时）、ctid、xmin、xmax等。
- 当用户对基表执行IUD（INSERT、UPDATE、DELETE）操作时，通过RETURNING对GSI进行IUD。
- 在查询计划生成阶段通过判断谓词条件中属性的组合，生成对应的GSI Scan（Index Only Scan）算子来对GSI进行查询。

特性原理

通过分布式计划构建、优化器模型构建、执行器新算子实现、以及分布式GBTree构建等，实现分布式全局二级索引功能，具体如下：

- 解析层：

  对于IUD与SELECT，在语义解析阶段设置hasGSI，标记当前Query Block是否含有GSI，即为Query结构体添加hasGSI成员变量，用于标记当前Query Block中的RangeTblEntry是否包含GSI。LP（Light Proxy）与FQS（Fast Query Shipping）需要根据查询树中的关系（表，RangeTblEntry）、属性（列，Var）等信息来判断一个查询是否适用。
- 优化器模块：
  - 通过GSI Hint来支持符合规则的查询走GSI Scan。
  - 对于涉及GSI的查询，借助Index Only Scan实现GSI Scan。其中，选择率、基数估计、代价模型、路径生成、计划生成模块在复用Index Only Scan逻辑的基础上，增加分布信息。
  - 对于涉及GSI的IUD，构建新的分布式执行计划。
- 执行器模块：
  - 对于IUD，新增GSI IUD算子。
  - 对于涉及GSI的查询，GSI Scan借助Index Only Scan实现，同时支持Bypass。
- 存储模块：
  - 创建GSI索引分为两个阶段：生成GSI元数据、插入索引数据。
  - GTM-LITE：在GTM-LITE模式下，各个DN维护自己的一套xid，因此在DN内部可通过复用UBtree可见性判断的逻辑，以保证对涉及GSI操作的正确性。目前版本只支持在GTM-LITE模式下创建GSI，暂不支持其他GTM模式。
  - UBTree支持GBTree：由于UBTree是一种local索引，并未体现DN信息。对于local索引，由ctid可唯一定位到堆表上的元组。但是对于GSI，由于涉及到跨DN，因此，兼顾到扩容、主备切换等场景，引入逻辑值xc_node_hash（基表分布键的hash值）来确定DN，进一步的，可通过ctid + xc_node_hash来唯一对应至堆表上的元组，从而使得UBTree支持GBTree。

  - Analyze：

    新增语法：analyze global index index_name for table table_name，正常analyze和analyze table时会跳过gsi。对GSI执行analyze时，对于DN，relpages可以直接得到，reltuples会通过采样的方式估计得到。同时，CN会读取DN的结果用于CN侧GSI统计信息的更新。

- 升级：

  添加升级脚本，以适配索引相关系统视图。修改pg_attribute系统表，插入xc_node_hash的信息。

- 工具：

  适配gs_dump工具，使其能够正常返回索引定义；适配gs_redis工具，在基表切换node group时，同步切换所有其GSI的node group分布信息；使用gds、copy导入建有GSI的基表时，将同步GSI数据。

特性增强

503.2.0版本增强：

- 创建性能优化：提供create_gsi_opt参数，利用GSI BUILD和STREAM算子提升创建GSI的性能。
- UPDATE/DELETE性能优化：将索引键置入WHERE子句中，提升GSI UPDATE/DELETE操作性能。
- 支持基表为USTORE。
- 支持SQLPATCH、STORAGE PARAMETER、CLUSTER、REINDEX TABLE 、VACUUM FULL、CN端获取DN端计划，ALTER PARTITION功能增强。
- 支持部分索引和表达式索引。
- 支持在线创建。
- 支持并行创建。

505.0.0版本增强：

- GSI支持回表查询。
- 支持使用gsitable hint指导生成GSI回表查询计划。
- COPY、GDS导入建有GSI的基表时，将同步GSI数据。

505.1.0版本增强：

- 支持插入GSI走STREAM模式。
- 支持COPY、GDS增量导入。
- 支持对HASHBUCKET表、段页式表创建GSI。
- 支持MERGE INTO功能

特性约束

- 同基表约束，GSI的分布列不支持更新（UPDATE、MERGE INTO）操作。
- 只支持GTM-LITE模式下创建GSI，不支持其他GTM模式，在其他模式下创建GSI会报错。
- Astore不支持创建GSI以外的UBTree，不支持对GSI创建分区。
- 创建与基表分布一致的GSI时，执行时会报错。

- 只支持对Ustore表执行CREATE GSI CONCURRENTLY，对Astore表执行CREATE GSI CONCURRENTLY会报语法错误；不支持表达式索引和部分索引CREATE GSI CONCURRENTLY，会报语法错误。不支持在线重建GSI。
- 支持对基表为hash分布的行存Astore表、Ustore表、分区表、HASHBUCKET、段页式、创建hash分布的GSI，不支持基表为复制表、list/range分布等，对于GSI本身不支持hash分布以外的分布。
- 不支持对基表列名或者ctid、xc_node_hash、xmin、xmax、tableoid（当基表为分区表时）、tablebucketid（当基表为HASHBUCKET表时）增加_new$$、_NEW$$后与自身列名重复的基表创建GSI。
- 如果在执行VACUUM FULL、CLUSTER或者REINDEX操作时中断，表上的GSI可能会变为UNUSABLE状态，此时查询语句走GSI会报错，建议执行REINDEX INDEX重建GSI。
- 当基表为分区非HASHBUCKET表时，GSI最多支持27列；当基表为HASHBUCKET非分区表时，GSI最多支持27列；当基表为HASHBUCKET分区表时，GSI最多支持26列；当基表为非分区非HASHBUCKET表最多支持28列（包括索引键和分布键）。
- 对于创建GSI（非在线）、重建GSI，以及涉及重建GSI的操作：比如分区表分区操作（包括DROP、TRUNCATE、MERGE、SPLIT、EXCHANGE PARTITION）指定UPDATE DISTRIBUTED GLOBAL INDEX，ALTER TABLE涉及重建数据的操作，HASHBUCKET表ALTER SET TABLESPACE操作、MOVE PARTITION操作，建议开启STREAM模式，以达到最优性能。（其中，STREAM模式指设置enable_stream_operator参数为ON，并设置create_gsi_opt参数置为build）
- 不支持UPSERT，建有GSI的基表上不支持IUD returning功能。
- 在对建有GSI的基表执行COPY、GDS数据导入时，需要开启enable_stream_operator参数，以达到最优数据导入性能。
- 当前会使GSI失效的操作：REINDEX数据库级、CLUSTER数据库级/分区级、ALTER TABLE PARTITION（DROP、TRUNCATE、MERGE、SPLIT、EXCHANGE PARTITION未指定UPDATE DISTRIBUTED GLOBAL INDEX将失效分区表上的所有GSI，其中，EXCHANGE PARTITITON未指定UPDATE DISTRIBUTED GLOBAL INDEX将同步失效普通表上的所有GSI）。
- 回表基于STREAM，继承STREAM相关约束。考虑到STREAM通信时延，当选择率过低或者谓词命中行数较少时，性能非最优，不建议使用回表计划，建议与普通索引配合使用。
- 对于Insert into select批量插入场景，建议打开enable_stream_operator，插入执行STREAM计划（当基表为段页式表、HASHBUCKET表，以及防篡改表时，不会执行STREAM计划，仍然走PGXC计划），如果关闭enable_stream_operator，执行计划采用回到CN的方式，性能较差（类比503.1.0版本创建GSI性能）。
- 对于INSERT、UPDATE、DELETE，执行计划采用分布式执行计划，会有性能损失，其中，UPDATE/DELETE批量场景，执行计划采用回到CN的方式，性能较差。
- GSI支持表达式索引，但存在以下约束：
  - 同基表约束，不支持分布键包含表达式（且无法创建索引列仅包含表达式的GSI，因为此时分布键必定为表达式），创建时会报语法错误。
  - 同普通索引约束，不支持CONTAINING列中包含表达式，创建时会报语法错误。
  - 若表上存在以"expr"为前缀的列名，不支持创建带有表达式的GSI，创建时会报语法错误。

依赖关系

无。

1.2.16 数据生命周期管理-OLTP 表压缩

可获得性

本特性自505.0.0版本开始引入。

特性简介

基于冷热分离的行存压缩。

客户价值

OLTP表压缩是GaussDB高级压缩中的一个特性，基于全新的压缩算法、细粒度的自动冷热判定和支持块内压缩等技术创新，可以在提供合理压缩率的同时大幅度降低对业务的影响、增加后台调度、增加查询Job执行状态以及节约空间，能够在支持关键在线业务的容量控制中发挥重要价值。

特性描述

用户可给数据对象指定ILM策略，策略分三部分：动作、条件、范围，本期仅支持行压缩动作、XX天未修改条件、行范围。指定策略的表会在后台定时调度、评估表中的每一行是否满足条件，若满足条件则执行行压缩动作。

特性增强

无。

特性约束

- 不支持系统表、内存表、全局临时表、本地临时表和序列表。
- 仅在ORA兼容模式与PG模式下有效。
- Ustore不支持编解码，压缩率小于2:1。
- 普通表开启压缩时，扩容空间预留需按照解压后的大小评估。
- HashBucket表不支持DBE_HEAT_MAP.ROW_HEAT_MAP和DBE_COMPRESSION.GET_COMPRESSION_TYPE。
- 扩容期间不支持压缩调度。
- 扩容前请确认当前是否有正在执行的压缩任务，如果有的话，要么等待压缩任务结束，要么执行DBE_ILM.STOP_ILM或DBE_ILM_ADMIN.DISABLE_ILM停掉，扩容完成后再执行

  DBE_ILM_ADMIN.ENABLE_ILM开启。

依赖关系

无。

1.2.17 ADIO 特性与去双写

可获得性

本特性自505.1.0版本开始引入。

特性简介

ADIO异步刷页使用异步直接I/O模式完成数据库的刷页操作。在多数场景下，主机不再记录双写文件。

客户价值

随着客户数据库中数据量的不断积累，客户数据库中存储的数据量相对于机器物理内存的比值将会越来越大。此时，刷页操作效率以及带来的I/O操作会限制数据库性能的发挥。因此，通过优化刷页模式和I/O操作来提升大容量场景下的性能便尤为重要。

特性描述

ADIO异步刷页：大容量场景下，I/O资源比较紧俏。当前数据库BIO模式对于整体I/O资源利用不充分，容易导致刷页速度落后于消耗页面的速度，导致缓冲区页面消耗完时产生性能震荡，进而影响性能。本特性通过ADIO(异步直接I/O模式)充分利用IO资源，从而提升整体数据库的性能。同时，提供从BIO模式到ADIO模式的在线切换（参见《管理员指南》中“配置运行参数 > GUC参数说明”章节的“enable_adio_function”参数说明），使用户可以在不影响业务的情况下切换到ADIO模式。

去双写：增量checkpoint开启后，由于没有full page write保护，因此采用双写文件方案（即写两次）来防止半写。拉起始遇到半写页面，便能通过双写文件方案恢复。但是写两次会导致整体I/O量多一倍，而在大容量场景下IO资源很紧俏，因此去双写可以有效降低I/O使用量。当开启去双写功能时，若所有备机都处于正常状态，则主机会停止记录双写文件。若主机由于宕机产生半写页面，则通过备机页面进行修复。（注意：由于当前版本存在特性约束，分布式模式下不支持去双写。）

特性增强

无。

特性约束

在当前版本，去双写与数据修复特性存在依赖关系，而数据修复特性不支持修复hashbucket，所以分布式模式下不支持去双写。除此之外，由于该版本数据修复接口的timeout为固定值，如果想让半写页面及时恢复，需要开启流控机制，以防止因主备之间页面版本差距过大而导致的频繁修复失败问题。

依赖关系

去双写依赖数据修复特性中提供的主备间相互修复的接口。

1.3 扩展性

1.3.1 支持在线扩容

可获得性

本特性自版本V300R002C00开始引入。

hashbucket表在线扩容特性自版本V500R002C00开始正式引入。

特性简介

在线扩容过程的本质是集群内众多本地表的NodeGroup之间的数据搬迁，扩容过程指较小的NodeGroup向较大的NodeGroup进行数据搬迁，反之则看成是缩容过程。对于每个扩容重分布的表而言可以看成是这3个过程：

1. 基线数据扩容重分布。
2. 增量数据重分布。
3. 表切换。

基于hashbucket表的在线扩容的扩容整体流程主要包含3个步骤：

1. 基线数据搬迁：生成扩容搬迁计划，根据搬迁计划中对涉及的库中的bucket文件进行跨节点文件搬迁。包含库级别的数据文件和实例级别的事务日志。
2. bucket日志流追增：识别bucket扩容过程中的增量修改，将对应的日志发送到目的节点并在目的端进行增量修改的日志回放。
3. bucket元数据切换：当bucket日志流追增完成后，原节点和目的节点的bucket数据达到一致状态，可以对原节点的bucket进行下线删除，对目的节点的bucket进行上线操作，同时修改CN上的bucket map映射使新的业务能够路由到正确的DN节点。

客户价值

随着数据量的增加，GaussDB需要进行扩容，但扩容的完成需要一定的时间窗口，可能会影响业务。业务中断将给用户带来重大损失，因此需要在线扩容的特性，使得用户业务在扩容重分布过程中无需中断，平滑过渡。

特性描述

普通表通过创建临时表和记录数据删除的增量表，在扩容过程中通过更新此轮数据重分布的范围来实现在线扩容。

扩容前提条件

- 集群状态必须为Normal或Degraded状态，集群Degraded状态只是实例异常引起的集群状态异常，不支持VM异常引起的集群状态异常（VM异常指非软件类异常，例如：磁盘故障、通信故障），请确保实例状态异常节点的网络正常。
- 实例状态：CN全部正常，GTM、CMS至少有主节点存活，ETCD实例状态满足多数正常。DN实例状态有以下三种情况：两副本集群单个分片内至少主DN实例正常（hashbucket扩容要求主备DN都正常）；两AZ集群DN实例单个分片至少存活一主一备；除了以上两种之外的其他情况DN实例单个分片异常数量小于一半。
- 如当前集群有异常但是不满足以上约束条件，请参见《工具参考》中的“服务端工具 > gs_replace”章节内容进行修复。

- 集群扩容要求整个集群没有被锁定，集群配置文件的配置信息正确并且和当前集群配置一致。
- 已按照扩容的集群配置文件执行过前置脚本gs_preinstall。
- 新增主机和现有集群之间通信已经建立，网络正常。
- 扩容前需保证集群所有CN可连接，在集群用户下执行gs_check -i CheckDBConnection命令检查CN是否可连接。
- 扩容前需做ANALYZE（保证扩容进度准确性时需要执行）。
- 重分布资源管控对象redisuser、redisrespool、redisclass、redisgrp不能被占用。
- 重分布过程中不能将其他用户绑定到重分布专用的资源池redisrespool上。不能用gs_cgroup -M等操作破坏当前Cgroups配置。若用户在Cgroups出现问题时使用gs_cgroup --revert恢复默认配置，则需重启集群才能保证新建立的redisclass和redisgrp生效。
- 如果当前集群有对hashbucket的表开启OLTP表压缩，需要等待全部压缩任务结束，并关闭自动调度任务，扩容期间不支持压缩调度（详情请参见《特性描述》“高性能 > 数据生命周期管理-OLTP表压缩 > 特性约束”）。
- 重分布前校验无效索引残留，需遍历库执行select count(*) from pg_index where indisvalid = 'f' and indisusable = 't' ，如果返回值大于1，用户需要手动效验失效索引，即执行ALTER INDEX index_name UNUSABLE。

表特性增强

- 扩容过程支持数据持续入库，业务不中断。
- 支持扩容过程中故障重入，支持表粒度的断点续传。
- 支持多表并行扩容，扩容性能高。
- 重分布过程中支持数据持续入库，业务不中断。

  重分布过程按database，表按顺序进行重分布（-j指定多线程并发时可以多个表同时进行）。重分布过程中区分外表和普通表。由于外表不涉及数据迁移，只涉及节点信息的更新，所有外表放在一个事务里，先进行重分布。重分布执行过程中，需要采用execute direct的形式在其他节点（其他CN、DN）上更新pgxc_class、pg_class等系统表里的元信息，主要包括表的重分布状态，交换relfilenode，变更nodegroup等相关信息。对于不涉及正在重分布表的DDL、DML等相关操作不受影响，用户可并发执行。对于正在重分布的表用户可以并发执行select、insert、insert/select操作。同时重分布的进度状态查询和表顺序设置都可通过直接连接CN查询。
- 扩容过程中，数据库支持DDL和DCL操作，如果用户并发事务块中包含DDL操作，用户DDL操作会报错，事务回滚。
- 重分布过程中，支持用户进行本地表的DROP、TRUNCATE、TRUNCATE-PARTITION业务，正在重分布的表支持更新、插入和删除数据，重分布阶段不允许执行性能统计查询。
- 重分布过程中用户可进行正在重分布的本地表跨节点组的关联查询业务，性能可能受到一定影响。
- 支持内核自适应等锁：即对于业务持续高峰且业务允许正常报错的场景，内核通过自动取消业务保证重分布线程顺利拿锁推进流程。
- 数据收敛到最后支持写报错模式完成追增。
- 支持重分布过程中对关键参数进行动态实时调整。

表特性约束

- 重分布过程，存在用户业务与重分布资源的争抢，不适合大业务背景下做在线数据重分布。对于资源有限的场景，需要对数据重分布做资源管控来降低对用户业务的影响，目前粗粒度的资源管控可能导致重分布无法正常结束，因此使用在线扩容前需要对资源使用情况进行评估，需要为重分布预留足够的资源。
- 数据重分布过程中，以表为粒度进行数据搬迁，在数据搬迁过程中需要拿高级别表锁来阻塞用户写业务，从而获取数据在页面上准确的起始位置。由于当前表锁采用排队机制，一旦遇到慢SQL导致重分布线程处于等锁状态，会阻塞用户业务，此时会造成用户业务时延上涨，对于不停重试的业务模型出现线程池满无法对外提供服务的严重影响，因此使用在线扩容前需要谨慎评估业务模型对等锁的容忍程度以及是否存在慢SQL情况来评估是否能够进行在线扩容。
- 对于用户业务有多表关联查询的场景（表间使用分布列join并且分布列的分布方式相同），在数据重分布过程中可能出现跨node group join导致性能劣化，对于有大量join查询的场景不能做在线扩容。
- 扩容过程中，不支持用户并发事务块中包含Temp表创建使用。
- 重分布过程，目前不适用于单表大数据量数据持续入库（频繁增删改）场景（用户单表系统资源占用，超过重分布追增系统资源使用），数据追增阶段会影响用户持续导入与查询失败。
- 重分布过程中，用户应当避免执行长时间的查询场景，否则可能导致重分布出现等待加锁超时失败。
- 重分布过程中，对于大分区表，重分布阶段时间较长，用户周期性删除分区操作会打断重分布。
- Unlogged表无xLog，在线数据重分布阶段，有数据少量丢失风险。
- 数据重分布开始后不支持回退，因此需要在进行在线扩容前做好风险评估。
- 多表扩容模式约束：
  - 为属于一个group的表预留足够的磁盘空间，即group中所有表（表+索引）总和的1倍；考虑到资源有限，逐个group进行重分布，因此需要预留最大group大小1倍即可。
  - 多表扩容需要配合写报错模式和业务快速失败模式进行，因此只支持用户业务允许报错重试的场景。写报错模式时间窗口需要等group中的所有表均完成才能进行元数据切换，相较于单张表，影响窗口增大，但仍然是秒级完成。写报错模式持续时长和用户配置的last_catchup_threshold参数有关，默认值5秒。切换元数据耗时：通常场景采用cancel模式拿锁加切换元数据，单表2秒，最长10秒（最多5张表）；最坏场景：(lockwait_timeout + 2)*group中表数量。

依赖关系

无。

1.3.2 支持线程池高并发

可获得性

本特性自V300R002C00版本开始引入。

特性简介

通过线程池化技术来支撑数据库大并发稳定运行。

客户价值

支撑客户大并发下，系统整体吞吐平稳。

特性描述

线程池技术的整体设计思想是线程资源池化、并且在不同连接之间复用。系统在启动之后会根据当前核数或者用户配置启动固定一批数量的工作线程，一个工作线程会服务一到多个连接session，这样把session和thread进行了解耦。因为工作线程数是固定的，因此在高并发下不会导致线程的频繁切换，而由数据库层来进行session的调度管理。

特性增强

V300R002C00版本引入该特性。

V500R001C00版本实现了线程池的动态扩缩容。

V500R001C00版本实现了stream线程池，用于解决存在跨节点数据交换类查询的高并发性能。

特性约束

无。

依赖关系

无。

1.3.3 支持数据分布式存储

可获得性

本特性自V300R002C00版本开始引入。

特性简介

支持将数据库内用户表的数据按指定的路由规则分布式存储在不同的数据分片实例上。在此基础之上，在同一个数据分片内，数据被分布式存储在不同副本实例上。

客户价值

通过数据分布式存储特性，可以达成以下目的：

- 提升整个数据库系统能够支持的总数据量。
- 路由规则对应用（用户）透明，易用性强。
- 支持基于路由规则的查询剪枝和优化，提升单分片查询的查询性能。
- 支持同一个分片内多个副本实例的自动同步和自动切换，提升系统的可用性和可靠性。

特性描述

作为分布式数据库，GaussDB提供两个层级的分布式存储能力。

- 分片间数据的分布式存储：支持Hash分布、复制分布、Range范围分布和List列表分布。其中，Hash分布主要应用于数据量较大的用户表，通过对用户指定的单个或者多个分布列进行Hash值计算，将同一张用户表的数据打散存储到不同的分片内，从而提升整个数据库能够支撑的总数据量，并提供了在此基础之上的分布式并行处理能力和剪枝处理能力。复制分布主要应用于数据量较小的用户表，在每个分片内都会保存复制分布表的全量数据，从而提升分布式多表关联查询的性能。范围分布和列表分布主要用于用户需要自定义数据分布规则的场景，根据分布列的范围或者具体值来确定数据最终存储分片，便于用户进行数据管理。
- 分片内数据的分布式存储：对于每个分片，支持基于Quorum协议的一主多备分布式多副本存储，从而保证数据库的高可靠和高可用，并提供了在此基础之上的主备自动切换、AZ切换、备机强制升主等功能，保证稳定的、符合预期的RPO和RTO指标。

特性增强

无。

特性约束

无。

依赖关系

无。

1.3.4 分布式事务处理

可获得性

本特性自V300R002C00版本开始支持分布式事务高扩展性。

特性简介

分布式事务的扩展性指的是随着数据节点的增加，性能也随着以相应的比例增加，高扩展性即增加一倍节点即可增加一倍性能。GaussDB提供线性扩展比超过0.9的分布式事务线性扩展能力。

客户价值

分布式集群随着节点的增加，性能随之增加，可以给客户提供高扩展性的服务。当性能无法满足客户要求时，可以通过增加节点来提高吞吐服务。给客户提供高性能及高扩展性的服务。

特性描述

分布式集群系统为了保持一致性，需要引入中心事务管理节点，该节点容易成为集群瓶颈，限制集群扩展性，即增加节点后集群性能无法有效提升。GaussDB分布式事务扩展性主要解决了单点瓶颈问题，最大限度优化中心节点处理能力，可对外提供线性扩展比高达0.9的线性扩展能力。

特性约束

无。

特性增强

V500R001C00版本提供GTM-Lite模式。

依赖关系

无。

1.3.5 在线 CN 缩容

可获得性

本特性自V500R002C10版本开始引入。

特性简介

通过将一个或多个CN依次从集群拓扑中去除，以此实现CN缩容。

客户价值

在独立部署环境下，CN独占一台设备，当业务方面暂时用不到这么多CN时，会造成资源浪费。使用在线CN缩容功能，删除多余的CN实例，释放机器，以便节省成本。

特性描述

在线CN缩容本质上是将一个或多个CN依次从集群拓扑中去除，在混合部署的集群模式下，只会删除CN实例，在独立部署的情况下，则会直接将CN节点的所有信息从集群配置文件中移除。

特性约束

- "cluster_state"为"Unavailable"时，将无法执行删除CN操作。
- 一次仅允许删除一个CN。
- 如果因CN故障造成集群处于Degraded状态，此时如果执行删除CN操作，必须先删除因故障被剔除的CN，之后才能删除其他CN。
- 若已开启CN自动剔除功能，CM会自动将故障CN剔除，即从pgxc_node中删掉，这样DDL可以正常执行。CN被自动剔除后，不会再被拉起，必须删除CN或通过实例替换、节点替换或温备修复，才能进行扩容、升级等其他操作。
- 删除CN前不能锁定集群，不能执行其他运维及变更类操作。
- 删除完成后集群中至少剩余一个正常的CN。
- 数据库安装用户有足够的权限将新xml文件分发到所有主机的相同目录下。
- 在执行删除CN操作时，建议不要进行数据增删改等DML操作以及DDL操作，以避免数据的丢失。
- 在删除CN操作时，执行删除命令的节点不能是要删除的CN节点。
- 单CN的集群不支持继续缩容操作。
- 3 CN以下的集群不建议进行缩容操作，避免缩容过程中或结束后因为CN故障导致集群功能不可用。
- 部署kerberos情况下，同时缩容kerberos server主备IP所在的CN会导致集群异常。

- CN缩容需保证节点网络正常。

特性增强

无

依赖关系

无。

1.3.6 在线 DN 分片缩容

可获得性

本特性自GaussDB 503.1.0版本开始引入。

特性简介

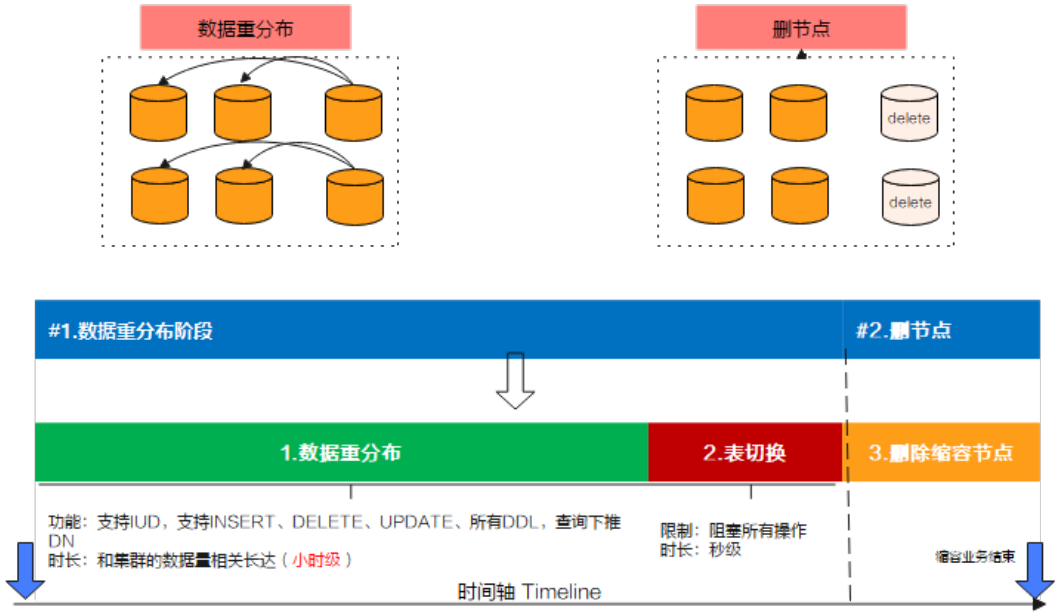
在线缩容过程的本质是集群内本地表的NodeGroup之间的数据搬迁，缩容过程指较大的NodeGroup向较小的NodeGroup进行数据搬迁，反之则看成是扩容过程。对于每个缩容重分布的表而言可以看成是这三个过程：

1. 基线数据重分布和增量数据重分布。
2. 表切换。
3. 删除缩容节点。

特性原理

随着客户业务的变化，原有集群在日常运行过程中存在资源富余情况，为降低使用成本需提供缩容能力，设计图如下：

图 1-3 在线 DN 分片缩容示意图



在线缩容过程的本质是集群内本地表的NodeGroup之间的数据搬迁，缩容过程指较大的NodeGroup向较小的NodeGroup进行数据搬迁，反之则看成是扩容过程。对于每个缩容重分布的表而言可以看成是这三个过程：

1. 基线数据重分布和增量数据重分布。
2. 表切换。
3. 删除缩容节点。

客户价值

随着客户业务的变化，原有集群在日常运行过程中存在资源富余情况，为降低使用成本需提供缩容能力，但缩容的完成需要一定的时间窗口，可能会影响业务。业务中断将给用户带来重大损失，因此需要在线缩容的特性，使得用户业务在缩容重分布过程中无需中断，平滑过渡。

特性描述

缩容以DN环为最小单元，根据用户预期缩容的DN分片数量计算出对应的DN环进行缩容，最终将DN环所包含的机器从当前集群中剔除。

重分布时普通表通过创建临时表和记录数据删除的增量表，在线缩容过程中通过更新此轮数据重分布的范围来实现在线缩容。

特性约束

- 集群状态为Normal，重分布状态为No。
- 不支持cms少数派故障场景下和gtmFree时gtm故障场景下的normal集群缩容。
- 集群不能处于锁定状态。
- 集群配置文件已经生成，配置的信息正确并且和当前集群状态一致。
- 缩容前用户需要确保default_storage_nodegroup参数值为installation。
- 缩容前需退出创建了临时表的客户端连接，因为在缩容过程中及缩容成功后临时表会失效，操作临时表也会失败。
- 仅支持大于等于两副本部署的集群。
- 缩容操作不可下发在被缩容的节点上。
- 缩容时集群不可处于read-only状态。
- 不支持表优先重分布、部分表不重分布。
- 不支持集群中存在list和range分布的表。
- 存在hashbucket表或绑定过group的数据库不支持缩容。

  通过以下SQL查看是否存在hashbucket表，查询结果为0表示存在hashbucket表，否则不存在。

  ```
  SELECT COUNT(*) FROM pg_catalog.pg_class where reloptions::text like '%%hashbucket=on%%';
  ```

  通过执行以下SQL语句查看数据库是否绑定group，查询结果为0表示未绑定过group，否则为绑定过group。

  ```
  SELECT count(*) FROM pg_catalog.pg_hashbucket;
  ```

- 集群缩容过程中，请勿改动$PGHOST/shrink_step/shrink_step.dat文件，否则影响缩容重入正确性。
- 缩容时，按照当前分片部署顺序从后向前缩容，不可跳跃。
- 缩容不支持包括CN的节点，如果包括CN，先使用增删CN工具，删除CN后再缩容。

- 缩容数据重分布失败，不影响业务，用户可选择合适的时间尽快完成重分布，否则会导致数据长期分布不均匀。
- 缩容的主机不能包含ETCD，GTM，CM Server。
- 重分布前，需要保证对应数据库下的data_redis为重分布预留schema，不允许用户操作该schema和其内部表。因为在重分布过程中，会使用到data_redis并且重分布结束后会删除该schema，如果存在用户表，则可能会出现数据误删。
- 缩容操作只支持集群中只有一个NodeGroup的缩容，不支持集群中包含多个NodeGroup的缩容。
- 缩容过程不支持gs_cgroup操作。
- 缩容完成以后，集群中唯一的NodeGroup名称为group_version1或者group_version2。
- 如果缩容前的Node Group下有依赖它的Child Node Group，缩容完成后，集群中新创建的Node Group下也会创建相应数量的Child Node Group，且和缩容前的Child Node Group一一对应。
- 缩容过程中，若因实例异常导致缩容失败，请参考《工具参考》中的“服务端工具 > gs_replace”章节内容在后台进行修复，故障修复后重入缩容。
- 执行缩容前，建议磁盘阈值满足以下条件：

  （全部表容量/缩容后DN个数 + 最大单表容量/缩容后DN个数）< 只读阈值（85%）*磁盘容量。
- 缩容失败后，不允许手动单独调用重分布，需要重入缩容。
- 缩容过程中系统将关闭“自动剔除故障CN”功能，在缩容完成后系统再次打开该功能。
- 缩容后若清理被缩容节点失败，缩容流程成功结束，此时如有需要需手动清理。
- 多表join功能约束：
  - 不能使用read-only模式。
  - 需要用户识别出带有join关系的用户表列表，利用文件的形式传递给内核。
  - 为属于一个group的表预留足够的磁盘空间，即group中所有表（表+索引）总和的1倍；考虑到资源有限，逐个group进行重分布，因此需要预留最大group大小1倍即可。
  - 每个group中的表数量上限为5张，下限为2张。
  - 表名中如果含有特殊字符，比如空格等需要使用双引号进行转义，否则代码可能无法正确读取表名，例如会使用空格做表名切分。
  - 多表join group中的表目前只支持普通行存表（包括普通表和分区表），range/list表、外表、物化视图、hashbucket表、复制表等均不支持。
  - 多表join需要配合写报错模式和业务快速失败模式进行，因此只支持用户业务允许报错重试的场景；写报错模式时间窗口需要等group中的所有表均完成才能进行元数据切换，相较于单张表，影响窗口增大，但仍然是秒级完成，写报错模式持续时长和用户配置的last_catchup_threshold参数有关，默认值5秒；切换元数据耗时：通常场景采用cancel模式拿锁加切换元数据，单表2秒，最长10秒（最多5张表）；最坏场景：(lockwait_timeout + 2)*group中表数量。
  - 重分布开始后支持修改group信息，但是由于重分布的顺序是先单表后group，一旦表已经完成重分布就无法回退且无法指定到group中。
  - 在重分布的重入场景中，如果修改了已经开始重分布的group中的表到两个不同的group中，下次重入表占用的空间不会自动释放，直到group完成元数据切换后才能进行资源清理。如果需要释放只能手动执行。

- 在IO、CPU资源使用较高的场景，不适合开启多表缩容功能。否则可能触发重分布工具主动资源管控，导致使用IO限制过严以及性能严重降低、追增不上的情况，使用前需谨慎评估。
- 对于group中的表写业务量都比较大的场景，开启多表缩容功能可能出现追增不上的风险，需要谨慎评估。

特性增强

无

依赖关系

无。

1.3.7 在线表类型转换

可获得性

本特性自505.0.0版本开始引入。

特性简介

支持普通表与hashbucket表在线本地进行表类型相互转换。

客户价值

随着hashbucket特性的演进，GaussDB提供将数据库中已有的普通表，在线本地转换为hashbucket表的能力，以便用户后续使用hashbucket特性。

特性描述

通过重分布的原理支持在线本地表类型转换。普通表与hashbucket表在线本地进行表类型相互转换，用户可以使用此特性将集群中已存在的普通表转换为hashbucket表，并在此基础上利用hashbucket扩容特性。

特性约束

- 集群状态必须为Normal或者Degraded状态，集群Degraded状态只是实例异常引起的集群状态异常，不支持VM异常引起的集群状态异常（VM异常指非软件类异常，例如：磁盘故障、通信故障），请确保实例状态异常节点的网络正常。
- 实例状态：CN全部正常，GTM、CMS至少有主节点存活。
- 重分布状态必须为No。
- 集群配置文件的配置信息正确并且和当前集群配置一致。
- 资源管控对象redisuser，redisrespool，redisclass，redisgrp不能被占用。
- 迁移失败后集群不回滚，需重入迁移操作。
- 当前仅支持普通astore行存非分区表或分区表，hashbucket非分区表或分区表之间的类型转换。不支持段页式表的转换。支持迁移的表类型继承hashbucket表约束，不支持ustore表、不支持range和list分布的表、不支持二级分区表、不支持临时表、不支持unlogged表、不支持透明加密表的转换。

- 迁移操作要求整个集群没有被锁定（没有业务持有集群锁禁用DDL）。
- 迁移操作仅支持分布式部署的集群。
- 迁移操作要求集群不能处于只读模式。
- 迁移失败后的待重入阶段禁止使用gs_om的managecn功能删除CN。
- 在线本地迁移磁盘空间要求预留最大单表的1.5倍、锁资源约束继承逻辑在线扩容的规格。
- 对业务的影响：本地迁移对在线业务的平均吞吐量和平均时延的影响，在用户设置的资源管控级别范围内，业务闪断控制在秒级。
- 资源使用情况：
  - 离线场景不控制I/O资源使用。
  - 在线场景I/O占用控制在用户下发的资源管控参数范围内。
  - 普通表转hashbucket表，磁盘膨胀与参数enable_segment_datafile_preallocate对磁盘预分配的影响相关。参数为on时，hashbucket表预分配文件占用额外磁盘空间。
- 性能规格：在f模式下，单DN 1TB在线本地数据迁移耗时是逻辑扩容的两倍。

具体场景：（此场景同在线扩容已商用特性的发布规格，未新增场景）

集群数据节点读带宽500MB/s、写带宽800MB/s，万兆网络，CPU和内存不是瓶颈，业务负载低峰且不做资源管控（资源管控级别为f，表内并发度为8，扫描并发度为4）时，单DN分片1TB在线本地数据迁移6小时完成。

- 转换前后hashbucket表的约束继承hashbucket特性本身的约束；hashbucket表不支持的表类型均不支持表类型迁移。
- 在线迁移正常运行过程中，用户修改转换文件无效。异常发生后，若重分布已开始不允许修改文件中表的转换类型及新增表，可删除未转换的表信息；重分布前允许用户更改文件内容（增、删、改）。

特性增强

无

依赖关系

无。

1.4 高可用性

1.4.1 主备机

可获得性

本特性自V300R002C00版本开始支持DN主备。

特性简介

为了保证故障的可恢复，需要将数据写多份，设置主备多个副本，通过日志进行数据同步，可以实现节点故障、停止后重启等情况下，GaussDB能够保证故障之前的数据无丢失，满足ACID特性。

客户价值

主备机功能可以支持主机故障时切换到备机，数据不丢失，业务可以快速恢复。

特性描述

主备环境可以支持一主多备两种模式。而在一主多备模式下，所有的备机都需要重做日志，都可以升主。一主多备提供更高的容灾能力，更加适合于大批量事务处理的OLTP系统。

主备之间可以通过switchover进行角色切换，主机故障后可以通过failover对备机进行升主。

初始化安装或者备份恢复等场景中，需要根据主机重建备机的数据，此时需要build功能，将主机的数据和WAL日志发送到备机。主机故障后重新以备机的角色加入时，也需要build功能将其数据和日志与新主机保持一致。另外，在在线扩容的场景中，需要通过build来同步元数据到新节点上的实例。build包含全量build和增量build，全量build要全部依赖主机数据进行重建，复制的数据量比较大，耗时比较长，而增量build只复制差异文件，复制的数据量比较小，耗时比较短。一般情况下，优先选择增量build来进行故障恢复，如果增量build失败，再继续执行全量build，直至故障恢复。

为了实现所有实例的高可用容灾能力，除了以上对DN设置主备多个副本，GaussDB还提供了其他一些主备容灾能力，比如CN（互为备份）、GTM（一主多备）、CM Sever（一主多备）以及ETCD（一主多备）等，使得实例故障后可以尽可能快地恢复，不中断业务，将因为硬件、软件和人为等因素造成的故障对业务的影响降到最低，以保证业务的连续性。

特性增强

无。

特性约束

无。

依赖关系

无。

1.4.2 AZ 容灾

可获得性

本特性自V300R002C00版本开始引入。

特性简介

客户往往会设置多个不同的可用区，将其分布在跨机房或者跨地域的不同位置。当某个可用区发生断电、断网等大范围故障时，其他可用区仍然可以继续提供服务。

客户价值

满足客户AZ级容灾能力的需求。

特性描述

GaussDB目前提供同城2AZ（均衡/亲和）4副本/6副本+Quorum+第三方仲裁和同城3AZ3副本+Quorum两种方案，分别如下所示：

图 1-4 2AZ4 副本+Quorum+第三方仲裁部署示意图

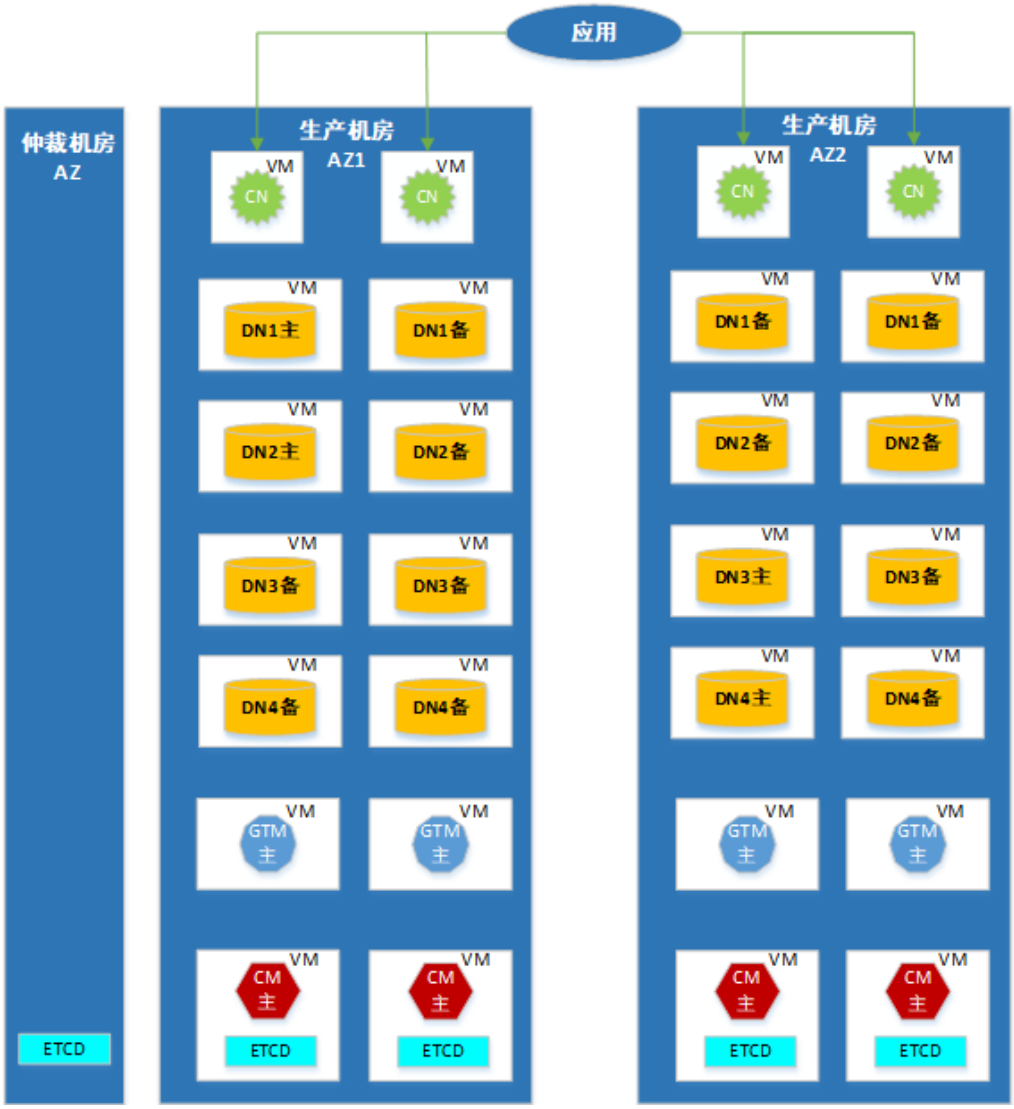


图 1-5 2AZ6 副本+Quorum+第三方仲裁部署示意图

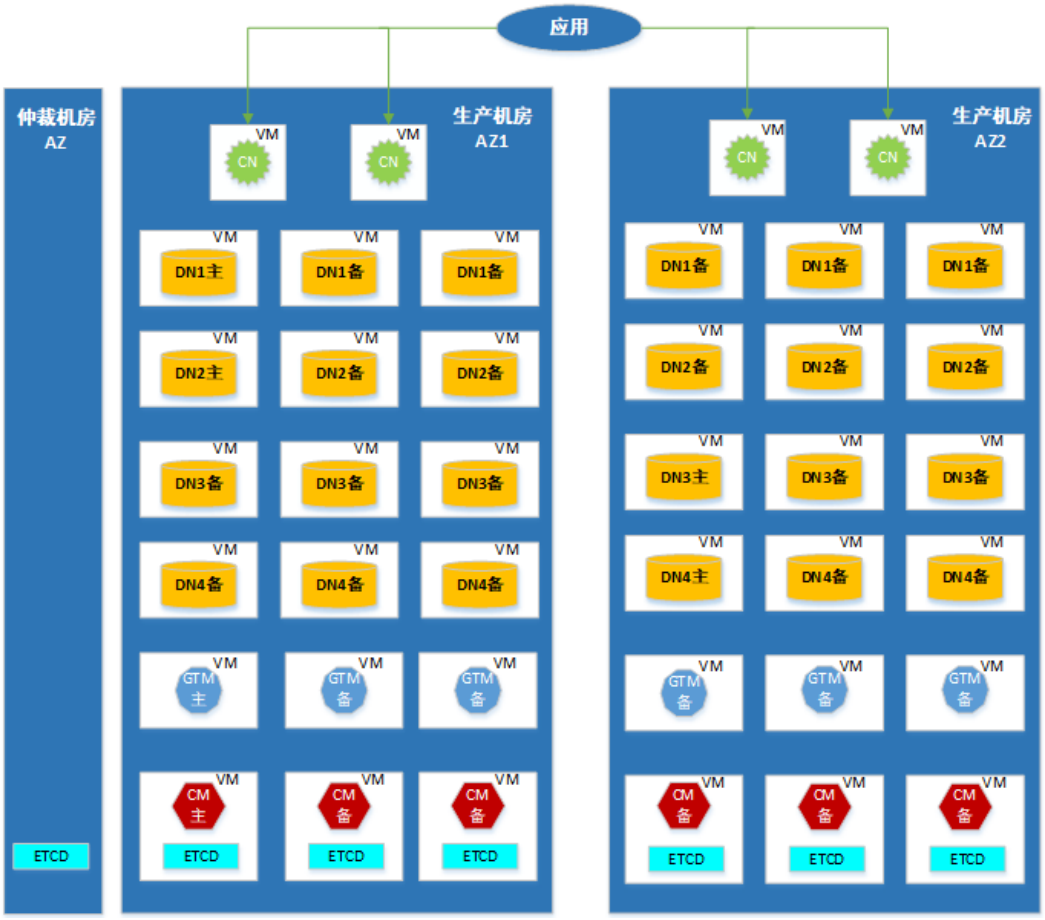


图 1-6 3AZ3 副本+Quorum 部署示意图

特性增强

无。

特性约束

无。

依赖关系

依赖主备机。

1.4.3 逻辑复制

可获得性

本特性自V500R001C00版本开始引入。

特性简介

GaussDB提供逻辑解码功能，将物理日志反解析为逻辑日志，通过DRS等逻辑复制工具将逻辑日志转化为SQL语句，到对端数据库回放，达到异构数据库同步数据的功能。目前支持GaussDB数据库与MySQL数据库、Oracle数据库之间的单向、双向逻辑复制。

客户价值

逻辑复制可以为集群数据实时迁移、双集群双活、支持滚动升级提供解决方案。

特性描述

DN通过物理日志反解析为逻辑日志，DRS等逻辑复制工具从CN或DN抽取逻辑日志转换为SQL语句，到对端数据库（如MySQL）回放。逻辑复制工具同时从对端数据库抽取逻辑日志，反解析为SQL语句之后回放到GaussDB，达到异构数据库同步数据的目的。

特性增强

GaussDB V500R001C00版本逻辑解码新增全量+增量抽取日志的方案。

GaussDB V500R002C00版本逻辑解码新增备机支持逻辑解码。

GaussDB V500R002C10版本逻辑解码的用户黑名单特性实现了输出逻辑解码日志的事务的用户粒度过滤，在解码阶段提前过滤非期望用户的事务操作，避免资源无效使用，进一步提升解码的性能。

GaussDB 503.1.0版本逻辑解码新增支持心跳日志，对于大事务解码过程中防止DRS长时间未收到GaussDB消息而误判。

GaussDB 503.1.0版本逻辑解码新增对分布式事务强一致的支持。

GaussDB 505.0.0版本备机逻辑解码新增对极致RTO的支持。

特性约束

不支持DDL复制，在线扩容后需要重新全量复制。

依赖关系

依赖于逻辑复制工具对逻辑日志进行解码。

1.4.4 高可靠事务处理

可获得性

本特性自V300R002C00版本开始支持分布式事务的高可用性。

特性简介

分布式事务满足ACID，在任何故障场景下无数据丢失，无数据错乱问题。

客户价值

- 提供可靠的数据一致性保障。
- 数据不丢失，数据不错乱。
- 分布式事务出现两阶段残留，无需客户手动处理。易用性高。

特性描述

分布式事务满足ACID，在任何故障场景下无数据丢失，无数据错乱问题。集群内部的两阶段提交如果出现中间故障，可以通过内部逻辑进行正确处理，保证数据不丢失不错乱。

特性增强

V300R002C00版本提供gs_clean并发清理性能增强。

V500R001C00版本提供GTM-Lite模式下异常残留事务清理机制。

特性约束

无。

依赖关系

依赖于HA相关高可用特性。

1.4.5 CN 自动剔除

可获得性

本特性自V300R002C00版本开始引入。

特性简介

数据库集群中有CN节点无法提供服务时快速感知并剔除，不影响业务的执行。

客户价值

可以保障数据库在各种异常场景下的持续高可用。在CN节点由于各种故障如磁盘损坏、网络隔离等情况下无法提供服务时，此时数据库能够做到CN自动剔除，不影响用户DDL语句。

特性描述

集群部署多个CN同时对外提供服务，CN的角色是对等的，即执行DML语句时连接到任何一个CN都可以得到一致的结果。而DDL语句需要在所有CN上都执行完成，保持相关定义一致，如果其中一个CN发生故障，整个集群将无法执行DDL语句，直到故障CN被修复。为了不影响用户业务的执行，CN故障自动剔除功能，系统检测到CN故障后在限定时间内将CN自动剔除，用户的DDL语句就可以继续执行。

特性增强

无。

特性约束

- 自动剔除故障CN功能默认开启，默认设置CN剔除时间为25秒。用户可根据自己实际场景和需求确定是否开启功能，以及开启后的剔除时间。

- 集群中部署的CN少于1个不会自动剔除。多CN场景下，共N个CN时，最多剔除N-1个CN。当正常CN数量小于等于1个时，不能进行自动剔除。如果开启了自动修复CN功能，在已剔除CN的故障消除时，系统可以自动修复或者用户执行实例替换命令手动修复。
- CN故障被剔除后，CN会处于Deleted状态， 集群处于Degraded状态，用户业务可以继续执行不受影响，但是物理集群的扩容、缩容、升级、增加CN、change IP操作将不能执行。

依赖关系

无。

1.4.6 在线节点替换

可获得性

本特性自V300R002C00版本开始引入。

特性简介

集群内某节点出现硬件故障造成节点不可用或者实例状态不正常，当集群没有加锁，通过节点替换或修复故障实例来恢复集群的过程中，支持用户DML操作，有限场景支持用户DDL操作。

客户价值

随着企业数据规模不断增大，节点数量急剧增加，硬件损坏概率相应增加，物理节点替换修复成为日常运维工作的常态。传统的离线节点替换方式无法满足客户业务不中断需求，日常运维操作中，经常的业务中断将给客户带来重大损失。而目前业界数据库产品在节点替换的过程中，或者需要中断业务，或者只允许部分操作，均不能满足大规模数据情况下，常态物理节点替换的需求。在线节点替换特性解决了以上问题，提升了数据库运行的可靠性，可为用户提供更加稳定的数据服务。

特性描述

如果数据库集群内某节点因为出现硬件故障而造成节点不可用或者实例不正常时，且集群未上锁的前提下，在通过节点替换或修复故障实例来恢复集群的过程中，支持用户DML操作，有限场景支持用户DDL操作。

特性增强

无。

特性约束

目前集群未上锁的前提下，节点替换已支持用户业务在线DDL：

- 在节点替换窗口期内，支持用户DML操作，有限场景支持用户DDL操作。

现有方案，所替换节点中包含CN时，存在如下约束：

- 在CN实例修复阶段，分为Base修复阶段与增量修复阶段，在增量修复阶段会短暂阻塞用户DDL操作（平均时长在1到5分钟内，最长为20分钟），DML不会阻塞。

- 节点修复阶段，用户应选择DDL业务相对不密集的阶段实施，可有效缩短增量修复阶段用户DDL阻塞时长。
- 增量修复阶段用户的DDL操作会被阻塞，如用户事务块跨CN数量变化窗口（增量修复阶段）会报错回滚，业务侧增加重试机制可解决此问题。

依赖关系

无。

1.4.7 物理备份

可获得性

本特性自V300R002C00版本开始引入。

特性简介

支持将整个分布式数据库集群的数据以内部格式备份到指定的存储介质中。

客户价值

通过物理备份特性，可以达成以下目的：

- 整个数据库集群的数据备份到可靠性更高的存储介质中，提升系统整体的可靠性。
- 通过采用数据库内部的数据格式，极大提升备份恢复性能。
- 通过分布式备份技术，极大提升备份性能。
- 可以用于冷数据的归档。

典型的物理备份策略和应用场景如下：

- 周一，执行数据库全量备份。
- 周二，以周一全量备份为基准点，执行增量备份。
- 周三，以周二增量备份为基准点，执行增量备份。
- ...
- 周日，以周六增量备份为基准点，执行增量备份。

上述备份策略以一个星期为周期。

特性描述

GaussDB提供物理备份能力，可以将整个集群的数据以数据库内部格式备份到本地磁盘文件、OBS对象、NAS对象或REMOTE对象中，并在同构数据库中恢复整个集群的数据。物理备份采用分布式并行技术，并行地对每个数据实例DN的数据文件进行物理备份，提供了极高的备份恢复性能。在基础之上，还提供压缩、流控、断点续备等高阶功能。

物理备份主要分为全量备份和增量备份，区别如下：全量备份包含备份时刻点上数据库的全量数据，耗时时间长（和数据库数据总量成正比），自身即可恢复出完整的数据库；增量备份只包含从指定时刻点之后的增量修改数据，耗时时间短（和增量数据成正比，和数据总量无关），但是必须要和全量备份数据一起才能恢复出完整的数据

库。除此之外，还支持从集群的备份数据中恢复单个或多个数据库、单个或多个表，全量、增量备份单个或多个表以及从表级备份数据中恢复单个或多个表的功能。

特性增强

支持全量备份和增量备份同时执行。

503.0.0版本支持OBS、NAS、DISK介质下的备机备份能力，即备份操作在备DN实例上执行。

特性约束

物理备份的约束条件请参见《管理员指南》中“备份与恢复 > Roach工具介绍 > 约束和限制”章节。

依赖关系

无。

1.4.8 负载均衡

可获得性

本特性自V300R002C00版本开始引入。

特性简介

在同一应用程序内为多CN提供一个统一的入口，将客户端的请求均匀的分发给集群中的各个CN服务器，使应用程序内所有请求负载均衡。

客户价值

当业务连接很多时，负载均衡作为JDBC中的一个重要功能，承担如下职责：

- 均衡各CN负载，充分利用多CN计算能力。
- 故障隔离，当CN故障后，负载均衡能感知故障，并自动停止向故障CN节点转发请求。

特性描述

- 连接参数配置autoBalance=true或autoBalance=balance或autoBalance=roundrobin时，支持同一业务的所有连接均匀分布到集群中的所有CN上。
- 连接参数配置autoBalance=priorityn时，支持同一业务的所有连接优先均匀分布到同AZ内的CN上，当同AZ内的CN均不可用时，连接分布到其他AZ内的CN上。
- 连接参数配置autoBalance=shuffle时，支持同一业务的所有连接随机分布到集群中的所有CN。
- 连接参数不配置autoBalance或者配置autoBalance=false，同一业务的所有连接会连接到同一CN上。

特性增强

无。

特性约束

- 应用层的IP与CN IP要处于同一网络地址空间。
- 用户初始配置的CN的IP，保证至少有一个CN可用，若CN全部故障，需要人工干预修复。
- 当前版本JDBC暂时无法统计各个CN的连接数、CPU和内存使用情况，不支持根据CN的负载动态调整到各个CN的连接，只支持将应用程序新建连接均匀的分发给各个可用CN。由于上述原因，在实际运行中当CN故障，之后再恢复，或者新增CN的场景，可能短时间内会存在负载不均衡的情况，这种情况在上层应用程序业务结束释放连接，新业务建立新连接后，会逐步重新平衡。

依赖关系

该特性依赖于数据库可连接，pgxc_node系统表可以正常查询。

1.4.9 极致 RTO

可获得性

本特性自V500R001C10版本开始引入，从503.1版本开始支持备机读。

特性简介

- 支撑数据库主机重启后快速恢复的场景。
- 支撑主机与同步备机通过日志同步，加速备机回放的场景。

客户价值

当业务压力过大时，备机的回放速度跟不上主机的速度。在系统长时间的运行后，备机上会出现日志累积。当主机故障后，数据恢复需要很长时间，数据库不可用，严重影响系统可用性。

在硬件资源充足的情况下，开启极致RTO（ Recovery Time Object，恢复时间目标 ）特性，可以减少备机的RTO，减少了主机故障后数据的恢复时间，提高了系统的可用性。

特性描述

极致RTO开关开启后，xLog日志回放建立多级流水线，提高并发度，提升日志回放速度。

采用page多版本的方式支持备机读，回放线程维护每一个page的日志链，读线程根据指定的LSN（ wal日志的位置 ）读取对应版本的page。当查询和回放冲突时，查询超时会被取消，报错信息是"canceling statement due to conflict with recovery"，错误码是40001，详细信息参见《 错误码参考 》的内核错误码>GAUSS-19501--GAUSS-20000章节。当出现这种类型的报错时，业务端可根据错误码进行重试。

造成查询和回放冲突的日志类型主要包含如下几种：

1. 删除文件

   触发条件：删除文件、reindex、truncate表等操作。

   处理方案：等待max_standby_streaming_delay时间后，发送cancel消息取消冲突的查询。

2. drop database

   触发条件：执行删除数据库操作。

   处理方案：等待max_standby_streaming_delay时间后，发送cancel消息取消冲突的查询。

3. drop tablespace

   触发条件：删除tablespace。

   处理方案：等待max_standby_streaming_delay时间后，发送cancel消息取消冲突的查询。

4. vacuum清理（仅在参数exrto_standby_read_opt开启下，会产生冲突）

   触发条件：vacuum操作。

   处理方案：等待max_standby_streaming_delay时间后，发送cancel消息取消冲突的查询。

5. reindex database

   触发条件：重建数据库索引。

   处理方案：在容灾GTM_LITE模式下，等待max_standby_streaming_delay时间后，发送cancel消息取消冲突的查询。

打开备机读之后，因为需要维护历史page版本，所以会占用更多I/O。

特性增强

为了充分发挥极致RTO基于多核CPU架构对回放性能的优化效果，建议将GUC参数redo_bind_cpu_attr（该参数用于控制回放线程的绑核操作）设置为cpuorderbind 类型，例如'cpuorderbind:16-32'。绑核区间应与通过GUC参数thread_pool_attr设置的线程池绑核区间以及通过GUC参数wal_rec_writer_bind_cpu、walwriteraux_bind_cpu、wal_receiver_bind_cpu设置绑定的cpu核号错开，区间大小根据线程数要调整，建议设置为大于等于recovery_parallelism（实际回放线程个数）+ 1。推荐将所有的回放线程绑定到一个numa组内，性能会更好。

特性约束

- 极致RTO采用了多个page redo线程并行加速回放进度。当备机回放追平主机，空载的情况下，单个page redo线程的CPU消耗大约在15%左右（实际值与具体硬件和参数配置相关），备机回放的总CPU消耗值 = 单个page redo线程的CPU消耗值 x page redo线程数。因为启动的更多的线程，CPU和内存的消耗都会比并行回放、串行回放要多。
- 极致RTO只关注同步备机的RTO是否满足需求。极致RTO去掉了自带的流控，统一使用recovery_time_target参数来做流控控制。
- 本特性支持备机读，由于增加了对数据页面历史版本的读取，备机上的查询性能会低于主机上的查询性能，低于并行回放备机读的查询性能，但是查询阻塞回放的情况有所缓解。
- DDL日志的回放速度远远慢于页面修改日志的回放，频繁DDL可能导致主备时延增大。

- 当节点的I/O和CPU使用过高时（建议不超过70%），回放和备机读性能会有明显下降。
- meta erp场景：

  硬件规格： Intel(R) Xeon(R) Gold 5220R CPU @ 2.20GHz，754G memory，nvme *2，10GE网卡*2

  业务模型： 使用erp 场景实例表cst_std_item_cost_t定义（表的个数1-20，以性能最优为准），行宽0.7k左右（以性能最优值为准）。使用jmeter等压测工具执行insert语句，单事务行数<4096（以性能最优值为准），并发85左右（以性能最优值为准），ustore表，无DDL。

  日志量：<=300MB/s

  主备时延： <=1s
- 极致RTO备机读在以下几种情况下会取消查询：
  a. 当查询时间超出了参数standby_max_query_time。
  b. 触发了备机读文件的强制回收。
  c. 当查询和回放有锁相关等冲突时，和并行回放备机读相同，取消查询由参数max_standby_streaming_delay控制。
  d. 在开启参数exrto_standby_read_opt的情况下，回放vacuum相关的清理日志时会发生冲突，和并行回放备机读相同，取消查询由参数max_standby_streaming_delay控制。
  e. 在容灾GTM_LITE模式下或单集群GTM_FREE模式、开启stream执行计划，查询和relmap类型的日志回放有冲突。
  f. 备机回放段页式（包括hashbucket表）物理空间收缩操作相关日志时会取消查询。

依赖关系

无。

1.4.10 基于 Paxos 协议的高可用

可获得性

本特性自503.1.0版本开始引入。

特性简介

DCF全称是Distributed Consensus Framework，即分布式一致性共识框架。它是一款自研的高性能、高度成熟可靠、易扩展、易使用的独立基础库，其他系统通过API接口可方便地集成DCF组件。DCF基于Paxos协议实现，解决分布式一致性问题，提升集群高可靠高可用能力。

当配置为DCF模式后，DN可以支持基于Paxos协议的复制与仲裁能力。DN基于Paxos的选主和日志复制，复制过程中支持压缩及流控，防止带宽占用过高。提供基于Paxos多种角色的节点类型，并能够进行调整。支持查询当前数据库实例的状态。

特性描述

- DCF进行日志复制时，支持对日志进行压缩后再传输，减小对网络带宽的占用。

- DCF支持SSL，包括TLS1.2和TLS1.3协议标准。当开启SSL时，DN默认将DCF配置为TLS1.2协议标准。
- DCF支持TLS1.2如下密码套件：TLS_ECDHE-ECDSA-AES256-GCM-SHA384、TLS_ECDHE-ECDSA-AES128-GCM-SHA256、TLS_ECDHE-RSA-AES256-GCM-SHA384、TLS_ECDHE-RSA-AES128-GCM-SHA256。
- DCF支持passive角色节点类型，passive节点不参与选举，只做日志的同步以及回放，该类型节点在高负载的情况下，日志同步会做流控。
- DCF支持logger角色节点，logger节点可以参与选举投票有投票权但无选举权，只复制DCF的日志，不进行redo。
- DCF的follower和passive角色可以在线互换，即不中断业务的情况下，follower角色的节点转化为passive角色，passive角色的节点转化为follower。
- DCF支持少数派强起能力，在数据库实例多数派故障的情况下，从正常的备DN中选择少数派模式强启成为主DN，其余正常的备DN从主DN复制日志。
- DCF支持自选主能力，在原主DN故障的场景下，在保证数据一致性的前提下，剩余备DN自动选出新的主DN。
- DCF支持策略化多数派能力，以多数派为前提，同时根据用户配置的AZ，保证AZ内至少有一个节点同步复制日志。
- DCF支持手动模式，在手动模式下不自动仲裁，此模式下对接上层CM等管理组件做仲裁适配，DCF进行日志复制功能。
- 支持DCF日志与DN日志合一存储，DCF多数派达成和DN仅存储一份日志，减少IO带宽占用，日志合一后日志刷盘的IO开销比两份日志下降20%+，优化性能。
- 支持从Quorum模式切换到DCF模式，以及从DCF模式恢复到Quorum模式。切换过程中不需要重启数据库，能做到数据不丢失。
- 支持级联备节点部署能力，级联备节点仅从备机同步日志，降低主机日志复制压力，仅支持容灾灾备集群部署级联备。
- 支持分布式安装部署，支持的功能包括安装、启停、增删副本、节点替换和修复、节点重建、switchover、failover、升级回滚、备份恢复、备机备份、自动降副本。
- DCF模式1主1备1logger组网下不支持增删副本、连接备机执行build、强切；强切功能适用于长时间无主的情况，在DCF模式下如果DCF层面出现leader，此时不需要再执行强切，因为主节点已经存在。
- 支持1主1备1 logger组网，logger节点辅助仲裁和保留日志、不可当主、无数据库，logger节点可以大大简配CPU和内存、轻量级部署。含logger节点集群环境在业务数据场景下，如果原主机故障且候选主机落后于日志节点情况下，由于候选主机需要同步完日志节点数据并回放升主，因此该场景不能保证RTO能力。
- 支持从Quorum到DCF模式切换，支持1主2备到1主1备1 logger替换。
- DCF默认工作在最大保护模式，支持通过配置工作在最大可用模式和最大性能模式。DCF支持强切功能。
- DCF支持分片自动升降副本，当一个分片故障了半数及以上节点时，为了降低分片故障对业务的影响，集群管理对分片上可用节点进行降副本操作，当检测到故障恢复后，自动触发升副本操作。
- DCF支持联合仲裁，即自动模式，在一些复杂场景会联合借助CM组件的全局视角实现自仲裁能力增强，解决复杂场景问题，例如半数节点故障自动降副本。支持安装为联合仲裁模式集群，管控需配置如下参数："dnParams": {"dcf_run_mode":0},"cmParams":{ "dn_arbitrate_mode":"paxos" }。

特性增强

无。

特性约束

- 若使用此功能，DN最少三节点，在安装部署阶段需要开启DCF开关。在DCF模式下通过多数派选举，安装过程中如果故障节点数加build节点数达到多数派会导致数据库实例安装失败，如在安装一主两备时，安装过程中一节点因内存不足导致安装失败，另外两节点正常启动，但随后备机会进行一次build，这时build节点加故障节点为2，达到多数派会导致数据库实例安装失败，请在安装过程中检查内存和磁盘等资源是否充足。
- 若某个AZ配置了策略化多数派参数，当AZ内所有的节点均故障时，在对节点做build相关的操作时，需要将该AZ配置从策略化多数派配置信息中移除。
- DCF支持手动模式是针对集群级的工作模式的设置，分布式集群仅支持手动模式选举，在此工作模式下不支持节点passive角色，集群安装部署也不支持passive角色。
- DCF在容灾场景下主备集群仅支持手动模式。
- 从Quorum模式切换到DCF模式，仅支持固定次数（3次）切换，如超过固定次数需再次切换，则需要重启数据库实例。模式切换过程中，数据库内核涉及到线程关闭和拉起，可能会短暂影响业务（一般小于1分钟），尤其是数据量大的时候。所以建议模式切换过程中尽量少或不要运行业务。
- 包含logger节点的集群典型组网是1主1备1 logger，其他组网不承诺支持。
- logger节点仅用来辅助仲裁和保留日志，不要在logger节点上启动roach备份等进程。logger节点不支持查询系统视图与系统表。logger节点部署实例最小支持4核8GB内存。

依赖关系

无。

1.4.11 两地三中心跨 Region 容灾

可获得性

本特性自V500R001C20版本开始引入。

基于流式复制的异地容灾解决方案，从V500R002C00版本开始提供该解决方案。

特性简介

支持两地三中心跨Region容灾。

客户价值

金融、银行等业务需要底层数据库提供跨地域的容灾能力，来保证极端灾难情况下数据的安全和可用性。

特性描述

金融、银行业对数据的安全有着较高的要求，当发生火灾，地震，战争等极端灾难情况下，需要保证数据的安全性，因此需要采取跨地域的容灾的方案。跨地域容灾通常

是指主备数据中心距离在200KM以上的情况，主机房在发生以上极端灾难的情况下，备机房的数据还具备能继续提供服务的能力。本特性的目的是提供一套支持gaussdb跨地域容灾的解决方案。

特性增强

V500R002C00版本开始针对两地三中心跨Region容灾特性新增基于流式复制的异地容灾解决方案：

- 支持容灾搭建。
- 支持灾备集群failover，以及灾备集群failover后主集群容灾解除。
- 支持容灾主备集群计划内switchover。
- 支持容灾状态查询。
- 支持容灾状态下主集群，灾备集群节点修复与节点替换。

503.0.0版本新增功能：

- 支持容灾主集群日志保持
- 支持容灾加回
- 支持容灾演练增强
- 支持结束容灾搭建等待状态的清理功能
- 支持灾备集群升主集群后手动升副本，恢复为灾备集群后手动降副本
- 支持容灾过程中修改容灾用户信息

特性约束

基于流式复制的异地容灾解决方案：

- 灾备集群可读不可写。灾备集群读不支持GTM FREE模式。
- 不支持不同GTM模式的集群搭建容灾，GTM FREE不支持容灾搭建。
- GTM FREE模式下如果所有GTM都损坏，容灾功能将不可用。
- 主集群和备集群应该具有相同的管理员用户名rdsAdminUser。
- 主集群和备集群应该具有相同的集群用户名。
- 灾备集群通过failover命令升主后，和原主集群灾备关系将失效，需要重新搭建容灾关系。
- 数据库实例状态对容灾操作的影响：
  - 在主集群和灾备集群处于normal状态且所有组件（CN、DN、ETCD、GTM、cm_agent、cm_server）状态正常时可进行容灾搭建；在主集群处于normal态所有组件状态正常并且灾备集群已经升主的情况下，主集群可执行容灾解除，其他集群状态不支持。
  - 在主集群和灾备集群处于normal状态且所有组件（CN、DN、ETCD、GTM、cm_agent、cm_server）状态正常时，通过计划内switchover命令，主集群可切换为灾备集群，灾备集群可切换为主集群。
  - 灾备集群处于非Normal且非Degraded状态时，无法升主，无法作为灾备集群继续提供容灾服务，需要修复或重建灾备集群。
  - 主集群分片存在多数派实例故障且没有打开最大可用模式时(参数most_available_sync)，不会向灾备集群进行日志发送，需要及时修复主集群故障实例。

- 当灾备集群为2副本时，需要确保打开最大可用模式(参数most_available_sync)。灾备集群在1个副本损坏时，仍可以升主对外提供服务，如果剩余的这个副本也损坏，将导致不可避免的数据丢失。
- 灾备集群支持单机1副本部署（1CN、多DN（单副本）、1ETCD、1CMS、1CMA、1GTM，单个DN资源不少于8核CPU），单机1副本集群当前版本仅支持安装、升级、参数设置、集群启停、集群状态查询、备份恢复，不支持扩容、节点修复、节点替换、修改端口、升降副本、灾备集群读等SLA功能，不支持单副本当做主集群搭建容灾。
- 主集群如果进行了强切操作(cm_ctl finishredo命令，参见《工具参考》系统内部使用的工具中的“统一集群管理工具”章节)，需要重建灾备集群。
- 主集群如果进行了少数派AZ强启，会出现数据丢失，需要重建灾备集群。
- 容灾关系搭建之后，灾备集群不支持全备和增备，主集群支持全备和增备。如果主集群要做恢复，需要先解除容灾关系，在完成备份恢复后重新搭建容灾关系。
- 容灾关系搭建之后，灾备集群不支持逻辑复制，主集群支持逻辑复制。
- 灾备集群支持节点替换和节点修复，继承节点替换和修复的约束。
- 建立容灾关系的主集群与灾备集群之间不支持GUC参数的同步。
- 容灾关系搭建之后，不支持CN/DN实例用于流式复制的端口修改。
- 容灾关系搭建之后，主备集群不支持扩容。主备集群需解除容灾关系后，各自完成扩容，并重新搭建容灾关系。
- 容灾状态中灾备集群支持降副本，灾备集群升主后正常支持升降副本。
- 该解决方案不支持DCF模式。
- 支持主集群与灾备集群CN个数不对等情况下搭建容灾，但主集群的CN配置数与灾备集群的CN配置比最大为8:1。
- 容灾搭建时需要在主集群和灾备集群下发容灾用户名和密码用于集群间鉴权：
  - 主备集群必须使用相同的容灾用户名和密码。
  - 不得使用已存在的数据库用户进行搭建。
- 搭建容灾的主备集群版本号必须相同。
- 容灾状态下主备集群升级：
  - 主备集群都要处于normal状态。
  - 主备集群升级时大版本升级会校验主备集群版本号是否相同，不相同不可升级。小版本升级不校验。
  - 不支持就地升级。
- 主集群频繁vacuum会导致灾备集群读报错或影响日志回放速度。
- 主集群频繁DDL会导致灾备集群读报错(ERROR: could not open relation with OID)。
- 主集群做DDL操作，备集群回放DDL日志的时候，可能DDL在备集群上的某些节点中还没回放成功，报错并结束查询(ERROR: current snapshot is invalid for this partition/relation)。
- 主集群频繁vacuum会导致灾备集群查询的数据被提前清理掉等问题，可能出现查询和回放冲突，报错并结束查询(ERROR: gtm csn small)。
- 容灾集群做COPY TO操作，由于容灾集群的CN是备机不能分配XID，COPY TO在CN上需要申请XID，执行时报错(ERROR: cannot assign TransactionIds during recovery)。

- 主集群启停或故障、极致RTO模式下由于某种原因(磁盘空间不足、修改GUC参数等)触发了强制回收，均可能导致灾备集群报错结束查询(ERROR: get_lsn_from_disaster_csn_info_list fail)。
- 主备集群异构、CN数量不一致时，进行灾备集群读之前，应先查询出与主集群存在映射关系的灾备集群CN，再根据查询出的CN进行相关业务操作（查询方式：select * from pgxc_disaster_read_status()）。
- 如果主集群与容灾集群之间出现断连，那么从容灾集群上查询到的数据可能是旧数据。
- 如果使用JDBC连接容灾集群，且开启了负载均衡功能，那么查询请求会发送到主集群上，导致查询性能会受到影响。
- 当主备集群查询所能占用到的资源相同时，对于并行回放下的astore表：
  - 在主机无写业务、备机的日志都回放完成的场景下，对于select count(*) from xxx语句的负载，从时延的维度上观察，备集群的查询性能达到主集群的80%。
  - 在带有写业务的场景下：主集群和容灾集群的查询性能可能都会受业务影响而有波动，单次查询不保证容灾集群的性能达到主集群的80%；在无DDL的场景下，对于sysbench和TPCC的负载，从TPS、第95百分位数和平均时延的维度上观察，容灾集群的性能达到主集群的80%。
  - 如果有多个查询的主节点或备节点部署在同一个机器上，因这个机器的资源不足导致查询性能下降的情况下，则上面两个场景不成立。
  - 当发现前两个场景不成立时，建议检查容灾集群中各节点是否存在资源(CPU、内存、磁盘)受限的情况。
  - 由于容灾集群在并行回放下的astore表不支持index only scan查询，因而对于select count(*) from xxx语句的负载可能需要读取更多的数据页面，建议通过调整shared_buffers参数进行缓存优化。

依赖关系

无。

1.4.12 按分片自动升降副本

可获得性

本特性自V500R001C20版本开始引入。

特性简介

两AZ+仲裁AZ集群部署方式支持自动升降副本功能。

客户价值

金融、银行等业务需要提供半数及以上节点数据故障业务快速恢复能力。

特性描述

金融、银行业务需要极高的容灾能力。当一个分片故障了半数及以上节点时，DN执行写操作会超时，主要是同步备中有节点故障了，无法执行写操作。为了降低分片故障对业务的影响，需要对分片上可用节点进行降副本操作，当检测到故障恢复后，自动触发升副本操作。

特性增强

无。

特性约束

- 基础保障(最多降至)：一主一备。
- 集群部署要求：两AZ+仲裁AZ集群，副本数大于3，分片总数（DN+CN+GTM）小于 64。
- 前提要求：DN主存在。
- 升级、扩容阶段，或者ETCD不可用时不会进行降副本操作。
- 只有半数以上DN发生故障（DN状态是down），且状态持续，才会进行降副本操作，升副本需要等待半数故障恢复后，且状态持续，才会自动升副本。
- 只有当上一轮降副本操作执行成功后，才能进行下一轮降副本操作，不支持二次故障。
- 不支持故障跳转，比如四个副本，第一次（3, 4）故障后，（1, 2）进行完降副本，第二次故障（1, 2），恢复（3, 4），此时集群不可用，无法选出主，且不能对（3, 4）进行降副本。
- 在升降副本结束后，才能执行switchover，且switchover 只能切换到同步列表中的备DN上。
- 升副本要求：故障节点恢复后，需要跟主机同步达到99%，才会被重新加入到主机的同步列表中。

依赖关系

无。

1.4.13 支持 global syscache

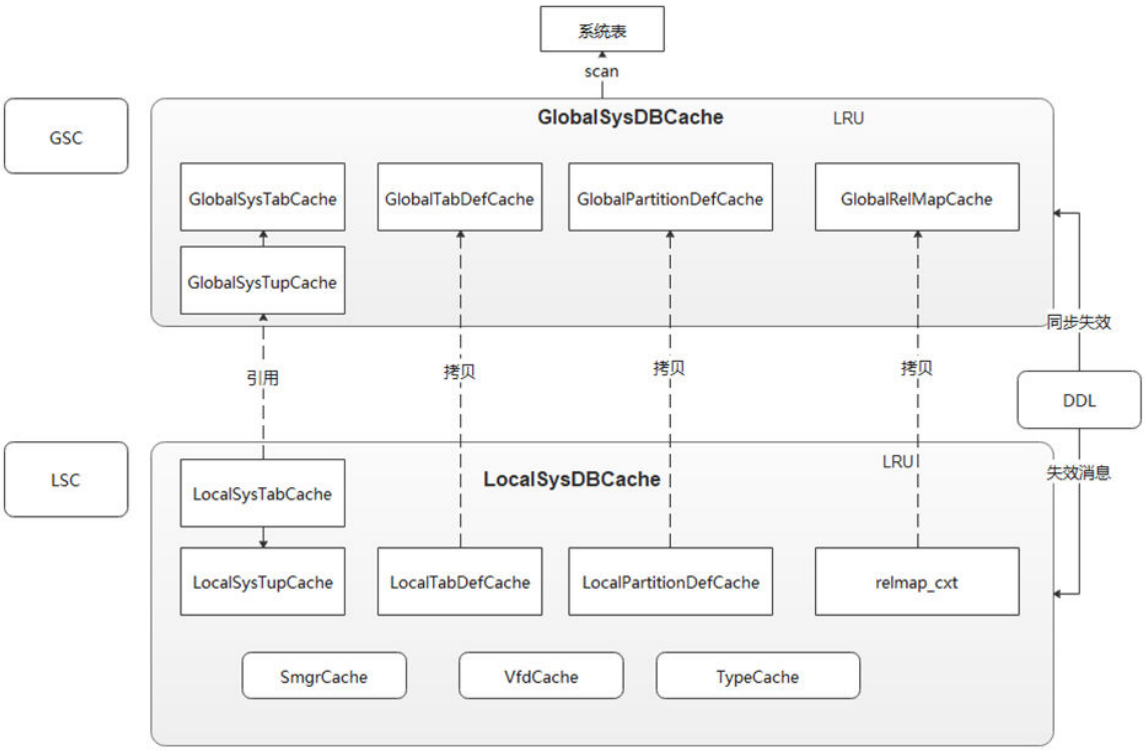
可获得性

本特性自V500R002C10版本开始引入。

特性简介

全局系统缓存（Global SysCache）是系统表数据的全局缓存和本地缓存。原理如图1-7所示。

图 1-7 Global SysCache 原理图



客户价值

全局系统缓存特性可以降低数据库进程的缓存内存占用，提升数据库的并发扩展能力。

特性描述

全局系统缓存特性指将系统缓存与会话解耦，绑定到线程上，结合线程池特性达到降低内存占用的目的，同时结合全局缓存，提升缓存命中率，保持性能稳定。

特性增强

支持更高的并发查询。

特性约束

- 当设置enable_global_syscache为on时，建议设置enable_thread_pool参数为on。
- 当DB数较多，且阈值global_syscache_threshold较小时，内存控制无法正常工作，性能会劣化。
- 不支持分布式时序相关的任务，这些任务的内存控制与性能不受GSC特性的影响。
- wal_level设置为minimal或者archive时，备机的查询性能会下降，会退化为短连接。

依赖关系

该特性降内存能力依赖于线程池特性。

1.4.14 并行逻辑解码

可获得性

本特性自V500R002C10版本开始引入。

特性简介

支持多线程并行解码。

客户价值

大幅提升逻辑解码性能，解码速度由3~5MBps可提升到标准场景（16核CPU、内存128G、网络带宽 > 200MBps、表的列数为10~100、单行数据量0.1KB~1KB、DML操作以insert为主、不涉及落盘事务即单个事务中语句数量小于4096）下的100MBps。

特性描述

在使用JDBC或pg_recvlogical解码时，设置配置选项parallel-decode-num为大于1且小于等于20的值，开启并行解码特性，使用一个读取线程、多个解码线程以及一个发送线程协同进行逻辑解码操作，显著提升解码速度。

特性增强

无。

特性约束

1. 当前的硬件和网络环境正常；由于逻辑日志一般为xLog的两倍，为保证xLog速度达到100MBps，I/O带宽至少保证200MBps；因为reader、decoder、sender线程均需预留资源，CPU需预留并发数+2的核数，如4并发场景需要预留6核。在实际场景中，使用备机解码即可保证需求，无需进行特殊的资源预留规划。为保证解码性能达标以及尽量降低对业务的影响，一台备机上应尽量仅建立一个并行解码连接，保证CPU、内存、带宽资源充足。
2. 日志级别的guc参数wal_level = logical。
3. guc参数max_replication_slots >= 每个DN所需的（物理流复制槽数+备份槽数+逻辑复制槽数）。
4. 解码配置选项parallel-decode-num > 1且<= 20，指定并行的解码线程数。
5. 逻辑解码不支持DDL。
6. 不支持数据页复制的DML解码。
7. 不支持解码分布式事务，当前机制为从DN解码，无法保证分布式事务一致性解码。
8. 单条元组大小不超过1GB，考虑解码结果可能大于插入数据，因此建议单条元组大小不超过500MB。
9. 不支持压缩表的DML语句解码。
10. GaussDB支持解码的数据类型为：INTEGER、BIGINT、SMALLILNT、TINYINT、SERIAL、SMALLSERIAL、BIGSERIAL、FLOAT、DOUBLE PRECISION、DATE、

TIME[WITHOUT TIME ZONE]、TIMESTAMP[WITHOUT TIME ZONE]、CHAR(n)、VARCHAR(n)、TEXT。

11. 在需要ssl连接的场景，需要前置条件保证guc参数ssl = on。
12. 不支持interval partition表DML复制。
13. 在事务中执行DDL语句时，该DDL语句之后的语句不会被解码。
14. 如需进行备机解码，需在对应主机上设置guc参数enable_slot_log = on。
15. 当前不支持超大CLOB解码。
16. 不允许主备，多个备机同时使用同一个复制槽解码，否则会产生数据不一致。
17. 禁止在使用逻辑复制槽时在其他节点对该复制槽进行操作，删除复制槽的操作需在该复制槽停止解码后执行。

依赖关系

依赖备机解码。

1.4.15 支持备机 build 备机

可获得性

本特性自V500R002C10版本开始引入。

特性简介

备机build备机加快备机故障的恢复， 减小主机I/O和带宽压力。

客户价值

当业务压力过大时，从主机build备机会对主机的资源造成影响，导致主机性能下降、build变慢的情况。使用备机build备机不会对主机业务造成影响。

特性描述

使用gs_ctl命令可以指定对应的备机去build需要修复的备机。具体操作可参考《 工具参考 》中的“系统内部调用的工具 > gs_ctl”章节。

特性增强

无。

特性约束

只支持备机build备机，只能使用指定ip和port的方式做build，同时在build前应确保需要修复备机的日志比发送数据的备机的日志落后。

依赖关系

无。

1.4.16 3AZ 多数派故障一键式强启及加回

可获得性

本特性自V500R001C20版本开始引入。

特性简介

AZ多数派故障场景下，一键式执行少数派强起命令，实现业务正常运行的目的。

客户价值

AZ多数派故障场景下，能够通过强起操作继续提供集群仲裁服务，使业务运行尽快恢复。

特性描述

部署同城双活集群（2AZ+仲裁AZ）时，在园区A部署一个生产中心AZ1，在同城园区B部署一个灾备中心AZ2。AZ1、AZ2都有完整的数据并且均部署第三方仲裁实例ETCD，ETCD数量相等， AZ3中只部署一个ETCD实例。ETCD多数派存活时，可正常工作，集群CMS实例可以正常选主。业务正常执行。当AZ2、AZ3发生故障时，AZ1中存活的ETCD数量少于总数的二分之一，即少数派存活，此时无法实现AZ的自动切换，因此要进行少数派强起。

特性增强

无。

特性约束

- 少数派强启命令属于高危操作，必须是满足多数派所有节点同时故障的情况下才能进行强启操作，即：少数派强启之前，一定要确认好强启AZ是否一直是集群内唯一一个没有故障的AZ，而不是先故障的AZ。

  说明

  场景举例：

  集群AZ1与AZ2、AZ3发生了网络隔离，AZ2,AZ3满足多数派，可以继续执行业务，AZ1由于网络隔离，数据都无法从多数派（AZ2、AZ3）同步。

  此后如果AZ1、AZ2又都发生了故障，不再满足多数派，这个时候，禁止少数派强启AZ1的原因：

  AZ1由于很早之前就故障了，数据不是最新的，少数派强启虽然会执行成功，但会发生数据丢失。
- 多数派故障后处理流程和多数派恢复后恢复流程处理过程中，需要将业务停止。
- 故障的CN没有剔除前，此时不允许执行DDL，否则会出现gsql连上做DDL，ctrl+c和pg_cancel_backend都无法停止。

  同样在恢复时，CN没有完全处理完成时，也不允许做DDL操作。
- 一键式强启执行之前，请务必认真确认当前是否确实是多数派故障场景，如果不满足多数派故障条件，强启后可能会产生数据不一致的问题。
- 一键式脚本限制执行命令节点所在AZ 到其他AZ 均无法正常连接时才会继续执行（防止用户对多数派故障的错误判断）。

- 少数派强启需求当前强启AZ内的节点不再有新的故障，否则在多数派故障的情况下再次叠加故障，可能会导致脚本无法正常执行。
- 不支持在logger节点执行一键式强启加回。
- 如果强启AZ中的DN数量小于半数，则强启无法保证RPO。
- 执行加回之前，如果集群current_az 不是 AZ_ALL，需要等到其为AZ_ALL的时候才能操作。
- 网络恢复后，由于ETCD的启动，会导致集群状态会变得不准确（如被剔除CN会变成normal），需要调用加回之后才能恢复。
- 强启运行在降级模式中，不支持扩容、升级、节点修复等工程能力。
- DCF自仲裁模式不支持一键式强启加回。

依赖关系

无。

1.4.17 分布式备机支持读

可获得性

本特性自503.0.0版本开始引入。

特性简介

提供分布式GTM-Free/GTM-Lite模式备机读能力，降低主机负载。

原理简介：分布式的场景下，通过CN中的收集线程不断向主DN与备DN收集csn和commit time，用于校验备机的csn和时延是否满足要求，进而选择一个备DN可以把部分读请求下发到备节点。

高可用机制：

- 一主多备场景下，当某个备机因网络、进程故障导致无法连接，CN后台收集线程会选取一个合适的DN备作为备机读节点，并且将备机读业务迁移至此备节点上。
- 一主一备一logger部署形态下，备机因网络、进程故障导致无法连接，无法提供备机读服务备机读业务报错，需要备DN故障恢复后恢复备机读业务。

客户价值

分布式集群中副本无法提供服务, 造成了极大浪费。通过提供副本可读的能力,来提升业务的吞吐量和资源利用率。

特性描述

- 支持分布式GTM FREE场景下的备机读：在与CN建立连接后，打开session级的guc参数enable_standby_read或JDBC参数enableStandbyRead，即可把后续的读请求发送给备机。
- 提供分布式GTM LITE场景下的备机读能力，可通过客户端直连备DN进行读操作。

特性增强

无。

特性约束

- GTM LITE模式下，需要配置参数listen_address_ext, 不支持同时开启极致RTO。
- 仅支持hot standby模式。
- 读写分离仅支持session级别，如果在备机读session执行写操作，会报错。不支持备机异常后，业务切换到主机上。
- 备机读到的数据与主机存在一定延时，主备差异超过指定数值（由standby_read_delay指定）时，会报错。
- 提供弱一致读，仅保证session内部的快照递增序，session之间不保证。
- 在串行回放或者并行回放的模式下，主机频繁vacuum会导致备机读报错或影响日志回放速度，需要调节备机上的参数选择哪种优先。具体调节参数为max_standby_archive_delay和max_standby_streaming_delay，分别表示备机回放本地日志和streamed日志时，取消查询的最大延时。
- 在极致RTO的回放模式下，查询最大允许的时长由参数standby_max_query_time调节。
- DDL操作会修改元数据，导致主备元数据不一致。此时备机读操作时会提示“relation \"<string>\" does not exist”、“column \"<string>\" does not exist”或其他错误信息，需等待备机元数据回放到与主机元数据一致时备机读才可用。
- 主备切换、备机异常重启、备机加回后或加副本后，需要回放到上次读的位置，才能支持备机读。
- 不支持stream计划，开启enable_stream_operator参数后，备机读会采用pgxc计划替代stream计划。
- 不支持备机间的负载均衡。
- 主机性能和RTO优先，备机读业务其次。在没有可用备机时均会报错，不会发送到主机上。
- 备机读业务会影响备机RTO，若所有备机的RTO都超过指定阈值（通过standby_read_rto指定），备机读业务报错，业务侧通过查看错误码进行熔断。
- enable_standby_read参数不支持在事务块中开启。
- 允许GTM-Free模式下gsql直接连接备机。
- 当开启极致RTO、进行删除文件的DDL操作时，备机会延迟删除相关文件，从而导致备机数据目录下的文件数目多于主机，此时若查询主备机数据库大小，则会显示备机数据库大于主机数据库。
- 分布式备机读不支持自治事务功能。

依赖关系

无。

1.4.18 ETCD 多数派故障一键修复

可获得性

本特性自505.1.0版本开始引入。

特性简介

ETCD实例多数派故障，并且不可恢复场景下，为了恢复集群健康可用，可调用ETCD多数派故障一键修复接口，完成集群的修复。

客户价值

ETCD实例多数派故障，并且不可恢复场景下，能够通过一键修复操作继续提供实例仲裁服务，使业务运行尽快恢复。

特性描述

部署多节点实例。ETCD多数派存活时，可正常工作，业务正常执行。当ETCD少数派存活，此时无法使用gs_replace修复ETCD，因此要进行ETCD一键修复。

特性增强

无

特性约束

- 需要确认ETCD已无法自恢复，否则会导致ETCD内数据丢失。
- ETCD少数派故障不能使用一键修复进行ETCD的修复。
- DCC模式不支持ETCD多数派故障一键修复。
- 集中式单节点不支持ETCD多数派故障一键修复。

依赖关系

无

1.5 可维护性

1.5.1 热补丁升级

可获得性

本特性自V300R002C00版本开始引入。

特性简介

热补丁将补丁以patch文件的形式加载到正在运行的集群进程中，达到零中断修复线上系统的目的。

客户价值

热补丁最大的优势是业务零中断加载补丁，他可以在不影响业务的前提下在线解决一部分数据库内核的紧急问题。

其价值主要体现在如下两点：

- 缩短版本发布时间，紧急问题从版本回归验证轻量化为补丁回归验证，提高了线上紧急问题的响应速度。
- 热补丁的加载，卸载对业务无感知，提高了客户满意度。

特性描述

热补丁基于发布的代码版本生成补丁文件，然后以模块的形式插入到数据库内核运行地址空间中，通过寻找热补丁目标函数的地址，并动态地，原子地替换入口地址，重定向函数代码段至补丁文件代码段达到修复线上系统缺陷的目的。

- 热补丁的制作通过修复特定缺陷函数，制作成模块，动态地加载到运行中的内核系统。
- 热补丁找到目标函数，并在目标函数的入口处加入跳转指令，当目标函数被调用时，跳转到补丁区执行补丁函数。
- 目标函数的替换和还原是原子操作CPU寄存器，热补丁可以随时随地加载和卸载，线上系统无需中断，即随时可运行最新的代码。

特性增强

V500R001C00版本增加了对ARM CPU的支持。

特性约束

无。

依赖关系

无。

1.5.2 灰度升级

可获得性

本特性自V300R002C00版本开始引入。

特性简介

支持按照用户定义的升级顺序，进行节点级滚动灰度升级。

客户价值

通过灰度升级，可以达成以下目的：

- 先升级部分备DN节点，即使升级失败，也不会对业务产生影响。
- 先升级部分CN节点，用户将业务路由到没有升级的其余CN节点，保证升级过程中业务不中断。
- 先升级业务影响小的组件节点（如ETCD、CMS），即使升级失败，也能将对业务的影响控制在最小范围。
- 每批节点升级完之后，均提供升级观察窗口，验证升级状态，动态评估升级的风险。

特性描述

灰度升级是一种支持优先升级部分节点的在线升级方式。灰度升级主要包含以下三个方面：

1. 对于大版本升级涉及的系统表变更，将不同版本的系统表结构和系统函数固化在二进制中，保证新、老版本二进制均能解析和使用新、老版本的系统表元组。
2. 对于大版本升级和二进制升级涉及的新、老二进制替换，先灰度替换指定节点上的二进制，待系统运行一定时间之后，再替换剩余节点的二进制
3. 在第2点的基础之上，如果升级亦涉及到节点的操作系统、硬件升级（且不能提前执行），那么在灰度升级部分节点之前，先将这些节点上的主实例全部切换到非灰度升级的节点上；如果升级只涉及数据库二进制的替换，为了尽可能降低对于业务的影响，采用同一节点两套二进制同时存在的方式，使用软连接切换的方式来进行进程版本的切换升级（闪断一次，10秒以内）

特性增强

无。

特性约束

灰度升级的约束条件请参见《升级指导书》中“升级工具介绍 > gs_upgradectl”章节。

依赖关系

无。

1.5.3 滚动升级

可获得性

本特性自V500R001C10版本开始引入。

特性简介

滚动升级支持全业务操作，采用滚动升级DN分片的方式对集群进行升级，用户可以指定部分DN分片先升级。

客户价值

- 支持数据库的大版本升级和小版本升级（内核版本号不变的升级方式为小版本升级，否则就是大版本升级）。
- 支持在线升级，允许优先升级部分节点。升级每个分片时，都会产生不超过5s的业务中断。升级最后一个分片，会中断2-3次业务：一次由于分片DN的切换，一次由于CN的切换；不开启gtm_free时，还会额外产生一次业务中断。

特性描述

1. 对于大版本升级涉及的系统表变更，将不同版本的系统表结构和系统函数固化在二进制中，保证新、老版本二进制均能解析和使用新、老版本的系统表元组；
2. 对于大版本升级和二进制升级涉及的新、老二进制替换，先替换指定分片上的二进制，待系统运行一定时间之后，再替换剩余分片的二进制。

3. 如果升级只涉及数据库二进制的替换，为了尽可能降低对于业务的影响，采用同一节点两套二进制同时存在的方式，使用软连接切换的方式来进行进程版本的切换升级。

特性增强

无。

特性约束

滚动升级的约束条件请参见《升级指导书》中“升级工具介绍 > gs_upgradectl”章节。

依赖关系

无。

1.5.4 就地升级

可获得性

本特性自V300R002C00版本开始引入。

特性简介

就地升级是一种离线升级方式。

客户价值

- 支持数据库的大版本升级和小版本升级（内核版本号不变的升级方式为小版本升级，否则就是大版本升级）。
- 提供一种相对稳定可靠的升级方式。

特性描述

就地升级是一种离线升级方式。升级过程中需要停止业务，不提供任何服务，会一次性升级集群中的所有节点。

特性增强

无。

特性约束

就地升级的约束条件请参见《升级指导书》中“升级工具介绍 > gs_upgradectl”章节。

依赖关系

无。

1.5.5 支持 WDR 诊断报告

可获得性

本特性自V300R002C00版本开始引入。

特性简介

WDR报告提供集群性能诊断报告，该报告基于基线性能数据和增量数据两个版本，从性能变化得到性能报告。

客户价值

- WDR报表是长期性能问题最主要的诊断手段。基于SNAPSHOT的性能基线，从多维度做性能分析，能帮助DBA掌握系统负载繁忙程度、各个组件的性能表现及性能瓶颈。
- SNAPSHOT也是后续性能问题自诊断和自优化建议的重要数据来源。

特性描述

WDR(Workload Diagnosis Report)基于两次不同时间点系统的性能快照数据，生成这两个时间点之间的性能表现报表，用于诊断数据库内核的性能故障。

使用generate_wdr_report(...) 可以生成基于两个性能快照的性能报告。

WDR性能快照数据存储在postgres库的snapshot schema下，默认的采集和保存策略为：

- 每小时采集一个快照(wdr_snapshot_interval=1h)。
- 每十二个快照中有一个全量快照（ wdr_snapshot_full_backup_interval=12 ）。
- 保留8天(wdr_snapshot_retention_days=8)。
- 不启用空间维度控制阈值（ wdr_snapshot_space_threshold=0 ）。

WDR主要依赖两个组件：

- SNAPSHOT性能快照：性能快照可以配置成按一定时间间隔从内核采集一定量的性能数据，持久化在用户表空间。任何一个SNAPSHOT可以作为一个性能基线，其他SNAPSHOT与之比较的结果，可以分析出与基线的性能表现。
- WDR Reporter：报表生成工具基于两个SNAPSHOT，分析系统总体性能表现，并能计算出更多项具体的性能指标在这两个时间段之间的变化量，生成SUMMARY和DETAIL两个不同级别的性能数据。如**表1-2**、**表1-3**所示。

表 1-2 SUMMARY 级别诊断报告

诊断类别	描述
Database Stat	主要用于评估当前数据库上的负载，I/O状况，负载和I/O是衡量TP系统最重要的特性。 包含当前连接到该数据库的session，提交、回滚的事务数，读取的磁盘块的数量，高速缓存中已经发现的磁盘块的次数，通过数据库查询返回、抓取、插入、更新、删除的行数，冲突、死锁发生的次数，临时文件的使用量，I/O读写时间等。

诊断类别	描述
Load Profile	从时间，I/O，事务，SQL几个维度评估当前系统负载的表现。 包含作业运行elapse time、CPU time，事务日志量，逻辑和物理读的量，读写I/O次数、大小，登录或者退出登录次数，SQL、事务执行量，SQL P80、P95响应时间等。
Instance Efficiency Percentages	用于评估当前系统的缓存的效率。 主要包含数据库缓存命中率。
Events	用于评估当前系统内核关键资源，关键事件的性能。 主要包含数据库内核关键事件的发生次数，事件的等待时间。
Wait Classes	用于评估当前系统关键事件类型的性能。 主要包含数据库内核在主要的等待事件的种类上的发布：STATUS、LWLOCK_EVENT、LOCK_EVENT、IO_EVENT。
CPU	主要包含CPU在用户态、内核态、Wait IO、空闲状态下的时间发布。
IO Profile	主要包含数据库Database IO次数、Database IO数据量、Redo IO次数、Redo IO量。
Memory Statistics	包含最大进程内存、进程已经使用内存、最大共享内存、已经使用共享内存大小等。

表 1-3 DETAIL 级别诊断报告

诊断类别	描述
Time Model	主要用于评估当前系统在时间维度的性能表现。 包含系统在各个阶段上消耗的时间：内核时间、CPU时间、执行时间、解析时间、编译时间、查询重写时间、计划生成时间、网络时间、IO时间。
SQL Statistics	主要用于SQL语句性能问题的诊断。 包含归一化的SQL的性能指标在多个维度上的排序：Elapsed Time、CPU Time、Rows Returned、Tuples Reads、Executions、Physical Reads、Logical Reads。这些指标的种类包括：执行时间，执行次数、行活动、Cache IO等。
Wait Events	主要用于系统关键资源，关键事件的详细性能诊断。 包含所有关键事件在一段时间内的表现，主要是事件发生的次数，消耗的时间。
Cache IO Stats	用于诊断用户表和索引的性能。 包含所有用户表、索引上的文件读写，缓存命中。
Utility status	用于诊断后台任务性能，包含复制等后台任务的性能。

诊断类别	描述
Object stats	用于诊断数据库对象的性能。 包含用户表、索引上的表、索引扫描活动，insert、update、delete活动，有效行数量，表维护操作的状态等。
Configuration settings	用于判断配置是否有变更。 包含当前所有配置参数的快照。
SQL detail	显示unique query text信息。

特性增强

无。

特性约束

- WDR snapshot性能快照会采集不同database的性能数据，如果集群中有大量的database或者大量表，做一次WDR snapshot会花费很长时间。
- 如果在大量DDL期间做WDR snapshot可能会造成WDR snapshot失败。
- 在drop database时，做WDR snapshot可能会造成WDR snapshot失败。
- 生成WDR报告的两次快照期间进行过降副本、节点重启和主备切换等操作，则无法生成WDR报告。
- 集群只读状态会造成WDR snapshot失败。

依赖关系

无。

1.5.6 支持 ASP 报告

可获得性

本特性自503.2.0版本开始引入。

特性简介

ASP报告提供会话级别的报告，报告基于活跃会话采样的样本进行分析，从SQL、Session、Wait event维度分析得到ASP报告。

客户价值

- ASP报告是针对短暂的性能抖动问题主要的诊断手段，基于活跃会话采样的样本，从SQL、Session、Wait event维度分析，能帮助DBA掌握系统负载变化情况以及性能瓶颈。
- ASP活跃样本数据也是后续性能问题自诊断和自优化建议的重要数据来源。

特性描述

ASP报告（Active Session Profile）基于活跃会话采样的样本进行分析，用于诊断数据库内核的短暂性能故障。

使用generate_asp_report(...) 可以生成基于活跃会话样本性能报告。

后台线程每秒会采集数据库中的所有活跃会话信息，并保存在内存（asp_sample_num控制内存中最大保留样本数）中，在内存达到上限后，ASP样本数据会被存储在postgres库的gs_asp表中，在从内存转储到表中时可以通过asp_flush_rate控制落盘时的二次采样率，默认的采集和保存策略为：

- 内存保留的最大样本数（asp_sample_num=100000）。
- 样本从内存中刷到磁盘上采样的比例默认是10：1（asp_flush_rate=10）。

特性增强

无。

特性约束

- ASP是按照1秒进行采样，内存达到阈值后并按照1：10采样的比例持久化到表中，所以对于表的数据的最小采样粒度为10秒，对于某些执行较快的SQL，ASP可能采不到，则ASP报告中不会展示。
- ASP中的SQL text信息来自于unique SQL，DN上unique SQL不会记录SQL text信息，所以DN的ASP报告SQL text为空，另外如果instr_unique_sql_count太小导致unique hash被占满时，新的语句就不会保存sql text，会导致ASP报告中部分SQL的text为空。
- 如果生成ASP报告的时间段过大、时间段内并发数过多、或者slot数过大，可能都会导致数据量过大，查询速度变慢，ASP报告生成较慢。
- 报告中显示slot的时间段的粒度默认只能到gs_asp表的10s的粒度，所以slot的个数可能比传入参数slots少。
- 目前报告中展示的数据都是去掉部分后台线程后分析的结果，屏蔽的后台线程有undo recycler，workload，Asp，PercentileJob，JobScheduler，Wal Writer，CheckPointer，WDR相关线程。
- ASP报告需要引入开源软件ECHARTS，该软件支持Chrome、edge浏览器，不支持IE浏览器。
- ASP报告生成不在一个事务内，如果传入的start time比ASP样本的最旧时间小，生成报告过程中ASP旧数据可能被回收，则整个ASP报告的sample count可能不一致。
- ASP报告不支持备机。

依赖关系

无。

1.5.7 慢 SQL 诊断

可获得性

本特性自V300R002C00版本开始引入，V500R001C20增强了该特性。

V500R001C10版本的慢SQL相关视图已废弃，包括：dbe_perf. gs_slow_query_info、dbe_perf.gs_slow_query_history、dbe_perf.global_slow_query_hisotry、dbe_perf.global_slow_query_info。

特性简介

慢SQL诊断提供诊断慢SQL所需要的必要信息，帮助开发者回溯执行时间超过阈值的SQL，诊断SQL性能瓶颈。

客户价值

慢SQL提供给用户对于慢SQL诊断所需的详细信息，用户无需通过复现就能离线诊断特定慢SQL的性能问题。表和函数接口方便用户统计慢SQL指标，对接第三方平台。

特性描述

慢SQL能根据用户提供的执行时间阈值(log_min_duration_statement)，记录所有超过阈值的执行完毕的作业信息。

慢SQL提供表和函数两种维度的查询接口，用户从接口中能查询到作业的执行计划、开始、结束执行时间、执行查询的语句、行活动、内核时间、CPU时间、执行时间、解析时间、编译时间、查询重写时间、计划生成时间、网络时间、I/O时间、网络开销、锁开销等。所有信息都是脱敏的。

特性增强

增加对慢SQL指标信息、安全性（脱敏）、执行计划、查询接口的增强。

特性约束

1. 目前的SQL跟踪信息，基于正常的执行逻辑。执行失败的SQL，其跟踪信息不具有准确的参考价值，例如：状态为cancelled等。
2. 节点重启，可能导致该节点的数据丢失。
3. 通过GUC参数设置收集SQL语句的数量，如果超过阈值，新的SQL语句执行信息不会被收集。
4. 通过GUC参数设置单条SQL语句收集的锁事件详细信息的最大字节数，如果超过阈值，新的锁事件详细信息不会被收集。
5. 当track_stmt_parameter为off时，query字段最大长度受track_activity_query_size控制。
6. 通过异步刷新方式刷新用户执行中的SQL信息，所以用户Query执行结束后，存在查询相关视图函数结果短暂时延。
7. 部分指标信息(行活动、Cache/IO、时间分布等)依赖于dbe_perf.statement视图收集，如果该视图对应记录数超过预定大小(依赖GUC:instr_unique_sql_count)，则本特性可能不收集相关指标。
8. statement_history表相关函数以及视图、备机dbe_perf.standby_statement_history中details字段为二进制格式，如果需要解析详细内容，请使用对应函数：pg_catalog.statement_detail_decode(details, 'plaintext', true)。
9. 通过GUC参数track_stmt_stat_level打开full SQL功能时会影响性能，并且可能会占用大量的磁盘空间。

10. statement_history表查询需要切换至postgres库，其它库中数据为空。
11. 备机dbe_perf.standby_statement_history函数查询需要切换至postgres库，其它库中查询会提示不可用。
12. statement_history表以及备机dbe_perf.standby_statement_history函数内容受track_stmt_stat_level控制，默认为'OFF,L0'，参数第一部分代表full SQL，第二部分是慢SQL；对于慢SQL,只有SQL运行时间超过log_min_duration_statement时才会被记录至statement_history表。
13. 当track_stmt_stat_level关闭full SQL时，SQL等锁超时可能会导致表中的query_plan信息为空，可通过detail字段内的wait event辅助定位分析。
14. 当track_stmt_flush_mode参数取值为"MEMORY,FILE"时，开启内核支持全量SQL功能，full SQL语句存储到内存中，slow SQL语句存储到磁盘文件中。开启功能后性能劣化不超过5%。由于全量SQL默认关闭，开启后仅占用固定内存大小，故当前版本全量SQL大小共享内存暂不受max_process_memory控制。在升级未提交期间，若版本过低，将无法开启全量SQL功能，并记录提示信息至日志。
15. 全量SQl采用共享内存方式存储full/slow SQL，gs_shared_mem_kpi.meta文件存储了共享内存访问信息，可通过System V共享内存的接口shmat访问，当前仅用于对接管控SQL全链路相关功能。

依赖关系

无。

1.5.8 Session 性能诊断

可获得性

本特性自V500R001C00版本开始引入。

特性简介

Session性能诊断提供给用户Session级别的性能问题诊断。

客户价值

- 查看最近用户Session最耗资源的事件。
- 查看最近比较占资源的SQL把资源都消耗在哪些等待事件上。
- 查看最近比较耗资源的Session把资源都花费在哪些等待事件上。
- 查看最近最耗资源的用户的信息。
- 查看过去Session相互阻塞的等待关系。

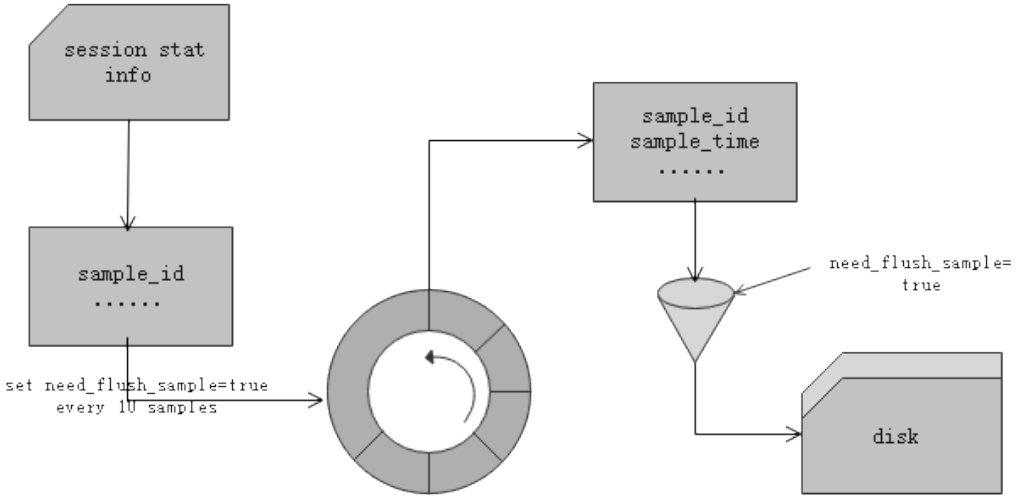
特性描述

Session性能诊断提供对当前系统所有活跃Session进行诊断的能力。由于实时采集所有活跃Session的指标对用户负载的影响加大，因此采取Session快照的技术对活跃Session的指标进行采样。从采样中统计出活跃Session的统计指标，这些统计指标从客户端信息、执行开始、结束时间，SQL文本，等待事件，当前数据库对象等维度，反映活跃Session的基本信息，状态，持有的资源。基于概率统计的活跃Session信息，可以帮助用户诊断系统中哪些Session消耗了更多的CPU、内存资源，哪些数据库对象是热对象，哪些SQL消耗了更多的关键事件资源等，从而定位出有问题Session，SQL，数据库设计。

Session采样数据分为两级，如图1-8所示：

1. 第一级为实时信息，存储在内存中，展示最近几分钟的活跃Session信息，具有最高的精度；
2. 第二级为持久化历史信息，存储在磁盘文件中，展示过去很长一段时间的历史活跃Session信息，从内存数据中抽样而来，适合长时间跨度的统计分析。

图 1-8 Session 性能诊断原理



部分使用场景如下所示：

1. 查看session之间的阻塞关系。
```
select sessionid, block_sessionid from pg_thread_wait_status;
```
2. 采样blocking session信息。
```
select sessionid, block_sessionid from DBE_PERF.local_active_session;
```
3. Final blocking session展示。
```
select sessionid, block_sessionid, final_block_sessionid from DBE_PERF.local_active_session;
```
4. 最耗资源的wait event。
```
SELECT s.type, s.event, t.count
FROM dbe_perf.wait_events s, (
SELECT event, COUNT(*)
FROM dbe_perf.local_active_session
WHERE sample_time > now() - 5 / (24 * 60)
GROUP BY event
)t WHERE s.event = t.event ORDER BY count DESC;
```
5. 查看最近五分钟较耗资源的session把资源都花费在哪些event上。
```
SELECT sessionid, start_time, event, count
FROM (
SELECT sessionid, start_time, event, COUNT(*)
FROM dbe_perf.local_active_session
WHERE sample_time > now() - 5 / (24 * 60)
GROUP BY sessionid, start_time, event
) as t ORDER BY SUM(t.count) OVER (PARTITION BY t. sessionid, start_time)DESC, t.event;
```
6. 最近五分钟比较占资源的SQL把资源都消耗在哪些event上。
```
SELECT query_id, event, count
FROM (
SELECT query_id, event, COUNT(*)
FROM dbe_perf.local_active_session
WHERE sample_time > now() - 5 / (24 * 60)
GROUP BY query_id, event
) t ORDER BY SUM(t.count) OVER (PARTITION BY t.query_id ) DESC, t.event DESC;
```

特性增强

无。

特性约束

无。

依赖关系

无。

1.5.9 系统 KPI 辅助诊断

可获得性

本特性自V300R002C00版本开始引入。

特性简介

KPI是内核组件或者整体性能关键指标的视图呈现，基于这些指标，用户可以了解到系统运行的实时或者历史状态。

客户价值

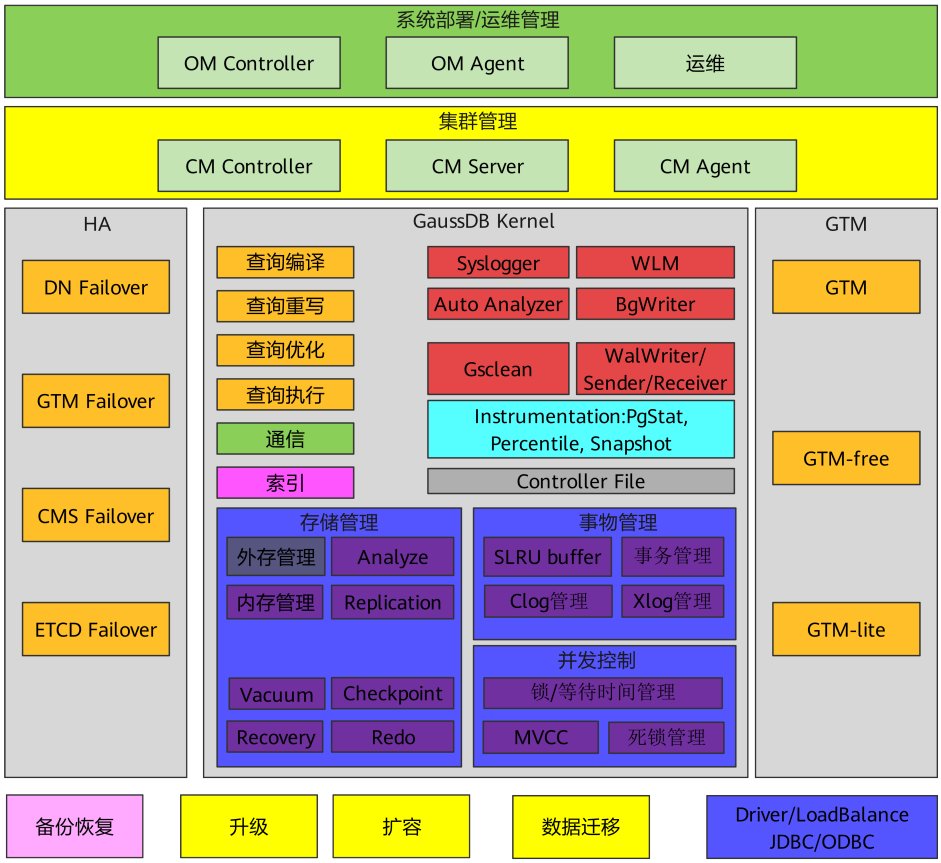
- 系统负载概要诊断
  系统负载异常（过载、失速、业务SLA）准确告警，系统负载精准画像。
- 系统时间模型概要诊断
  Instance和Query级别时间模型细分，诊断Instance和Query性能问题根因。
- Query性能诊断
  集群级Query概要信息，TopSQL，SQL CPU，I/O消耗，执行计划，硬解析过多。
- 磁盘I/O、索引、buffer性能问题
- 连接池，线程池异常
- Checkpoint，Redo（RTO）性能问题
- 系统I/O、LWLock、Waits性能问题诊断
  诊断60+模块，240+关键操作性能问题。
- 函数级性能看护诊断(GSTRACE)，功能诊断
  50+存储和执行层函数trace。

特性描述

GaussDB提供涵盖11大类，26个子类的KPI，包括：Instance、File、Object、Workload、Communication、Session、Thread、Cache IO、Lock、Wait Event、Cluster。

KPI指标内核的分布如**图1-9**所示。

图 1-9 KPI 指标内核分布



特性增强

无。

特性约束

- 对于utility语句不支持归一化，主要体现为不支持非DML语句，比如：create/drop/copy/vacuum等。
- 当前归一化SQL仅记录顶层SQL，对于存储过程语句，不对存储过程内部的SQL进行归一化处理，只记录调用存储过程的SQL。
- EXECUTE DIRECT ON语句不记录到unique sql中。

依赖关系

无。

1.5.10 支持一键式收集诊断信息

可获得性

本特性自V300R002C00版本开始引入。

特性简介

一键式诊断信息收集提供一种收集集群诊断信息的手段，用户无需逐个登录到节点，即可按需收集所有节点的诊断信息。

客户价值

无需登录集群所有节点，即可一键式收集所有节点的诊断日志。

特性描述

提供多种套件用于捕获、收集、分析诊断数据，使问题可以诊断，加速诊断过程。能根据开发和定位人员的需要，从生产环境中将必要的数据库日志、集群管理日志、堆栈信息等提取出来，定位人员根据获得信息进行问题的定界定位。

一键式收集工具，根据生产环境中问题的不同，从生产环境中获取不同的信息，从而提高问题定位定界的效率。用户可以通过改写配置文件，收集自己想要的信息：

- 通过Linux命令收集操作系统相关的信息。
- 通过查询系统表或者视图获得数据库系统相关的信息。
- 数据库系统运行日志和集群管理相关的日志。
- 数据库系统的配置信息。
- 数据库相关进程产生的Core文件。
- 数据库相关进程的堆栈信息。
- 数据库进程产生的trace信息。
- 数据库产生的redo日志文件xLog。
- 计划复现信息。

可以使用如下命令进行收集：

gs_collector --begin-time="BEGINTIME" --end-time="ENDTIME" [-h HOSTNAME | -f HOSTFILE] [-C CONFIGFILE] [--keyword=KEYWORD] [--speed-limit=SPEED] [-o OUTPUT] [-l LOGFILE]

特性增强

无。

特性约束

无。

依赖关系

无。

1.5.11 支持热点 key 快速检测

可获得性

本特性自V500R001C20版本开始引入。

特性简介

热点key检测为分布式数据库提供一种直观的判断热点分布情况的方法，可以通过查询系统函数和系统视图查询当前节点和整个集群内热点键值统计信息。

客户价值

热点key统计能够提供被频繁访问的键值所在的database、schema、table以及被访问次数等详细信息，从而帮助客户定位当前时段热点的分布详情，进而进行相应的业务调整。

特性描述

开关开启后，用户通过PBE或者gsql执行业务时，数据库会在优化器生成计划阶段将能够下推到单一DN上的键值进行收集上报，所有被访问的有效键值将通过pgstat线程进行异步处理，通过LRU算法对键值进行更新和淘汰。

GaussDB提供两种维度的接口来进行热点key的检测：

```
当前节点：
gaussdb=# select * from gs_stat_get_hotkeys_info() order by count, hash_value;
 database_name | schema_name |     table_name      | key_value | hash_value | count
---------------+-------------+---------------------+-----------+------------+-------
 regression    | public      | hotkey_single_col | {22}      | 1858004829 |     2
 regression    | public      | hotkey_single_col | {11}      | 2011968649 |     2
(2 rows)
整个集群：
gaussdb=# select * from global_stat_get_hotkeys_info() order by count, hash_value;
 database_name | schema_name |     table_name      | key_value | hash_value | count
---------------+-------------+---------------------+-----------+------------+-------
 regression    | public      | hotkey_single_col | {22}      | 1858004829 |     2
 regression    | public      | hotkey_single_col | {11}      | 2011968649 |     2
(2 rows)
```

此外还提供对应维度的清理接口来清理历史统计信息，避免历史遗留信息导致的统计误差。

特性增强

无。

特性约束

- 只收集分布列以及分布列上有等值过滤条件的，join的复杂查询暂不支持（比如select * from t where a = 1; 其中a 为分布列，由column - a, column_type - int和value – 1组成的结构体作为一个键值，而且相同的键值结构被访问至少2次才被当做热点key处理）。
- 只统计value长度小于1M的键值。
- 只支持能够下发到单一DN上执行的语句。
- 支持表的类型：hash分布表、List/Range分布表、hashbucket表，不支持系统表、复制表、单DN表、物化视图、临时表、unlogged表。
- 支持语句类型：Select、Update。
- 正确结果（统计误差小于5%）。
- 不支持节点故障重启后数据恢复（pgstat约束）。
- 当前只支持统计当前节点Top 16的热key（查询整个集群热点key时没有数量限制；查询系统函数时无排序，只有查询视图时才根据count排序）。

- 键值被访问超过两次才会进入LRU热key队列中。
- 开关打开时才能进行查询（清理接口不受开关影响）。
- 只有系统管理员和监控管理员才能调用查询和清理热key接口。
- 只能在CN上进行热点查询和清理。

说明

增加了键值收集逻辑，打开enable_hotkeys_collection开关开启热key功能会导致一定程度的性能劣化。

依赖关系

无。

1.5.12 内置 stack 工具

可获得性

本特性自V500R002C10版本开始引入。

特性简介

stack工具是获取数据库中各线程的调用栈的工具，用于辅助数据库运维人员定位死锁、hang等问题。

客户价值

提供函数级别的调用栈信息，提升数据库内核运维人员分析、定位死锁、hang等问题的效率。

特性描述

可以通过函数gs_stack()或者工具gs_ctl stack两种方式获取数据库中线程的调用栈。

1. gs_stack()函数方式详见《开发者指南》的“SQL参考 > 函数和操作符 > 统计信息函数”章节。
2. gs_ctl stack方式获取调用栈，详见《工具参考》的“系统内部调用的工具 > gs_ctl”章节。

特性增强

无。

特性约束

1. 仅用于gaussdb进程，其他进程，如cms、gtm等不支持。
2. 如果使用SQL的方式执行，则需要CN、DN进程处于正常状态，可连接和执行SQL。
3. 如果使用gs_ctl的方式执行，则需要CN、DN进程处于可响应信号的状态。
4. 不支持并发，在获取全线程栈的场景，各个线程的调用栈不处于同一时间点。
5. 最多支持128层调用栈，如果实际情况超过128层，则仅保留栈顶的128层。

6. 符号表没有被trip（当前release版本，使用的是strip –d，仅去掉了debug信息，符号表没有被trip，如果改为strip –s，则仅能显示指针，无法显示出符号名）。
7. SQL执行方式仅支持monadmin、sysadmin用户。
8. 注册了SIGURG信号的线程，才能获取调用栈。
9. 对于屏蔽操作系统SIGUSR2的代码段，无法获取调用栈 ，如果线程没有注册signal_slot，同样无法获取调用栈。

依赖关系

无。

1.5.13 支持 SQL PATCH

可获得性

本特性自503.1.0版本开始引入。

特性简介

SQL PATCH能够在避免直接修改用户业务语句的前提下对查询执行的方式做一定调整。在发现查询语句的执行计划、执行方式未达预期的场景下，可以通过创建查询补丁的方式，使用Hint对查询计划进行调优或对特定的语句进行报错短路处理。

客户价值

在业务产生查询计划不优导致的性能问题或系统内部错误导致服务不可用问题时，可以在数据库内通过运维函数调用对特定的场景进行调优或提前报错，以规避更严重的问题，能够大幅降低上述问题的运维成本。

特性描述

SQL PATCH主要设计给DBA、运维人员及其他需要对SQL进行调优的角色使用，用户通过其他运维视图或定位手段识别到业务语句存在计划不优导致的性能问题时，可以通过创建SQL PATCH对业务语句进行基于Hint的调优。目前支持行数、扫描方式、连接方式、连接顺序、PBE custom/generic计划选择、语句级参数设置、参数化路径的Hint。此外，对于部分由特定语句触发系统内部问题导致系统可服务性受损的语句，在不对业务语句变更的情况下，也可以通过创建用于单点规避的SQL PATCH，对问题场景提前报错处理，避免更大的损失。

SQL PATCH的实现基于Unique SQL ID，所以需要打开相关的运维参数才可以生效（enable_resource_track = on，instr_unique_sql_count > 0），Unique SQL ID在WDR报告和慢SQL视图中都可以获取到，在创建SQL PATCH时需要指定Unique SQL ID，对于存储过程内的SQL则需要设置参数 instr_unique_sql_track_type = 'all' 后在dbe_perf.statement_history视图中查询Unique SQL ID。

特性增强

无。

特性约束

1. 仅支持针对Unique SQL ID添加补丁，如果存在Unique SQL ID冲突，用于Hint调优的SQL PATCH可能影响性能，但不影响语义正确性。

2. 仅支持不改变SQL语义的Hint作为PATCH，不支持SQL改写。
3. 不支持逻辑备份、恢复。
4. 不支持在DN上创建SQL PATCH。
5. 仅初始用户、运维管理员、监控管理员、系统管理员用户有权限执行。
6. 库之间不共享，创建SQL PATCH时需要连接目标库。如果创建SQL PATCH的CN被剔除并触发全量Build，则会继承全量Build的目标CN中的SQL PATCH，因此建议在各个CN上尽量都创建对应的SQL PATCH。
7. CN之间由于Unique SQL ID不同，不共享SQL PATCH，需要用户手动在不同的CN上创建对应的SQL PATCH。
8. 限制在存储过程内的SQL PATCH和全局的SQL PATCH不允许同时存在。
9. 使用PREPARE + EXECUTE语法执行的预编译语句执行不支持使用SQL PATCH。
10. SQL PATCH不建议在数据库中长期使用，只应该作为临时规避方法。遇到内核问题所导致的特定语句触发数据库服务不可用问题，以及使用Hint进行调优的场景，需要尽快修改业务或升级内核版本解决问题。并且升级后由于Unique SQL ID生成方法可能变化，可能导致规避方法失效。
11. 当前，除DML语句之外，其他SQL语句（如CREATE TABLE等）的Unique SQL ID是对语句文本直接哈希生成的，所以对于此类语句，SQL PATCH对大小写、空格、换行等敏感，即不同的文本的语句，即使语义相同，仍然需要对应不同的SQL PATCH。对于DML，则同一个SQL PATCH可以对不同入参的语句生效，并且忽略大小写和空格。

依赖关系

本特性依赖于资源实时监控功能。对于不同的语句，数据库无法保证生产的Unique SQL ID哈希值全局唯一，如果不同的语句生成的Unique SQL ID冲突，会导致SQL PATCH命中预期外的其他语句。其中使用DBE_SQL_UTIL.create_hint_sql_patch/DBE_SQL_UTIL.create_remote_hint_sql_patch接口创建的用于调优的Hint PATCH可能会影响错误命中语句的性能，使用DBE_SQL_UTIL.create_abort_sql_patch/DBE_SQL_UTIL.create_remote_abort_sql_patch接口创建的用于避险的Abort PATCH需要谨慎使用。

1.5.14 支持设置云服务产品版本号

可获得性

本特性自505.0.0版本开始引入。

特性简介

云服务等第三方运维平台，可通过GUC工具设置相关的产品版本号和热补丁版本号信息，以便于用户查看。

客户价值

数据库产品为云服务的产品版本号管理提供便利。

特性描述

数据库提供product_version和hotpatch_version参数，云服务等运维集成平台可以将其对应的产品版本号和热补丁版本号信息设置到数据库系统，具体参数说明请参见

《管理员指南》中“配置运行参数 > GUC参数说明 > 版本和平台兼容性 > 云服务产品版本号”章节。参数设置完成后，可通过SQL接口查看相关参数信息：

```
gaussdb=#  show product_version;
gaussdb=#  show hotpatch_version;
```

特性增强

无。

特性约束

1. 版本号字符串长度限制，product_version 限制长度不超过50个字符，hotpatch_version限制长度不超过1500个字符。
2. 非法字符限制，参数值不能包含以下字符：“|”、“;”、“&”、“$”、“>”、“<”、“`”、“\”、“!”和换行符。

依赖关系

无。

1.5.15 内置 perf 工具

可获得性

本特性自505.0.0版本开始引入。

特性简介

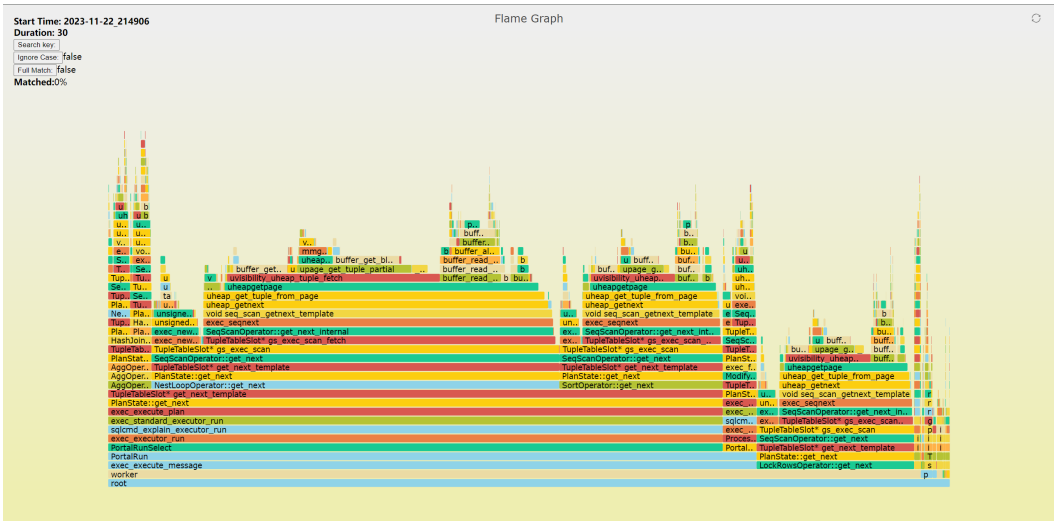
perf工具是在一定时间范围内，采集数据库中各线程的调用栈及其时间占比的工具，用于辅助数据库运维人员定位性能问题。该工具分为自动采集和手动采集两个功能。

自动采集堆栈功能，会定时采集数据库运行时活跃线程的函数调用栈及时间占比，并生成图形化火焰图报告。火焰图报告存储在$GAUSSLOG/gs_flamegraph/{datanode}路径下，下载该路径中的echarts.min.js文件和.html文件到同一目录下，用浏览器打开.html文件，即可展示采集到的堆栈调用火焰图。采集一次堆栈并生成火焰图的时间间隔通过GUC参数gs_perf_interval控制，范围为0或5-60，单位为min，默认为5min。自动采集堆栈功能功能默认开启，可以通过设置GUC参数gs_perf_interval=0关闭该功能。火焰图报告的保留时长由GUC参数gs_perf_retention_days控制，范围为1~8，单位为天，默认为3天。在gs_perf_interval不为0时，超过gs_perf_retention_days保存天数之外的火焰图文件会被回收。

手动采集堆栈功能，需要执行gs_perf_start()函数，采集一段时间内的堆栈调用情况，之后执行gs_perf_query()函数查询文字版堆栈调用信息。如需生成图形化火焰图报告，请执行gs_perf_report()函数。

堆栈调用火焰图是一种用于可视化性能分析的工具，它能够帮助开发者快速地识别程序中的性能瓶颈。堆栈调用火焰图y轴表示调用栈，每一层都是一个函数，顶部是当前正在执行的函数，下方是它的父函数，火焰的高度代表调用栈的深度；x轴表示采样执行时间占比，函数在x轴上的宽度越宽，表示它被采样的次数越多，即执行时间越长。火焰图展示的”平顶”，即占据宽度较大的函数，往往被用来定位数据库运行时的性能问题。

图 1-10 火焰图样例



客户价值

提供函数级别的调用栈信息，提供视图支持查询当前节点下所有活跃线程的函数调用栈及时间占比，提供生成火焰图文件的能力，提升数据库内核运维人员分析、定位性能问题的效率。

特性描述

可以通过gs_perf_start()，gs_perf_query()、gs_perf_clean()分别采集，查询和删除堆栈数据。通过dbe_perf.perf_query视图查询堆栈数据。

1. gs_perf_start()函数详见《开发者指南》中的“SQL参考 > 函数和操作符 > 统计信息函数”章节。
2. gs_perf_query()函数详见《开发者指南》中的“SQL参考 > 函数和操作符 > 统计信息函数”章节。
3. gs_perf_report()函数详见《开发者指南》中的“SQL参考 > 函数和操作符 > 统计信息函数”章节。
4. gs_perf_clean()函数详见《开发者指南》中的“SQL参考 > 函数和操作符 > 统计信息函数”章节。
5. dbe_perf.perf_query视图详见《开发者指南》中的“Schema > DBE_PERF Schema > OS > PERF_QUERY”章节。

特性增强

无。

特性约束

1. 执行性能采集时，需要数据库在normal、degrade状态，unavailable状态不支持。在degrade状态，异常cn/dn不支持采集。
2. 仅支持CN、DN、DN备、不支持logger节点，不支持cm_server、cm_agent、gtm、UDF等组件。
3. 仅支持on cpu采集，不支持off cpu采集。

4. 手动采集支持的采集时间范围为1s-60s。
5. 手动采集支持的采集频率为10HZ-1000HZ。
6. 自动采集堆栈功能一次采集5s，采集频率为99HZ。
7. 手动采集结果不落盘，存储在内存中。如果执行gs_perf_start后重启进程，则采集结果丢失。
8. 不支持并发采集。同一进程，同一时间，最多有一个session可以执行gs_perf_start操作。且同一进程内，在执行gs_perf_start操作期间，不支持执行gs_perf_query、gs_perf_report或gs_perf_clean操作。在自动采集堆栈期间，若手动执行采集堆栈函数gs_perf_start，会打断自动采集堆栈进程。
9. 自动或手动采集堆栈期间，不支持使用Linux perf工具操作同一进程。
10. 该特性依赖操作系统内核参数/proc/sys/kernel/perf_event_mlock_kb，该参数用来配置内置perf工具允许使用的最大内存值，操作系统中该参数的默认值为516KB。当操作系统中/proc/sys/kernel/perf_event_mlock_kb参数可修改时，在集群安装或升级的预安装阶段，默认将该参数调整为100MB。若该参数过小可能会导致手动和自动采集失败。

依赖关系

无。

1.6 数据库安全

1.6.1 访问控制模型

可获得性

本特性自V300R002C00版本开始引入。

特性简介

管理用户访问权限，为用户分配完成任务所需要的最小权限。

客户价值

客户依据自身需求创建对应的数据库用户并赋予相应的权限给操作人员，将数据库使用风险降到最低。

特性描述

数据库提供了基于角色的访问控制模型和基于三权分立的访问控制模型。在基于角色的访问控制模型下，数据库用户可分为系统管理员用户、监控管理员用户、运维管理员用户、安全策略管理员用户以及普通用户。系统管理员创建角色或者用户组，并为角色分配对应的权限；监控管理员查看dbe_perf模式下的监控视图或函数；运维管理员使用Roach工具执行数据库备份恢复操作；安全策略管理员创建资源标签、脱敏策略、统一审计策略。用户通过绑定不同的角色获得角色所拥有的对应的操作权限。

在基于三权分立的访问控制模型下，数据库用户可分为系统管理员、安全管理员、审计管理员、监控管理员用户、运维管理员用户、安全策略管理员用户以及普通用户。安全管理员负责创建用户，系统管理员负责为用户赋权，审计管理员负责审计所有用户的行为。

默认情况下，使用基于角色的访问控制模型。客户可通过设置GUC参数enableSeparationOfDuty为on来切换。

特性增强

无。

特性约束

系统管理员的具体权限受GUC参数enableSeparationOfDuty控制。

三权分立开启和关闭切换时需要重启数据库，且无法对新模型下不合理的用户权限进行自主识别，需要DBA识别并修正。

依赖关系

无。

1.6.2 数据库认证机制

可获得性

本特性自V300R002C00版本开始引入。

特性简介

提供基于客户端/服务端(C/S)模式的客户端连接认证机制。

客户价值

加密认证过程中采用单向Hash不可逆加密算法PBKDF2，有效防止彩虹攻击。

特性描述

GaussDB采用基本的客户端连接认证机制，客户端发起连接请求后，由服务端完成信息校验并依据校验结果发送认证所需信息给客户端（认证信息包括盐值，token以及服务端签名信息）。客户端响应请求发送认证信息给服务端，由服务端调用认证模块完成对客户端认证信息的认证。用户的密码被加密存储在内存中。整个过程中口令加密存储和传输。当用户下次登录时通过计算相应的hash值并与服务端存储的key值比较来进行正确性校验。

特性增强

统一加密认证过程中的消息处理流程，可有效防止攻击者通过抓取报文猜解用户名或者密码的正确性。

特性约束

无。

依赖关系

无。

1.6.3 数据加密存储

可获得性

本特性自V300R002C00版本开始引入。

特性简介

提供对导入数据的加密存储。

客户价值

为客户提供加密导入接口，对客户认为是敏感信息的数据进行加密后存储在表内。

特性描述

GaussDB提供加密函数gs_encrypt_aes128()、gs_encrypt ()、gs_encrypt_bytea ()和解密函数gs_decrypt_aes128()、gs_decrypt()、gs_decrypt_bytea()接口。通过加密函数，可以对需要输入到表内的某列数据进行加密后再存储到表格内。调用格式为:

```
gs_encrypt_aes128(column, key),gs_encrypt (encryptstr, keystr,encrypttype),gs_encrypt_bytea(encryptstr, keystr, encrypttype)
```

其中key为用户指定的初始口令，用于派生加密密钥。当客户需要对整张表进行加密处理时，则需要为每一列单独书写加密函数。

当具有对应权限的用户需要查看具体的数据时，可通过解密函数接口对相应的属性列进行解密处理，调用格式为：

```
gs_decrypt_aes128(column, key),gs_decrypt(decryptstr, keystr,decrypttype),gs_decrypt_bytea(decryptstr, keystr, decrypttype)
```

参数	类型	描述	取值范围
encryptstr/ decryptstr	text	需要加解密的数据	-
keystr	text	密钥	8~16字节，至少包含3种字符（大写字母、小写字母、数字、特殊字符）
encrypttype/decryptyt	text	加解密类型（不区分大小写）	aes128_cbc_sha256、aes256_cbc_sha256、aes128_gcm_sha256,、aes256_gcm_sha256、sm4_ctr_sm3

说明

gs_encrypt_aes128、gs_decrypt_aes128使用默认加解密参数。

gs_encrypt、gs_decrypt除支持表格中参数外，兼容原有参数aes128、sm4。

特性增强

无。

特性约束

无。

依赖关系

无。

1.6.4 数据库审计

可获得性

本特性自V300R002C00版本开始引入。

特性简介

审计日志记录用户对数据库的启停、连接、DDL、DML、DCL等操作。

客户价值

审计日志机制主要增强数据库系统对非法操作的追溯及举证能力。

特性描述

数据库审计功能对数据库系统的安全性至关重要。数据库安全管理员可以利用审计日志信息，重现导致数据库现状的一系列事件，找出非法操作的用户、时间和内容等。

审计功能包括传统审计和统一审计两种审计模式。传统审计通过参数配置各个审计项开关，管理员可以通过参数配置对哪些语句或操作记录审计日志。传统审计采用记录到OS文件的方式来保存审计日志，支持审计管理员通过SQL函数接口审计日志查询和删除。统一审计机制是一种通过定制化制定审计策略而实现高效安全审计管理的一种技术。当管理员定义审计对象和审计行为后，用户执行的任务如果关联到对应的审计策略，则生成对应的审计行为，并记录审计日志。定制化审计策略可涵盖常见的用户管理活动，DDL和DML行为，满足日常审计诉求。

特性增强

503.1.0版本在传统审计功能基础上，新增支持用户级别审计功能。新增GUC参数full_audit_users配置全量审计用户列表，对列表中的用户执行的所有可被审计的操作记录审计日志；新增GUC参数no_audit_client配置无需记录审计的客户端列表；新增GUC参数audit_system_function_exec配置系统函数审计开关。

特性约束

无。

依赖关系

无。

1.6.5 网络通信安全

可获得性

本特性自V300R002C00版本开始引入。

特性简介

为保护敏感数据在Internet上传输的安全性，GaussDB支持通过SSL加密客户端和服务器之间的通讯。

客户价值

保证客户的客户端与服务器通讯安全。

特性描述

GaussDB支持SSL协议标准。SSL(Secure Socket Layer)协议是一种安全性更高的应用层通信协议，主要用于Web安全传输，SSL包含记录层和传输层，记录层协议确定传输层数据的封装格式，传输层安全协议使用X.509认证。SSL协议利用非对称加密演算来对通信方做身份认证，之后交换对称密钥作为会谈密钥。通过SSL协议可以有效保障两个应用间通信的保密性和可靠性，使客户与服务器之间的通信不被攻击者窃听。

GaussDB支持TLS 1.2协议标准。TLS 1.2协议是一种安全性更高的传输层通信协议，它包括两个协议组，TLS记录协议和TLS握手协议，每一组协议具有很多不同格式的信息。TLS协议是独立于应用协议的，高层协议可以透明地分布在TLS协议上面。通过TLS协议可保证通信双方的数据保密性和数据完整性。

GaussDB支持国密TLS加密传输，支持“ECC-SM4-SM3”和“ECDHE-SM4-SM3”国密加密算法套件，使用国密TLS需要配置国密双证书文件。

特性增强

证书签名算法强度检查：对于一些强度较低的签名算法，给出告警信息，提醒客户更换包含高强度签名算法的证书。

证书超时时间检查：如果距离超期日期小于7天则给出告警信息，提醒客户端更换证书。

证书权限检查：在建连阶段对证书的权限进行校验。

特性约束

- 从CA认证中心申请到正式的服务器、客户端的证书和密钥。
  - 使用国际证书认证，其服务器的私钥为server.key，证书为server.crt，客户端的私钥为client.key，证书为client.crt，CA根证书名称为cacert.pem。
  - 使用国密证书认证，其服务器的签名证书私钥为server.key，签名证书为server.crt，加密证书私钥为server_enc.key，加密证书为server_enc.crt，客户端的签名证书私钥为client.key，签名证书为client.crt，加密证书私钥为client_enc.key，加密证书为client_enc.crt，CA根证书名称为cacert.pem。
- 使用该功能需要打开SSL开关，并且配置证书和连接方式。
- 国密TLS加密传输当前只支持gsql客户端。

依赖关系

该特性依赖OpenSSL开源软件，国密TLS加密传输需要依赖支持国密TLS的OpenSSL版本。

1.6.6 资源标签机制

可获得性

本特性自V500R001C00版本开始引入。

特性简介

数据库资源是指数据库所记录的各类对象，包括数据库、模式、表、列、视图、trigger等，数据库对象越多，数据库资源的分类管理就越繁琐。资源标签机制是一种通过对具有某类相同“特征”的数据库资源进行分类标记而实现资源分类管理的一种技术。当管理员对数据库内某些资源“打上”标签后，可以基于该标签进行如审计或数据脱敏的管理操作，从而实现对标签所包含的所有数据库资源进行安全管理。

客户价值

合理的制定资源标签能够有效的进行数据对象分类，提高对象管理效率，降低安全策略配置的复杂性。当管理员需要对某组数据库资源对象做统一审计或数据脱敏等安全管理动作时，可将这些资源划分到一个资源标签，该标签即包含了具有某类特征或需要统一配置某种策略的数据库资源，管理员可直接对资源标签执行管理操作，大大降低了策略配置的复杂性和信息冗余程度，提高了管理效率。

特性描述

资源标签机制是将当前数据库内包含的各种资源进行“有选择性的”分类，管理员可以使用如下SQL语法进行资源标签的创建，从而将一组数据库资源打上标签：

CREATE RESOURCE LABEL schm_lb ADD SCHEMA(schema_for_label);

CREATE RESOURCE LABEL tb_lb ADD TABLE(schema_for_label.table_for_label);

CREATE RESOURCE LABEL col_lb ADD COLUMN(schema_for_label.table_for_label.column_for_label);

CREATE RESOURCE LABEL multi_lb ADD SCHEMA(schema_for_label), TABLE(table_for_label);

其中，schema_for_label、table_for_label、column_for_label分别为待标记模式、表、列。schm_lb标签包含了模式schm_for_label；tb_lb包含了表table_for_label；col_lb包含了列column_for_label；multi_lb包含模式schm_for_label和列table_for_label。对这些已配置的资源标签进行如统一审计或动态数据脱敏也即是对标签所包含的每一个数据库资源进行管理。

当前，资源标签所支持的数据库资源类型包括：SCHEMA、TABLE、COLUMN、VIEW、FUNCTION。

特性增强

无。

特性约束

- 资源标签需要由具备POLADMIN和SYSADMIN属性的用户或初始用户创建。
- 不支持对临时表创建资源标签。
- 同一个基本表的列只可能属于一个资源标签。
- 不支持通过gs_dump导出资源标签。系统管理员或安全策略管理员可以访问GS_POLICY_LABEL系统表查询已创建的资源标签。

依赖关系

无。

1.6.7 统一审计机制

可获得性

本特性自V500R001C00版本开始引入。

特性简介

审计机制是行之有效的安全管理方案，可有效解决攻击者抵赖，审计的范围越大，可监控的行为就越多，而产生的审计日志就越多，影响实际审计效率。统一审计机制是一种通过定制化制定审计策略而实现高效安全审计管理的一种技术。当管理员定义审计对象和审计行为后，用户执行的任务如果关联到对应的审计策略，则生成对应的审计行为，并记录审计日志。定制化审计策略可涵盖常见的用户管理活动，DDL和DML行为，满足日常审计诉求。

客户价值

审计是日常安全管理中必不可少的行为，当使用传统审计机制审计某种行为时，如SELECT，会导致产生大量的审计日志，进而增加整个系统的I/O，影响系统的性能；另一方面，大量的审计日志会影响管理员的审计效率。统一审计机制使得客户可以定制化生成审计日志的策略，如只审计数据库账户A查询某个表table的行为。通过定制化审计，可以大大减少生成审计日志的数量，从而在保障审计行为的同时降低对系统性能的影响。而定制化审计策略可以提升管理员的审计效率。

特性描述

统一审计机制基于资源标签进行审计行为定制化，且将当前所支持的审计行为划分为access类和privileges类。一个完整的审计策略创建的SQL语法如下所示：

CREATE RESOURCE LABEL auditlabel add table(table_for_audit1, table_for_audit2);

CREATE AUDIT POLICY audit_select_policy ACCESS SELECT ON LABEL(auditlabel) FILTER ON ROLES(usera);

CREATE AUDIT POLICY audit_admin_policy PRIVILEGES ALTER, DROP ON LABEL(auditlabel) FILTER ON IP(local);

其中，auditlabel为本轮计划审计的资源标签，该资源标签中包含了两个表对象；audit_select_policy定义了用户usera对auditlabel对象的SELECT行为的审计策略，不区分访问源；audit_admin_policy定义了从本地对auditlabel对象进行ALTER和DROP操作

行为的审计策略，不区分执行用户；当不指定ACCESS和PRIVILEGES的具体行为时，表示审计针对某一资源标签的所有支持的DDL和DML行为。当不指定具体的审计对象时，表示审计针对所有对象的操作行为。统一审计策略的增删改也会记录在统一审计日志中。

当前，统一审计支持的审计行为包括：

SQL类型	支持操作和对象类型
DDL	操作：ALL、ALTER、ANALYZE/VACUUM、COMMENT、CREATE、DROP、GRANT、REVOKE、SET、SHOW 对象：DATABASE、SCHEMA、FUNCTION/PROCEDURE、TRIGGER、TABLE、SEQUENCE、FOREIGN_SERVER、FOREIGN_TABLE、TABLESPACE、ROLE/USER/GROUP、INDEX、VIEW、DATA_SOURCE、WEAK PASSWORD DICTIONARY、AUDIT POLICY、MASKING POLICY、RESOURCE LABEL、MATERIALIZED VIEW/INCREMENTAL MATERIALIZED VIEW 注：对不支持的对象类型统一审计日记均标记为UNKNOWN
DML	操作：ALL、COPY、DEALLOCATE、DELETE_P、EXECUTE、REINDEX、INSERT、PREPARE、SELECT、TRUNCATE、UPDATE

说明

ALL指的是上述DDL或DML中支持的所有对数据库的操作。当形式为{ DDL | ALL }时，ALL指所有DDL操作；当形式为{ DML | ALL }时，ALL指所有DML操作。

其中EXECUTE是指执行预备语句的EXECUTE操作，并非存储过程中动态调用匿名块EXECUTE IMMEDIATE…USING语句。

特性增强

无。

特性约束

- 统一审计策略需要由具备POLADMIN或SYSADMIN属性的用户或初始用户创建，普通用户无访问安全策略系统表和系统视图的权限。
- 统一审计策略语法要么针对DDL行为，要么针对DML语法行为，同一个审计策略不可同时包含DDL和DML行为；统一审计策略目前支持最多设置98个。
- 统一审计监控用户通过客户端在CN节点上执行的SQL语句，而不会记录数据库内部SQL语句。
- 同一个审计策略下，相同资源标签可以绑定不同的审计行为，相同行为可以绑定不同的资源标签，操作"ALL"类型包括DDL或者DML下支持的所有操作。
- 同一个资源标签可以关联不同的统一审计策略，统一审计会按照SQL语句匹配的策略依次打印审计信息。
- 统一审计策略的审计日志单独记录，暂不提供可视化查询接口，整个日志依赖于操作系统自带rsyslog服务，通过配置完成日志归档。
- 。

- FILTER中的APP项建议仅在同一信任域内使用，由于客户端不可避免的可能出现伪造名称的情况，该选项使用时需要与客户端联合形成一套安全机制，减少误用风险。一般情况下不建议使用，使用时需要注意客户端仿冒的风险。
- FILTER中的IP地址以ipv4为例支持如下格式：

ip地址格式	示例
单ip	127.0.0.1
掩码表示ip	127.0.0.1|255.255.255.0
cidr表示ip	127.0.0.1/24
ip区间	127.0.0.1-127.0.0.5

- 不支持通过gs_dump导出统一审计策略。系统管理员或安全策略管理员可以访问GS_AUDITING_POLICY、GS_AUDITING_POLICY_ACCESS、GS_AUDITING_POLICY_PRIVILEGES、GS_AUDITING_POLICY_FILTERS系统表查询已创建的统一审计策略。
- 由于GROUP是ROLE的别名，当统一审计的对象为GROUP时，统一审计日志中会将相应操作对象统一记录为ROLE类型。
- 统一审计日志中不区分存储过程和函数，当数据库对象是存储过程PROCEDURE时，日志中也会将其记录为FUNCTION类型。
- 统一审计策略中的ANALYZE对应VACUUM和ANALYZE两种SQL操作，审计日志中VACUUM操作也会被记录为ANALYZE。
- 由于语法解析机制，ALTER INDEX ... REBUILD语句会被审计为REINDEX语句。GRANT ALL PRIVILEGES TO user、REVOKE ALL PRIVILEGES FROM user语句会被审计为ALTER ROLE语句。
- WITH res1 AS (UPDATE ...) INSERT INTO ... VALUES (SELECT * from res1, ...)语句中不审计UPDATE语句，只审计主语句中的INSERT操作。
- 对于提升子查询语句不记录审计日志，例如INSERT INTO ... SELECT * FROM ...语句中只记录INSERT操作，不记录SELECT操作。
- 某些执行失败的DML语句不审计，例如唯一键约束导致执行失败、对只读子查询进行DML操作执行失败等。

⚠ 注意

使用统一审计功能时，强烈建议明确需要被审计的对象、需要审计的操作以及客户端、用户信息，根据场景创建准确的审计策略。对不必要的对象和操作进行审计会产生大量的审计日志引起数据库性能劣化、磁盘空间膨胀，也会影响管理员查询审计日志的效率。

1.6.8 动态数据脱敏机制

可获得性

本特性自V500R001C00版本开始引入。

特性简介

数据脱敏是行之有效的数据库隐私保护方案之一，可以在一定程度上限制非授权用户对隐私数据的窥探。动态数据脱敏机制是一种通过定制化制定脱敏策略从而实现对隐私数据保护的一种技术，可以有效地在保留原始数据的前提下解决非授权用户对敏感信息的访问问题。当管理员指定待脱敏对象和定制数据脱敏策略后，用户所查询的数据库资源如果关联到对应的脱敏策略时，则会根据用户身份和脱敏策略进行数据脱敏，从而限制非授权用户对隐私数据的访问。

客户价值

数据隐私保护是数据库安全所需要具备的安全能力之一，可以在一定程度上限制非授权用户对隐私数据的访问，保证隐私数据安全。动态数据脱敏机制可以通过配置脱敏策略实现对指定数据库资源信息的隐私保护，另一方面，脱敏策略的配置也具有一定的灵活性，可以仅针对特定用户场景实现有针对性的隐私保护能力。

特性描述

动态数据脱敏机制基于资源标签进行脱敏策略的定制化，可根据实际场景选择特定的脱敏方式，也可以针对某些特定用户制定脱敏策略。一个完整的脱敏策略创建的SQL语法如下所示：

```
CREATE RESOURCE LABEL label_for_creditcard ADD
COLUMN(user1.table1.creditcard);

CREATE RESOURCE LABEL label_for_name ADD COLUMN(user1.table1.name);

CREATE MASKING POLICY msk_creditcard creditcardmasking ON
LABEL(label_for_creditcard);

CREATE MASKING POLICY msk_name randommasking ON LABEL(label_for_name)
FILTER ON IP(local), ROLES(dev);
```

其中，label_for_creditcard和label_for_name为本轮计划脱敏的资源标签，分别包含了两个列对象；creditcardmasking、randommasking为预置的脱敏函数；msk_creditcard定义了所有用户对label_for_creditcard标签所包含的资源访问时做creditcardmasking的脱敏策略，不区分访问源；msk_name定义了本地用户dev对label_for_name标签所包含的资源访问时做randommasking的脱敏策略；当不指定FILTER对象时则表示对所有用户生效，否则仅对标识场景的用户生效。

当前，预置的脱敏函数包括：

脱敏函数名	示例
creditcardmasking	'4880-9898-4545-2525' 将会被脱敏为 'xxxx-xxxx-xxxx-2525'，该函数仅对后4位之前的数字进行脱敏
basicemailmasking	'abcd@gmail.com' 将会被脱敏为'xxxx@gmail.com'， 对出现第一个'@'之前的文本进行脱敏
fullemailmasking	'abcd@gmail.com' 将会被脱敏为 'xxxx@xxxxx.com'，对出现最后一个'.'之前的文本（除'@'符外）进行脱敏
alldigitsmasking	'alex123alex' 将会被脱敏为 'alex000alex'， 仅对文本中的数字进行脱敏

shufflemasking	'hello word' 将会被随机打乱顺序脱敏为 'hlwoeor dl'， 该函数通过字符乱序排列的方式实现，属于弱脱敏函数，语义较强的字符串不建议使用该函数脱敏。
randommasking	'hello word' 将会被脱敏为 'ad5f5ghdf5'，将文本按字符随机脱敏
regexpmasking	需要用户顺序输入四个参数，reg为被替换的字符串，replace_text为替换后的字符串，pos为目标字符串开始替换的初始位置，为整数类型，reg_len为替换长度，为整数类型。reg、replace_text可以用正则表达，pos如果不指定则默认为0，reg_len如果不指定则默认为-1，即pos后所有字符串。如果用户输入参数与参数类型不一致，则会使用maskall方式脱敏。 `CREATE MASKING POLICY msk_creditcard regexpmasking('[\d+]', 'x', 5, 9 ) ON LABEL(label_for_creditcard);` '4880-9898-4545-2525' 将会被脱敏为 '4880-xxxx-xxxx-2525'。
maskall	'4880-9898-4545-2525' 将会被脱敏为 'xxxxxxxxxxxxxxxxxxx'

每个脱敏函数规格如下：

脱敏函数名	支持的数据类型
creditcardmasking	定长类型：CHAR, BPCHAR；变长类型：VARCHAR, NVARCHAR2, TEXT（注：仅针对信用卡格式的文本类数据）
basicemailmasking	定长类型：CHAR, BPCHAR；变长类型：VARCHAR, NVARCHAR2, TEXT（注：仅针对email格式的文本类型数据）
fullemailmasking	定长类型：CHAR, BPCHAR；变长类型：VARCHAR, NVARCHAR2, TEXT (注：仅针对email格式的文本类型数据)
alldigitsmasking	定长类型：CHAR, BPCHAR；变长类型：VARCHAR, NVARCHAR2, TEXT (注：仅针对包含数字的文本类型数据)
shufflemasking	定长类型：CHAR, BPCHAR；变长类型：VARCHAR, NVARCHAR2, TEXT (注：仅针对文本类型数据)
randommasking	定长类型：CHAR, BPCHAR；变长类型：VARCHAR, NVARCHAR2, TEXT (注：仅针对文本类型数据)
regexpmasking	定长类型：CHAR, BPCHAR；变长类型：VARCHAR, NVARCHAR2, TEXT (注：仅针对文本类型数据)
maskall	BOOL、RELTIME、TIME、TIMETZ、INTERVAL、TIMESTAMP、TIMESTAMPTZ、SMALLDATETIME、ABSTIME、 TEXT、CHAR、BPCHAR、VARCHAR、NVARCHAR2、NAME、INT8、INT4、INT2、INT1、NUMRIC、FLOAT4、FLOAT8

对于不支持的数据类型，默认使用maskall函数进行数据脱敏。BOOL类型脱敏成'0'；RELTIME类型脱敏成'1970'；TIME, TIMETZ, INTERVAL类型脱敏成'00:00:00.0000+00'；TIMESTAMP, TIMESTAMPTZ, SMALLDATETIME, ABSTIME类型

脱敏成'1970-01-01 00:00:00.0000'；TEXT, CHAR, BPCHAR, VARCHAR, NVARCHAR2, NAME类型脱敏成'x'；INT8, INT4, INT2, INT1, NUMERIC, FLOAT4, FLOAT8类型脱敏成'0'。若数据类型不属于maskall支持的类型，则不支持创建脱敏策略。如果脱敏列涉及隐式转换，则结果以隐式转换后的数据类型为基础进行脱敏。另外需要说明的是，如果脱敏策略应用到数据列并生效，此时对该列数据的操作将以脱敏后的结果为基础而进行。

动态数据脱敏适用于和实际业务紧密相关的场景，根据业务需要为用户提供合理的脱敏查询接口以及报错处理逻辑，以避免通过撞库而获取原始数据。

动态数据脱敏配置脱敏策略时，对用户创建的自定义函数进行支持适配。

示例如下。

1. Poladmin权限用户创建一般函数，将传入的字符串中间8位替换成'xxxx-xxxx'后返回字符串。
```
create or replace function msk_creditcard(col text) returns TEXT as $$
declare
result TEXT;
begin
result := overlay(col placing 'xxxx-xxxx' from 6);
return result;
end;
$$ language plpgsql;
```
2. Poladmin权限用户使用上述函数创建脱敏策略。
```
CREATE MASKING POLICY msk_creditcard msk_creditcard ON LABEL(label_for_creditcard);
```
3. 查询信用卡号时，脱敏函数生效。

   '4880-9898-4545-2525' 将会被脱敏为 '4880-xxxx-xxxx-2525'。

应用于动态数据脱敏的UDF规格如下。

UDF参数类型	支持的数据类型
输入参数	CHAR、BPCHAR、VARCHAR、NVARCHAR2、TEXT、INT8、INT4、INT2、INT1、FLOAT4、FLOAT8、NUMERIC

应用于动态数据脱敏的UDF入参参数不在字符类型（定长类型：char, bpchar；变长类型：varchar, nvarchar2, text），数字类型（int8，int4，int2，int1，numeric，float4，float8）中时，用maskall脱敏。应用于动态数据脱敏的UDF入参参数和返回参数需保持数据类型一致，字符类型可以兼容，否则用maskall脱敏。如果列类型为maskall不支持的类型，则报错。

特性增强

GaussDB V500R002C00版本加入动态数据脱敏支持UDF，预置函数支持正则表达脱敏函数。

特性约束

- 动态数据脱敏策略需要由具备POLADMIN或SYSADMIN属性的用户或初始用户创建，普通用户没有访问安全策略系统表和系统视图的权限。
- 动态数据脱敏只在配置了脱敏策略的数据表上生效，而审计日志不在脱敏策略的生效范围内。

- 在一个脱敏策略中，对于同一个资源标签仅可指定一种脱敏方式，不可重复指定。
- 不允许多个脱敏策略对同一个资源标签进行脱敏，除以下脱敏场景外：使用FILTER指定策略生效的用户场景，包含相同资源标签的脱敏策略间FILTER生效场景无交集，此时可以根据用户场景明确辨别资源标签被哪种策略脱敏。
- Filter中的APP项建议仅在同一信任域内使用，由于客户端不可避免的可能出现伪造名称的情况，该选项使用时需要与客户端联合形成一套安全机制，减少误用风险。一般情况下不建议使用，使用时需要注意客户端仿冒的风险。
- 对于带有query子句的INSERT或MERGE INTO操作，如果源表中包含脱敏列，则上述两种操作中插入或更新的结果为脱敏后的值，且不可还原。
- 在内置安全策略开关开启的情况下，执行ALTER TABLE EXCHANGE PARTITION操作的源表若在脱敏列则执行失败。
- 对于设置了动态数据脱敏策略的表，需要谨慎授予其他用户对该表的trigger权限，以免其他用户利用触发器绕过脱敏策略。
- 最多支持创建98个动态数据脱敏策略。
- 仅支持对只包含COLUMN属性的资源标签做脱敏。
- 仅支持对普通表且为永久表的列进行数据脱敏。
- 仅支持对SELECT直接查询到的数据进行脱敏，对已脱敏结果进行二次处理会导致脱敏策略失效或不符合预期。
- 应用于动态数据脱敏的UDF只支持标准数据库SQL、PL/SQL function。
- 应用于动态数据脱敏的UDF中，如果包含访问数据库资源的语句如（select，insert），使用该UDF的动态数据脱敏结果可能会不符合预期或导致安全风险。
- 应用于动态数据脱敏的UDF创建脱敏策略成功后，如果对该脱敏列进行alter或者drop，会导致脱敏策略失效或不符合预期。
- 动态数据脱敏的UDF函数不支持使用SECURITY INVOKER函数。应用于动态数据脱敏的UDF创建脱敏策略成功后，不允许对该function进行create、alter或drop。
- 应用于动态数据脱敏的UDF只能由具有poladmin权限用户创建。由具有poladmin权限的用户将访问schema的usage权限赋予public，如果因为grant/revoke操作，导致用户不能访问UDF，则使用maskall脱敏。
- 应用于动态数据脱敏的UDF应为幂等，即多次执行结果一样。如果设置UDF为非幂等，在分布式场景下使用UDF的动态数据脱敏结果可能会不符合预期。
- 不支持在系统表上应用动态数据脱敏的UDF创建脱敏策略。
- FILTER中的IP地址以ipv4为例支持如下格式：

ip地址格式	示例
单ip	127.0.0.1
掩码表示ip	127.0.0.1|255.255.255.0
cidr表示ip	127.0.0.1/24
ip区间	127.0.0.1-127.0.0.5

- 不支持通过gs_dump导出动态数据脱敏策略。系统管理员或安全策略管理员可以访问GS_MASKING_POLICY、GS_MASKING_POLICY_ACTIONS、GS_MASKING_POLICY_FILTERS系统表查询已创建的动态数据脱敏策略。

依赖关系

无。

1.6.9 行级访问控制

可获得性

本特性自V300R002C00版本开始引入。

特性简介

行级访问控制特性将数据库访问控制精确到数据表行级别，使数据库达到行级访问控制的能力。不同用户执行相同的SQL查询操作，读取到的结果是不同的。

客户价值

不同用户执行相同的SQL查询操作，读取到的结果是不同的。

特性描述

用户可以在数据表创建行访问控制(Row Level Security)策略，该策略是指针对特定数据库用户、特定SQL操作生效的表达式。当数据库用户对数据表访问时，若SQL满足数据表特定的Row Level Security策略，在查询优化阶段将满足条件的表达式，按照属性(PERMISSIVE | RESTRICTIVE)类型，通过AND或OR方式拼接，应用到执行计划上。

行级访问控制的目的是控制表中行级数据可见性，通过在数据表上预定义Filter，在查询优化阶段将满足条件的表达式应用到执行计划上，影响最终的执行结果。当前受影响的SQL语句包括SELECT，UPDATE，DELETE。

特性增强

505.1版本新增GUC参数enable_rls_match_index；该参数打开后，对于查询谓词包含unleakproof类型系统函数或like操作符的目标场景，允许基于该谓词条件执行索引扫描。

特性约束

- 行级访问控制策略仅可以应用到SELECT、UPDATE和DELETE操作，不支持应用到INSERT和MERGE操作。
- 支持对行存表、行存分区表、复制表、unlogged表、HASH分布表定义行级访问控制策略，不支持对外表、临时表定义行级访问控制策略。
- 不支持对视图定义行级访问控制策略。
- 同一张表上可以创建多个行级访问控制策略，一张表最多允许创建100个行级访问控制策略。
- 初始用户和系统管理员不受行级访问控制策略的影响。
- 对于设置了行级访问控制策略的表，需要谨慎授予其他用户对该表的trigger权限，以免其他用户利用触发器绕过行级访问控制策略。
- 行级访问控制策略不支持使用SECURITY INVOKER函数。对于已经使用了SECURITY INVOKER函数的策略，不允许对该FUNCTION进行CREATE、ALTER或DROP。

- 对于设置了行级访问控制策略、且具有函数表达式索引的表，仅当该函数为leakproof类型时，该表达式索引生效。

依赖关系

无。

1.6.10 用户口令强度校验机制

可获得性

本特性自V300R002C00版本开始引入。

特性简介

对用户访问数据库所设置的口令强度进行校验。

客户价值

用户无法设置过低强度的口令，加固客户数据安全。

特性描述

初始化数据库、创建用户、修改用户时需要指定密码。密码必须满足强度校验，否则会提示用户重新输入密码。账户密码复杂度要求如下：

- 包含大写字母（A-Z）的最少个数（password_min_uppercase）
- 包含小写字母（a-z）的最少个数（password_min_lowercase）
- 包含数字（0-9）的最少个数（password_min_digital）
- 包含特殊字符的最少个数（password_min_special）
- 密码的最小长度（password_min_length）
- 密码的最大长度（password_max_length）
- 至少包含上述四类字符中的三类。
- 不能和用户名、用户名倒写相同，本要求为非大小写敏感。
- 不能和当前密码、当前密码的倒写相同。
- 不能是弱口令。
  - 弱口令指的是强度较低，容易被破解的密码，对于不同的用户或群体，弱口令的定义可能会有所区别，用户需要自己添加定制化的弱口令。

参数password_policy设置为1时表示采用密码复杂度校验，默认值为1。

弱口令字典中的口令存放在gs_global_config系统表中（name字段为weak_password的记录是储存的弱口令），当创建用户、修改用户需要设置密码时，将会把用户设置口令和弱口令字典中存放的口令进行对比，如果命中，则会提示用户该口令为弱口令，设置密码失败。

弱口令字典默认为空，用户通过以下语法可以对弱口令字典进行增加和删除，示例如下：

```
CREATE WEAK PASSWORD DICTIONARY WITH VALUES ('password1'), ('password2');
DROP WEAK PASSWORD DICTIONARY;
```


其中"password1", "password2"是用户事先准备的弱口令，该语句执行成功后将会存入弱口令系统表中。

当用户尝试通过CREATE WEAK PASSWORD DICTIONARY 插入表中已存在的弱口令时，会只在表中保留一条该弱口令。

DROP WEAK PASSWORD DICTIONARY语句会清空整张系统表弱口令相关部分。

gs_global_config系统表没有唯一索引，不建议用户通过COPY FROM命令重复用相同数据对该表进行操作。

若用户需要对弱口令相关操作进行审计，应设置audit_system_object参数中的第三位为1。

特性增强

V5R001C10版本实现了弱口令字典功能。

特性约束

初始用户、系统管理员和安全管理员可以查看、新增、删除弱口令字典。

依赖关系

无。

1.6.11 口令脱敏机制

可获得性

本特性自V300R002C00版本开始引入。

特性简介

在数据库审计日志或运行日志中打印SQL语句，或者在系统表或系统视图中记录SQL语句时，对SQL语句中包含的可识别的口令或密钥进行掩码脱敏处理，以防止敏感信息泄露。

客户价值

防止口令或密钥等敏感信息泄露，提升数据安全性。

特性描述

用户无需进行参数配置和执行其他操作，在打印SQL语句前，系统自动将识别到的口令或密钥信息进行脱敏。

支持口令或密钥脱敏的SQL语句场景如下：

- 对CREATE ROLE、CREATE USER、CREATE GROUP语句中的PASSWORD或IDENTIFIED BY参数进行脱敏。
- 对ALTER ROLE、ALTER USER语句中的PASSWORD、IDENTIFIED BY或REPLACE参数进行脱敏。

- 对SET ROLE、SET SESSION AUTHORIZATION语句中的PASSWORD参数进行脱敏。
- 对CREATE/ALTER SERVER语句中的secret_access_key参数进行脱敏。
- 对CREATE/ALTER TEXT SEARCH DICTIONARY语句中的FILEPATH参数如果为OBS目录则进行脱敏。
- 对DBE_SCHEDULER.CREATE_CREDENTIAL函数的password参数进行脱敏。
- 对gs_encrypt_aes128、gs_decrypt_aes128、gs_encrypt、gs_decrypt、gs_encrypt_bytea、gs_decrypt_bytea、aes_encrypt、aes_decrypt加解密函数中的密钥参数进行脱敏。
- 对pg_create_physical_replication_slot_extern函数中OBS归档槽相关敏感信息进行脱敏。

特性增强

无。

特性约束

- 只针对SQL语法中明确定义的口令或密钥信息进行脱敏，不支持用户自定义参数场景。
- 如果执行SQL语句中存在语法错误场景可能会导致脱敏失效。

依赖关系

无。

1.6.12 全密态数据库等值查询

可获得性

本特性自V500R001C20版本开始引入。

特性简介

密态数据库意在解决数据全生命周期的隐私保护问题，使得系统无论在何种业务场景和环境下，数据在传输、运算以及存储的各个环节始终都处于密文状态。当数据拥有者在客户端完成数据加密并发送给服务端后，在攻击者借助系统脆弱点窃取用户数据的状态下仍然无法获得有效信息，从而起到保护数据隐私的能力。

客户价值

由于整个业务数据流在数据处理过程中都是以密文形态存在，通过全密态数据库，可以实现：

1. 保护数据在云上全生命周期的隐私安全，无论数据处于何种状态，攻击者都无法从数据库服务端获取有效信息。
2. 帮助云服务提供商获取第三方信任，无论是企业服务场景下的业务管理员、运维管理员，还是消费者云业务下的应用开发者，用户通过将密钥掌握在自己手上，使得高权限用户无法获取数据有效信息。
3. 让云数据库服务借助全密态能力更好的遵守个人隐私保护方面的法律法规。

特性描述

在加解密阶段，为保证客户端能够自动化地对数据进行加解密，用户需在数据定义阶段定义加密方案。同时，GaussDB新增密钥管理语法，并支持第三方密钥管理工具，保证用户自主和灵活地选择加密方案。使用全密态数据库的整体流程分为如下四个阶段。

1. 密钥实体管理阶段：通过独立的密钥管理工具/服务管理密钥实体。目前，GaussDB支持通过Huawei KMS对密钥进行独立管理。
   - Huawei KMS：由华为云提供的在线密钥管理服务，提供创建、删除、查询和备份密钥等功能，并支持在线使用密钥对数据进行加解密。
   - user_token：由用户提供的密码在客户端派生密钥，或者直接对接密钥。
   - third_kms：在加载第三方加密库后，由第三方加密库提供密钥管理服务。

   须知

   使用third_kms时，密钥由第三方加密库管理。第三方加密库并非华为提供，需要保证该动态库功能执行正常与安全。

   如果第三方动态库异常或其他不可控因素，可能会导致数据库进程异常、进程崩溃、内存泄露等，需要联系第三方动态库进行解决，请谨慎使用。
2. 密钥对象管理阶段：通过新增的密钥管理SQL语法管理密钥对象，新增语法如下。
   - CREATE COLUMN ENCRYPTION KEY：支持用户定义用于加密表中指定列的密钥对象，同时，该对象中存储了列加密密钥实体的密文。
   - CREATE CLIENT MASTER KEY：支持用户定义用于加密CEK的CMK对象，该CMK对象不存储CMK密钥实体，而是存储从独立密钥管理工具/服务中读取CMK密钥实体的方法。
3. 数据定义阶段：新增对表中指定列进行加密定义的语法，新增语法如下。
   - CREATE TABLE ... (column DATE_TYPE ENCRYPTED WITH ...)：支持用户指定CEK来加密指定的列。
4. 数据加解密阶段：完成数据定义后，客户端便能够基于用户定义的加解密方案，自动地对表中数据进行加解密。

具体的语法可参考《开发者指南》中的“SQL参考 > SQL语法”章节。

特性增强

无。

特性约束

- 密钥实体管理约束。
  - 仅支持使用密钥管理服务Huawei KMS管理CMK密钥实体。
- 密钥对象管理约束。
  - CREATE CLIENT MASTER语法中，KEY_PATH字段仅能指向外部密钥管理工具/服务中已存在的密钥；由Huawei KMS生成的密钥，仅能用于AES_256和SM4算法。
  - CREATE COLUMN ENCRYPTION KEY语法中，ALGORITHM仅支持AEAD_AES_256_CBC_HMAC_SHA256、

AEAD_AES_128_CBC_HMAC_SHA256、AEAD_AES_256_CTR_HMAC_SHA256、AES_256_GCM和SM4_SM3。

- 如果使用由Huawei KMS生成CMK来加密CEK，在CREATE COLUMN ENCRYPTION KEY语法中，如果使用ENCRYPTED_VALUE字段，则该字段的长度需为16字节的整数倍。

- 数据以列级别进行加密，而无法按照行级别区分加密策略。
- 除Rename操作外，不支持通过Alter Table语法实现对加密表列的更改(包括加密列和非加密列之间的互转换)，支持添加(Add)和删除(Drop)对应的加密列。
- 不支持对加密列设置大部分check限制性语法，但是支持check(column is not null)语法。
- 当support_extended_features = off时，不支持通过DISTRIBUTE BY子句指定加密列为分布列。当support_extended_features = on时，仅支持通过DISTRIBUTE BY子句指定确定性加密列为哈希分布列。
- 当support_extended_features = off时，不支持对加密列使用primary key、unique。当support_extended_features = on时，仅支持确定性加密列使用primary key、unique。
- 仅支持对加密列建btree索引以及ubtree索引，不支持建索引的时候使用加密列做过滤操作。
- 不支持不同数据类型之间的隐式转换。不支持转义字符。
- 不支持不同数据类型密文间的集合操作。
- 不支持加密列为数组类型。
- 不支持加密列创建分区。
- 加密列仅支持repeat和empty_blob()函数。
- 当前版本只支持gsql和JDBC（部署在linux操作系统）客户端，暂不支持ODBC等其他客户端实现密态等值查询。
- 只支持通过客户端执行copy from stdin的方式、\copy命令的方式以及insert into values(…)的方式往密态表中导入数据。
- 不支持从加密表到文件之间的copy操作。
- 不支持包括范围查询以及模糊查询等在内的除等值以外的各种密态查询。
- 支持部分函数存储过程密态语法，密态支持函数存储过程具体约束查看《开发者指南》的“设置密态等值查询 > 密态支持函数/存储过程”章节。
- 不支持通过insert into…select…，merge into语法将非加密表数据插入到加密表数据中。
- 仅JDBC客户端支持调用decryptData接口，将通过非密态连接、逻辑解码等其他方式获得的密文，对密文进行解密。调用方法查看《特性指南》中“设置密态等值查询 > 使用JDBC操作密态数据库 > 执行密态等值密文解密”。
- 对于处于连接状态的连接请求，只有触发更新缓存的操作（更改用户，解密加密列失败等）和重新建连后才能感知服务端CEK信息变更。
- 不支持在由随机加密算法加密的列上进行密态等值查询，仅支持简单插入语法及全表查询。
- 不支持不同精度、不同原始数据类型或使用不同列加密密钥加密的密文数据进行数据导入或等值比较。
- 密态等值查询不支持外表。
- 不支持针对包含加密列的密态表创建物化视图。

- 不支持针对包含加密列的密态表及基于密态表的视图、函数、存储过程创建同义词。
- 对于数据库服务侧配置变更(pg_settings系统表、权限、密钥和加密列等信息), 需要重新建立JDBC连接保证配置变更生效。
- 不支持多条SQL语句一起执行，insert into语句多批次执行场景不受此条约束限制。
- 密态数据库对长度为零的空字符串不进行加密。
- 确定性加密存在频率攻击的潜在风险，不建议在明文频率分布明显的场景下使用。
- 密态表不支持闪回drop，闪回查询和闪回表。
- 密态等值查询采用客户端默认精度，与服务端精度设置不相关。
- COLLATE子句指定列的排序规则（该列必须是可排列的数据类型），加密列类型为非可排序的数据类型。
- JDBC不支持加密列使用setBlob接口。
- 不支持使用prepare执行DDL操作。
- 当update语句有临时表时，where条件不支持加密列做查询条件。
- 创建预编译语句时，同一个参数请勿同时用于预编译语句中的加密列和非加密列参数。
- 使用密态数据库过程中，请勿将数据从一个加密列插入到另一个密钥加密的加密列中，如insert into t1 values(select * from t2)，否则当该数据的密钥删除后，另一个密钥加密的加密列中有数据会获取不到该密钥。
- 当单事务中的加密字段多或者单次数据量过大，可能造成加密时间过长，产生超时等异常，建议拆分为子语句进行处理。
- 当使用third_kms时，密钥由第三方加解密库的KMS管理，数据库仅记录加解密所需密钥的ID。
- 当使用third_kms时，不支持密钥轮转语句轮转密钥。
- 原始类型为varchar、text、varchar2、clob、bytea的密文在数据导入时视为同一数据类型允许互相导入。
- 密态等值查询支持的数据类型包括：

数据类	类型	描述
整型	tinyint/tinyint(n)	微整数，同int1。
	smallint	小整数，同int2。
	int4	常用整数。
	binary_integer	Oracle兼容类型，常用整数。
	bigint/bigint(n)	大整数，同int8。
数值类型	numeric(p,s)	精度为p的准确数值类型。
	number	Oracle兼容类型，等同numeric(p,s)。
浮点类型	float4	单精度浮点数。
	float8	双精度浮点数。

	double precision	双精度浮点数。
字符类型	char/char(n)	定长字符串，不足补空格，默认精度为1。
	varchar(n)	变长字符串，n是指允许的最大字节长度。
	text	文本类型。
	varchar2(n)	Oracle兼容类型，等同varchar(n)。
	clob	大文本类型。
二进制类型	bytea	变长的二进制字符串。
	blob	二进制大对象。该类型按照字符串处理，不支持其他转换操作。

依赖关系

使用全密态相关特性建议更新至相同版本的libpq_ce客户端驱动及JDBC客户端。

1.6.13 账本数据库机制

可获得性

本特性自V500R002C00版本开始引入。

特性简介

账本数据库特性，对用户指定的防篡改表增加校验信息，并记录用户对防篡改表中数据的修改历史，通过数据和修改历史的一致性校验来识别用户数据是否被恶意篡改。在用户对防篡改表执行DML操作时，系统对防篡改表追加少量行级校验信息，同时记录操作的SQL语句和数据的变化历史。通过特性提供的校验接口，用户可以方便的校验防篡改表中的数据是否与系统记录的操作信息是否一致。

客户价值

账本数据库提供对用户数据的操作记录、数据历史变化记录以及易用的一致性的校验接口，方便用户随时校验数据库中的敏感信息是否被恶意篡改，有效提高数据库防篡改能力。

特性描述

账本数据库采用账本Schema对普通表和防篡改用户表进行隔离。用户在账本Schema中创建的行存表具有防篡改属性，即为防篡改用户表。用户向防篡改用户表中插入数据时，系统会自动生成少量行级校验信息。当用户执行DML时，系统会在全局区块表(GS_GLOBAL_CHAIN)中记录用户的操作、同时在用户表对应的历史表中记录数据的更改等信息，操作记录、数据变化记录和用户表中的数据三者严格保持一致。账本数据库提供高性能校验接口，能够供用户方便的校验数据的一致性，如果一致性校验失败，则说明数据可能发生篡改，需要及时联系审计管理员回溯操作记录历史。

> **说明**
>
> 防篡改表在创建时，支持以下数据类型：
>
> char, abstime, bigint, boolean, bytea, character varying, character, date, double precision, int2vector, integer, interval, money, name, numeric, nvarchar2, oid, oidvector, raw, real, reltime, smalldatetime, smallint, text, time with time zone, time without time zone, timestamp with time zone, timestamp without time zone, tinyint, uuid, clob。

特性增强

无。

特性约束

- 防篡改模式下的行存表具有防篡改属性，而临时表、UNLOGGED表等均不具有防篡改属性。升级过程中需要在特定schema中创建表的情况下，不建议在升级前将该schema设置为防篡改属性。
- 不允许修改防篡改用户表的结构，不允许truncate防篡改相关表，不允许将防篡改用户表切换到普通的Schema中，不允许将非防篡改表切换到防篡改Schema中。
- 防篡改表如果为分区表，则不支持exchange partition、drop partition、truncate partition等操作。
- 分布式场景下，不支持使用函数、存储过程、TRIGGER修改防篡改用户表数据，允许操作类型为SELECT的存储过程执行，不支持通过EXECUTE DIRECT ON的方式直接修改DN节点的防篡改表。
- 防篡改用户表使用CREATE TABLE LIKE语句，源表为防篡改用户表，新表继承源表中所有字段名（包含检验列），会执行失败，不支持用户主动修改校验列的值以及对该列创建索引。
- 普通用户调用篡改校验接口只能校验自己有权查询的表。
- 只允许审计管理员、系统管理员和初始用户查询全局区块表和BLOCKCHAIN模式中的表，普通用户无权访问，所有用户均无权修改。
- 扩容后，校验数据可能会不一致，此时可执行账本数据库在线修复功能保证后续篡改校验功能的正常执行。
- CN剔除修复后，会丢失部分全局区块表中的数据，需要使用全局区块表修复功能，保证后续的账本数据库的正常使用。
- 根据用户历史表命名规则，若待创建表的Schema或表名以'_'结尾或开头，可能会出现对应历史表名与已有表名冲突的情况，需要重新命名。
- 账本数据库目前针对用户行级数据的hash摘要仅用来保证数据的一致性，无法保证密码学完整性，且当前能力暂时无法防止攻击用户直接对数据文件的篡改。
- 创建防篡改schema以及更改普通schema为防篡改schema，需设置enable_ledger参数为on。enable_ledger参数默认值为off。
- 基于防篡改表创建的定时任务对防篡改表有数据更改操作，将执行失败。
- 不允许基于防篡改表创建RULE。
- 对防篡改表执行闪回表操作，历史表和全局区块表无法自动进行数据同步，为保持数据一致，需对历史表执行闪回表操作，使用ledger_gchain_repair修复全局区块表。
- 由于在CTE或者子计划中修改表，会存在一个嵌套插入，导致无法记录整体的防篡改表修改，因此当前不支持使用CTE或者子计划修改防篡改表。

- 集群运行异常造成账本数据库校验函数返回不一致时，需要使用账本数据库修复函数对不一致的历史表或全局区块表进行修复。

依赖关系

无。

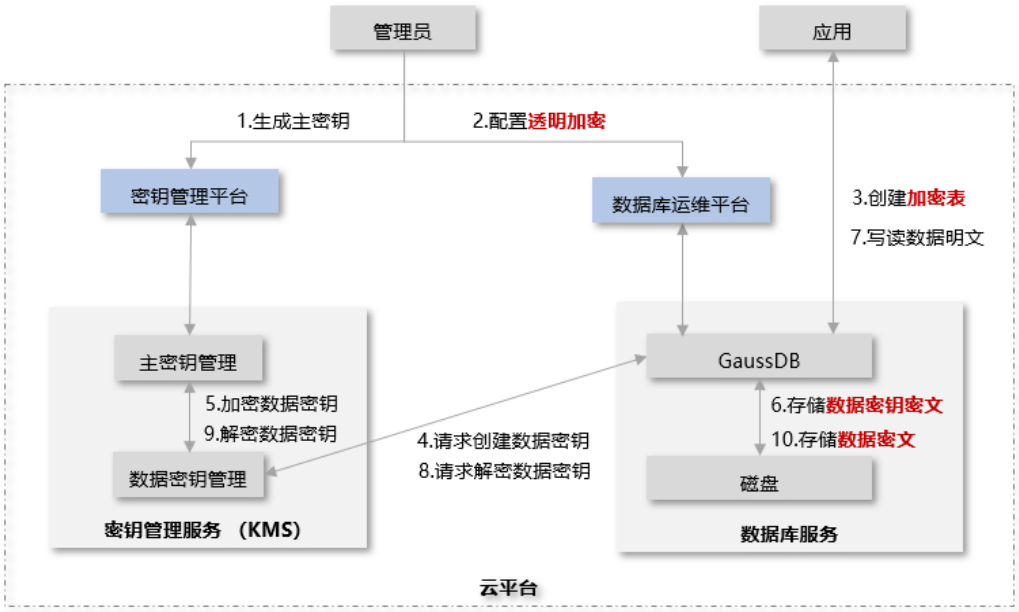
1.6.14 透明数据加密

可获得性

本特性自V500R002C00版本开始引入。

特性简介

透明数据加密（Transparent Data Encryption，TDE）提供表级数据加密存储功能。当用户使用本特性提供的语法创建加密表后，数据库向磁盘写入加密表数据前，会自动将其加密；同时，数据库从磁盘读取加密表数据后，会自动将其解密。



客户价值

安全：本特性可有效解决攻击者绕过数据库认证机制直接读取数据库文件引起静态数据泄露的问题。

易用：用户在创建表时，通过语法指定表是否需要被加密，数据库可自动对数据进行加密存储。

性能：本特性仅在数据写入磁盘时加密，并以数据页为粒度对数据进行加密，对数据库性能影响较小。

特性描述

多级加密：本特性使用多级加密模型，密钥分为主密钥（CMK）和数据加密密钥（DEK），数据由DEK加密，而DEK由CMK加密，CMK由外部密钥管理服务（Key

Management Service，KMS）加密存储。数据库在运行过程中，需要通过网络或其他途径访问外部密钥管理者，以实现对DEK进行加解密。

表级加密：在创建表时，通过ENABLE_TDE语法指定表为加密表，通过ENCRYPT_ALGO语法指定使用何种加密算法，同时，数据库会自动为每个加密表生成1个DEK。对于每个加密表，数据库会在系统表和数据文件中存储加密信息以及DEK密文。

密钥管理：CMK由外部密钥管理者生成并存储。目前的外部密钥管理者主要指密钥管理服务（Key Management Service，KMS），大部分云服务提供商均提供KMS。

密钥轮转：本特性提供DEK轮转语法，加密表与非加密表转换语法。

特性增强

503.2.0：

- 密钥管理支持。
- 存储引擎支持Ustore。

505.1.0：

- 支持段页式表。
- 支持hashbucket表。
- 支持对索引加密，支持直接将非加密表转换为加密表。

特性约束

规格

- 加密规格
  - 加密对象：支持对astore表、ustore表、临时表、unlog表、段页式表等表加密，不支持对压缩表、物化视图、toast表等其他表加密。支持对btree索引、ubtree索引加密。
  - 加密算法：支持AES_128_CTR（默认算法）、SM4_CTR。
- 密钥规格
  - 密钥管理：主密钥由单独的密钥服务管理，支持以下密钥服务：华为云密钥管理服务（ ）、第三方密钥管理服务（third_kms）。
  - 密钥隔离：每个加密表都使用单独的数据密钥。
  - 密钥复用：如果对索引进行加密，则索引与基表使用相同的数据密钥与加密算法，不支持单独指定索引的加密算法。
  - 密钥轮转机制：进行密钥轮转时，表中新数据页将使用新密钥，旧数据页仍使用旧密钥，执行VACUUM FULL等重建数据文件的操作后，旧数据页才会使用新密钥。
- 使用规格
  - 加密表转换：

    1. 将非加密表转换为加密表时，表中新数据页将会被加密，旧数据页不会立刻被加密，执行VACUUM FULL等重建数据文件的操作后，旧数据页才会被加密。

    2. 将加密表转换为非加密表时，表中新数据页将不会被加密，旧数据页仍处于加密状态，执行VACUUM FULL等重建数据文件的操作后，旧数据页才会被解密。

须知

使用third_kms时，密钥由第三方加密库管理。第三方加密库并非华为提供，需要保证该动态库功能执行正常与安全。

如果第三方动态库异常或其他不可控因素，可能会导致数据库进程异常、进程崩溃、内存泄露等，需要联系第三方动态库进行解决，请谨慎使用。

约束

- 运行环境约束

  网络约束：需保证每个数据库节点与KMS之间网络通畅。

- 配置约束

  特性开关：如果开启透明加密，创建加密表并向加密表中写入数据，在关闭透明加密功能后，无法对加密表进行读写操作。

- 功能约束
  - 索引加密：只支持对基表为加密表的索引进行加密。
  - 加密范围：支持对表和索引的数据文件中的数据进行加密，不支持对网络传输、xlog文件、pg_static系统表文件和core文件等其他介质中的数据进行加密。
  - 废弃数据：不对数据库清理机制产生的废弃数据进行加密。
  - 超长数据：由于加密表的数据页容量小于非加密表，如果非加密表中含长度接近8000字节的单条非toast数据，请谨慎该表转换为加密表，否则可能出现加密表可能无法存储超长数据的异常。解决异常的方案是通过ALTER .. SET (enable_tde=off)语法将加密表还原为非加密表，异常场景示例如下：
    - 异常示例1：将表转换为加密表后，执行VACUUM FULL tablename语法语法失败。
    - 异常示例2：将表转换为加密表后，UPDATE旧数据失败。
- 特性交互约束
  - 备份恢复：不支持细粒度备份恢复。
  - repair：调用repair函数修复加密表的数据页时，不对生成的临时文件中数据进行加密。
- 其他约束
  - 性能劣化：与非加密表相比，在加密表上进行DML操作时，性能会较小劣化。
  - 数据膨胀：与非加密表相比，加密表数据文件中存储了加密信息，存储空间缩小5%以内。

依赖关系

本特性依赖外部密钥管理服务提供密钥管理功能。

1.6.15 基于标签的强制访问控制

可获得性

本特性自503.2.0版本开始引入。

特性简介

基于标签的强制访问控制特性支持用户对主体和客体设置安全标签，并基于系统设定好的强制访问控制策略规则执行访问控制。通过给主体（用户或角色）和客体（表或表的列）设置合适的安全标签来控制用户/角色可以操作数据库的哪些表或表的列。

系统新增语法和系统表支持安全标签的创建删除和记录，然后在针对数据库表或表的列进行权限校验的位置，增加主体和客体安全标签的比较逻辑，通过比较安全标签级别和范围是否符合强制访问控制策略规则来决定校验是否通过，不通过则拒绝访问。

客户价值

基于标签的强制访问控制是由系统管理人员设置主体和客体的安全标签，系统会按照严格的强制访问控制策略进行访问控制，规则由系统决定，不能更改，从而对数据库中的敏感信息提供更严格的权限控制。

特性描述

首先用户根据业务需要创建由等级和范围组成的安全标签，并将安全标签分别应用到主体（用户或角色）和客体（表或表的列）上。然后当强制访问控制检查开关打开（enable_mac_check=on），用户执行DML操作时，系统会自动根据内置的基于安全标签的强制访问控制策略校验主体是否被允许访问客体，如果校验不通过，则访问失败。

- 安全标签由等级和范围两部分组成，两者中间用冒号分隔，形式如：等级类别:范围类别，其中等级类别有且仅由一个等级组成，范围类别可由多个范围组成，但至少需要有一个范围，例如“L1:G2,G41,G6-G27”。
- 等级分类中有1024个等级，命名为Li，其中1≤i≤1024，等级满足偏序关系（若i ≤ j，则Li ≤ Lj），例如等级L1小于等级L3。
- 范围分类中有1024个范围，命名为Gi，其中1≤i≤1024，范围之间无法比较大小，但可以进行集合运算，多个范围之间用逗号分隔，连字符表示区间，例如{G2-G5}表示{G2,G3,G4,G5}，集合{G1}是集合{G1,G6}的子集。
- 等级和范围的首字母L和G均为大写；L和G之后至少要有一个数字字符，且第一位非零，不允许出现其他非数字字符；{Gxxx-Gyyy}形式中数字yyy必须大于等于xxx。
- 不符合要求的等级和范围均为非法输入，系统会报错。

基于安全标签的强制访问控制策略规则由系统内置设定，用户不能更改：

- 插入（INSERT）策略：只有主体安全标记等级小于等于客体安全标记等级且主体安全标记范围是客体安全标记范围的子集时才允许插入数据。
- 查询（SELECT）策略：只有主体安全标记等级大于等于客体安全标记等级且主体安全标记范围是客体安全标记范围的超集时才允许查询数据。
- 修改（UPDATE）和删除（DELETE）策略：只有主体安全标记等级等于客体安全标记等级且主体安全标记范围等于客体安全标记范围时才允许修改和删除数据。
- 若客体未标记，则强制访问控制策略对该客体不生效。
- 若主体未标记，则不能访问任意带有标记的客体。

特性增强

无。

特性约束

- 初始用户、具有SYSADMIN权限的用户或者继承了内置角色gs_role_seclabel权限的用户有权限创建、删除和应用安全标签。
- 只有初始用户才能应用或取消初始用户和具有persistence属性的用户的安全标签。
- 对主体和客体设置安全标签，主体支持用户和角色，客体支持普通表（pg_class中的relkind='r'）和普通表的列，不支持在系统表和系统表的列上应用安全标签。
- 打开强制访问控制开关后，基于安全标签强制访问控制与原有的自主访问控制（ACL权限和ANY权限）是“与”的关系，需要都满足才能访问成功。
- 出于防呆考虑，初始用户和系统管理员默认也受强制访问控制限制，但是可以随时更改或取消安全标签来使自己满足强制访问控制策略规则。
- 对于设置了安全标签的表，需要谨慎授予其他用户对该表的trigger权限，以免其他用户利用触发器绕过强制访问控制策略规则。

依赖关系

无。

1.6.16 敏感数据发现

可获得性

本特性自505.1.0版本开始引入。

特性简介

敏感数据发现功能提供函数gs_sensitive_data_discovery()和gs_sensitive_data_discovery_detail()，通过调用的不同函数，指定扫描对象和敏感数据分类器，得到对应扫描对象不同明细级别的敏感数据信息。

客户价值

敏感数据发现功能配合数据库内其他数据标记（如**动态数据脱敏机制**，**基于标签的强制访问控制**等）和保护特性（如**透明数据加密**，**行级访问控制**等）能够帮助客户：

1. 有效识别敏感信息和资产。
2. 提供更全面的数据保护能力。
3. 满足客户隐私数据保护以及符合监管的需求。

特性描述

本特性以函数调用的形式实现功能。分为两个部分：

1. 敏感数据发现函数框架的实现：包含内置函数的添加、参数校验、采样扫描、对采样数据应用分类器、扫描结果处理、异常处理、函数执行动作记录审计日志等。
2. 敏感数据分类器的具体算法实现：本次支持内置以下分类器（不区分大小写），分别为Email（电子邮件）/ PhoneNumber（电话号码）/ CreditCard（信用卡）/ ChineseName（中文姓名）/ EncryptedContent （加密数据）/ all（以上五种分类器全部选择）。

具体分类原理及命中规则如下表1-4所示

表 1-4 敏感数据分类器及分类器说明

分类器	分类器说明
电子邮件	匹配形如“example@example.com”、“example@example.co.uk”、“example@example.com.org”等的电子邮件地址。
电话号码（中国）	匹配手机号码，必须是（+国家代码）手机号格式，手机号长度必须是11位数字，形如：+86 xxxxxxxxxxx，xxxxxxxxxxx，xxx xxxx xxxx，xxx-xxxx-xxxx。
信用卡	信用卡卡号位数需满足16位并且满足**Luhn（mod 10）**算法要求。 **说明** Luhn（mod 10）算法： 1. 从卡号最后一位数字开始，逆向将奇数位（1、3、5等）相加； 2. 从卡号最后一位数字开始，逆向将偶数位数字，先乘以2（如果乘积为两位数，则将其减去9），再求和； 3. 将奇数位总和加上偶数位总和，结果能被10整除则认为是合法的信用卡号，否则不是。
中文姓名	使用中国常见的姓氏进行判断。需要满足以下约束： ● 仅支持简体中文。 ● 对于某些少数民族姓名，可能存在漏报场景。 ● 仅对姓名长度为2-4的数据进行判断，超出此范围的数据不认为满足姓名规则。 ● 姓或名拆开为单独的列不会被识别成敏感。 ● 仅姓在前名在后的数据会被认为是符合规则的，姓在后名在前的数据也不会被识别为敏感。
加密数据	通过计算数据信息熵，并与设置的**阈值**相比较，信息熵高于该**阈值**，则认为数据为敏感。该**阈值**保证数据库中所有加解密算法得到的密文均能识别为密文敏感数据。

特性增强

无。

特性约束

- 仅支持对扫描范围内可访问对象中的列应用分类器进行判断，无权限对象则跳过，函数执行结果中回显NOTICE提示存在无权限访问的对象。
- 仅支持对用户创建的模式、普通表、分区表、二级分区表进行扫描，不支持对snapshot、AI、系统schema等模式进行扫描，不支持对索引、物化视图等进行扫描。如果指定的扫描范围本身是规格范围外的对象，则会提示不支持；如果指定的扫描范围本身支持，但其中包含的某些对象不支持，则不会提示。
- 中文姓名分类器中只支持简体中文。对于少数民族姓名可能无法识别，会有漏报出现。

- 不建议在数据库业务繁忙阶段调用敏感数据发现函数、不建议多个客户端并发执行敏感数据发现函数、不建议在短时间内多次调用敏感数据发现函数，性能敏感场景下，可以开启IO管控和流控后再调用敏感数据发现函数。
- 由于敏感数据发现是在堆表层进行扫描，不会受到行级访问控制策略的影响。添加了行级访问控制的行数据也可以被正常扫描。

依赖关系

无。

1.7 资源管理

1.7.1 资源管控

可获得性

本特性自505.1.0版本开始引入。

特性简介

资源管控作为企业级应用场景的关键诉求，在MetaERP等场景下有着重要的作用。当进程或者某一单一线程占用大量的资源时，需要资源管控能力来限制单一进程/线程使用的资源，避免出现单一进程/线程占用大量资源导致资源的情况。

GaussDB目前已经识别的可管控的资源有：CPU、内存、I/O、最大并发数、磁盘空间等。

客户价值

提升数据库资源管理能力，针对不同用户限制不同的资源使用，避免出现某一用户占用全部资源导致其他用户无法使用资源的场景。

特性描述

- 支持进程级CPU管控和调整。
- 支持资源池粒度的CPU资源管控，用户和资源池可绑定，间接管控用户的CPU资源使用。
- 支持资源池级CPU使用的资源监控能力。
- 支持session/线程粒度的CPU管控能力。
- 支持资源池粒度的最大连接数的限制能力。
- SMP计划支持预占stream线程执行。
- 支持实例级别、资源池级别、session级别、SQL级别的内存管控和监控能力。
- 支持资源池级别、作业级别的I/O管控和监控能力。
- 支持资源池粒度的最大并发数管控能力。
- 支持进程级别的最大并发数管控能力。

特性增强

无

特性约束

- 资源管控仅在use_workload_manager=on和enable_control_group=on时生效。
- 需要初始化control group文件系统。

依赖关系

无

1.7.2 支持 I 层高时延逃生能力

可获得性

本特性自V500R002C10SPC500版本开始引入。

特性简介

I层异常会导致数据库SQL执行时延升高，进而导致内存或者线程池出现过载问题，针对此场景GaussDB提供自动逃生能力。

客户价值

当数据库由于I层异常导致SQL执行时延升高，会话堆积，内存或线程池过载无法对外提供服务时，能够快速实现逃生，短时间内恢复对外提供服务的能力。

特性描述

- 数据库内存出现过载问题时，快速kill会话并禁止新连接接入，待内存恢复正常状态后恢复对外服务能力。内存过载和恢复正常的内存阈值通过GUC参数resilience_memory_reject_percent设置，默认关闭该功能。
- 数据库线程池使用率过载时，快速kill会话并禁止新连接接入，待会话数降低到线程池可承受能力时恢复对外服务能力。线程池使用率过载和恢复正常的阈值通过GUC参数resilience_thread_reject_cond设置，默认关闭该功能。

特性增强

无。

特性约束

- 内存或者线程池过载触发逃生能力时，默认不对sysadmin或monitoradmin权限的用户的session做清理操作，若想对sysadmin或monitoradmin权限的用户的session做清理操作，请设置参数resilience_escape_user_permissions，具体请参考resilience_escape_user_permissions的详细描述和使用方法。
- 升级模式下，不触发该特性功能。

依赖关系

无。

1.7.3 并发场景支持抗过载逃生能力

可获得性

本特性自503.1.0版本开始引入。

特性简介

慢SQL导致的过载问题在现有的多个测试场景中经常出现，通常应用层在业务上需要保证对外提供的服务具备稳定可靠的SLA，每当出现慢SQL以后应用层会通过增加对数据库的连接请求数来确保SLA目标达成，因此对于数据库来说则是连接请求数增多导致的并发陡增问题，在现有的实现机制上由于连接数增多会导致CPU、内存等资源消耗增加，同时由于慢SQL无法执行完成导致执行slot被长时间占用，新的业务请求无法进入，最终导致业务吞吐量托底并且无有效恢复手段。针对这一慢SQL入侵场景，本特性针对该场景下提升过载逃生的韧性能力，通过韧性的增强，能够在大并发场景下数据库服务端保持一定能力的稳定业务输出。

客户价值

当数据库由于慢SQL无法快速执行完成导致执行slot被长时间占用，新的业务请求无法进入，最终导致数据库无法对外提供服务时，能够提升过载逃生的韧性能力，通过韧性的增强，能够在大并发场景下数据库服务端保持一定能力的稳定业务输出。

特性描述

- 支持韧性检测能力：通过定义慢SQL的执行时间来实现慢SQL检测，一旦检测出慢SQL后则启动承受能力。
- 支持韧性承受能力：慢SQL入侵被认定以后，慢SQL在总量上受限于预先设定值，避免所有的执行slot都被慢SQL所占用，能够在整体上承受慢SQL的入侵。
- 支持韧性调整能力：慢SQL入侵被认定以后，在慢车道的管控态运行，系统对其IO资源使用加以限制和隔离以降低对其他作业的影响，确保系统整体KPI不会全面恶化。
- 支持实时观测：慢SQL进入管控态后，可以通过视图查询慢SQL的具体运行状态。
- 支持灵活容错：对于资源充足或者偶发慢SQL需要具有一定包容性，零星异常SQL仍然有机会执行完毕。
- 支持灵活配置：对于预期执行时间基于规则设置，后续可自适应的选择合理的预期执行时间。

特性增强

无。

特性约束

- 仅对非sysadmin或monitoradmin权限的用户执行的select类型的语句进行慢SQL管控。
- 仅在线程池开启模式下生效。
- 资源管控仅在use_workload_manager=on时生效。

依赖关系

无。

1.7.4 SQL 限流能力

可获得性

本特性自505.0.0版本开始引入。

特性简介

在数据库系统中，时常会出现某类SQL执行异常，占用较多系统资源，或者出现某类SQL因异常或业务需求并发激增，影响其他业务执行，甚至整个数据库系统无法响应其他业务请求的情况。为了解决该问题，GaussDB实现了SQL限流的能力，可以从多维度限制某类SQL执行的并发数。

客户价值

当数据库由于某类SQL并发数突增或者长时间执行，导致其他的业务请求无法执行，最终导致数据库无法对外提供服务时，通过本特性能够限制异常SQL的并发数，让正常的业务得到保障，提升系统韧性。

特性描述

本特性可以实现多种规则的SQL限流能力，限制某类SQL或实例的最大并发数，包括：

1. 根据Unique SQL ID进行限流：在明确某条SQL为慢SQL或者占用资源较高的SQL时，可以通过Unique SQL ID对该SQL进行限流，避免业务大量执行此SQL而影响其他业务；
2. 根据SQL类型及关键字进行限流：在明确某类SQL请求可能会随业务量增长而增长的时候，使用SQL类型和关键字对此类SQL进行限流；
3. 由于某些业务高峰是可以预知的，在仅希望在业务高峰时段对SQL请求进行限制的情况下，可以设置SQL限流的生效时间，避免限流规则常驻系统；
4. 当只希望限流规则作用于业务库，而对系统库的SQL不做控制，可以按不同库的维度进行限流；
5. 除了对某类SQL进行限流，本特性还提供实例级别的限流能力；
6. 对于限流规则，提供查询统计的能力，可以查询所有的限流规则，并根据规则列表对限流规则进行管理。此外，还提供查询限流规则限制访问的SQL次数。

特性增强

无

特性约束

- 只支持在CN上限流。
- 对于Unique SQL ID限流，需要设置GUC参数enable_resource_track = on，instr_unique_sql_count > 0。
- 对于关键字限流，按照并发度排序，并发度越低优先级越高。关键字不区分大小写，支持模糊匹配。

- 数据库名称区分大小写。删除某个数据库，再创建同名的数据库，会导致所有已录入的针对这个数据库的限流规则失效。
- 基于资源的实例级最大活跃并发数限流，按照并发度排序，并发度越低优先级越高。当前无论用户设置的和实际的cpu使用率和内存使用率是多少，超过设置的限制并发数都会限流。
- 对于限流周期结束的规则，会在下次限流规则触发时将is_valid标记为false，后续不再检查。
- 对于SQL限流次数的统计没有落盘，是数据库从启动到当前的累计次数。
- 游标、存储过程中的SQL语句不会被限流。
- 管理员用户执行的SQL语句不会被限流。
- 限流规则数据库间不共享，创建限流规则时需要连接目标库。如果创建限流规则的CN被剔除并触发全量Build，则会继承全量Build目标CN中的限流规则，因此建议在各个CN上尽量都创建对应的限流规则。
- CN之间由于Unique SQL ID不同，不共享限流规则，需要用户手动在不同的CN上创建对应的限流规则。

依赖关系

无

1.8 AI 能力

1.8.1 ABO 优化器

1.8.1.1 智能基数估计

可获得性

本特性自GaussDB 503.0.0版本开始引入。

特性简介

智能基数估计利用库内轻量级算法进行多列数据分布建模，并且提供多列等值基数估计的能力。在数据分布倾斜并且列之间相关性强的数据场景下能够提供更准确的估计结果，从而给优化器提供准确的代价参考，提高计划生成准确率，提高数据库查询执行效率。

客户价值

用户可以通过创建智能统计信息改善多列统计的准确率，从而提升查询优化性能。

特性描述

智能基数估计首先利用数据库内数据样本进行数据分布建模，并且将模型压缩存储在数据库中。优化器在执行计划生成阶段触发智能估计，实现对代价更精确的估计，并且生成更优的计划。

特性增强

本版本增加了包含等值和范围条件复合查询的支持。

特性约束

- 数据库运行状态良好，无资源紧张状况。
- 仅支持FLOAT8、Double Precision、FlOAT4、REAL、INT16、BIGINT、INTEGER、VARCHAR、CHARACTER VARYING、CHAR、CHARACTER、NUMERIC、TIMESTAMP以及TIMESTAMP WITH TIMEZONE数据类型。
- 支持不超过64列的查询基数估计，由于同时受到ANALYZE语法对于列数的约束，因此当前仅支持32列基数估计模型创建。
- 为了保证系统性能，模型创建只利用一定量的数据样本（最多1024*1024），如果数据过于稀疏，估计结果可能不准确。
- 为了能够充分利用有限的内存进行模型访问加速，建议创建的AI统计列数量不超过100个，否则可能会触发内存替换，可以通过参数进行调整。
- 如果出现过长的变长字符串类型数据，可能会影响基数估计模型创建和估计的性能。

依赖关系

依赖于数据库内的多列统计信息创建语法和数据采样算法。

2 主备版

2.1 面向应用开发的基本功能

2.1.1 支持标准 SQL

可获得性

本特性自V500R001C20版本开始引入。

特性简介

SQL是用于访问和处理数据库的标准计算机语言。SQL标准的定义分成核心特性以及可选特性，绝大部分的数据库都没有100%支撑SQL标准。

GaussDB数据库支持SQL:2011大部分的核心特性，同时还支持部分的可选特性，为使用者提供统一的SQL界面。

客户价值

标准SQL的引入为所有的数据库厂商提供统一的SQL界面，减少使用者的学习成本和应用程序的迁移代价。

特性描述

具体的特性列表请参见《开发者指南》中“SQL参考 > SQL语法”章节。

特性增强

建表SQL语法支持分区表、外部表特性。

503.1.0引入DATABASE LINK特性，仅在A兼容模式下可用，扩展标准SQL语法。

特性约束

无。

依赖关系

无。

2.1.2 支持标准开发接口

可获得性

ODBC和JDBC特性自V500R001C20版本开始引入，Go特性自V500R002C10版本开始引入。

本特性自GaussDB 503.0.0版本开始引入ECPG。

特性简介

支持ODBC 3.5、JDBC 4.0及Go 1.13标准接口。

支持ECPG、ECPG用来处理嵌入式SQL-C程序。

客户价值

提供业界标准的ODBC、JDBC及Go接口，保证用户业务快速迁移至数据库。

提供ECPG常用标准接口，保证用户嵌入式SQL-C业务快速迁移至数据库。

特性描述

目前支持标准的ODBC 3.5、JDBC 4.0及Go 1.13接口，其中ODBC支持SUSE、Win32、Win64平台，JDBC无平台差异，Go只支持Linux版本，无平台差异。

ECPG具体的特性列表请参见《开发者指南》中“应用程序开发教程 > 基于ecpg开发”章节。

特性增强

增加JDBC对接第三方日志框架功能。JDBC对接第三方日志框架功能可满足用户对日志管控的需求。

特性约束

无。

依赖关系

无。

2.1.3 函数及存储过程支持

可获得性

本特性自V500R001C20版本开始引入。

特性简介

函数和存储过程是数据库中的一种重要对象，主要功能将用户特定功能的SQL语句集进行封装，并方便调用。

客户价值

1. 允许客户模块化程序设计，对SQL语句集进行封装，方便调用。
2. 存储过程会进行编译缓存，可以提升用户执行SQL语句集的速度。
3. 系统管理员通过执行某一存储过程的权限进行限制，能够实现对相应的数据的访问权限的限制，避免了非授权用户对数据的访问，保证了数据的安全。

特性描述

支持SQL标准中的函数及存储过程，其中存储过程兼容了部分主流数据库存储过程的语法，增强了存储过程的易用性。

特性增强

无。

特性约束

无。

依赖关系

无。

2.1.4 支持 SQL hint

可获得性

本特性自V500R001C20版本开始引入。

特性简介

支持SQL hint影响执行计划生成。

客户价值

提升SQL查询性能。

特性描述

Plan Hint为用户提供了直接影响执行计划生成的手段，用户可以通过指定join顺序，join、stream、scan方法，指定结果行数，指定重分布过程中的倾斜信息等多个手段来进行执行计划的调优，以提升查询的性能。

特性增强

无。

特性约束

无。

依赖关系

无。

2.1.5 Copy 接口支持容错机制

可获得性

本特性自V500R001C20版本开始引入。

特性简介

支持将Copy过程中的部分错误导入到指定的错误表中，并且保持Copy过程不被中断。

客户价值

提升Copy功能的可用性和易用性，提升对于源数据格式异常等常见错误的容忍性和鲁棒性。

特性描述

GaussDB提供用户封装好的Copy错误表创建函数，并允许用户在使用Copy From指令时指定容错选项，使得Copy From语句在执行过程中部分解析、数据格式、字符集等相关的报错不会报错中断事务，而是被记录至错误表中，使得在Copy From的目标文件即使有少量数据错误也可以完成入库操作。用户随后可以在错误表中对相关的错误进行定位以及进一步排查。

支持容错的具体错误种类请参见《 管理员指南 》中“导入数据 > 使用COPY FROM STDIN导入数据 > 处理错误表”章节。

特性增强

无。

特性约束

无。

依赖关系

无。

2.2 高性能

2.2.1 CBO 优化器

可获得性

本特性自V500R001C20版本开始引入。

特性简介

GaussDB优化器是基于代价的优化 (Cost-Based Optimization，简称CBO)。

客户价值

GaussDB CBO优化器能够在众多计划中依据代价选出最高效的执行计划，最大限度的满足客户业务要求。

特性描述

在CBO优化器模型下，数据库根据表的元组数、字段宽度、NULL记录比率、distinct值、MCV值、HB值等表的特征值，以及一定的代价计算模型，计算出每一个执行步骤的不同执行方式的输出元组数和执行代价(cost)，进而选出整体执行代价最小/首元组返回代价最小的执行方式进行执行。

特性增强

无。

特性约束

无。

依赖关系

无。

2.2.2 Ustore 存储引擎

可获得性

本特性自V500R002C00版本开始引入。

特性简介

In-place Update（原地更新）行存储引擎，简称Ustore。相比于Append Update（追加更新）行存储引擎，Ustore存储引擎可以提高数据页面内更新的HOT UPDATE的垃圾回收效率，有效降低多次更新元组后存储空间占用的问题。

Ustore存储引擎采用NUMA-aware的Undo子系统设计，使得Undo子系统可以在多核平台上有效扩展；同时采用多版本索引技术，解决索引清理问题，有效提升了存储空间的回收复用效率。

Ustore存储引擎结合Undo空间，可以实现更高效、更全面的闪回查询和回收站机制，能快速回退人为“误操作”，为GaussDB提供了更丰富的企业级功能。Ustore基于

Undo回滚段技术、页面并行回放技术、多版本索引技术、xLog无锁落盘技术等实现了高可用高可靠的行存储引擎。

Ustore完全支持ACID特性：

- 原子性（Atomicity）：原子事务是一系列不可分割的数据库操作。在事务完成（分别提交或中止）之后，这些操作要么全部发生，要么全部不发生。
- 一致性（Consistency）：事务结束后，数据库处于一致状态，保留数据完整性。
- 隔离性（Isolation）：事务之间不能相互干扰。Ustore支持读已提交隔离级别，事务只能读到已提交的数据而不会读到未提交的数据，这是缺省值。
- 持久性（Durability）：即使发生崩溃和失败，成功完成（提交）的事务效果持久保存。

客户价值

- 针对OLTP场景，实现Inplace-update，利用Undo实现新旧版本分离存储；降低类似于AStore存储引擎由于频繁更新或闪回功能开启导致的数据页空间膨胀，以及相应引起的索引空间膨胀。
- 通过在DML操作过程中执行动态页面清理，去除VACUUM依赖，减少由于异步数据清理产生的大量读写IO。通过Undo子系统，实现事务级的空间管控，旧版本集中回收。

特性描述

Ustore的关键特性如下：

- In-place Update存储模式：Ustore存储引擎将最新版本的“有效数据”和历史版本的“垃圾数据”分离存储。将最新版本的“有效数据”存储在数据页面上，并单独开辟一段Undo空间，用于统一管理历史版本的“垃圾数据”，因此数据空间不会由于频繁更新而膨胀，“垃圾数据”集中回收效率更高。
- 回滚段设计：回滚段简称Undo，负责历史记录的插入、查询以及Undo空间的分配与释放等操作，北向对接Ustore，南向对接Buffer Pool。基于历史版本直接进行回收，实现了自治式的空间管理机制，减少了I/O时的性能抖动。同时实现了多个后台线程的并发访问，降低并发业务冲突竞争，从而提高性能。
- 基于页面的并行回放技术：Ustore利用多线程技术加速日志回放，Startup线程从磁盘中读取xLog日志，把组装好的xLog记录通过Dispatcher线程分配给多个回放线程进行回放。Dispatcher线程基于页面号进行xLog记录的分配，分配更加均匀，各个回放线程并行执行回放，提高了回放的速度。
- 闪回：数据库恢复技术的一环，能够使得DBA有选择性地高效撤销一个已提交事务的影响，将数据从人为的不正确的操作中进行恢复。在采用闪回技术之前，只能通过备份恢复、PITR等手段找回已提交的数据库修改，恢复时长需要数分钟甚至数小时。采用闪回技术后，通过闪回Drop和闪回Truncate恢复已提交的数据库Drop/Truncate的数据，只需要秒级，而且恢复时间和数据库大小无关。Ustore支持闪回表、闪回查询、闪回TRUNCATE、闪回DROP，而且适用于分区表。
- UBtree：与原有的Btree索引相比，索引页面增加了事务信息，使得UBtree索引具备MVCC能力以及独立过期旧版本回收能力。In-place Update引擎支持 UBtree索引，UBtree也是In-place Update引擎的默认索引类型。支持并行创建索引、索引空间管理算法优化，索引空间进一步压缩。

特性增强

无。

特性约束

Ustore设计几乎能够覆盖SQL和未来特性集，支持大多数的SQL标准，也支持常见的数据库特性。下面介绍Ustore的各种约束。

Ustore不支持以下特性：

- 不支持串行化隔离级别。
- 对于支持row movement的分区表，不支持并发更新或删除同一行操作。
- 不支持的DDL功能：在线vacuum full/cluster、在线alter table（除新增字段、重命名等无需全量重写数据的操作外）、table sampling。
- 不支持GiST索引、SP-GiST索引、BRIN索引。
- 不支持批量访存接口。不支持rowid语义。
- 不支持创建、使用物化视图。
- 不支持单事务块或语句中既包含Astore表又包含Ustore表。
- 数据表与回滚段要同为页式。

依赖关系

无。

2.2.3 自适应压缩

可获得性

本特性自V500R001C20版本开始引入。

特性简介

数据压缩是当前数据库采用的主要技术。数据类型不同，适用于它的压缩算法不同。对于相同类型的数据，其数据特征不同，采用不同的压缩算法达到的效果也不相同。自适应压缩正是从数据类型和数据特征出发，采用相应的压缩算法，实现了良好的压缩比、快速的入库性能以及良好的查询性能。

客户价值

数据入库和频繁的海量数据查询是用户的主要应用场景。 在数据入库场景中，自适应压缩可以大幅度地减少数据量，成倍提高IO操作效率，将数据簇集存储，从而获得快速的入库性能。当用户进行数据查询时，少量的IO操作和快速的数据解压可以加快数据获取的速率，从而在更短的时间内得到查询结果。

特性描述

目前，数据库已实现了RLE、DELTA、BYTEPACK/BITPACK、LZ4、ZLIB、LOCAL DICTIONARY等多种压缩算法。数据库支持的数据类型与压缩算法的映射关系如下表所示。

-	RLE	DELTA	BITPACK/BYTEPACK	LZ4	ZLIB	LOCAL DICTIONARY
Smallint/Int/Bigint/Oid Decimal/Real/Double Money/Time/Date/ Timestamp	√	√	√	√	√	-
Tinterval/Interval/Time with time zone/	-	-	-	-	√	-
Numeric/Char/Varchar/ Text/Nvarchar2 以及其他支持数据类型	√	√	√	√	√	√

特性增强

支持对压缩算法进行不同压缩水平的调整。

特性约束

仅支持列存。

依赖关系

开源压缩软件LZ4/ZLIB。

2.2.4 分区

可获得性

本特性自V500R001C20版本开始引入。

特性简介

数据分区是在一个节点内部按照用户指定的策略对数据做进一步的水平分表，将表按照指定范围划分为多个数据互不重叠的部分。

客户价值

对于大多数用户使用场景，分区表和普通表相比具有以下优点：

- 改善查询性能：对分区对象的查询可以仅搜索自己关心的分区，提高检索效率。
- 增强可用性：如果分区表的某个分区出现故障，表在其他分区的数据仍然可用。

特性描述

目前支持范围分区表、列表分区表、哈希分区表、间隔分区表和二级分区表：

- 范围分区表：将数据基于范围映射到每一个分区，这个范围是由创建分区表时指定的分区键决定的。这种分区方式是最为常用的。

  范围分区功能，即根据表的一列或者多列，将要插入表的记录分为若干个范围（这些范围在不同的分区里没有重叠），然后为每个范围创建一个分区，用来存储相应的数据。

- 列表分区表：将数据基于各个分区内包含的键值映射到每一个分区，分区包含的键值在创建分区时指定。

  列表分区功能，即根据表的一列或者多列，将要插入表的记录中出现的键值分为若干个列表（这些列表在不同的分区里没有重叠），然后为每个列表创建一个分区，用来存储相应的数据。

- 哈希分区表：将数据通过哈希映射到每一个分区，每一个分区中存储了具有相同哈希值的记录。

  哈希分区功能，即根据表的一列，通过内部哈希算法将要插入表的记录划分到对应的分区中。

- 间隔分区表：间隔分区是一种特殊的范围分区，相比范围分区，新增间隔值定义，当插入记录找不到匹配的分区时，可以根据间隔值自动创建分区。
- 二级分区表：二级分区表是在一级分区的基础上再进行分区，分区方案是由两个一级分区的分区方案组合而来的，目前二级分区表支持范围分区、列表分区和哈希分区交叉组合的9种分区策略。

用户在CREATE TABLE时增加PARTITION参数，即表示针对此表应用数据分区功能。用户可以在实际使用中根据需要调整建表时的分区键，使每次查询结果尽可能存储在相同或者最少的分区内（称为“分区剪枝”），通过获取连续I/O大幅度提升查询性能。

实际业务中，时间经常被作为查询对象的过滤条件。因此，用户可考虑选择时间列为分区键，键值范围可根据总数据量、一次查询数据量调整。

特性增强

支持范围分区表的合并功能。

特性约束

无。

依赖关系

无。

2.2.5 高级分析函数支持

可获得性

本特性自V500R001C20版本开始引入。

特性简介

无。

客户价值

GaussDB提供窗口函数来进行数据高级分析处理。窗口函数将一个表中的数据进行预先分组，每一行属于一个特定的组，然后在这个组上进行一系列的关联分析计算。这样可以挖掘出每一个元组在这个集合里的一些属性和与其他元组的关联信息。

特性描述

简单举例说明窗口分析功能：

分析某一部门内每个人的薪水和部门平均薪水的对比。

```
SELECT depname, empno, salary, avg(salary) OVER (PARTITION BY depname) FROM empsalary;
depname | empno | salary | avg
-----------+-------+--------+-----------------------
develop | 11 | 5200 | 5020.0000000000000000
develop | 7 | 4200 | 5020.0000000000000000
develop | 9 | 4500 | 5020.0000000000000000
develop | 8 | 6000 | 5020.0000000000000000
develop | 10 | 5200 | 5020.0000000000000000
personnel | 5 | 3500 | 3700.0000000000000000
personnel | 2 | 3900 | 3700.0000000000000000
sales | 3 | 4800 | 4866.6666666666666667
sales | 1 | 5000 | 4866.6666666666666667
sales | 4 | 4800 | 4866.6666666666666667
(10 rows)
```

可以看到，通过这个avg(salary) OVER (PARTITION BY depname)分析函数，每一个人的薪水和部门的平均薪水很容易计算出来。

目前，系统支持row_number(), rank(), dense_rank(), percent_rank(), cume_dist(), ntile(),lag(), lead(),first_value(), last_value(), nth_value()分析函数。具体的函数用法和语句请参见《开发者指南》中“SQL参考 > 函数和操作符 > 窗口函数”章节。

特性增强

无。

特性约束

无。

依赖关系

无。

2.2.6 SQLBypass

可获得性

本特性自V300R002C00版本开始引入。

特性简介

通过对TP场景典型查询的定制化执行方案来提高查询性能。

客户价值

提升TP类查询的性能。

特性描述

在典型的OLTP场景中，简单查询占了很大一部分比例。这种查询的特征是只涉及单表和简单表达式的查询，因此为了加速这类查询，提出了SQLBypass框架，在parse层对这类查询做简单的模式判别后，进入到特殊的执行路径里，跳过经典的执行器执行框架，包括算子的初始化与执行、表达式与投影等经典框架，重写一套简洁的执行路径，并且直接调用存储接口，这样可以大大加速简单查询的执行速度。

特性增强

存储过程支持该特性。

特性约束

该特性的约束条件请参见《管理员指南》中“配置运行参数 > 查询规划 > 其他优化器选项 > enable_opfusion”章节。

依赖关系

无。

2.2.7 支持 HyperLogLog

可获得性

本特性自V500R001C20版本开始引入。

特性简介

通过使用HyperLogLog相关函数，计算唯一值个数Count(Distinct)，提升性能。

客户价值

提升AP/TP类查询的性能。

特性描述

HLL（HyperLoglog）是统计数据集中唯一值个数的高效近似算法。它有着计算速度快，节省空间的特点，不需要直接存储集合本身，而是存储一种名为HLL的数据结构。每当有新数据加入进行统计时，只需要把数据经过哈希计算并插入到HLL中，最后根据HLL就可以得到结果。

HLL在计算速度和所占存储空间上都占优势。在时间复杂度上，Sort算法需要排序至少O(nlogn)的时间，虽说Hash算法和HLL一样扫描一次全表O(n)的时间就可以得出结果，但是存储空间上，Sort算法和Hash算法都需要先把原始数据存起来再进行统计，会导致存储空间消耗巨大。而对HLL来说不需要存原始数据，只需要维护HLL数据结构，所以占用空间始终是1280字节常数级别。

特性增强

无。

特性约束

无。

依赖关系

无。

2.2.8 NUMA 架构优化

可获得性

本特性自V500R001C20版本开始引入。

特性简介

NUMA架构优化，主要面向ARM处理器架构特点、ARMv8指令集等，进行相应的系统优化，涉及到从操作系统、软件架构、锁并发、日志、原子操作、Cache访问等一系列的多层次优化，从而大幅提升了GaussDB数据库在ARM平台上的处理性能。

客户价值

数据库的处理性能，如每分钟处理交易量（Transaction Per Minute），是数据库竞争力的关键性能指标，在同等硬件成本的条件下，数据库能提供的处理性能越高，那么就可以提供给用户更多的业务处理能力，从而降低客户的使用成本。

特性描述

- GaussDB根据ARM处理器的多核NUMA架构特点，进行了一系列的架构相关优化。一方面尽量减少跨核内存访问的时延问题，另一方面充分发挥ARM多核算力优势，所提供的关键技术包括重做日志批量插入、热点数据NUMA分布、CLog分区等，大幅提升TP系统的处理性能。
- GaussDB基于鲲鹏芯片所使用的ARMv8.1架构，利用LSE扩展指令集实现高效的原子操作，有效提升CPU利用率，从而提升多线程间同步性能、xLog写入性能等。
- GaussDB基于鲲鹏芯片提供的更宽的L3缓存cacheline，针对热点数据访问进行优化，有效提高缓存访问命中率，降低Cache缓存一致性维护开销，大幅提升系统整体的数据访问性能。
- 鲲鹏920，2P服务器（64cores*2，内存768 GB），网络10 GE，I/O为4块NVME PCIE SSD时，TPCC为1000warehouse，性能是150万 tpmC。

特性增强

- 支持重做日志批量插入，分区CLog，提升ARM平台下的数据库处理性能。
- 支持LSE扩展指令集的原子操作，提升多线程同步性能。

特性约束

无。

依赖关系

无。

2.2.9 物化视图

可获得性

本特性自V500R001C20版本开始引入。

特性简介

物化视图实际上就是一种特殊的物理表，物化视图是相对普通视图而言的。普通视图是虚拟表，应用的局限性较大，任何对视图的查询实际上都是转换为对SQL语句的查询，性能并没有实际上提高。而物化视图实际上就是存储SQL所执行语句的结果，起到缓存的效果。

客户价值

使用物化视图功能提升查询效率。

特性描述

支持全量物化视图和增量物化视图。全量物化视图只支持全量更新；增量物化视图同时还支持增量更新功能，用户可通过执行语句把新增数据刷新到物化视图中。

特性增强

无。

特性约束

全量物化视图支持的场景与CREATE TABLE AS语句基本一致，增量物化视图支持基表简单过滤查询和UNION ALL语句。

依赖关系

无。

2.2.10 Parallel Page-based Redo For Ustore

可获得性

本特性自V500R002C00版本开始引入。

特性简介

优化Ustore Inplace Update WAL log写入，Ustore DML Operation回放提高并行度。

客户价值

对于Update的WAL使用空间减少,Ustore DML Operation回放提高并行度。

特性描述

通过利用Prefix和suffix来减少update WAL log的写入，通过把回放线程分多个类型来解决Ustore DML WAL大多都是多页面回放问题；同时把Ustore的数据页面回放按照blkno去分发。

特性增强

无。

特性约束

无。

依赖关系

依赖于Ustore引擎。

2.2.11 xLog no Lock Flush

可获得性

本特性自V500R002C00版本开始引入。

特性简介

取消WalInsertLock争抢及WalWriter专用磁盘写入线程。

客户价值

在保持原有xLog功能不变的基础上，进一步提升系统性能。

特性描述

对WalInsertLock进行优化，利用LSN（Log Sequence Number）及LRC（Log Record Count）记录了每个backend的复制进度，取消WalInsertLock机制。在backend将日志复制至WalBuffer时，不用对WalInsertLock进行争抢，可直接进行日志复制操作。并利用专用的WalWriter写日志线程，不需要backend线程自身来保证xLog的Flush。

特性增强

无。

特性约束

无。

依赖关系

无。

2.2.12 SMP 并行执行

可获得性

本特性自V500R002C00版本开始引入。

特性简介

GaussDB的SMP并行技术是一种利用计算机多核CPU架构来实现多线程并行计算，以充分利用CPU资源来提高查询性能的技术。

客户价值

SMP并行技术充分利用了系统多核的能力，来提高重查询的性能。

特性描述

在复杂查询场景中，单个查询的执行较长，系统并发度低，通过SMP并行执行技术实现算子级的并行，能够有效减少查询执行时间，提升查询性能及资源利用率。SMP并行技术的整体实现思想是对于能够并行的查询算子，将数据分片，启动若干个工作线程分别计算，最后将结果汇总，返回前端。SMP并行执行增加数据交互算子（Stream），实现多个工作线程之间的数据交互，确保查询的正确性，完成整体的查询。

特性增强

无。

特性约束

- 不满足条件的索引扫描不支持并行执行，具体情况如下：
  - 不支持hash、psort、gist索引类型；
  - 不支持bitmapscan；
  - QUERY_DOP等于1。
- MergeJoin不支持并行执行。
- WindowAgg order by不支持并行执行。
- 不支持子查询subplan和initplan的并行，以及包含子查询的算子的并行。
- 查询语句中带有median操作的查询不支持并行执行。
- 物化视图的更新不支持并行执行。
- 会触发trigger的查询不支持并行执行；特别的，对包含外键的表执行INSERT/UPDATE/DELETE操作会触发trigger。
- 包含rownum的查询不支持并行执行。
- 涉及拼接大于1G的LOB的语句不支持并行执行。

依赖关系

无。

2.2.13 CLOB/BLOB 字段长度拓展

可获得性

本特性自V500R002C10版本开始引入。

特性简介

- 支持CLOB字段超过1GB，最大32TB。
- 通过高级DBE_LOB接口，对超过1GB数据进行操作。

客户价值

兼容O，lob字段可以可超过1GB，兼容能力增强，支持超大超字段的读写。

特性描述

lob字段支持大于1GB，DBE_LOB相关接口，可以对大于1GB的数据进行读取、写入。

特性增强

无。

特性约束

1. 大于1GB数据只能通过高级包函数读取和处理、系统函数传入大于1GB数据报错。
2. 操作符、字符串函数不支持大于1GB数据。
3. 存储过程中buffer最大32KB。
4. lob列不支持distinct、group by、order by操作。
5. 高级包最大支持32TB数据。
6. lob_write接口不加update不能更新表。

依赖关系

无。

2.2.14 数据生命周期管理-OLTP 表压缩

可获得性

本特性自505.0.0版本开始引入。

特性简介

基于冷热分离的行存压缩。

客户价值

OLTP表压缩是GaussDB高级压缩中的一个特性，基于全新的压缩算法、细粒度的自动冷热判定和支持块内压缩等技术创新，可以在提供合理压缩率的同时大幅度降低对业

务的影响、增加后台调度、增加查询Job执行状态以及节约空间，能够在支持关键在线业务的容量控制中发挥重要价值。

特性描述

用户可给数据对象指定ILM策略，策略分三部分：动作、条件、范围，本期仅支持行压缩动作、XX天未修改条件、行范围。指定策略的表会在后台定时调度、评估表中的每一行是否满足条件，若满足条件则执行行压缩动作。

特性增强

无。

特性约束

- 不支持系统表、内存表、全局临时表、本地临时表和序列表。
- 支持用户为普通表、分区、二级分区设置ILM策略。
- 支持普通表、分区、二级分区的Astore/Ustore用户表和透明加密表。
- 特性仅在A兼容模式与PG模式下有效。
- Ustore不支持编解码，压缩率小于2:1。

依赖关系

无。

2.2.15 ADIO 特性与去双写

可获得性

本特性自505.1.0版本开始引入。

特性简介

ADIO异步刷页使用异步直接I/O模式完成数据库的刷页操作。在多数场景下，主机不再记录双写文件。

客户价值

随着客户数据库中数据量的不断积累，客户数据库中存储的数据量相对于机器物理内存的比值将会越来越大。此时，刷页操作效率以及带来的I/O操作会限制数据库性能的发挥。因此，通过优化刷页模式和I/O操作来提升大容量场景下的性能便尤为重要。

特性描述

ADIO异步刷页：大容量场景下，I/O资源比较紧俏。当前数据库BIO模式对于整体I/O资源利用不充分，容易导致刷页速度落后于消耗页面的速度，导致缓冲区页面消耗完时产生性能震荡，进而影响性能。本特性通过ADIO(异步直接I/O模式)充分利用IO资源，从而提升整体数据库的性能。同时，提供从BIO模式到ADIO模式的在线切换（参见《管理员指南》中“配置运行参数 > GUC参数说明”章节的“enable_adio_function”参数说明），使用户可以在不影响业务的情况下切换到ADIO模式。

去双写：增量checkpoint开启后，由于没有full page write保护，因此采用双写文件方案（即写两次）来防止半写。拉起始遇到半写页面，便能通过双写文件方案恢复。但是写两次会导致整体I/O量多一倍，而在大容量场景下IO资源很紧俏，因此去双写可以有效降低I/O使用量。当开启去双写功能时，若所有备机都处于正常状态，则主机会停止记录双写文件。若主机由于宕机产生半写页面，则通过备机页面进行修复。

特性增强

无。

特性约束

在当前版本，去双写与数据修复特性存在依赖关系，由于该版本数据修复接口的timeout为固定值，如果想让半写页面及时恢复，需要开启流控机制，以防止因主备之间页面版本差距过大而导致的频繁修复失败问题。

依赖关系

去双写依赖数据修复特性中提供的主备间相互修复的接口。

2.3 扩展性

2.3.1 支持线程池高并发

可获得性

本特性自V500R001C20版本开始引入。

特性简介

通过线程池化技术来支撑数据库大并发稳定运行。

客户价值

支撑客户大并发下，系统整体吞吐平稳。

特性描述

线程池技术的整体设计思想是线程资源池化、并且在不同连接之间复用。系统在启动之后会根据当前核数或者用户配置启动固定一批数量的工作线程，一个工作线程会服务一到多个连接session，这样把session和thread进行了解耦。因为工作线程数是固定的，因此在高并发下不会导致线程的频繁切换，而由数据库层来进行session的调度管理。

特性增强

支持线程池的动态扩缩容。

支持stream线程池，用于解决存在跨节点数据交换类查询的高并发性能。

特性约束

无。

依赖关系

无。

2.4 高可用性

2.4.1 主备机

可获得性

本特性自V300R002C00版本开始支持DN主备。

特性简介

为了保证故障的可恢复，需要将数据写多份，设置主备多个副本，通过日志进行数据同步，可以实现节点故障、停止后重启等情况下，GaussDB能够保证故障之前的数据无丢失，满足ACID特性。

客户价值

主备机功能可以支持主机故障时切换到备机，数据不丢失，业务可以快速恢复。

特性描述

主备环境支持一主多备模式。在一主多备模式下，所有的备机都需要重做日志，都可以升主。一主多备提供更高的容灾能力，更加适合于大批量事务处理的OLTP系统。

主备之间可以通过switchover进行角色切换，主机故障后可以通过failover对备机进行升主。

初始化安装或者备份恢复等场景中，需要根据主机重建备机的数据，此时需要build功能，将主机的数据和WAL日志发送到备机。主机故障后重新以备机的角色加入时，也需要build功能将其数据和日志与新主保持一致。另外，在在线扩容的场景中，需要通过build来同步元数据到新节点上的实例。build包含全量build和增量build，全量build要全部依赖主机数据进行重建，复制的数据量比较大，耗时比较长，而增量build只复制差异文件，复制的数据量比较小，耗时比较短。一般情况下，优先选择增量build来进行故障恢复，如果增量build失败，再继续执行全量build，直至故障恢复。

为了实现所有实例的高可用容灾能力，除了以上对DN设置主备多个副本，GaussDB还提供了其他一些主备容灾能力，比如CM Sever（一主多备）以及ETCD（一主多备）等，使得实例故障后可以尽可能快地恢复，不中断业务，将因为硬件、软件和人为等因素造成的故障对业务的影响降到最低，以保证业务的连续性。

特性增强

无。

特性约束

无。

依赖关系

无。

2.4.2 逻辑复制

可获得性

本特性自V500R001C00版本开始引入。

特性简介

GaussDB提供逻辑解码功能，将物理日志反解析为逻辑日志，通过DRS等逻辑复制工具将逻辑日志转化为SQL语句，到对端数据库回放，达到异构数据库同步数据的功能。目前支持GaussDB数据库与MySQL数据库、Oracle数据库之间的单向、双向逻辑复制。

客户价值

逻辑复制可以为数据库数据实时迁移、双库双活、支持滚动升级提供解决方案。

特性描述

DN通过物理日志反解析为逻辑日志，DRS等逻辑复制工具从DN抽取逻辑日志转换为SQL语句，到对端数据库（如MySQL）回放。逻辑复制工具同时从对端数据库抽取逻辑日志，反解析为SQL语句之后回放到GaussDB，达到异构数据库同步数据的目的。

特性增强

GaussDB V500R001C00版本逻辑解码新增全量+增量抽取日志的方案。

GaussDB V500R002C00版本逻辑解码新增备机支持逻辑解码。

GaussDB V500R002C10版本逻辑解码的用户黑名单特性实现了输出逻辑解码日志的事务的用户粒度过滤，在解码阶段提前过滤非期望用户的事务操作，避免资源无效使用，进一步提升解码的性能。

GaussDB 503.1.0版本逻辑解码新增支持心跳日志，对于大事务解码过程中防止DRS长时间未收到GaussDB消息而误判。

GaussDB 503.1.0版本逻辑解码用户支持对逻辑解码过程中性能指标的统计，从而对性能瓶颈能够进行定位定界。

GaussDB 503.1.0版本逻辑解码新增速率监控功能，用于进行逻辑解码过程中各模块的运行速率监控。

GaussDB 505.0.0版本逻辑解码新增对DDL语句的支持。

GaussDB 505.0.0版本备机逻辑解码新增对极致RTO的支持。

特性约束

无。

依赖关系

依赖于逻辑复制工具对逻辑日志进行解码。

2.4.3 在线节点替换

可获得性

本特性自V500R001C20版本开始引入。

特性简介

数据库内某节点出现硬件故障造成节点不可用或者实例状态不正常，当数据库没有加锁，通过节点替换或修复故障实例来恢复数据库的过程中，支持用户DML操作，有限场景支持用户DDL操作。

客户价值

随着企业数据规模不断增大，节点数量急剧增加，硬件损坏概率相应增加，物理节点替换修复成为日常运维工作的常态。传统的离线节点替换方式无法满足客户业务不中断需求，日常运维操作中，经常的业务中断将给客户带来重大损失。而目前业界数据库产品在节点替换的过程中，或者需要中断业务，或者只允许部分操作，均不能满足大规模数据情况下，常态物理节点替换的需求。在线节点替换特性解决了以上问题，提升了数据库运行的可靠性，可为用户提供更加稳定的数据服务。

特性描述

如果数据库集群内某节点因为出现硬件故障而造成节点不可用或者实例不正常时，且集群未上锁的前提下，在通过节点替换或修复故障实例来恢复集群的过程中，支持用户DML操作，有限场景支持用户DDL操作。

特性增强

无。

特性约束

目前集群未上锁的前提下，节点替换已支持用户业务在线DDL：

- 在节点替换窗口期内，支持用户DML操作，有限场景支持用户DDL操作。

依赖关系

无。

2.4.4 物理备份

可获得性

本特性自V500R001C20版本开始引入。

特性简介

支持将整个数据库的数据以内部格式备份到指定的存储介质中。

客户价值

通过物理备份特性，可以达成以下目的：

- 整个数据库的数据备份到可靠性更高的存储介质中，提升系统整体的可靠性。
- 通过采用数据库内部的数据格式，极大提升备份恢复性能。
- 可以用于冷数据的归档。

典型的物理备份策略和应用场景如下：

- 周一，执行数据库全量备份。
- 周二，以周一全量备份为基准点，执行增量备份。
- 周三，以周二增量备份为基准点，执行增量备份。
- ...
- 周日，以周六增量备份为基准点，执行增量备份。

上述备份策略以一个星期为周期。

特性描述

GaussDB提供物理备份能力，可以将整个数据库的数据以数据库内部格式备份到本地磁盘文件、OBS对象、NAS对象或REMOTE对象中，并在同构数据库中恢复整个数据库的数据。在基础之上，还提供压缩、流控、断点续备等高阶功能。

物理备份主要分为全量备份和增量备份，区别如下：全量备份包含备份时刻点上数据库的全量数据，耗时时间长（和数据库数据总量成正比），自身即可恢复出完整的数据库；增量备份只包含从指定时刻点之后的增量修改数据，耗时时间短（和增量数据成正比，和数据总量无关），但是必须要和全量备份数据一起才能恢复出完整的数据库。除此之外，还支持从数据库实例的备份数据中恢复单个或多个数据库、单个或多个表，全量、增量备份单个或多个表以及从表级备份数据中恢复单个或多个表的功能。

特性增强

支持全量备份和增量备份同时执行。

503.0.0之后支持OBS、NAS、DISK介质下的备机备份能力，即备份操作在备DN实例上执行。

特性约束

物理备份的约束条件请参见《管理员指南》中“备份与恢复 > Roach工具介绍 > 约束和限制”章节。

依赖关系

无。

2.4.5 极致 RTO

可获得性

本特性自V500R001C20版本开始引入，从503.1.0版本开始支持备机读。

特性简介

- 支撑数据库主机重启后快速恢复的场景。
- 支撑主机与同步备机通过日志同步，加速备机回放的场景。

客户价值

当业务压力过大时，备机的回放速度跟不上主机的速度。在系统长时间的运行后，备机上会出现日志累积。当主机故障后，数据恢复需要很长时间，数据库不可用，严重影响系统可用性。

在硬件资源充足的情况下，开启极致RTO（ Recovery Time Object，恢复时间目标 ）特性，可以减少备机的RTO，减少了主机故障后数据的恢复时间，提高了系统的可用性。

特性描述

极致RTO开关开启后，xLog日志回放建立多级流水线，提高并发度，提升日志回放速度。

采用page多版本的方式支持备机读，回放线程维护每一个page的日志链，读线程根据指定的LSN（ wal日志的位置 ）读取对应版本的page。当查询和回放冲突时，查询超时会被取消，报错信息是"canceling statement due to conflict with recovery"，错误码是40001，详细信息参见《 错误码参考 》的内核错误码>GAUSS-19501--GAUSS-20000章节。当出现这种类型的报错时，业务端可根据错误码进行重试。

造成查询和回放冲突的日志类型主要包含如下几种：

1. 删除文件

   触发条件：删除文件、reindex、truncate表等操作。

   处理方案：等待max_standby_streaming_delay时间后，发送cancel消息取消冲突的查询。
2. drop database

   触发条件：执行删除数据库操作。

   处理方案：等待max_standby_streaming_delay时间后，发送cancel消息取消冲突的查询。
3. drop tablespace

   触发条件：删除tablespace。

   处理方案：等待max_standby_streaming_delay时间后，发送cancel消息取消冲突的查询。
4. vacuum清理（ 仅在参数exrto_standby_read_opt开启下，会产生冲突 ）

触发条件：vacuum操作。

处理方案：等待max_standby_streaming_delay时间后，发送cancel消息取消冲突的查询。

打开备机读之后，因为需要维护历史page版本，所以会占用更多I/O。

特性增强

为了充分发挥极致RTO基于多核CPU架构对回放性能的优化效果，建议将GUC参数redo_bind_cpu_attr（该参数用于控制回放线程的绑核操作）设置为cpuorderbind 类型，例如'cpuorderbind:16-32'。绑核区间应与通过GUC参数thread_pool_attr设置的线程池绑核区间以及通过GUC参数wal_rec_writer_bind_cpu、walwriteraux_bind_cpu、wal_receiver_bind_cpu设置绑定的cpu核号错开，区间大小根据线程数要调整，建议设置为大于等于recovery_parallelism（实际回放线程个数）+ 1。推荐将所有的回放线程绑定到一个numa组内，性能会更好。

特性约束

- 极致RTO采用了多个page redo线程并行加速回放进度。当备机回放追平主机，空载的情况下，单个page redo线程的CPU消耗大约在15%左右（实际值与具体硬件和参数配置相关），备机回放的总CPU消耗值 = 单个page redo线程的CPU消耗值 x page redo线程数。因为启动的更多的线程，CPU和内存的消耗都会比并行回放、串行回放要多。
- 极致RTO只关注同步备机的RTO是否满足需求。极致RTO去掉了自带的流控，统一使用recovery_time_target参数来做流控控制。
- 本特性支持备机读，由于增加了对数据页面历史版本的读取，备机上的查询性能会低于主机上的查询性能，低于并行回放备机读的查询性能，但是查询阻塞回放的情况有所缓解。
- DDL日志的回放速度远远慢于页面修改日志的回放，频繁DDL可能导致主备时延增大。
- 当节点的I/O和CPU使用过高时（建议不超过70%），回放和备机读性能会有明显下降。
- meta erp场景：

  硬件规格： Intel(R) Xeon(R) Gold 5220R CPU @ 2.20GHz，754G memory，nvme *2，10GE网卡*2

  业务模型： 使用erp 场景实例表cst_std_item_cost_t定义（表的个数1-20，以性能最优为准），行宽0.7k左右（以性能最优值为准）。使用jmeter等压测工具执行insert语句，单事务行数<4096（以性能最优值为准），并发85左右（以性能最优值为准），ustore表，无DDL。

  日志量：<=300MB/s

  主备时延： <=1s
- 极致RTO备机读在以下几种情况下会取消查询：
  a. 当查询时间超出了参数standby_max_query_time。
  b. 触发了备机读文件的强制回收。
  c. 当查询和回放有锁相关等冲突时，和并行回放备机读相同，取消查询由参数max_standby_streaming_delay控制。
  d. 在开启参数exrto_standby_read_opt的情况下，回放vacuum相关的清理日志时会发生冲突，和并行回放备机读相同，取消查询由参数max_standby_streaming_delay控制。

e. 备机回放段页式物理空间收缩操作相关日志时会取消查询。
f. 开启stream执行计划，查询和relmap类型日志回放有冲突。

依赖关系

无。

2.4.6 基于 Paxos 协议的高可用

可获得性

本特性自V500R002C00版本开始引入。

特性简介

DCF全称是Distributed Consensus Framework，即分布式一致性共识框架。它是一款自研的高性能、高度成熟可靠、易扩展、易使用的独立基础库，其他系统通过API接口可方便地集成DCF组件。DCF基于Paxos协议实现，解决分布式一致性问题，提升集群高可靠高可用能力。

当配置为DCF模式后，DN可以支持基于Paxos协议的复制与仲裁能力。DN基于Paxos的自选主及日志复制，复制过程中支持压缩及流控，防止带宽占用过高。提供基于Paxos多种角色的节点类型，并能够进行调整。支持查询当前数据库实例的状态。

特性描述

- DCF进行日志复制时，支持对日志进行压缩后再传输，减小对网络带宽的占用。
- DCF支持SSL，包括TLS1.2和TLS1.3协议标准。当开启SSL时，DN默认将DCF配置为TLS1.2协议标准。
- DCF支持TLS1.2如下密码套件：TLS_ECDHE-ECDSA-AES256-GCM-SHA384、TLS_ECDHE-ECDSA-AES128-GCM-SHA256、TLS_ECDHE-RSA-AES256-GCM-SHA384、TLS_ECDHE-RSA-AES128-GCM-SHA256。
- DCF支持passive角色节点类型，passive节点不参与选举，只做日志的同步以及回放，该类型节点在高负载的情况下，日志同步会做流控。
- DCF支持logger角色节点，logger节点可以参与选举投票有投票权但无选举权，只复制DCF的日志，不复制xlog，不进行redo。
- DCF的follower和passive角色可以在线互换，即不中断业务的情况下，follower角色的节点转化为passive角色，passive角色的节点转化为follower。
- DCF支持少数派强起能力，在数据库实例多数派故障的情况下，从正常的备DN中选择少数派模式强启成为主DN，其余正常的备DN从主DN复制日志。
- DCF支持自选主能力，在原主DN故障的场景下，在保证数据一致性的前提下，剩余备DN自动选出新的主DN。
- DCF支持策略化多数派能力，以多数派为前提，同时根据用户配置的AZ，保证AZ内至少有一个节点同步复制日志。
- DCF支持手动模式，在手动模式下不自动仲裁，此模式下对接上层CM等管理组件做仲裁适配，DCF进行日志复制功能。
- 支持DCF日志与DN日志合一存储，DCF多数派达成和DN仅存储一份日志，减少IO带宽占用，日志合一后日志刷盘的IO开销比两份日志下降20%+，优化性能。

- 支持从Quorum模式切换到DCF模式，以及从DCF模式恢复到Quorum模式。切换过程中不需要重启数据库，能做到数据不丢失。
- 支持级联备节点部署能力，级联备节点仅从备机同步日志，降低主机日志复制压力，仅支持容灾灾备实例部署级联备。
- 支持1主1备1 logger组网，logger节点辅助仲裁和保留日志、不可当主、无数据库，logger节点可以大大简配CPU和内存、轻量级部署；含logger节点实例环境在业务数据场景下，如果原主故障且候选主机落后于日志节点情况下，由于候选主机需要同步完日志节点数据并回放升主，因此该场景不能保证RTO能力。
- DCF模式1主1备1logger组网下不支持增删副本、连接备机build、强切；强切功能适用于长时间无主的情况，在DCF模式下如果DCF层面出现leader，此时不需要再执行强切，因为主节点已经存在。
- 支持1主2备到1主1备1 logger替换。
- DCF默认工作在最大保护模式，支持通过配置工作在最大可用模式和最大性能模式。
- DCF支持强切功能、备机备份功能、备份恢复功能、自动降副本、增删副本功能。
- DCF支持自动升降副本，当数据库实例故障了半数及以上节点时，为了降低故障对业务的影响，数据库实例管理对可用节点进行降副本操作，当检测到故障恢复后，自动触发升副本操作。
- DCF支持联合仲裁，即自动模式，在一些复杂场景会联合借助CM组件的全局视角实现自仲裁能力增强，解决复杂场景问题，例如半数节点故障自动降副本。支持安装为联合仲裁模式集群，管控需配置如下参数："dnParams": {"dcf_run_mode":0},"cmParams":{ "dn_arbitrate_mode":"paxos" }。

特性增强

无。

特性约束

- 若使用此功能，DN最少三节点，在安装部署阶段需要开启DCF开关。在DCF模式下通过多数派选举，安装过程中如果故障节点数加build节点数达到多数派会导致数据库实例安装失败，如在安装一主两备时，安装过程中一节点因内存不足导致安装失败，另外两节点正常启动，但随后备机会进行一次build，这时build节点加故障节点为2，达到多数派会导致数据库实例安装失败，请在安装过程中检查内存和磁盘等资源是否充足。
- 若某个AZ配置了策略化多数派参数，当AZ内所有的节点均故障时，在对节点做build相关的操作时，需要将该AZ配置从策略化多数派配置信息中移除。
- DCF支持手动模式是针对实例级的工作模式的设置，在此工作模式下不支持节点passive角色，集群安装部署也不支持passive角色。
- 支持DCF日志与DN日志合一存储后，503.0.0之前版本的DCF两份日志模式升级到503.0.0版本的DCF一份日志模式只支持就地升级，不支持灰度和滚动升级。
- 从Quorum模式切换到DCF模式，仅支持固定次数（3次）切换，如超过固定次数需再次切换，则需要重启数据库实例。模式切换过程中，数据库内核涉及到线程关闭和拉起，可能会短暂影响业务（一般小于1分钟），尤其是数据量大的时候。所以建议模式切换过程中尽量少或不要运行业务。
- 包含logger节点的实例典型组网是1主1备1 logger，可扩展场景为1主1备1 logger+级联备（级联备实例个数为1~5），其他组网不承诺支持。

- logger节点仅用来辅助仲裁和保留日志，不要在logger节点上启动roach备份等进程。logger节点不支持查询系统视图与系统表。logger节点部署实例最小支持4核8GB内存。

依赖关系

无。

2.4.7 两地三中心跨 Region 容灾

可获得性

本特性自V500R001C20版本开始引入。

特性简介

支持两地三中心跨Region容灾。

客户价值

金融、银行等业务需要底层数据库提供跨地域的容灾能力，来保证极端灾难情况下数据的安全和可用性。

特性描述

金融、银行业对数据的安全有着较高的要求，当发生火灾，地震，战争等极端灾难情况下，需要保证数据的安全性，因此需要采取跨地域的容灾的方案。跨地域容灾通常是指主备数据中心距离在200KM以上的情况，主机房在发生以上极端灾难的情况下，备机房的数据还具备能继续提供服务的能力。本特性的目的是提供一套支持gaussdb跨地域容灾的解决方案。

特性增强

V500R002C00版本为解除两地三中心跨Region容灾当前存在的关键约束，主要包括灾备数据库实例无运维能力、主备数据库实例不支持来回切等，新增如下特性：

- 支持灾备数据库实例节点替换、节点修复。
- 支持容灾主备数据库实例计划内switchover。
- 支持投票副本。
- 支持容灾数据库实例副本增加。

V500R002C00版本针对两地三中心跨Region容灾特性新增基于流式复制的异地容灾解决方案：

- 支持灾备数据库实例节点修复。
- 支持灾备数据库failover。
- 支持容灾主备数据库实例计划内switchover。

503.0.0版本新增功能：

- 支持容灾主实例日志保持。

- 支持容灾加回。
- 支持容灾演练增强。
- 支持结束容灾搭建等待状态的清理功能。
- 支持灾备实例升主实例后手动升副本，恢复为灾备实例后手动降副本。
- 支持容灾过程中修改容灾用户信息。
- 异地容灾解决方案支持Paxos协议。

特性约束

基于流式复制的异地容灾解决方案：

- 灾备数据库实例可读不可写。
- 灾备数据库实例通过failover命令升主后，和原主数据库实例灾备关系将失效，需要重新搭建容灾关系。
- 主数据库实例和备数据库实例应该具有相同的管理员用户名。
- 主数据库实例和备数据库实例应该具有相同的数据库实例用户名。
- 数据库实例状态对容灾操作的影响：
  - 在主数据库实例和灾备数据库实例处于normal状态且所有组件(datanode、etcd、cm_agent、cm_server)状态正常时可进行容灾搭建；在主数据库实例处于normal态所有组件状态正常并且灾备数据库实例已经升主的情况下，主数据库实例可执行容灾解除，其他数据库实例状态不支持。
  - 在主数据库实例和灾备数据库实例处于normal且所有组件(datanode、etcd、cm_agent、cm_server)状态正常状态时，通过计划内switchover命令，主数据库实例可切换为灾备数据库实例，灾备数据库实例可切换为主数据库实例。
  - 灾备数据库实例处于非Normal且非Degraded状态时，无法升主，无法作为灾备数据库实例继续提供容灾服务，需要修复或重建灾备数据库实例。
  - 主数据库实例存在多数派实例故障且没有打开最大可用模式时（参数most_available_sync），不会向灾备数据库实例进行日志发送，需要及时修复主数据库故障实例。
- 灾备数据库实例节点替换和节点修复的约束，继承节点替换和修复的约束。
- 容灾实例主数据库实例仅支持增删级联备副本，增删副本后灾备数据库实例需要刷新容灾信息。基于paxos协议实现不支持主实例增删级联备副本。
- 不支持基于quorum协议实现高可用的数据库实例与基于paxos协议实现高可用的数据库实例搭建容灾关系，容灾过程中的主备数据库实例不支持基于quorum协议实现高可用与基于paxos协议实现高可用的切换。
- 容灾状态中灾备数据库实例支持降副本，灾备数据库实例升主并且删除容灾信息后正常支持升降副本。
- 当灾备数据库实例为2副本时，需要确保打开最大可用模式(参数most_available_sync)。灾备数据库实例在1个副本损坏时，仍可以升主对外提供服务，如果剩余的这个副本也损坏，将导致不可避免的数据丢失。
- 灾备数据库实例支持1副本部署，1副本单数据库实例当前版本仅支持安装、升级、参数设置、数据库实例启停、数据库实例状态查询、备份恢复，不支持扩容、节点修复、节点替换、修改端口、升降副本等SLA功能，不支持单副本当做主数据库实例搭建容灾，不支持非OBS/NAS介质下的单副本备份恢复到多副本。

- 主数据库实例如果进行了强切操作（cm_ctl finishredo命令，请参见《工具参考》中“统一数据库管理工具 > cm_ctl工具介绍”章节），需要重建灾备数据库实例。
- 主数据库实例如果进行了少数派AZ强启（请参见《OM服务化》中“OM Agent接口 > 数据库实例管理相关接口 > 单AZ一键强启”章节），会出现数据丢失，需要重建灾备数据库实例。
- 灾备数据库实例不支持全备和增备，主数据库实例支持全备和增备。灾备数据库实例会关闭DN实例上的enable_cbm_tracking参数。如果主数据库实例要做恢复，需要先解除容灾关系，在完成备份恢复后重新搭建容灾关系。
- 容灾关系搭建之后，灾备数据库实例不支持逻辑复制，主数据库实例支持逻辑复制。
- 建立容灾关系的主数据库实例与灾备数据库实例之间不支持GUC参数的同步。
- 容灾关系搭建之后，不支持DN实例端口修改。
- 容灾关系搭建后，支持在主数据库实例添加级联备作为只读节点，只读节点因硬件规格较低，未开放升备升主能力；容灾倒换后该只读节点也不会升为灾备数据库实例的首备。因此，添加只读节点，对端集群可不刷新容灾关系。
- 容灾搭建时需要在主数据库实例和灾备数据库实例下发容灾用户名和密码用于数据库实例间鉴权：
  - 主备数据库实例必须使用相同的容灾用户名和密码。
  - 不得使用已存在的数据库用户进行搭建。
- 搭建容灾的主备数据库实例版本号必须相同。
- 容灾状态下主备数据库实例升级：
  - 主备数据库实例都要处于normal状态。
  - 主备数据库实例升级时大版本升级会校验主备数据库实例版本号是否相同，不相同不可升级。小版本升级不校验。
  - 不支持就地升级。
- 在带有写业务的场景下：使用sysbench进行测试，主数据库实例执行50并发的update类型业务，主数据库实例备机和容灾数据库实例同时执行200并发的读类型业务，在I/O和CPU不受限的条件下，串行回放容灾读的性能不低于主数据库实例备机读性能的80%，极致RTO容灾读的性能和串行回放容灾读的性能相比劣化不超过10%。

依赖关系

无。

2.4.8 按分片自动升降副本

可获得性

本特性自V500R001C20版本开始引入。

特性简介

两AZ+仲裁AZ数据库实例部署方式支持自动升降副本功能。

客户价值

金融、银行等业务需要提供半数及以上节点数据故障业务快速恢复能力。

特性描述

金融、银行业务需要极高的容灾能力。当一个分片故障了半数及以上节点时，DN执行写操作会超时，主要是同步备中有节点故障了，无法执行写操作。为了降低分片故障对业务的影响，需要对分片上可用节点进行降副本操作，当检测到故障恢复后，自动触发升副本操作。

特性增强

无。

特性约束

- 基础保障(最多降至)：一主一备。
- 实例部署要求：两AZ+仲裁AZ实例，副本数大于3，分片总数（DN+CN+GTM）小于 64。
- 前提要求：DN主存在。
- 升级、扩容阶段，或者ETCD不可用时不会进行降副本操作。
- 只有半数以上DN发生故障（DN状态是down），且状态持续，才会进行降副本操作，升副本需要等待半数故障恢复后，且状态持续，才会自动升副本。
- 只有当上一轮降副本操作执行成功后，才能进行下一轮降副本操作，不支持二次故障。
- 不支持故障跳转，比如四个副本，第一次（3, 4）故障后，（1, 2）进行完降副本，第二次故障（1, 2），恢复（3, 4），此时实例不可用，无法选出主，且不能对（3, 4）进行降副本。
- 在升降副本结束后，才能执行switchover，且switchover 只能切换到同步列表中的备DN上。
- 升副本要求：故障节点恢复后，需要跟主机同步达到99%，才会被重新加入到主机的同步列表中。

依赖关系

无。

2.4.9 支持 global syscache

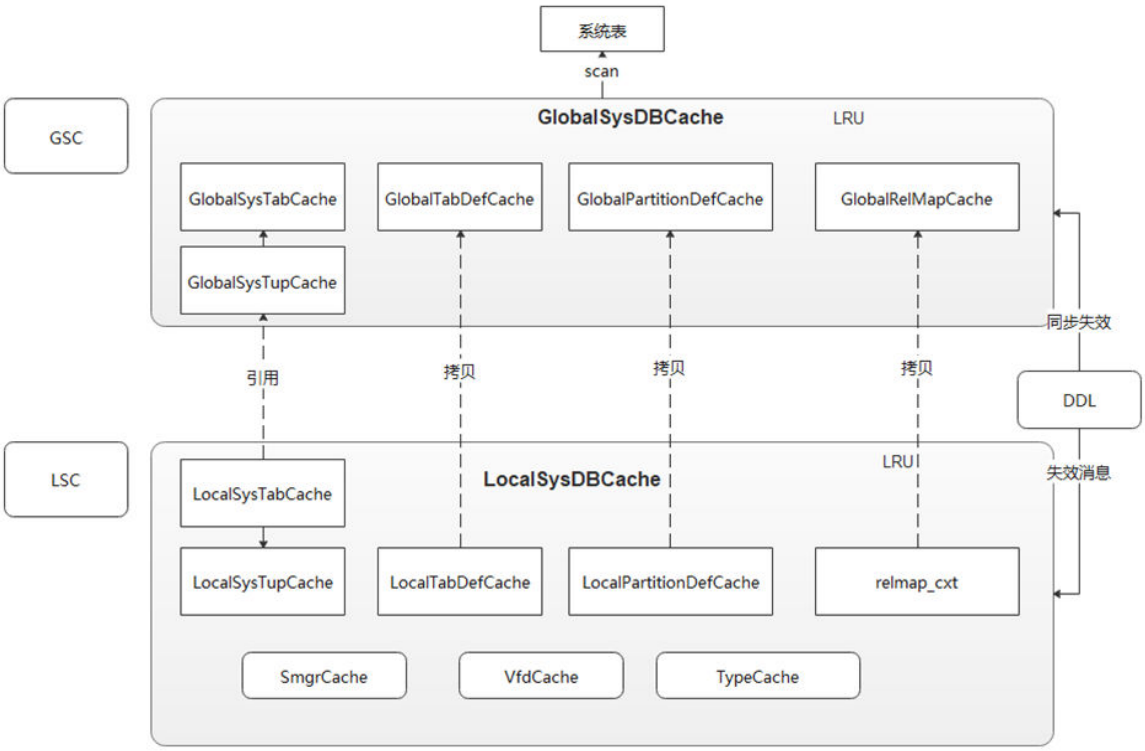
可获得性

本特性自V500R002C10版本开始引入。

特性简介

全局系统缓存（Global SysCache）是系统表数据的全局缓存和本地缓存。原理如图2-1所示。

图 2-1 Global SysCache 原理图



客户价值

全局系统缓存特性可以降低数据库进程的缓存内存占用，提升数据库的并发扩展能力。

特性描述

全局系统缓存特性指将系统缓存与会话解耦，绑定到线程上，结合线程池特性达到降低内存占用的目的，同时结合全局缓存，提升缓存命中率，保持性能稳定。

特性增强

支持更高的并发查询。

特性约束

- 设置enable_global_syscache为on。建议设置enable_thread_pool参数为on。
- 当DB数较多，且阈值global_syscache_threshold较小时，内存控制无法正常工作，性能会劣化。
- 不支持分布式时序相关的任务，这些任务的内存控制与性能不受GSC特性的影响。
- wal_level设置为minimal或者archive时，备机的查询性能会下降，会退化为短连接。

依赖关系

该特性降内存能力依赖于线程池特性。

2.4.10 并行逻辑解码

可获得性

本特性自V500R002C10版本开始引入。

特性简介

支持多线程并行解码。

客户价值

大幅提升逻辑解码性能，解码速度由3~5MBps可提升到标准场景（16核CPU、内存128G、网络带宽 > 200MBps、表的列数为10~100、单行数据量0.1KB~1KB、DML操作以insert为主、不涉及落盘事务即单个事务中语句数量小于4096）下的100MBps。

特性描述

在使用JDBC或pg_recvlogical解码时，设置配置选项parallel-decode-num为大于1且小于等于20的值，开启并行解码特性，使用一个读取线程、多个解码线程以及一个发送线程协同进行逻辑解码操作，显著提升解码速度。

特性增强

无。

特性约束

1. 当前的硬件和网络环境正常；由于逻辑日志一般为xLog的两倍，为保证xLog速度达到100MBps，I/O带宽至少保证200MBps；因为reader、decoder、sender线程均需预留资源，CPU需预留并发数+2的核数，如4并发场景需要预留6核。在实际场景中，使用备机解码即可保证需求，无需进行特殊的资源预留规划。为保证解码性能达标以及尽量降低对业务的影响，一台备机上应尽量仅建立一个并行解码连接，保证CPU、内存、带宽资源充足。
2. 日志级别的guc参数wal_level = logical。
3. guc参数max_replication_slots >= 每个DN所需的（物理流复制槽数+备份槽数+逻辑复制槽数）。
4. 解码配置选项parallel-decode-num > 1且<= 20，指定并行的解码线程数。
5. 逻辑解码支持DDL约束，参见《特性指南》中“逻辑复制 > 逻辑解码 > 逻辑解码支持DDL > 规格约束”。
6. 不支持数据页复制的DML解码。
7. 单条元组大小不超过1GB，考虑解码结果可能大于插入数据，因此建议单条元组大小不超过500MB。
8. 不支持压缩表的DML语句解码。
9. GaussDB支持解码的数据类型为：INTEGER、BIGINT、SMALLILNT、TINYINT、SERIAL、SMALLSERIAL、BIGSERIAL、FLOAT、DOUBLE PRECISION、DATE、

TIME[WITHOUT TIME ZONE]、TIMESTAMP[WITHOUT TIME ZONE]、CHAR(n)、VARCHAR(n)、TEXT。

10. 在需要ssl连接的场景，需要前置条件保证guc参数ssl = on。
11. 不支持interval partition表DML复制。
12. 在事务中执行DDL语句时，约束请参见《特性指南》中“逻辑复制 > 逻辑解码 > 逻辑解码支持DDL > 规格约束”。
13. 如需进行备机解码，需在对应主机上设置guc参数enable_slot_log = on。
14. 当前不支持超大CLOB解码。
15. 不允许主备，多个备机同时使用同一个复制槽解码，否则会产生数据不一致。
16. 禁止在使用逻辑复制槽时在其他节点对该复制槽进行操作，删除复制槽的操作需在该复制槽停止解码后执行。

依赖关系

依赖备机解码。

2.4.11 支持备机 build 备机

可获得性

本特性自V500R002C10版本开始引入。

特性简介

备机build备机加快备机故障的恢复， 减小主机I/O和带宽压力。

客户价值

当业务压力过大时，从主机build备机会对主机的资源造成影响，导致主机性能下降、build变慢的情况。使用备机build备机不会对主机业务造成影响。

特性描述

使用gs_ctl命令可以指定对应的备机去build需要修复的备机。具体操作可参考《工具参考》中的“系统内部调用的工具 > gs_ctl”章节。

特性增强

无。

特性约束

只支持备机build备机，只能使用指定ip和port的方式做build，同时在build前应确保需要修复备机的日志比发送数据的备机的日志落后。

依赖关系

无。

2.4.12 3AZ 多数派故障一键式强启及加回

可获得性

本特性自V500R001C20版本开始引入。

特性简介

AZ多数派故障场景下，一键式执行少数派强起命令，实现业务正常运行的目的。

客户价值

AZ多数派故障场景下，能够通过强起操作继续提供实例仲裁服务，使业务运行尽快恢复。

特性描述

部署同城双活实例（2AZ+仲裁AZ）时，在园区A部署一个生产中心AZ1，在同城园区B部署一个灾备中心AZ2。AZ1、AZ2都有完整的数据并且均部署第三方仲裁实例ETCD，ETCD数量相等， AZ3中只部署一个ETCD实例。ETCD多数派存活时，可正常工作，实例CMS实例可以正常选主。业务正常执行。当AZ2、AZ3发生故障时，AZ1中存活的ETCD数量少于总数的二分之一，即少数派存活，此时无法实现AZ的自动切换，因此要进行少数派强起。

特性增强

无。

特性约束

- 少数派强启命令属于高危操作，必须是满足多数派所有节点同时故障的情况下才能进行强启操作，即：少数派强启之前，一定要确认好强启AZ是否一直是实例内唯一一个没有故障的AZ，而不是先故障的AZ。

  说明

  场景举例：

  实例AZ1与AZ2、AZ3发生了网络隔离，AZ2,AZ3满足多数派，可以继续执行业务，AZ1由于网络隔离，数据都无法从多数派（AZ2、AZ3）同步。

  此后如果AZ1、AZ2又都发生了故障，不再满足多数派，这个时候，禁止少数派强启AZ1的原因：

  AZ1由于很早之前就故障了，数据不是最新的，少数派强启虽然会执行成功，但会发生数据丢失。
- 多数派故障后处理流程和多数派恢复后恢复流程处理过程中，需要将业务停止。
- 一键式强启执行之前，请务必认真确认当前是否确实是多数派故障场景，如果不满足多数派故障条件，强启后可能会产生数据不一致的问题。
- 一键式脚本限制执行命令节点所在AZ 到其他AZ 均无法正常连接时才会继续执行（防止用户对多数派故障的错误判断）。
- 少数派强启需当前强启AZ内的节点不再有新的故障，否则在多数派故障的情况下再次叠加故障，可能会导致脚本无法正常执行。
- 不支持在logger节点执行一键式强启加回。

- 如果强启AZ中的DN数量小于半数，则强启无法保证RPO。
- 执行加回之前，如果实例current_az 不是 AZ_ALL，需要等到其为AZ_ALL的时候才能操作。
- 强启运行在降级模式中，不支持扩容、升级、节点修复等工程能力。
- DCF自仲裁模式不支持一键式强启加回。

依赖关系

无。

2.4.13 浮动 IP 安装/部署升级

可获得性

本特性自503.1.0版本开始引入。

特性简介

浮动IP为业界比较通用的主备数据库连接方案，GaussDB通过VIP（Virtual IP）来实现该功能。

CM在集中式一主多备部署场景下，提供对VIP进行监控与仲裁，当检测到主DN变化时，对老VIP卸载，在新的主DN上进行绑定。

客户价值

- 通过VIP可以直接找到主机，连接重连更准更快（毫秒级别）。
- 当出现双主时，依然可以通过VIP访问到唯一一个主机，降低了双主丢数据的风险。

特性描述

CM在集中式一主多备部署场景下，提供对VIP进行监控与仲裁，当检测到主DN变化时，对老VIP卸载，在新的主DN上进行绑定。

特性原理

- CMA对DN状态角色和DN上挂载的VIP状态进行监控，由CMA上报给CMS进行仲裁，执行绑定VIP操作。
- 提供VIP状态查询命令，展示VIP绑定状态。

特性增强

无。

特性约束

- 不支持单副本且不支持容灾。
- 需用户配置ifconfig sudo权限。
- 需合理规划float IP，避免float IP和部署环境的网段冲突，并保证其对外可用。

依赖关系

无。

2.4.14 一主一备一 logger+级联备

可获得性

本特性自505.0.0版本开始引入。

特性简介

支持paxos协议下的一主一备一logger+级联备部署。

客户价值

降本增效，使用更少的硬件资源来完成业务目标。

特性描述

当前版本支持一主一备一logger和一主两备+级联备的部署形态，两种部署方式结合，保障可靠性的同时兼顾性能要求，还可以有效降低成本。

特性增强

无。

特性约束

- 只读备不支持直接升主。
- 不支持多备DN（一主两备一logger一级联备）、无备DN（一主一logger一级联备）、多logger实例的部署形态（一主一备两logger一级联备）。
- 不支持将级联备替换为备机或者logger。
- 不支持Dorado部署。
- 不支持Dcf切换为Quorum。

依赖关系

无。

2.4.15 计划内应用无损透明

可获得性

本特性自505.1.0版本开始引入。

特性简介

通过部署GNS（GaussDB Notification Service）组件，支持计划内应用无损透明功能。

客户价值

- 提供了一种数据库状态变化的主动消息通知机制，可以方便地在业务侧实现数据库状态变化的处理逻辑。
- 在数据库运维平台界面上进行DN主备倒换、重启DN、重启节点、重启实例操作场景下，应用层不需要显式执行连接重连和事务重试。

特性描述

应用无损透明（ALT，Application Lossless and Transparent）提供了一种数据库状态变化的主动消息通知机制。JDBC驱动也向应用程序提供了数据库实例状态变化的回调函数注册接口。应用程序可以针对某些数据库连接、向JDBC驱动注册状态变化的回调函数。当数据库实例状态发生变化时，JDBC驱动会对注册的函数进行调用，通过注册回调函数可以很方便地在业务侧实现数据库状态变化的邮件通知、告警平台上报等运维管理操作。

针对使用JDBC驱动连接GaussDB数据库的应用程序，当数据库进行计划内维护时（即：在数据库运维平台界面上进行DN主备倒换、重启DN、重启节点、重启实例操作），可以选择是否支持“计划内ALT”功能，如果选择支持，数据库的计划内维护操作不会导致业务的中断（表现为业务卡顿一段时间），如果选择不支持，则和原功能保持一致（业务收到连接报错、事务执行失败）。

特性增强

无

特性约束

1. 数据库切换或重启等操作会导致业务一段时间内不可用，业务的等待超时时间要设置大于数据库的切换或重启时间，即在数据库切换或重启等操作期间，业务不能因为超时设置问题导致主动断链。
2. 仅适用于短事务业务场景（事务执行时间不能超过计划内ALT超时时间，建议：单个事务执行总时长为秒级），例如tpcc场景。如果存在长事务，当长事务在配置的计划内ALT超时时间内没有执行结束，则会执行报错，连接断开。
3. 仅支持使用JDBC连接主节点的情况下计划内ALT生效，不支持备机连接。
4. 支持事务内游标，不支持跨事务的游标。
5. 不支持临时表，临时函数恢复。
6. 不支持SET SESSION 【 ROLE | AUTHORIZATION 】语句。
7. 不支持ALTER 【 USER | ROLE | DATABASE 】【 SET | RESET 】语句。
8. 不支持Session级的咨询锁，支持事务级咨询锁。
9. 容灾切换不支持计划内ALT，切换后新主数据库实例正常支持计划内ALT。
10. 升级过程中不支持计划内ALT。
11. GNS组件异常不影响数据库实例升级，若升级后出现GNS实例异常，请参考《故障处理》进行修复。

须知

使用ALT特性前一定要确认不含有上述4~8约束问题，否则可能引起严重问题。

依赖关系

无

2.4.16 ETCD 多数派故障一键修复

可获得性

本特性自505.1.0版本开始引入。

特性简介

ETCD实例多数派故障，并且不可恢复场景下，为了恢复集群健康可用，可调用ETCD多数派故障一键修复接口，完成集群的修复。

客户价值

ETCD实例多数派故障，并且不可恢复场景下，能够通过一键修复操作继续提供实例仲裁服务，使业务运行尽快恢复。

特性描述

部署多节点实例。ETCD多数派存活时，可正常工作，业务正常执行。当ETCD少数派存活，此时无法使用gs_replace修复ETCD，因此要进行ETCD一键修复。

特性增强

无

特性约束

- 需要确认ETCD已无法自恢复，否则会导致ETCD内数据丢失。
- ETCD少数派故障不能使用一键修复进行ETCD的修复。
- DCC模式不支持ETCD多数派故障一键修复。
- 集中式单节点不支持ETCD多数派故障一键修复。

依赖关系

无

2.5 可维护性

2.5.1 热补丁升级

可获得性

本特性自V500R001C20版本开始引入。

特性简介

热补丁将补丁以patch文件的形式加载到正在运行的数据库进程中，达到零中断修复线上系统的目的。

客户价值

热补丁最大的优势是业务零中断加载补丁，他可以在不影响业务的前提下在线解决一部分数据库内核的紧急问题。

其价值主要体现在如下两点：

- 缩短版本发布时间，紧急问题从版本回归验证轻量化为补丁回归验证，提高了线上紧急问题的响应速度。
- 热补丁的加载，卸载对业务无感知，提高了客户满意度。

特性描述

热补丁基于发布的代码版本生成补丁文件，然后以模块的形式插入到数据库内核运行地址空间中，通过寻找热补丁目标函数的地址，并动态地，原子地替换入口地址，重定向函数代码段至补丁文件代码段达到修复线上系统缺陷的目的。

- 热补丁的制作通过修复特定缺陷函数，制作成模块，动态地加载到运行中的内核系统。
- 热补丁找到目标函数，并在目标函数的入口处加入跳转指令，当目标函数被调用时，跳转到补丁区执行补丁函数。
- 目标函数的替换和还原是原子操作CPU寄存器，热补丁可以随时随地加载和卸载，线上系统无需中断，即随时可运行最新的代码。

特性增强

支持ARM CPU。

特性约束

无。

依赖关系

无。

2.5.2 灰度升级

可获得性

本特性自V500R002C00版本开始引入。

特性简介

支持按照用户定义的升级顺序，进行节点级滚动灰度升级。

客户价值

通过灰度升级，可以达成以下目的：

- 先升级部分备DN节点，即使升级失败，也不会对业务产生影响。
- 先升级业务影响小的组件节点（如ETCD、CMS），即使升级失败，也能将对业务的影响控制在最小范围。
- 每批节点升级完之后，均提供升级观察窗口，验证升级状态，动态评估升级的风险。

特性描述

灰度升级是一种支持优先升级部分节点的在线升级方式。灰度升级主要包含以下三个方面：

1. 对于大版本升级涉及的系统表变更，将不同版本的系统表结构和系统函数固化在二进制中，保证新、老版本二进制均能解析和使用新、老版本的系统表元组。
2. 对于大版本升级和二进制升级涉及的新、老二进制替换，先灰度替换指定节点上的二进制，待系统运行一定时间之后，再替换剩余节点的二进制
3. 在第2点的基础之上，如果升级亦涉及到节点的操作系统、硬件升级（且不能提前执行），那么在灰度升级部分节点之前，先将这些节点上的主实例全部切换到非灰度升级的节点上；如果升级只涉及数据库二进制的替换，为了尽可能降低对于业务的影响，采用同一节点两套二进制同时存在的方式，使用软连接切换的方式来进行进程版本的切换升级（闪断一次，10秒以内）

特性增强

无。

特性约束

灰度升级的约束条件请参见《升级指导书》中“升级影响和升级约束”章节。

依赖关系

无。

2.5.3 就地升级

可获得性

本特性自V500R001C20版本开始引入。

特性简介

就地升级是一种离线升级方式。

客户价值

- 支持数据库的大版本升级和小版本升级（内核版本号不变的升级方式为小版本升级，否则就是大版本升级）。

- 提供一种相对稳定可靠的升级方式。

特性描述

就地升级过程中需要停止业务，不提供任何服务，会一次性升级数据库实例中的所有节点。

特性增强

无。

特性约束

就地升级的约束条件请参见《升级指导书》中“升级影响和升级约束”章节。

依赖关系

无。

2.5.4 支持 WDR 诊断报告

可获得性

本特性自V500R001C20版本开始引入。

特性简介

WDR报告提供数据库性能诊断报告，该报告基于基线性能数据和增量数据两个版本，从性能变化得到性能报告。

客户价值

- WDR报表是长期性能问题最主要的诊断手段。基于SNAPSHOT的性能基线，从多维度做性能分析，能帮助DBA掌握系统负载繁忙程度、各个组件的性能表现及性能瓶颈。
- SNAPSHOT也是后续性能问题自诊断和自优化建议的重要数据来源。

特性描述

WDR(Workload Diagnosis Report)基于两次不同时间点系统的性能快照数据，生成这两个时间点之间的性能表现报表，用于诊断数据库内核的性能故障。

使用generate_wdr_report(...)可以生成基于两个性能快照的性能报告。

WDR性能快照数据存储在postgres库的snapshot schema下，默认的采集和保存策略为：

- 每小时采集一个快照(wdr_snapshot_interval=1h)。
- 每十二个快照中有一个全量快照（wdr_snapshot_full_backup_interval=12）。
- 保留8天(wdr_snapshot_retention_days=8)。
- 不启用空间维度控制阈值（wdr_snapshot_space_threshold=0）。

WDR主要依赖两个组件：

- SNAPSHOT性能快照：性能快照可以配置成按一定时间间隔从内核采集一定量的性能数据，持久化在用户表空间。任何一个SNAPSHOT可以作为一个性能基线，其他SNAPSHOT与之比较的结果，可以分析出与基线的性能表现。
- WDR Reporter：报表生成工具基于两个SNAPSHOT，分析系统总体性能表现，并能计算出更多项具体的性能指标在这两个时间段之间的变化量，生成SUMMARY和DETAIL两个不同级别的性能数据。如**表2-1**、**表2-2**所示。

表 2-1 SUMMARY 级别诊断报告

诊断类别	描述
Database Stat	主要用于评估当前数据库上的负载，I/O状况，负载和I/O是衡量TP系统最重要的特性。 包含当前连接到该数据库的session，提交、回滚的事务数，读取的磁盘块的数量，高速缓存中已经发现的磁盘块的次数，通过数据库查询返回、抓取、插入、更新、删除的行数，冲突、死锁发生的次数，临时文件的使用量，I/O读写时间等。
Load Profile	从时间，I/O，事务，SQL几个维度评估当前系统负载的表现。 包含作业运行elapse time、CPU time，事务日志量，逻辑和物理读的量，读写I/O次数、大小，登录登出次数，SQL、事务执行量，SQL P80、P95响应时间等。
Instance Efficiency Percentages	用于评估当前系统的缓存的效率。 主要包含数据库缓存命中率。
Events	用于评估当前系统内核关键资源，关键事件的性能。 主要包含数据库内核关键事件的发生次数，事件的等待时间。
Wait Classes	用于评估当前系统关键事件类型的性能。 主要包含数据库内核在主要的等待事件的种类上的发布：STATUS、LWLOCK_EVENT、LOCK_EVENT、IO_EVENT。
CPU	主要包含CPU在用户态、内核态、Wait IO、空闲状态下的时间发布。
IO Profile	主要包含数据库Database I/O次数、Database I/O数据量、Redo I/O次数、Redo I/O量。
Memory Statistics	包含最大进程内存、进程已经使用内存、最大共享内存、已经使用共享内存大小等。

表 2-2 DETAIL 级别诊断报告

诊断类别	描述
Time Model	主要用于评估当前系统在时间维度的性能表现。 包含系统在各个阶段上消耗的时间：内核时间、CPU时间、执行时间、解析时间、编译时间、查询重写时间、计划生成时间、网络时间、I/O时间。

诊断类别	描述
SQL Statistics	主要用于SQL语句性能问题的诊断。 包含归一化的SQL的性能指标在多个维度上的排序：Elapsed Time、CPU Time、Rows Returned、Tuples Reads、Executions、Physical Reads、Logical Reads。这些指标的种类包括：执行时间，执行次数、行活动、Cache IO等。
Wait Events	主要用于系统关键资源，关键事件的详细性能诊断。 包含所有关键事件在一段时间内的表现，主要是事件发生的次数，消耗的时间。
Cache IO Stats	用于诊断用户表和索引的性能。 包含所有用户表、索引上的文件读写，缓存命中。
Utility status	用于诊断后台任务性能，包含复制等后台任务的性能。
Object stats	用于诊断数据库对象的性能。 包含用户表、索引上的表、索引扫描活动，insert、update、delete活动，有效行数量，表维护操作的状态等。
Configuration settings	用于判断配置是否有变更。 包含当前所有配置参数的快照。
SQL detail	显示unique query text信息。

特性增强

无。

特性约束

- WDR snapshot性能快照会采集不同database的性能数据，如果数据库实例中有大量的database或者大量表，做一次WDR snapshot会花费很长时间。
- 如果在大量DDL期间做WDR snapshot可能造成WDR snapshot失败。
- 在drop database时，做WDR snapshot可能造成WDR snapshot失败。
- 如果生成WDR报告的两次快照期间进行过降副本、节点重启和主备切换等操作，则无法生成WDR报告。
- 数据库实例只读状态会造成WDR snapshot失败。

依赖关系

无。

2.5.5 支持 ASP 报告

可获得性

本特性自503.2.0版本开始引入。

特性简介

ASP报告提供会话级别的报告，报告基于活跃会话采样的样本进行分析，从SQL、Session、Wait event维度分析得到ASP报告。

客户价值

- ASP报告是针对短暂的性能抖动问题主要的诊断手段，基于活跃会话采样的样本，从SQL、Session、Wait event维度分析，能帮助DBA掌握系统负载变化情况以及性能瓶颈。
- ASP活跃样本数据也是后续性能问题自诊断和自优化建议的重要数据来源。

特性描述

ASP报告（Active Session Profile）基于活跃会话采样的样本进行分析，用于诊断数据库内核的短暂性能故障。

使用generate_asp_report(...) 可以生成基于活跃会话样本性能报告。

后台线程每秒会采集数据库中的所有活跃会话信息，并保存在内存（asp_sample_num控制内存中最大保留样本数）中，在内存达到上限后，ASP样本数据会被存储在postgres库的gs_asp表中，在从内存转储到表中时可以通过asp_flush_rate控制落盘时的二次采样率，默认的采集和保存策略为：

- 内存保留的最大样本数（asp_sample_num=100000）。
- 样本从内存中刷到磁盘上采样的比例默认是10：1（asp_flush_rate=10）。

特性增强

无。

特性约束

- ASP是按照1s进行采样，内存达到阈值后并按照1：10采样的比例持久化到表中，所以对于表的数据的最小采样粒度为10s，对于某些执行较快的SQL，ASP可能采不到，则ASP报告中不会展示。
- ASP中的SQL text信息来自于unique SQL，如果instr_unique_sql_count太小导致unique hash被占满时，新的语句就不会保存sql text，会导致ASP报告中部分SQL的text为空。
- 如果生成ASP报告的时间段过大、时间段内并发数过多、或者slot数过大，可能都会导致数据量过大，查询速度变慢，ASP报告生成较慢。
- 报告中显示slot的时间段的粒度默认只能到gs_asp表的10s的粒度，所以slot的个数可能比传入参数slots少。
- 目前报告中展示的数据都是去掉部分后台线程后分析的结果，屏蔽的后台线程有undo recycler，workload，Asp，PercentileJob，JobScheduler，Wal Writer，CheckPointer，WDR相关线程。
- ASP报告需要引入开源软件ECHARTS，该软件支持Chrome、edge浏览器，不支持IE浏览器。
- ASP报告生成不在一个事务内，如果传入的start time比ASP样本的最老时间小，生成报告过程中ASP旧的数据可能被回收，则整个ASP报告的sample count可能不一致。

- ASP报告不支持备机。

依赖关系

无。

2.5.6 慢 SQL 诊断

可获得性

本特性自V500R001C20版本开始引入。重构前慢SQL相关视图已废弃，包括：dbe_perf. gs_slow_query_info、dbe_perf.gs_slow_query_history、dbe_perf.global_slow_query_hisotry、dbe_perf.global_slow_query_info。

特性简介

慢SQL诊断提供诊断慢SQL所需要的必要信息，帮助开发者回溯执行时间超过阈值的SQL，诊断SQL性能瓶颈。

客户价值

慢SQL提供给用户对于慢SQL诊断所需的详细信息，用户无需通过复现就能离线诊断特定慢SQL的性能问题。表和函数接口方便用户统计慢SQL指标，对接第三方平台。

特性描述

慢SQL能根据用户提供的执行时间阈值(log_min_duration_statement)，记录所有超过阈值的执行完毕的作业信息。

慢SQL提供表和函数两种维度的查询接口，用户从接口中能查询到作业的执行计划、开始、结束执行时间、执行查询的语句、行活动、内核时间、CPU时间、执行时间、解析时间、编译时间、查询重写时间、计划生成时间、网络时间、IO时间、网络开销、锁开销等。所有信息都是脱敏的。

特性增强

支持对慢SQL指标信息，安全性（脱敏）、执行计划、查询接口的增强。

特性约束

1. 目前的SQL跟踪信息，基于正常的执行逻辑。执行失败的SQL，其跟踪信息不具有准确的参考价值，例如：状态为cancelled等。
2. 节点重启，可能导致该节点的数据丢失。
3. 通过GUC参数设置收集SQL语句的数量，如果超过阈值，新的SQL语句执行信息不会被收集。
4. 通过GUC参数设置单条SQL语句收集的锁事件详细信息的最大字节数，如果超过阈值，新的锁事件详细信息不会被收集。
5. 当track_stmt_parameter为off时，query字段最大长度受track_activity_query_size控制。
6. 通过异步刷新方式刷新用户执行中的SQL信息，所以用户Query执行结束后，存在查询相关视图函数结果短暂时延。

7. 部分指标信息(行活动、Cache/IO、时间分布等)依赖于dbe_perf.statement视图收集，如果该视图对应记录数超过预定大小(依赖GUC:instr_unique_sql_count)，则本特性可能不收集相关指标。
8. statement_history表相关函数以及视图、备机dbe_perf.standby_statement_history函数中的details字段为二进制格式，如果需要解析详细内容，请使用对应函数：pg_catalog.statement_detail_decode(details, 'plaintext', true)。
9. statement_history表查询需要切换至postgres库，其它库中数据为空。
10. 备机dbe_perf.standby_statement_history函数查询需要切换至postgres库，其它库中查询会提示不可用。
11. statement_history表以及备机dbe_perf.standby_statement_history函数的内容受track_stmt_stat_level控制，默认为'OFF,L0'，参数第一部分代表Full SQL，第二部分是慢SQL；对于慢SQL,只有SQL运行时间超过log_min_duration_statement时才会被记录至statement_history表。
12. 当track_stmt_stat_level关闭Full SQL时，SQL等锁超时可能会导致表中的query_plan信息为空，可通过detail字段内的wait event辅助定位分析。
13. 当track_stmt_flush_mode参数取值为"MEMORY,FILE"时，开启内核支持全量SQL功能，Full SQL语句存储到内存中，Slow SQL语句存储到磁盘文件中。开启功能后性能劣化不超过5%。由于全量SQL默认关闭，开启后仅占用固定内存大小，故当前版本全量SQL共享内存暂不受max_process_memory控制。在升级未提交期间，若版本过低，开启全量SQL功能将失效，并记录提示信息至日志。
14. 全量SQl采用共享内存方式存储Full/Slow SQL，gs_shared_mem_kpi.meta文件存储了共享内存访问信息，可通过System V共享内存的接口shmat访问，当前仅用于对接管控SQL全链路相关功能。

依赖关系

无。

2.5.7 Session 性能诊断

可获得性

本特性自V500R001C20版本开始引入。

特性简介

Session性能诊断提供给用户Session级别的性能问题诊断。

客户价值

- 查看最近用户Session最耗资源的事件。
- 查看最近比较占资源的SQL把资源都消耗在哪些等待事件上。
- 查看最近比较耗资源的Session把资源都花费在哪些等待事件上。
- 查看最近最耗资源的用户的信息。
- 查看过去Session相互阻塞的等待关系。

特性描述

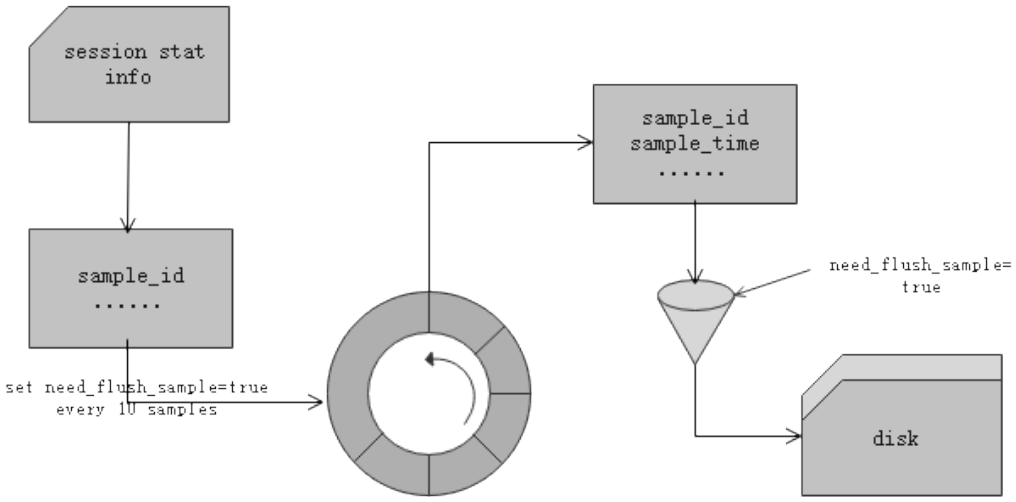
Session性能诊断提供对当前系统所有活跃Session进行诊断的能力。由于实时采集所有活跃Session的指标对用户负载的影响加大，因此采取Session快照的技术对活跃

Session的指标进行采样。从采样中统计出活跃Session的统计指标，这些统计指标从客户端信息、执行开始、结束时间，SQL文本，等待事件，当前数据库对象等维度，反映活跃Session的基本信息，状态，持有的资源。基于概率统计的活跃Session信息，可以帮助用户诊断系统中哪些Session消耗了更多的CPU、内存资源，哪些数据库对象是热对象，哪些SQL消耗了更多的关键事件资源等，从而定位出有问题Session，SQL，数据库设计。

Session采样数据分为两级，如图2-2所示：

1. 第一级为实时信息，存储在内存中，展示最近几分钟的活跃Session信息，具有最高的精度；
2. 第二级为持久化历史信息，存储在磁盘文件中，展示过去很长一段时间的历史活跃Session信息，从内存数据中抽样而来，适合长时间跨度的统计分析。

图 2-2 Session 性能诊断原理



部分使用场景如下所示：

1. 查看session之间的阻塞关系

   select sessionid, block_sessionid from pg_thread_wait_status;
2. 采样blocking session信息

   select sessionid, block_sessionid from DBE_PERF.local_active_session;
3. Final blocking session展示

   select sessionid, block_sessionid, final_block_sessionid from DBE_PERF.local_active_session;
4. 最耗资源的wait event

   SELECT s.type, s.event, t.count

   FROM dbe_perf.wait_events s, (

   SELECT event, COUNT(*)

   FROM dbe_perf.local_active_session

   WHERE sample_time > now() - 5 / (24 * 60)

   GROUP BY event)t WHERE s.event = t.event ORDER BY count DESC;
5. 查看最近五分钟较耗资源的session把资源都花费在哪些event上。

   SELECT sessionid, start_time, event, count

```
FROM (
SELECT sessionid, start_time, event, COUNT(*)
FROM dbe_perf.local_active_session
WHERE sample_time > now() - 5 / (24 * 60)
GROUP BY sessionid, start_time, event) as t ORDER BY SUM(t.count) OVER (PARTITION BY t. sessionid, start_time)DESC, t.event;
```

6. 最近五分钟比较占资源的SQL把资源都消耗在哪些event上

```
SELECT query_id, event, count
FROM (
SELECT query_id, event, COUNT(*)
FROM dbe_perf.local_active_session
WHERE sample_time > now() - 5 / (24 * 60)
GROUP BY query_id, event) t ORDER BY SUM(t.count) OVER (PARTITION BY t.query_id ) DESC, t.event DESC;
```

特性增强

无。

特性约束

无。

依赖关系

无。

2.5.8 系统 KPI 辅助诊断

可获得性

本特性自V500R001C20版本开始引入。

特性简介

KPI是内核组件或者整体性能关键指标的视图呈现，基于这些指标，用户可以了解到系统运行的实时或者历史状态。

客户价值

- 系统负载概要诊断
  系统负载异常（过载、失速、业务SLA）准确告警，系统负载精准画像。
- 系统时间模型概要诊断
  Instance和Query级别时间模型细分，诊断Instance和Query性能问题根因。
- Query性能诊断
  数据库级Query概要信息，TopSQL，SQL CPU，I/O消耗，执行计划，硬解析过多。

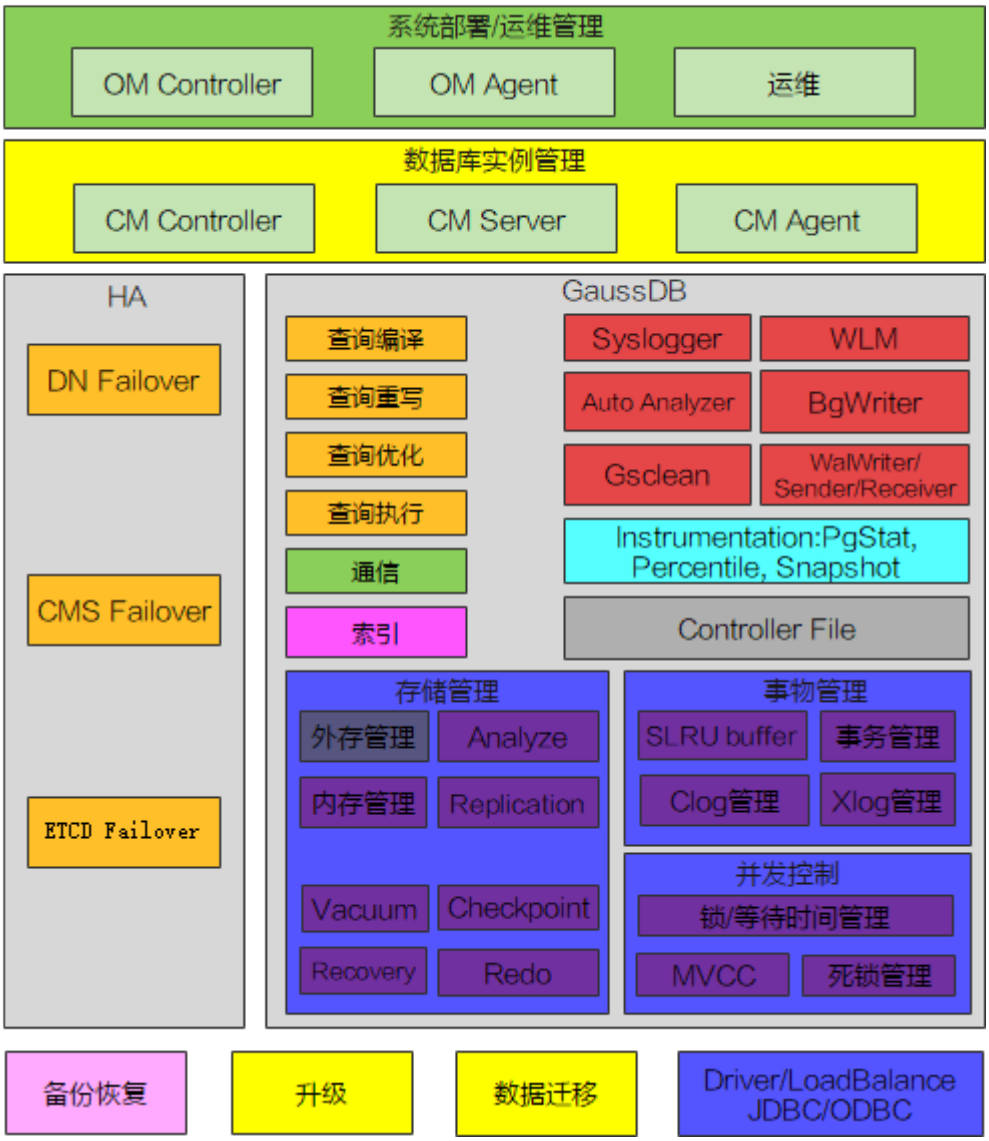
- 磁盘I/O、索引、buffer性能问题
- 连接池，线程池异常
- Checkpoint，Redo（RTO）性能问题
- 系统I/O、LWLock、Waits性能问题诊断
  诊断60+模块，240+关键操作性能问题。
- 函数级性能看护诊断(GSTRACE)，功能诊断
  50+存储和执行层函数trace。

特性描述

GaussDB提供涵盖11大类，26个子类的KPI，包括：Instance、File、Object、Workload、Communication、Session、Thread、Cache IO、Lock、Wait Event、Cluster。

KPI指标内核的分布如图2-3所示。

图 2-3 KPI 指标内核分布



特性增强

无。

特性约束

- 对于utility语句不支持归一化，主要体现为非DML语句，比如：create/drop/copy/vacuum等语句。
- 当前归一化SQL仅记录顶层SQL，对于存储过程语句，不对存储过程内部的SQL进行归一化处理，只记录调用存储过程的SQL。

依赖关系

无。

2.5.9 内置 stack 工具

可获得性

本特性自V500R002C10版本开始引入。

特性简介

stack工具是获取数据库中各线程的调用栈的工具，用于辅助数据库运维人员定位死锁、hang等问题。

客户价值

提供函数级别的调用栈信息，提升数据库内核运维人员分析、定位死锁、hang等问题的效率。

特性描述

可以通过函数gs_stack()或者工具gs_ctl stack两种方式获取数据库中线程的调用栈。

1. gs_stack()函数方式详见《开发者指南》的“SQL参考 > 函数和操作符 > 统计信息函数”章节。
2. gs_ctl stack方式获取调用栈，详见《工具参考》的“系统内部调用的工具 > gs_ctl”章节。

特性增强

无。

特性约束

1. 仅用于gaussdb进程，其他进程，如cms、gtm等不支持。
2. 如果使用SQL的方式执行，则需要CN、DN进程处于正常状态，可连接和执行SQL。
3. 如果使用gs_ctl的方式执行，则需要CN、DN进程处于可响应信号的状态。
4. 不支持并发，在获取全线程栈的场景，各个线程的调用栈不处于同一时间点。

5. 最多支持128层调用栈，如果实际情况超过128层，则仅保留栈顶的128层。
6. 符号表没有被trip（当前release版本，使用的是strip –d，仅去掉了debug信息，符号表没有被trip，如果改为strip –s，则仅能显示指针，无法显示出符号名）。
7. SQL执行方式仅支持monadmin、sysadmin用户。
8. 注册了SIGURG信号的线程，才能获取调用栈。
9. 对于屏蔽操作系统SIGUSR2的代码段，无法获取调用栈 ，如果线程没有注册signal_slot，同样无法获取调用栈。

依赖关系

无。

2.5.10 支持 SQL PATCH

可获得性

本特性自V500R002C10版本开始引入。

特性简介

SQL PATCH能够在避免直接修改用户业务语句的前提下对查询执行的方式做一定调整。在发现查询语句的执行计划、执行方式未达预期的场景下，可以通过创建查询补丁的方式，使用Hint对查询计划进行调优或对特定的语句进行报错短路处理。

客户价值

在业务产生查询计划不优导致的性能问题或系统内部错误导致服务不可用问题时，可以在数据库内通过运维函数调用对特定的场景进行调优或提前报错，以规避更严重的问题，能够大幅降低上述问题的运维成本。

特性描述

SQL PATCH主要设计给DBA、运维人员及其他需要对SQL进行调优的角色使用，用户通过其他运维视图或定位手段识别到业务语句存在计划不优导致的性能问题时，可以通过创建SQL PATCH对业务语句进行基于Hint的调优。目前支持行数、扫描方式、连接方式、连接顺序、PBE custom/generic计划选择、语句级参数设置、参数化路径的Hint。此外，对于部分由特定语句触发系统内部问题导致系统可服务性受损的语句，在不对业务语句变更的情况下，也可以通过创建用于单点规避的SQL PATCH，对问题场景提前报错处理，避免更大的损失。

SQL PATCH的实现当前基于Unique SQL ID，所以需要打开相关的运维参数才可以生效（enable_resource_track = on，instr_unique_sql_count > 0），Unique SQL ID在WDR报告和慢SQL视图中都可以获取到，在创建SQL PATCH时需要指定Unique SQL ID，对于存储过程内的SQL则需要设置参数 instr_unique_sql_track_type = 'all' 后在dbe_perf.statement_history视图中查询Unique SQL ID。

特性增强

无。

特性约束

1. 仅支持针对Unique SQL ID添加补丁，如果存在Unique SQL ID冲突，用于Hint调优的SQL PATCH可能影响性能，但不影响语义正确性。
2. 仅支持不改变SQL语义的Hint作为PATCH，不支持SQL改写。
3. 不支持逻辑备份、恢复。
4. 不支持创建时校验PATCH合法性，如果PATCH的Hint存在语法或语义错误，不影响查询正确执行。
5. 仅初始用户、运维管理员、监控管理员、系统管理员用户有权限执行。
6. 库之间不共享，创建SQL PATCH时需要连接目标库。
7. 配置集中式备机可读时，需要指定主机执行SQL PATCH创建/修改/删除函数调用，备机执行报错。
8. SQL PATCH同步给备机存在一定延迟，待备机回放相关日志后PATCH生效。
9. 限制在存储过程内的SQL PATCH和全局的SQL PATCH不允许同时存在。
10. 使用PREPARE + EXECUTE语法执行的预编译语句执行不支持使用SQL PATCH。
11. SQL PATCH不建议在数据库中长期使用，只应该作为临时规避方法。遇到内核问题所导致的特定语句触发数据库服务不可用问题，以及使用Hint进行调优的场景，需要尽快修改业务或升级内核版本解决问题。并且升级后由于Unique SQL ID生成方法可能变化，可能导致规避方法失效。
12. 当前，除DML语句之外，其他SQL语句（如CREATE TABLE等）的Unique SQL ID是对语句文本直接哈希生成的，所以对于此类语句，SQL PATCH对大小写、空格、换行等敏感，即不同的文本的语句，即使语义相同，仍然需要对应不同的SQL PATCH。对于DML，则同一个SQL PATCH可以对不同入参的语句生效，并且忽略大小写和空格。

依赖关系

本特性依赖于资源实时监控功能。对于不同的语句，数据库无法保证生产的Unique SQL ID哈希值全局唯一，如果不同的语句生成的Unique SQL ID冲突，会导致SQL PATCH命中预期外的其他语句。其中使用DBE_SQL_UTIL.create_hint_sql_patch接口创建的用于调优的Hint PATCH可能会影响错误命中语句的性能，使用DBE_SQL_UTIL.create_abort_sql_patch接口创建的用于避险的Abort PATCH需要谨慎使用。

2.5.11 SPM 计划管理

可获得性

本特性自GaussDB Kernel 505.0.0版本开始引入。

特性简介

SPM计划管理（SQL Plan Management）是一种数据库自动管理执行计划的预防机制，它确保数据库仅使用已知的、经过验证的执行计划。这种预防机制被称为baseline（全称SQL Plan Baseline），主要依托于Outline（一组用于固定计划的优化器Hint）来实现。

客户价值

解决计划跳变的问题：本特性会将SQL的baseline落盘存储，因此数据库重启前后均会使用之前预期的计划，从而有效地防止了计划跳变。

主动评估计划优劣：本特性提供了计划演进功能，用户可以主动评估SQL下计划的优劣，根据自动生成的演进报告，用户可决定该计划是否被数据库使用。

特性描述

SPM计划管理提供计划固定和计划演进两部分功能。计划固定可以有效地防止计划跳变，计划演进可以帮助用户发现执行效率更好的计划，并将其接受为可以被使用的计划。计划固定包含两个组件：计划捕获和计划选择。SPM计划管理的主要组件如下：

- **计划捕获：**此组件存储SQL语句和计划相关的信息。
- **计划选择：**此组件是基于已存储的当前SQL所有计划，在baseline中选择适当的计划以避免潜在的性能降低。
- **计划演进：**该组件功能是将目标计划手动演进到现有baseline。建议用户只有在验证计划的性能表现良好后，才接受计划到baseline中。

主要可以为如下场景带来显著收益：

1. 数据库升级安装新的优化器版本的过程中，通常会导致少量SQL语句的计划更改。SPM计划管理可以让数据库在升级前后，仅使用baseline中的计划，来有效避免计划跳变带来的性能劣化。
2. 持续的系统和数据变化可能会影响某些SQL语句的计划，存在导致性能劣化的风险，SPM计划管理可以让数据库仅使用baseline中的计划，来有效避免计划跳变带来的性能劣化。

特性约束

规格

- 支持DML、DQL语句（IUD、SELECT）的计划管理。
- 所有cplan默认都会被捕获为UNACC状态（受限于复用性问题），只有第一次被捕获的gplan会被标记为ACC状态（无复用性问题）。
- 关闭SPM时，性能不受影响。

约束

- 不支持对包含系统表的SQL语句进行计划管理。
- INSERT语句如果没有需要固化的算子，则SPM不会对其管理。
- SPM计划的并行行为依赖于数据库当时运行的参数query_dop。对每个算子的stream方式（local broadcast和local redistribute）会依照当前版本的并发能力，由优化器生成并发信息。
- 不支持存储过程中包含变量的SQL。

依赖关系

无

2.5.12 单节点支持磁盘只读告警

可获得性

本特性自GaussDB Kernel 503.1.0开始引入。

特性简介

数据库系统在运行过程中，运维人员需要监控系统运行状态，在问题出现后，及时提供运维服务，解决或者规避问题，保障业务正常可靠运行。该特性支持数据磁盘使用达到设定阈值后，实例DN状态变为只读状态（ ReadOnly ），实例上报告警。

客户价值

保障数据的可靠性，定时备份数据，在异常情况下恢复数据，保障数据一致性和完整性。

特性描述

为提升集中式单节点部署运维健康能力，在已支持安装部署、升级、备份恢复等运维能力基础上，实现磁盘只读告警上报功能。

特性增强

无

特性约束

- 数据库状态正常、客户端能够正常连接。
- CM Server参数enable_transaction_read_only为on并且ddb_type参数等于1。

2.5.13 支持设置云服务产品版本号

可获得性

本特性自505.0.0版本开始引入。

特性简介

云服务等第三方运维平台，可通过GUC工具设置相关的产品版本号和热补丁版本号信息，以便于用户查看。

客户价值

数据库产品为云服务的产品版本号管理提供便利。

特性描述

数据库提供product_version和hotpatch_version参数，云服务等运维集成平台可以将其对应的产品版本号和热补丁版本号信息设置到数据库系统，具体参数说明参考《管理员指南》中“配置运行参数>GUC参数说明>版本和平台兼容性>云服务产品版本号”章节。参数设置完成后，可通过SQL接口查看相关参数信息：


```
gaussdb=#  show product_version;
gaussdb=#  show hotpatch_version;
```

特性增强

无

特性约束

1. 版本号字符串长度限制，product_version 限制长度不超过50个字符，hotpatch_version限制长度不超过1500个字符。
2. 非法字符限制，参数值中不能包含以下字符：“|”、“;”、“&”、“$”、“>”、“<”、“`”、“\”、“!”和换行符。

依赖关系

无

2.5.14 内置 perf 工具

可获得性

本特性自505.0.0版本开始引入。

特性简介

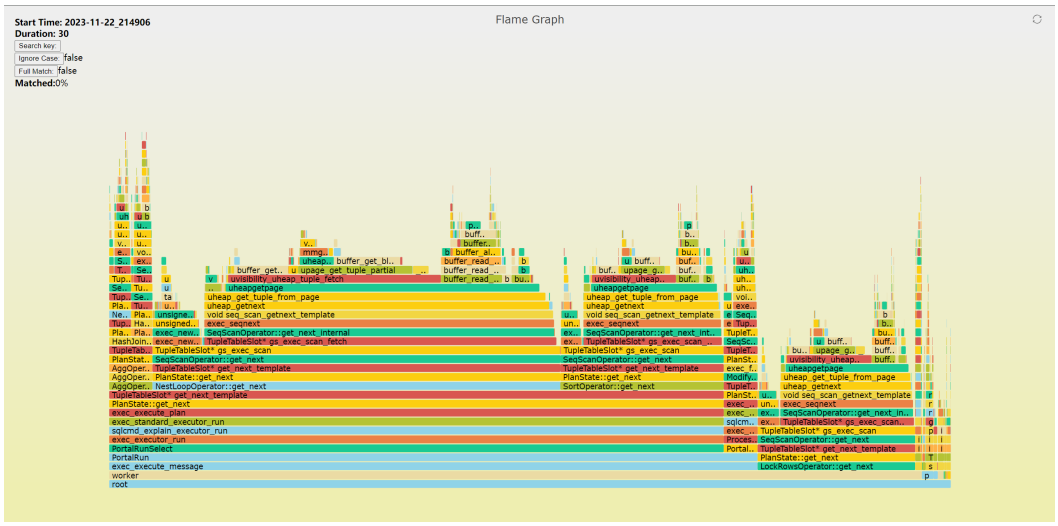
perf工具是在一定时间范围内，采集数据库中各线程的调用栈及其时间占比的工具，用于辅助数据库运维人员定位性能问题。该工具分为自动采集和手动采集两个功能。

自动采集堆栈功能，会定时采集数据库运行时活跃线程的函数调用栈及时间占比，并生成图形化火焰图报告。火焰图报告存储在$GAUSSLOG/gs_flamegraph/{datanode}路径下，下载该路径中的echarts.min.js文件和.html文件到同一目录下，用浏览器打开.html文件，即可展示采集到的堆栈调用火焰图。采集一次堆栈并生成火焰图的时间间隔通过GUC参数gs_perf_interval控制，范围为0或5-60，单位为min，默认为5min。自动采集堆栈功能功能默认开启，可以通过设置GUC参数gs_perf_interval=0关闭该功能。火焰图报告的保留时长由GUC参数gs_perf_retention_days控制，范围为1~8，单位为天，默认为3天。在gs_perf_interval不为0时，超过gs_perf_retention_days保存天数之外的火焰图文件会被回收。

手动采集堆栈功能，需要执行gs_perf_start()函数，采集一段时间内的堆栈调用情况，之后执行gs_perf_query()函数查询文字版堆栈调用信息。如需生成图形化火焰图报告，请执行gs_perf_report()函数。

堆栈调用火焰图是一种用于可视化性能分析的工具，它能够帮助开发者快速地识别程序中的性能瓶颈。堆栈调用火焰图y轴表示调用栈，每一层都是一个函数，顶部是当前正在执行的函数，下方是它的父函数，火焰的高度代表调用栈的深度；x轴表示采样执行时间占比，函数在x轴上的宽度越宽，表示它被采样的次数越多，即执行时间越长。火焰图展示的”平顶”，即占据宽度较大的函数，往往被用来定位数据库运行时的性能问题。

图 2-4 火焰图样例



客户价值

提供函数级别的调用栈信息，提供视图支持查询当前节点下所有活跃线程的函数调用栈及时间占比，提供生成火焰图文件的能力，提升数据库内核运维人员分析、定位性能问题的效率。

特性描述

可以通过gs_perf_start()，gs_perf_query()、gs_perf_clean()分别采集，查询和删除堆栈数据。通过dbe_perf.perf_query视图查询堆栈数据。

1. gs_perf_start()函数详见《开发者指南》中的“SQL参考 > 函数和操作符 > 统计信息函数”章节。
2. gs_perf_query()函数详见《开发者指南》中的“SQL参考 > 函数和操作符 > 统计信息函数”章节。
3. gs_perf_report()函数详见《开发者指南》中的“SQL参考 > 函数和操作符 > 统计信息函数”章节。
4. gs_perf_clean()函数详见《开发者指南》中的“SQL参考 > 函数和操作符 > 统计信息函数”章节。
5. dbe_perf.perf_query视图详见《开发者指南》中的“Schema > DBE_PERF Schema > OS > PERF_QUERY”章节。

特性增强

无。

特性约束

1. 执行性能采集时，需要数据库在normal、degrade状态，unavailable状态不支持。在degrade状态，异常cn/dn不支持采集。
2. 仅支持CN、DN、DN备、不支持logger节点，不支持cm_server、cm_agent、gtm、UDF等组件。
3. 仅支持on cpu采集，不支持off cpu采集。

4. 手动采集支持的采集时间范围为1s-60s。
5. 手动采集支持的采集频率为10HZ-1000HZ。
6. 自动采集堆栈功能一次采集5s，采集频率为99HZ。
7. 手动采集结果不落盘，存储在内存中。如果执行gs_perf_start后重启进程，则采集结果丢失。
8. 不支持并发采集。同一进程，同一时间，最多有一个session可以执行gs_perf_start操作。且同一进程内，在执行gs_perf_start操作期间，不支持执行gs_perf_query、gs_perf_report或gs_perf_clean操作。在自动采集堆栈期间，若手动执行采集堆栈函数gs_perf_start，会打断自动采集堆栈线程。
9. 自动或手动采集堆栈期间，不支持使用Linux perf工具操作同一进程。
10. 该特性依赖操作系统内核参数/proc/sys/kernel/perf_event_mlock_kb，该参数用来配置内置perf工具允许使用的最大内存值，操作系统中该参数的默认值为516KB。当操作系统中/proc/sys/kernel/perf_event_mlock_kb参数可修改时，在集群安装或升级的预安装阶段，默认将该参数调整为100MB。若该参数过小可能会导致手动和自动采集失败。

依赖关系

无。

2.6 数据库安全

2.6.1 访问控制模型

可获得性

本特性自V500R001C20版本开始引入。

特性简介

管理用户访问权限，为用户分配完成任务所需要的最小权限。

客户价值

客户依据自身需求创建对应的数据库用户并赋予相应的权限给操作人员，将数据库使用风险降到最低。

特性描述

数据库提供了基于角色的访问控制模型和基于三权分立的访问控制模型。在基于角色的访问控制模型下，数据库用户可分为系统管理员用户、监控管理员用户、运维管理员用户、安全策略管理员用户以及普通用户。系统管理员创建角色或者用户组，并为角色分配对应的权限；监控管理员查看dbe_perf模式下的监控视图或函数；运维管理员使用Roach工具执行数据库备份恢复操作；安全策略管理员创建资源标签、脱敏策略、统一审计策略。用户通过绑定不同的角色获得角色所拥有的对应的操作权限。

在基于三权分立的访问控制模型下，数据库用户可分为系统管理员、安全管理员、审计管理员、监控管理员用户、运维管理员用户、安全策略管理员用户以及普通用户。安全管理员负责创建用户，系统管理员负责为用户赋权，审计管理员负责审计所有用户的行为。

默认情况下，使用基于角色的访问控制模型。客户可通过设置GUC参数enableSeparationOfDuty为on来切换。

特性增强

无。

特性约束

系统管理员的具体权限受GUC参数enableSeparationOfDuty控制；

三权分立开启和关闭切换时需要重启数据库，且无法对新模型下不合理的用户权限进行自主识别，需要DBA识别并修正；

依赖关系

无。

2.6.2 数据库认证机制

可获得性

本特性自V500R001C20版本开始引入。

特性简介

提供基于客户端/服务端(C/S)模式的客户端连接认证机制。

客户价值

加密认证过程中采用单向Hash不可逆加密算法PBKDF2，有效防止彩虹攻击。

特性描述

GaussDB采用基本的客户端连接认证机制，客户端发起连接请求后，由服务端完成信息校验并依据校验结果发送认证所需信息给客户端（认证信息包括盐值，token以及服务端签名信息）。客户端响应请求发送认证信息给服务端，由服务端调用认证模块完成对客户端认证信息的认证。用户的密码被加密存储在内存中。整个过程中口令加密存储和传输。当用户下次登录时通过计算相应的hash值并与服务端存储的key值比较来进行正确性校验。

特性增强

统一加密认证过程中的消息处理流程，可有效防止攻击者通过抓取报文猜解用户名或者密码的正确性。

特性约束

无。

依赖关系

无。

2.6.3 数据加密存储

可获得性

本特性自V500R001C20版本开始引入。

特性简介

提供对导入数据的加密存储。

客户价值

为客户提供加密导入接口，对客户认为是敏感信息的数据进行加密后存储在表内。

特性描述

提供加密函数gs_encrypt_aes128()、gs_encrypt ()、gs_encrypt_bytea ()和解密函数gs_decrypt_aes128()、gs_decrypt()、gs_decrypt_bytea()接口。通过加密函数，可以对需要输入到表内的某列数据进行加密后再存储到表格内。调用格式为:

```
gs_encrypt_aes128(column, key),gs_encrypt (decryptstr,keystr,decrypttype),gs_encrypt_bytea(encryptstr, keystr, encrypttype)
```

其中key为用户指定的初始密码，用于派生加密密钥。当客户需要对整张表进行加密处理时，则需要为每一列单独书写加密函数。

当具有对应权限的用户需要查看具体的数据时，可通过解密函数接口对相应的属性列进行解密处理，调用格式为：

```
gs_decrypt_aes128(column, key),gs_decrypt(decryptstr,keystr,decrypttype),gs_decrypt_bytea(decryptstr, keystr, decrypttype)
```

参数	类型	描述	取值范围
encryptstr/ decryptstr	text	需要加解密的数据	-
keystr	text	密钥	8~16字节，至少包含3种字符（大写字母、小写字母、数字、特殊字符）
encrypttype/decryptyt	text	加解密类型（不区分大小写）	aes128_cbc_sha256、aes256_cbc_sha256、aes128_gcm_sha256,、aes256_gcm_sha256、sm4_ctr_sm3

说明

gs_encrypt_aes128、gs_decrypt_aes128使用默认加解密参数。

gs_encrypt、gs_decrypt除支持表格中参数外，兼容原有参数aes128、sm4。

特性增强

无。

特性约束

无。

依赖关系

无。

2.6.4 数据库审计

可获得性

本特性自V500R001C20版本开始引入。

特性简介

审计日志记录用户对数据库的启停、连接、DDL、DML、DCL等操作。

客户价值

审计日志机制主要增强数据库系统对非法操作的追溯及举证能力。

特性描述

数据库审计功能对数据库系统的安全性至关重要。数据库安全管理员可以利用审计日志信息，重现导致数据库现状的一系列事件，找出非法操作的用户、时间和内容等。

审计功能包括传统审计和统一审计两种审计模式。传统审计通过参数配置各个审计项开关，管理员可以通过参数配置对哪些语句或操作记录审计日志。传统审计采用记录到OS文件的方式来保存审计日志，支持审计管理员通过SQL函数接口审计日志查询和删除。统一审计机制是一种通过定制化制定审计策略而实现高效安全审计管理的一种技术。当管理员定义审计对象和审计行为后，用户执行的任务如果关联到对应的审计策略，则生成对应的审计行为，并记录审计日志。定制化审计策略可涵盖常见的用户管理活动，DDL和DML行为，满足日常审计诉求。

特性增强

503.1.0版本在传统审计功能基础上，新增支持用户级别审计功能。新增GUC参数full_audit_users配置全量审计用户列表，对列表中的用户执行的所有可被审计的操作记录审计日志；新增GUC参数no_audit_client配置无需记录审计的客户端列表；新增GUC参数audit_system_function_exec配置系统函数审计开关。

特性约束

无。

依赖关系

无。

2.6.5 网络通信安全

可获得性

本特性自V500R001C20版本开始引入。

特性简介

为保护敏感数据在Internet上传输的安全性，GaussDB支持通过SSL加密客户端和服务器之间的通讯。

客户价值

保证客户的客户端与服务器通讯安全。

特性描述

支持SSL协议标准。SSL(Secure Socket Layer)协议是一种安全性更高的应用层通信协议，主要用于Web安全传输，SSL包含记录层和传输层，记录层协议确定传输层数据的封装格式，传输层安全协议使用X.509认证。SSL协议利用非对称加密演算来对通信方做身份认证，之后交换对称密钥作为会谈密钥。通过SSL协议可以有效保障两个应用间通信的保密性和可靠性，使客户与服务器之间的通信不被攻击者窃听。

支持TLS 1.2协议标准。TLS 1.2协议是一种安全性更高的传输层通信协议，它包括两个协议组，TLS记录协议和TLS握手协议，每一组协议具有很多不同格式的信息。TLS协议是独立于应用协议的，高层协议可以透明地分布在TLS协议上面。通过TLS协议可保证通信双方的数据保密性和数据完整性。

GaussDB支持国密TLS加密传输，支持“ECC-SM4-SM3”和“ECDHE-SM4-SM3”国密加密算法套件，使用国密TLS需要配置国密双证书文件。

特性增强

证书签名算法强度检查：对于一些强度较低的签名算法，给出告警信息，提醒客户更换包含高强度签名算法的证书。

证书超时时间检查：如果距离超期日期小于7天则给出告警信息，提醒客户端更换证书。

证书权限检查：在建连阶段对证书的权限进行校验。

特性约束

- 从CA认证中心申请到正式的服务器、客户端的证书和密钥。
  - 使用国际证书认证，其服务器的私钥为server.key，证书为server.crt，客户端的私钥为client.key，证书为client.crt，CA根证书名称为cacert.pem。
  - 使用国密证书认证，其服务器的签名证书私钥为server.key，签名证书为server.crt，加密证书私钥为server_enc.key，加密证书为server_enc.crt，客户端的签名证书私钥为client.key，签名证书为client.crt，加密证书私钥为client_enc.key，加密证书为client_enc.crt，CA根证书名称为cacert.pem。
- 使用该功能需要打开SSL开关，并且配置证书和连接方式。
- 国密TLS加密传输当前只支持gsql客户端。

依赖关系

该特性依赖OpenSSL开源软件，国密TLS加密传输需要依赖支持国密TLS的OpenSSL版本。

2.6.6 资源标签机制

可获得性

本特性自V500R001C20版本开始引入。

特性简介

数据库资源是指数据库所记录的各类对象，包括数据库、模式、表、列、视图、trigger等，数据库对象越多，数据库资源的分类管理就越繁琐。资源标签机制是一种通过对具有某类相同“特征”的数据库资源进行分类标记而实现资源分类管理的一种技术。当管理员对数据库内某些资源“打上”标签后，可以基于该标签进行如审计或数据脱敏的管理操作，从而实现对标签所包含的所有数据库资源进行安全管理。

客户价值

合理的制定资源标签能够有效的进行数据对象分类，提高对象管理效率，降低安全策略配置的复杂性。当管理员需要对某组数据库资源对象做统一审计或数据脱敏等安全管理动作时，可将这些资源划分到一个资源标签，该标签即包含了具有某类特征或需要统一配置某种策略的数据库资源，管理员可直接对资源标签执行管理操作，大大降低了策略配置的复杂性和信息冗余程度，提高了管理效率。

特性描述

资源标签机制是将当前数据库内包含的各种资源进行“有选择性的”分类，管理员可以使用如下SQL语法进行资源标签的创建，从而将一组数据库资源打上标签：

CREATE RESOURCE LABEL schm_lb ADD SCHEMA(schema_for_label);

CREATE RESOURCE LABEL tb_lb ADD TABLE(schema_for_label.table_for_label);

CREATE RESOURCE LABEL col_lb ADD COLUMN(schema_for_label.table_for_label.column_for_label);

CREATE RESOURCE LABEL multi_lb ADD SCHEMA(schema_for_label), TABLE(table_for_label);

其中，schema_for_label、table_for_label、column_for_label分别为待标记模式、表、列。schm_lb标签包含了模式schm_for_label；tb_lb包含了表table_for_label；col_lb包含了列column_for_label；multi_lb包含模式schm_for_label和列table_for_label。对这些已配置的资源标签进行如统一审计或动态数据脱敏也即是对标签所包含的每一个数据库资源进行管理。

当前，资源标签所支持的数据库资源类型包括：SCHEMA、TABLE、COLUMN、VIEW、FUNCTION。

特性增强

无。

特性约束

- 资源标签需要由具备POLADMIN和SYSADMIN属性的用户或初始用户创建。
- 不支持对临时表创建资源标签。
- 同一个基本表的列只可能属于一个资源标签。
- 不支持通过gs_dump导出资源标签。系统管理员或安全策略管理员可以访问GS_POLICY_LABEL系统表查询已创建的资源标签。

依赖关系

无。

2.6.7 统一审计机制

可获得性

本特性自V500R001C20版本开始引入。

特性简介

审计机制是行之有效的安全管理方案，可有效解决攻击者抵赖，审计的范围越大，可监控的行为就越多，而产生的审计日志就越多，影响实际审计效率。统一审计机制是一种通过定制化制定审计策略而实现高效安全审计管理的一种技术。当管理员定义审计对象和审计行为后，用户执行的任务如果关联到对应的审计策略，则生成对应的审计行为，并记录审计日志。定制化审计策略可涵盖常见的用户管理活动，DDL和DML行为，满足日常审计诉求。

客户价值

审计是日常安全管理中必不可少的行为，当使用传统审计机制审计某种行为时，如SELECT，会导致产生大量的审计日志，进而增加整个系统的I/O，影响系统的性能；另一方面，大量的审计日志会影响管理员的审计效率。统一审计机制使得客户可以定制化生成审计日志的策略，如只审计数据库账户A查询某个表table的行为。通过定制化审计，可以大大减少生成审计日志的数量，从而在保障审计行为的同时降低对系统性能的影响。而定制化审计策略可以提升管理员的审计效率。

特性描述

统一审计机制基于资源标签进行审计行为定制化，且将当前所支持的审计行为划分为access类和privileges类。一个完整的审计策略创建的SQL语法如下所示：

CREATE RESOURCE LABEL auditlabel add table(table_for_audit1, table_for_audit2);

CREATE AUDIT POLICY audit_select_policy ACCESS SELECT ON LABEL(auditlabel) FILTER ON ROLES(usera);

CREATE AUDIT POLICY audit_admin_policy PRIVILEGES ALTER, DROP ON LABEL(auditlabel) FILTER ON IP(local);

其中，auditlabel为本轮计划审计的资源标签，该资源标签中包含了两个表对象；audit_select_policy定义了用户usera对auditlabel对象的SELECT行为的审计策略，不区分访问源；audit_admin_policy定义了从本地对auditlabel对象进行ALTER和DROP操作

行为的审计策略，不区分执行用户；当不指定ACCESS和PRIVILEGES的具体行为时，表示审计针对某一资源标签的所有支持的DDL和DML行为。当不指定具体的审计对象时，表示审计针对所有对象的操作行为。统一审计策略的增删改也会记录在统一审计日志中。

当前，统一审计支持的审计行为包括：

SQL类型	支持操作和对象类型
DDL	操作：ALL、ALTER、ANALYZE/VACUUM、COMMENT、CREATE、DROP、GRANT、REVOKE、SET、SHOW 对象：DATABASE、SCHEMA、FUNCTION/PROCEDURE、TRIGGER、TABLE、SEQUENCE、FOREIGN_SERVER、FOREIGN_TABLE、TABLESPACE、ROLE/USER/GROUP、INDEX、VIEW、DATA_SOURCE、WEAK PASSWORD DICTIONARY、AUDIT POLICY、MASKING POLICY、RESOURCE LABEL、MATERIALIZED VIEW/INCREMENTAL MATERIALIZED VIEW 注：对不支持的对象类型统一审计日记均标记为UNKNOWN
DML	操作：ALL、COPY、DEALLOCATE、DELETE_P、EXECUTE、REINDEX、INSERT、PREPARE、SELECT、TRUNCATE、UPDATE

说明

ALL指的是上述DDL或DML中支持的所有对数据库的操作。当形式为{ DDL | ALL }时，ALL指所有DDL操作；当形式为{ DML | ALL }时，ALL指所有DML操作。

其中EXECUTE是指执行预备语句的EXECUTE操作，并非存储过程中动态调用匿名块EXECUTE IMMEDIATE…USING语句。

特性增强

无。

特性约束

- 统一审计策略需要由具备POLADMIN或SYSADMIN属性的用户或初始用户创建，普通用户无访问安全策略系统表和系统视图的权限。
- 统一审计策略语法要么针对DDL行为，要么针对DML语法行为，同一个审计策略不可同时包含DDL和DML行为；统一审计策略目前支持最多设置98个。
- 统一审计监控用户通过客户端执行的SQL语句, 而不会记录数据库内部SQL语句。
- 同一个审计策略下，相同资源标签可以绑定不同的审计行为, 相同行为可以绑定不同的资源标签， 操作"ALL"类型包括DDL或者DML下支持的所有操作。
- 同一个资源标签可以关联不同的统一审计策略，统一审计会按照SQL语句匹配的策略依次打印审计信息。
- 统一审计策略的审计日志单独记录，暂不提供可视化查询接口，整个日志依赖于操作系统自带rsyslog服务，通过配置完成日志归档。
- 。
- FILTER中的APP项建议仅在同一信任域内使用，由于客户端不可避免的可能出现伪造名称的情况，该选项使用时需要与客户端联合形成一套安全机制，减少误用风险。一般情况下不建议使用，使用时需要注意客户端仿冒的风险。

- FILTER中的IP地址以ipv4为例支持如下格式:

ip地址格式	示例
单ip	127.0.0.1
掩码表示ip	127.0.0.1|255.255.255.0
cidr表示ip	127.0.0.1/24
ip区间	127.0.0.1-127.0.0.5

- 不支持通过gs_dump导出统一审计策略。系统管理员或安全策略管理员可以访问GS_AUDITING_POLICY、GS_AUDITING_POLICY_ACCESS、GS_AUDITING_POLICY_PRIVILEGES、GS_AUDITING_POLICY_FILTERS系统表查询已创建的统一审计策略。
- 由于GROUP是ROLE的别名，当统一审计的对象为GROUP时，统一审计日志中会将相应操作对象记录为ROLE类型。
- 统一审计日志中不区分存储过程和函数，当数据库对象是存储过程PROCEDURE时，日志中也会将其记录为FUNCTION类型。
- 统一审计策略中的ANALYZE对应VACUUM和ANALYZE两种SQL操作，审计日志中VACUUM操作也会被记录为ANALYZE。
- 由于语法解析机制，ALTER INDEX xxx REBUILD 语句会被审计为REINDEX语句。GRANT ALL PRIVILEGES TO user，REVOKE ALL PRIVILEGES FROM user语句会被审计为ALTER ROLE语句。
- WITH res1 AS (UPDATE ...) INSERT INTO ... VALUES (SELECT * from res1, ...)语句中不审计UPDATE语句，只审计主语句中的INSERT操作。
- 对于提升子查询语句不记录审计日志，例如INSERT INTO ... SELECT * FROM ...语句中只记录INSERT操作，不记录SELECT操作。
- 某些执行失败的DML语句不审计，例如唯一键约束导致执行失败、对只读子查询进行DML操作执行失败等。

⚠ 注意

使用统一审计功能时，强烈建议明确需要被审计的对象、需要审计的操作以及客户端、用户信息，根据场景创建准确的审计策略。对不必要的对象和操作进行审计会产生大量的审计日志引起数据库性能劣化、磁盘空间膨胀，也会影响管理员查询审计日志的效率。

2.6.8 动态数据脱敏机制

可获得性

本特性自V500R001C20版本开始引入。

特性简介

数据脱敏是行之有效的数据库隐私保护方案之一，可以在一定程度上限制非授权用户对隐私数据的窥探。动态数据脱敏机制是一种通过定制化制定脱敏策略从而实现对隐

私数据保护的一种技术，可以有效地在保留原始数据的前提下解决非授权用户对敏感信息的访问问题。当管理员指定待脱敏对象和定制数据脱敏策略后，用户所查询的数据库资源如果关联到对应的脱敏策略时，则会根据用户身份和脱敏策略进行数据脱敏，从而限制非授权用户对隐私数据的访问。

客户价值

数据隐私保护是数据库安全所需要具备的安全能力之一，可以在一定程度上限制非授权用户对隐私数据的访问，保证隐私数据安全。动态数据脱敏机制可以通过配置脱敏策略实现对指定数据库资源信息的隐私保护，另一方面，脱敏策略的配置也具有一定的灵活性，可以仅针对特定用户场景实现有针对性的隐私保护能力。

特性描述

动态数据脱敏机制基于资源标签进行脱敏策略的定制化，可根据实际场景选择特定的脱敏方式，也可以针对某些特定用户制定脱敏策略。一个完整的脱敏策略创建的SQL语法如下所示：

```
CREATE RESOURCE LABEL label_for_creditcard ADD
COLUMN(user1.table1.creditcard);

CREATE RESOURCE LABEL label_for_name ADD COLUMN(user1.table1.name);

CREATE MASKING POLICY msk_creditcard creditcardmasking ON
LABEL(label_for_creditcard);

CREATE MASKING POLICY msk_name randommasking ON LABEL(label_for_name)
FILTER ON IP(local), ROLES(dev);
```

其中，label_for_creditcard和label_for_name为本轮计划脱敏的资源标签，分别包含了两个列对象；creditcardmasking、randommasking为预置的脱敏函数；msk_creditcard定义了所有用户对label_for_creditcard标签所包含的资源访问时做creditcardmasking的脱敏策略，不区分访问源；msk_name定义了本地用户dev对label_for_name标签所包含的资源访问时做randommasking的脱敏策略；当不指定FILTER对象时则表示对所有用户生效，否则仅对标识场景的用户生效。

当前，预置的脱敏函数包括：

脱敏函数名	示例
creditcardmasking	'4880-9898-4545-2525' 将会被脱敏为 'xxxx-xxxx-xxxx-2525'，该函数仅对后4位之前的数字进行脱敏
basicemailmasking	'abcd@gmail.com' 将会被脱敏为'xxxx@gmail.com'， 对出现第一个'@'之前的文本进行脱敏
fullemailmasking	'abcd@gmail.com' 将会被脱敏为 'xxxx@xxxxx.com'，对出现最后一个'.'之前的文本（除'@'符外）进行脱敏
alldigitsmasking	'alex123alex' 将会被脱敏为 'alex000alex'， 仅对文本中的数字进行脱敏
shufflemasking	'hello word' 将会被随机打乱顺序脱敏为 'hlwoeor dl'， 该函数通过字符乱序排列的方式实现，属于弱脱敏函数，语义较强的字符串不建议使用该函数脱敏。

randommasking	'hello word' 将会被脱敏为 'ad5f5ghdf5'，将文本按字符随机脱敏
regexpmasking	需要用户顺序输入四个参数，reg为被替换的字符串，replace_text为替换后的字符串，pos为目标字符串开始替换的初始位置，为整数类型，reg_len为替换长度，为整数类型。reg、replace_text可以用正则表达，pos如果不指定则默认为0，reg_len如果不指定则默认为-1，即pos后所有字符串。如果用户输入参数与参数类型不一致，则会使用maskall方式脱敏。 `CREATE MASKING POLICY msk_creditcard regexpmasking('[\d+]', 'x', 5, 9 ) ON LABEL(label_for_creditcard);` '4880-9898-4545-2525' 将会被脱敏为 '4880-xxxx-xxxx-2525'
maskall	'4880-9898-4545-2525' 将会被脱敏为 'xxxxxxxxxxxxxxxxxxx'

每个脱敏函数规格如下：

脱敏函数名	支持的数据类型
creditcardmasking	定长类型：CHAR, BPCHAR；变长类型：VARCHAR, NVARCHAR2, TEXT（注：仅针对信用卡格式的文本类数据）
basicemailmasking	定长类型：CHAR, BPCHAR；变长类型：VARCHAR, NVARCHAR2, TEXT (注：仅针对email格式的文本类型数据)
fullemailmasking	定长类型：CHAR, BPCHAR；变长类型：VARCHAR, NVARCHAR2, TEXT (注：仅针对email格式的文本类型数据)
alldigitsmasking	定长类型：CHAR, BPCHAR；变长类型：VARCHAR, NVARCHAR2, TEXT (注：仅针对包含数字的文本类型数据)
shufflemasking	定长类型：CHAR, BPCHAR；变长类型：VARCHAR, NVARCHAR2, TEXT (注：仅针对文本类型数据)
randommasking	定长类型：CHAR, BPCHAR；变长类型：VARCHAR, NVARCHAR2, TEXT (注：仅针对文本类型数据)
regexpmasking	定长类型：CHAR, BPCHAR；变长类型：VARCHAR, NVARCHAR2, TEXT (注：仅针对文本类型数据)
maskall	BOOL、RELTIME、TIME、TIMETZ、INTERVAL、TIMESTAMP、TIMESTAMPTZ、SMALLDATETIME、ABSTIME、 TEXT、CHAR、BPCHAR、VARCHAR、NVARCHAR2、NAME、INT8、INT4、INT2、INT1、NUMRIC、FLOAT4、FLOAT8

对于不支持的数据类型，默认使用maskall函数进行数据脱敏。BOOL类型脱敏成'0'；RELTIME类型脱敏成'1970'；TIME, TIMETZ, INTERVAL类型脱敏成'00:00:00.0000+00'；TIMESTAMP, TIMESTAMPTZ, SMALLDATETIME, ABSTIME类型脱敏成'1970-01-01 00:00:00.0000'；TEXT, CHAR, BPCHAR, VARCHAR, NVARCHAR2, NAME类型脱敏成'x'；INT8, INT4, INT2, INT1, NUMERIC, FLOAT4, FLOAT8类型脱敏成'0'。若数据类型不属于maskall支持的类型则不支持创建脱敏策略。如果脱敏列涉及隐式转换，则结果以隐式转换后的数据类型为基础进行脱敏。另外需要说明的是，如

果脱敏策略应用到数据列并生效，此时对该列数据的操作将以脱敏后的结果为基础而进行。

动态数据脱敏适用于和实际业务紧密相关的场景，根据业务需要为用户提供合理的脱敏查询接口以及报错处理逻辑，以避免通过撞库而获取原始数据。

动态数据脱敏配置脱敏策略时，对用户创建的自定义函数进行支持适配。示例如下：

1. Poladmin权限用户创建一般函数，将传入的字符串中间8位替换成'xxxx-xxxx'后返回字符串:
```
create or replace function msk_creditcard(col text) returns TEXT as $$
declare
result TEXT;
begin
result := overlay(col placing 'xxxx-xxxx' from 6);
return result;
end;
$$ language plpgsql;
```
2. Poladmin权限用户使用上述函数创建脱敏策略：
```
CREATE MASKING POLICY msk_creditcard msk_creditcard ON LABEL(label_for_creditcard);
```
3. 查询信用卡号时，脱敏函数生效：

   '4880-9898-4545-2525' 将会被脱敏为 '4880-xxxx-xxxx-2525'

应用于动态数据脱敏的UDF规格如下：

UDF参数类型	支持的数据类型
输入参数	CHAR、BPCHAR、VARCHAR、NVARCHAR2、TEXT、INT8、INT4、INT2、INT1、FLOAT4、FLOAT8、NUMERIC

- 应用于动态数据脱敏的UDF入参参数不在字符类型(定长类型：char, bpchar；变长类型：varchar, nvarchar2, text)，数字类型(int8, int4, int2, int1, numeric, float4, float8)中时，用maskall脱敏。
- 应用于动态数据脱敏的UDF入参参数和返回参数需保持数据类型一致，字符类型可以兼容，否则用maskall脱敏。
- 如果列类型为maskall不支持的类型，则报错

特性增强

V500R002C00版本加入动态数据脱敏支持UDF，预置函数支持正则表达脱敏函数。

特性约束

- 动态数据脱敏策略需要由具备POLADMIN或SYSADMIN属性的用户或初始用户创建，普通用户没有访问安全策略系统表和系统视图的权限。
- 动态数据脱敏只在配置了脱敏策略的数据表上生效，而审计日志不在脱敏策略的生效范围内。
- 在一个脱敏策略中，对于同一个资源标签仅可指定一种脱敏方式，不可重复指定。
- 不允许多个脱敏策略对同一个资源标签进行脱敏，除以下脱敏场景外：使用FILTER指定策略生效的用户场景，包含相同资源标签的脱敏策略间FILTER生效场景无交集，此时可以根据用户场景明确辨别资源标签被哪种策略脱敏。

- Filter中的APP项建议仅在同一信任域内使用，由于客户端不可避免的可能出现伪造名称的情况，该选项使用时需要与客户端联合形成一套安全机制，减少误用风险。一般情况下不建议使用，使用时需要注意客户端仿冒的风险。
- 对于带有query子句的INSERT或MERGE INTO操作，如果源表中包含脱敏列，则上述两种操作中插入或更新的结果为脱敏后的值，且不可还原。
- 在内置安全策略开关开启的情况下，执行ALTER TABLE EXCHANGE PARTITION操作的源表若在脱敏列则执行失败。
- 对于设置了动态数据脱敏策略的表，需要谨慎授予其他用户对该表的trigger权限，以免其他用户利用触发器绕过脱敏策略。
- 最多支持创建98个动态数据脱敏策略。
- 仅支持对只包含COLUMN属性的资源标签做脱敏。
- 仅支持对普通表且为永久表的列进行数据脱敏。
- 仅支持对SELECT直接查询到的数据进行脱敏，对已脱敏结果进行二次处理会导致脱敏策略失效或不符合预期。
- 应用于动态数据脱敏的UDF只支持标准数据库SQL、PL/SQL function。
- 应用于动态数据脱敏的UDF中，如果包含访问数据库资源的语句如(select, insert)，使用该UDF的动态数据脱敏结果可能会不符合预期或导致安全风险。
- 应用于动态数据脱敏的UDF创建脱敏策略成功后，如果对该脱敏列进行alter或者drop，会导致脱敏策略失效或不符合预期。
- 动态数据脱敏的UDF函数不支持使用SECURITY INVOKER函数。应用于动态数据脱敏的UDF创建脱敏策略成功后，不允许对该function进行create、alter或drop。
- 应用于动态数据脱敏的UDF只能由具有poladmin权限用户创建。由具有poladmin权限的用户将访问schema的usage权限赋予public，如果因为grant/revoke操作，导致用户不能访问UDF，则使用maskall脱敏。
- 应用于动态数据脱敏的UDF应为幂等，即多次执行结果一样。如果设置UDF为非幂等，在分布式场景下使用UDF的动态数据脱敏结果可能会不符合预期。
- 不支持在系统表上应用动态数据脱敏的UDF创建脱敏策略。
- FILTER中的IP地址以ipv4为例支持如下格式：

ip地址格式	示例
单ip	127.0.0.1
掩码表示ip	127.0.0.1|255.255.255.0
cidr表示ip	127.0.0.1/24
ip区间	127.0.0.1-127.0.0.5

- 不支持通过gs_dump导出动态数据脱敏策略。系统管理员或安全策略管理员可以访问GS_MASKING_POLICY、GS_MASKING_POLICY_ACTIONS、GS_MASKING_POLICY_FILTERS系统表查询已创建的动态数据脱敏策略。

依赖关系

无。

2.6.9 行级访问控制

可获得性

本特性自V500R001C20版本开始引入。

特性简介

行级访问控制特性将数据库访问控制精确到数据表行级别，使数据库达到行级访问控制的能力。不同用户执行相同的SQL查询操作，读取到的结果是不同的。

客户价值

不同用户执行相同的SQL查询操作，读取到的结果是不同的。

特性描述

用户可以在数据表创建行访问控制(Row Level Security)策略，该策略是指针对特定数据库用户、特定SQL操作生效的表达式。当数据库用户对数据表访问时，若SQL满足数据表特定的Row Level Security策略，在查询优化阶段将满足条件的表达式，按照属性(PERMISSIVE | RESTRICTIVE)类型，通过AND或OR方式拼接，应用到执行计划上。

行级访问控制的目的是控制表中行级数据可见性，通过在数据表上预定义Filter，在查询优化阶段将满足条件的表达式应用到执行计划上，影响最终的执行结果。当前受影响的SQL语句包括SELECT，UPDATE，DELETE。

特性增强

- 505.1版本新增GUC参数enable_rls_match_index；该参数打开后，对于查询谓词包含unleakproof类型系统函数或like操作符的目标场景，允许基于该谓词条件执行索引扫描。

特性约束

- 行级访问控制策略仅可以应用到SELECT、UPDATE和DELETE操作，不支持应用到INSERT和MERGE操作。
- 支持对行存表、行存分区表、unlogged表定义行级访问控制策略，不支持对外表、本地临时表定义行级访问控制策略。
- 不支持对视图定义行级访问控制策略。
- 同一张表上可以创建多个行级访问控制策略，一张表最多允许创建100个行级访问控制策略。
- 初始用户和系统管理员不受行级访问控制策略的影响。
- 对于设置了行级访问控制策略的表，需要谨慎授予其他用户对该表的trigger权限，以免其他用户利用触发器绕过行级访问控制策略。
- 行级访问控制策略不支持使用SECURITY INVOKER的函数。对于已经使用了SECURITY INVOKER函数的策略，不允许对该FUNCTION进行CREATE、ALTER或DROP。
- 对于设置了行级访问控制策略、且具有函数表达式索引的表，仅当该函数为leakproof类型时，该表达式索引生效。

依赖关系

无。

2.6.10 用户口令强度校验机制

可获得性

本特性自V300R002C00版本开始引入。

特性简介

对用户访问数据库所设置的口令强度进行校验。

客户价值

用户无法设置过低强度的口令，加固客户数据安全。

特性描述

初始化数据库、创建用户、修改用户时需要指定密码。密码必须满足强度校验，否则会提示用户重新输入密码。账户密码复杂度要求如下：

- 包含大写字母（A-Z）的最少个数（password_min_uppercase）
- 包含小写字母（a-z）的最少个数（password_min_lowercase）
- 包含数字（0-9）的最少个数（password_min_digital）
- 包含特殊字符的最少个数（password_min_special）
- 密码的最小长度（password_min_length）
- 密码的最大长度（password_max_length）
- 至少包含上述四类字符中的三类。
- 不能和用户名、用户名倒写相同，本要求为非大小写敏感。
- 不能和当前密码、当前密码的倒写相同。
- 不能是弱口令。
  - 弱口令指的是强度较低，容易被破解的密码，对于不同的用户或群体，弱口令的定义可能会有所区别，用户需要自己添加定制化的弱口令。

参数password_policy设置为1时表示采用密码复杂度校验，默认值为1。

弱口令字典中的口令存放在gs_global_config系统表中（name字段为weak_password的记录是储存的弱口令），当创建用户、修改用户需要设置密码时，将会把用户设置口令和弱口令字典中存放的口令进行对比，如果命中，则会提示用户该口令为弱口令，设置密码失败。

弱口令字典默认为空，用户通过以下语法可以对弱口令字典进行增加和删除，示例如下：

```
CREATE WEAK PASSWORD DICTIONARY WITH VALUES ('password1'), ('password2');
DROP WEAK PASSWORD DICTIONARY;
```

其中"password1", "password2"是用户事先准备的弱口令，该语句执行成功后将会存入弱口令系统表中。

当用户尝试通过CREATE WEAK PASSWORD DICTIONARY 插入表中已存在的弱口令时，会只在表中保留一条该弱口令。

DROP WEAK PASSWORD DICTIONARY语句会清空整张系统表弱口令相关部分。

gs_global_config系统表没有唯一索引，不建议用户通过COPY FROM命令重复用相同数据对该表进行操作。

若用户需要对弱口令相关操作进行审计，应设置audit_system_object参数中的第三位为1。

特性增强

支持弱口令字典功能。

特性约束

初始用户、系统管理员和安全管理员可以查看、新增、删除弱口令字典。

依赖关系

无。

2.6.11 口令脱敏机制

可获得性

本特性自V300R002C00版本开始引入。

特性简介

在数据库审计日志或运行日志中打印SQL语句，或者在系统表或系统视图中记录SQL语句时，对SQL语句中包含的可识别的口令或密钥进行掩码脱敏处理，以防止敏感信息泄露。

客户价值

防止口令或密钥等敏感信息泄露，提升数据安全性。

特性描述

用户无需进行参数配置和执行其他操作，在打印SQL语句前，系统自动将识别到的口令或密钥信息进行脱敏。

支持口令或密钥脱敏的SQL语句场景如下：

- 对CREATE ROLE、CREATE USER、CREATE GROUP语句中的PASSWORD或IDENTIFIED BY参数进行脱敏。
- 对ALTER ROLE、ALTER USER语句中的PASSWORD、IDENTIFIED BY或REPLACE参数进行脱敏。
- 对SET ROLE、SET SESSION AUTHORIZATION语句中的PASSWORD参数进行脱敏。
- 对CREATE/ALTER SERVER语句中的secret_access_key参数进行脱敏。

- 对CREATE/ALTER TEXT SEARCH DICTIONARY语句中的FILEPATH参数如果为OBS目录则进行脱敏。
- 对CREATE/ALTER USER MAPPING语句中的password参数进行脱敏。
- 对DBE_SCHEDULER.CREATE_CREDENTIAL函数的password参数进行脱敏。
- 对gs_encrypt_aes128、gs_decrypt_aes128、gs_encrypt、gs_decrypt、gs_encrypt_bytea、gs_decrypt_bytea、aes_encrypt、aes_decrypt加解密函数中的密钥参数进行脱敏。
- 对pg_create_physical_replication_slot_extern函数中OBS归档槽相关敏感信息进行脱敏。

特性增强

无。

特性约束

- 只针对SQL语法中明确定义的口令或密钥信息进行脱敏，不支持用户自定义参数场景。
- 如果执行SQL语句中存在语法错误场景可能会导致脱敏失效。

依赖关系

无。

2.6.12 全密态数据库等值查询

可获得性

本特性自V500R001C20版本开始引入。

特性简介

密态数据库意在解决数据全生命周期的隐私保护问题，使得系统无论在何种业务场景和环境下，数据在传输、运算以及存储的各个环节始终都处于密文状态。当数据拥有者在客户端完成数据加密并发送给服务端后，在攻击者借助系统脆弱点窃取用户数据的状态下仍然无法获得有效信息，从而起到保护数据隐私的能力。

客户价值

由于整个业务数据流在数据处理过程中都是以密文形态存在，通过全密态数据库，可以实现：

1. 保护数据在云上全生命周期的隐私安全，无论数据处于何种状态，攻击者都无法从数据库服务端获取有效信息。
2. 帮助云服务提供商获取第三方信任，无论是企业服务场景下的业务管理员、运维管理员，还是消费者云业务下的应用开发者，用户通过将密钥掌握在自己手上，使得高权限用户无法获取数据有效信息。
3. 让云数据库借助全密态能力更好的遵守个人隐私保护方面的法律法规。

特性描述

加解密阶段，为保证客户端能够自动化地对数据进行加解密，用户需在数据定义阶段定义加密方案。同时，新增密钥管理语法，并支持第三方密钥管理工具，保证用户自主和灵活地选择加密方案。使用全密态数据库的整体流程分为如下四个阶段。

1. 密钥实体管理阶段：通过独立的密钥管理工具/服务管理密钥实体。目前，支持通过Huawei KMS对密钥进行独立管理。
   - Huawei KMS：由华为云提供的在线密钥管理服务，提供创建、删除、查询和备份密钥等功能，并支持在线使用密钥对数据进行加解密。
   - user_token：由用户提供的密码在客户端派生密钥，或者直接对接密钥。
   - third_kms：在加载三方加密库后，由三方加密库提供密钥管理服务。

   ⚠ 注意

   使用third_kms时密钥由第三方加密库管理。第三方加密库并非华为提供，需要保证该动态库功能执行正确。因第三方动态库的不正确实现，可能导致数据库进程异常、进程崩溃、内存泄露等等，需要联系第三方动态库解决，请谨慎使用。
2. 密钥对象管理阶段：通过新增的密钥管理SQL语法管理密钥对象，新增语法如下。
   - CREATE COLUMN ENCRYPTION KEY：支持用户定义用于加密表中指定列的密钥对象，同时，该对象中存储了列加密密钥实体的密文。
   - CREATE CLIENT MASTER KEY：支持用户定义用于加密CEK的CMK对象，该CMK对象不存储CMK密钥实体，而是存储从独立密钥管理工具/服务中读取CMK密钥实体的方法。
3. 数据定义阶段：新增对表中指定列进行加密定义的语法，新增语法如下。
   - CREATE TABLE ... (column DATE_TYPE ENCRYPTED WITH ...)：支持用户指定CEK来加密指定的列。
4. 数据加解密阶段：完成数据定义后，客户端便能够基于用户定义的加解密方案，自动地对表中数据进行加解密。

具体的语法可参考《开发者指南》中的“SQL参考 > SQL语法”章节。

特性增强

无。

特性约束

- 密钥实体管理约束。
  - 仅支持使用l和密钥管理服务Huawei KMS管理CMK密钥实体。
- 密钥对象管理约束。
  - CREATE CLIENT MASTER语法中，KEY_PATH字段仅能指向外部密钥管理工具/服务中已存在的密钥；由Huawei KMS生成的密钥，仅能用于AES_256和SM4算法。
  - CREATE COLUMN ENCRYPTION KEY语法中，ALGORITHM仅支持AEAD_AES_256_CBC_HMAC_SHA256、

AEAD_AES_128_CBC_HMAC_SHA256、AEAD_AES_256_CTR_HMAC_SHA256、AES_256_GCM和SM4_SM3。

- 如果使用由Huawei KMS生成CMK来加密CEK，在CREATE COLUMN ENCRYPTION KEY语法中，如果使用ENCRYPTED_VALUE字段，则该字段的长度需为16字节的整数倍。

- 数据以列级别进行加密，而无法按照行级别区分加密策略。
- 除Rename操作外，不支持通过Alter Table语法实现对加密表列的更改(包括加密列和非加密列之间的互转换)，支持添加(Add)和删除(Drop)对应的加密列。
- 不支持对加密列设置大部分check限制性语法，但是支持check(column is not null)语法。
- 仅支持对加密列建btree索引以及ubtree索引，不支持建索引的时候使用加密列做过滤操作。
- 不支持不同数据类型之间的隐式转换。不支持转义字符。
- 不支持不同数据类型密文间的集合操作。
- 不支持加密列为数组类型。
- 不支持加密列创建分区。
- 加密列仅支持repeat和empty_blob()函数。
- 当前版本只支持gsql、JDBC（部署在linux操作系统）和Go（部署在linux操作系统）客户端，暂不支持ODBC等其他客户端实现密态等值查询。
- 只支持通过客户端执行copy from stdin的方式、\copy命令的方式以及insert into values(…)的方式往密态表中导入数据。
- 不支持从加密表到文件之间的copy操作。
- 不支持包括范围查询以及模糊查询等在内的除等值以外的各种密态查询。
- 支持部分函数存储过程密态语法，密态支持函数存储过程具体约束查看《特性指南》中“设置密态等值查询 > 密态支持函数/存储过程章节。
- 不支持通过insert into…select…，merge into语法将非加密表数据插入到加密表数据中。
- 仅JDBC客户端支持调用decryptData接口，将通过非密态连接、逻辑解码等其他方式获得的密文，对密文进行解密。调用方法查看《特性指南》中“设置密态等值查询 > 使用JDBC操作密态数据库 > 执行密态等值密文解密”。
- 对于处于连接状态的连接请求，只有触发更新缓存的操作（更改用户，解密加密列失败等）和重新建连后才能感知服务端CEK信息变更。
- 不支持在由随机加密算法加密的列上进行密态等值查询，仅支持简单插入语法及全表查询。
- 不支持不同精度、不同原始数据类型或使用不同列加密密钥加密的密文数据进行数据导入或等值比较。
- 密态等值查询不支持外表。
- 不支持针对包含加密列的密态表创建物化视图。
- 不支持针对包含加密列的密态表及基于密态表的视图、函数、存储过程创建同义词。
- 对于数据库服务侧配置变更(pg_settings系统表、权限、密钥和加密列等信息), 需要重新建立JDBC连接保证配置变更生效。
- 不支持多条SQL语句一起执行，insert into语句多批次执行场景不受此条约束限制。

- 密态数据库对长度为零的空字符串不进行加密。
- 确定性加密存在频率攻击的潜在风险，不建议在明文频率分布明显的场景下使用。
- 密态表不支持闪回drop，闪回查询和闪回表。
- 密态等值查询采用客户端默认精度，与服务端精度设置不相关。
- COLLATE子句指定列的排序规则（该列必须是可排列的数据类型），加密列类型为非可排序的数据类型。
- JDBC不支持加密列使用setBlob接口。
- 不支持使用prepare执行DDL操作。
- 当update语句有临时表时，where条件不支持加密列做查询条件。
- 创建预编译语句时，同一个参数请勿同时用于预编译语句中的加密列和非加密列参数。
- 使用密态数据库过程中，请勿将数据从一个加密列插入到另一个密钥加密的加密列中，如insert into t1 values(select * from t2)，否则当该数据的密钥删除后，另一个密钥加密的加密列中有数据会获取不到该密钥。
- 当单事务中的加密字段多或者单次数据量过大，可能造成加密时间过长，产生超时等异常，建议拆分为子语句进行处理。
- 当使用third_kms时，密钥由第三方加解密库的KMS管理，数据库仅记录加解密所需密钥的ID。
- 当使用third_kms时，不支持密钥轮转语句轮转密钥。
- 原始类型为varchar、text、varchar2、clob、bytea的密文在数据导入时视为同一数据类型允许互相导入。
- 密态等值查询支持的数据类型包括：

数据类	类型	描述
整型	tinyint/tinyint(n)	微整数，同int1。
	smallint	小整数，同int2。
	int4	常用整数。
	binary_integer	A数据库兼容类型，常用整数。
	bigint/bigint(n)	大整数，同int8。
数值类型	numeric(p,s)	精度为p的准确数值类型。
	number	A数据库兼容类型，等同numeric(p,s)。
浮点类型	float4	单精度浮点数。
	float8	双精度浮点数。
	double precision	双精度浮点数。
字符类型	char/char(n)	定长字符串，不足补空格，默认精度为1。
	varchar(n)	变长字符串，n是指允许的最大字节长度。

	text	文本类型。
	varchar2(n)	A数据库兼容类型，等同varchar(n)。
	clob	大文本类型。
二进制类型	bytea	变长的二进制字符串。
	blob	二进制大对象。该类型按照字符串处理，不支持其他转换操作。

依赖关系

使用全密态相关特性建议更新至相同版本的libpq_ce客户端驱动、Go客户端驱动及JDBC客户端。

2.6.13 内存解密逃生通道

可获得性

本特性自503.1.0版本开始引入。

特性简介

内存解密作为密态等值查询的一个逃生通道使用。在该逃生通道中，会将密钥传输到数据库内存中，对数据进行解密，以实现密文字段的特殊计算或查询功能，包括范围查询、排序；其他涉及密文操作、隐式或显式类型转换时，进行自动加解密。

⚠ 注意

内存解密逃生通道，会将客户端密钥发送到数据库服务端内存缓存，实际的解密和数据明文运算都是在数据库内存中进行，断开连接或session中断后会对密钥和数据进行清理。在使用该特性前，请充分考虑该风险。

客户价值

作为密态等值查询的一个逃生通道使用，可通过服务端明文加密和密文解密，实现数据迁移以及密文字段的多种计算、查询。

特性描述

1. 密钥传输安全通道：在客户端和服务端基于RSA签名和ECDH密钥协商建立加密传输通道，由客户端命令触发将数据加密密钥（CEK）使用安全通道传到服务端。
   - gsql使用元命令“\send_token”或者“\st”传输密钥；使用元命令“\clear_token”或者“\ct”销毁密钥。
   - JDBC使用setClientInfo("send_token", null)传输密钥；使用setClientInfo("clear_token", null)销毁密钥。
   - Go使用db.Exec("send_token")传输密钥、db.Exec("clear_token")销毁密钥，或在连接时配置auto_sendtoken=yes开启密钥自动传输/销毁模式。

  - libpq使用函数PQenableTrustedDomain(conn, true)传输密钥，使用PQenableTrustedDomain(conn, false)销毁密钥。
2. 内存解密计算：通过类型转换函数，将密文解密成明文数据进行计算。
3. 隐式/显式加解密：支持在涉及密文操作时，通过自动类型转换函数进行加解密：
  - 支持在以密文数据为原始类型，明文数据为目标类型的显示或隐式转换时，进行自动解密。
  - 支持在以明文数据为原始类型，由insert/update的目标字段指定的密文类型的隐式转换时，进行自动加密。
4. 加解密函数：支持服务端密态等值的明文加密和密文解密，从而支持服务端对明密文列进行数据迁移。
  - 服务端支持明文加密至密态等值的密文的功能。
  - 服务端支持密态等值的密文解密至明文的功能，该操作返回明文至非可信执行环境，谨慎使用。

特性增强

无。

特性约束

- 内存解密逃生通道继承等值查询的特性约束。
- 密钥成功传输后开启内存解密逃生通道。
- 密文运算只支持密文数据使用确定性加密算法。
- 支持default语法中填表达式，但使用需注意该表达式会被直接落盘。
- 密文排序，只支持行存表。
- 密文范围查询和排序，不支持精确的代价估计。
- 密文排序规则，与服务端本地排序规则一致。
- 密钥传输和密文运算，在gsql中不支持使用\parallel on。
- 服务端缓存的密钥在session级别的变量中，和session生命周期一致，当session切换或断连、重连，或者切换用户后，密钥会自动清零，如需使用需要重新传输密钥。
- 密文范围查询和排序不支持使用索引加速，默认使用顺序扫描，不影响等值查询使用索引加速查询。
- 加解密函数仅支持通过INSERT INTO…SELECT…语法将非加密表数据插入到加密表数据中，或者将加密表数据插入到非加密表数据中。
- 服务端加解密函数，是在类型转换自动加解密失效时的手动逃生通道，仅配合INSERT INTO…SELECT…语法使用，且要求原始数据类型与加密或解密的目标字段类型匹配，其中解密函数仅支持原始类型为字符型数据类型。建议优先使用隐式或显示转换方式进行自动加解密。
- 显式/隐式加解密，只支持目标类型为全密态支持类型的隐式类型转换进行自动解密；未触发隐式类型转换的场景（例如密文作为字符串，可以直接用于处理时），以及触发隐式类型转换时，密文原始类型无法获取的场景（密文原始类型存储在typemod中），建议使用显式类型转换函数进行解密；只支持密态支持类型为原始类型，目标类型由insert或update字段指定的自动加密，不支持显式类型转换进行加密；不支持在存储过程中隐式转换进行自动解密；在密态算子（如等值、范围）可直接计算时，不会通过隐式类型转换进行解密。

- 如果涉及不同原始类型的加密列相互导入，需要将源加密列显式转换为目标列的原始类型后再导入数据。
- 请谨慎在密文列的插入或更新语句中使用明文常量参与计算的表达式，否则在服务端开启日志打印的情况下会在日志中打印。
- 密文范围查询和排序支持的数据类型如下：

数据类	类型	描述
整型	tinyint/tinyint(n)	微整数，同int1。
	smallint	小整数，同int2。
	int4	常用整数。
	binary_integer	A数据库兼容类型，常用整数。
	bigint/bigint(n)	大整数，同int8。
数值类型	numeric(p,s)	精度为p的准确数值类型。
	number	A数据库兼容类型，等同numeric(p,s)。
浮点类型	float4	单精度浮点数。
	float8	双精度浮点数。
	double precision	双精度浮点数。
字符类型	char/char(n)	定长字符串，不足补空格，默认精度为1。
	varchar(n)	变长字符串，n是指允许的最大字节长度。
	text	文本类型。

依赖关系

使用全密态相关特性建议更新至相同版本的libpq_ce客户端驱动、Go客户端驱动以及JDBC客户端驱动。

2.6.14 账本数据库机制

可获得性

本特性自V500R002C00版本开始引入。

特性简介

账本数据库特性，对用户指定的防篡改表增加校验信息，并记录用户对其数据的操作历史，通过数据和操作历史的一致性校验来保证用户数据无法被恶意篡改。在用户对防篡改表执行DML操作时，系统对防篡改表增加少量额外的行级校验信息，同时记录操作的SQL语句和数据的变化历史。通过特性提供的校验接口，用户可以方便的校验防篡改表中的数据是否与系统记录的操作信息是否一致。

客户价值

账本数据库通过提供对用户数据的操作记录、数据历史变化记录以及易用的一致性的校验接口，方便用户随时校验数据库中的敏感信息是否发生恶意篡改，有效提高数据库防篡改能力。

特性描述

账本数据库采用账本Schema对普通表和防篡改用户表进行隔离。用户在账本Schema中创建的行存表具有防篡改属性，即为防篡改用户表。用户向防篡改用户表中插入数据时，系统会自动生成少量行级校验信息。在用户执行DML时，系统会在全局区块表(GS_GLOBAL_CHAIN)中记录用户的操作、在用户表相应的历史表中记录数据的更改等信息，操作记录、数据变化记录和用户表中的数据三者严格保持一致。账本数据库提供高性能校验接口，能够供用户方便的校验数据的一致性，如果一致性校验失败，则说明数据可能发生篡改，需要及时联系审计管理员回溯操作记录历史。

说明

防篡改表在创建时，支持以下数据类型：

char, abstime, bigint, boolean, bytea, character varying, character, date, double precision, int2vector, integer, interval, money, name, numeric, nvarchar2, oid, oidvector, raw, real, reltime, smalldatetime, smallint, text, time with time zone, time without time zone, timestamp with time zone, timestamp without time zone, tinyint, uuid, clob。

特性增强

无。

特性约束

- 防篡改模式下的行存表具有防篡改属性，而临时表、UNLOGGED表等均不具有防篡改属性。升级过程中需要在特定schema中创建表的情况下，不建议在升级前将该schema设置为防篡改属性。
- 不允许修改防篡改用户表的结构，不允许truncate防篡改相关表，不允许将防篡改用户表切换到普通的Schema中，不允许将非防篡改表切换到防篡改Schema中。
- 不支持指定二级分区表为防篡改表。防篡改表如果为分区表，则不支持exchange partition、drop partition、truncate partition等操作。
- 不支持使用函数、TRIGGER修改防篡改用户表数据，允许操作类型为SELECT的存储过程执行。
- 不支持用户主动修改校验列的值以及对该列创建索引。
- 普通用户调用篡改校验接口只能校验自己有权查询的表。
- 只允许审计管理员、系统管理员和初始用户查询全局区块表和BLOCKCHAIN模式中的表，普通用户无权访问，所有用户均无权修改。
- 根据用户历史表命名规则，若待创建表的Schema或表名以'_'结尾或开头，可能会出现对应历史表名与已有表名冲突的情况，需要重新命名。
- 账本数据库目前针对用户行级数据的hash摘要仅用来保证数据的一致性，无法保证密码学完整性，且当前能力暂时无法防止攻击用户直接对数据文件的篡改。
- 创建防篡改schema以及更改普通schema为防篡改schema，需设置enable_ledger参数为on。enable_ledger参数默认值为off。

- 基于防篡改表创建的定时任务对防篡改表有数据更改操作，将执行失败。
- 不允许基于防篡改表创建RULE。
- 不允许基于防篡改表创建发布，基于防篡改表创建for all table 的发布将会过滤掉防篡改表。
- 不允许对防篡改表执行多表操作。
- 对防篡改表执行闪回表操作，历史表和全局区块表无法自动进行数据同步，为保持数据一致，需对历史表执行闪回表操作，使用ledger_gchain_repair修复全局区块表。
- 由于在CTE或者子计划中修改表，会存在一个嵌套插入，导致无法记录整体的防篡改表修改，因此当前不支持使用CTE或者子计划修改防篡改表。
- 集群运行异常造成账本数据库校验函数返回不一致时，需要使用账本数据库修复函数对不一致的历史表或全局区块表进行修复。

依赖关系

无。

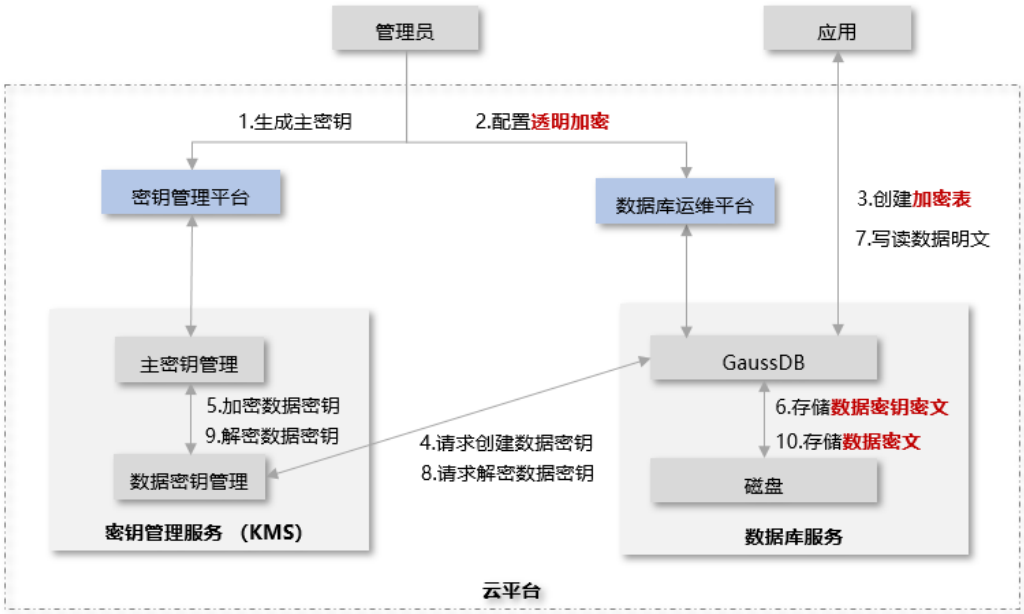
2.6.15 透明数据加密

可获得性

本特性自V500R002C00版本开始引入。

特性简介

透明数据加密（Transparent Data Encryption，TDE）提供表级数据加密存储功能。当用户使用本特性提供的语法创建加密表后，数据库向磁盘写入加密表数据前，会自动将其加密；同时，数据库从磁盘读取加密表数据后，会自动将其解密。



客户价值

安全：本特性可有效解决攻击者绕过数据库认证机制直接读取数据库文件引起静态数据泄露的问题。

易用：用户在创建表时，通过语法指定表是否需要被加密，数据库可自动对数据进行加密存储。

性能：本特性仅在数据写入磁盘时加密，并以数据页为粒度对数据进行加密，对数据库性能影响较小。

特性描述

多级加密：本特性使用多级加密模型，密钥分为主密钥（CMK）和数据加密密钥（DEK），数据由DEK加密，而DEK由CMK加密，CMK由外部密钥管理服务（Key Management Service，KMS）加密存储。数据库在运行过程中，需要通过网络或其他途径访问外部密钥管理者，以实现对DEK进行加解密。

表级加密：在创建表时，通过ENABLE_TDE语法指定表为加密表，通过ENCRYPT_ALGO语法指定使用何种加密算法，同时，数据库会自动为每个加密表生成1个DEK。对于每个加密表，数据库会在系统表和数据文件中存储加密信息以及DEK密文。

密钥管理：CMK由外部密钥管理者生成并存储。目前的外部密钥管理者主要指密钥管理服务（Key Management Service，KMS），大部分云服务提供商均提供KMS。

密钥轮转：本特性提供DEK轮转语法，加密表与非加密表转换语法。

特性增强

503.2.0：

- 密钥管理支持。
- 存储引擎支持Ustore。

505.1.0：

- 支持段页式表。
- 支持对索引加密，支持直接将非加密表转换为加密表。

特性约束

规格

- 加密规格
  - 加密对象：支持对astore表、ustore表、临时表、unlog表、段页式表等表加密，不支持对压缩表、物化视图、toast表等其他表加密。支持对btree索引、ubtree索引加密。
  - 加密算法：支持AES_128_CTR（默认算法），SM4_CTR。
- 密钥规格
  - 密钥管理：主密钥由单独的密钥服务管理，支持以下密钥服务：华为云密钥管理服务（ ）、第三方密钥管理服务（third_kms）。
  - 密钥隔离：每个加密表都使用单独的数据密钥加密算法，不支持单独指定索引的加密算法。

  - 密钥复用：如果对索引进行加密，则索引与基表使用相同的数据密钥。
  - 密钥轮转机制：进行密钥轮转时，表中新数据页将使用新密钥，旧数据页仍使用旧密钥，执行VACUUM FULL等重建数据文件的操作后，旧数据页才会使用新密钥。
- 使用规格
  - 加密表转换：

    1. 将非加密表转换为加密表时，表中新数据页将会被加密，旧数据页不会立刻被加密，执行VACUUM FULL等重建数据文件的操作后，旧数据页才会被加密。

    2. 将加密表转换为非加密表时，表中新数据页将不会被加密，旧数据页仍处于加密状态，执行VACUUM FULL等重建数据文件的操作后，旧数据页才会被解密。

    ⚠ 注意

    使用third_kms时密钥由第三方加密库管理。第三方加密库并非华为提供，需要保证该动态库功能执行正确。因第三方动态库的不正确实现，可能导致数据库进程异常、进程崩溃、内存泄露等等，需要联系第三方动态库解决，请谨慎使用。

约束

- 运行环境约束

  网络约束：需保证每个数据库节点与KMS之间网络通畅。
- 配置约束

  特性开关：如果开启透明加密，创建加密表并向加密表中写入数据，在关闭透明加密功能后，无法对加密表进行读写操作。
- 功能约束
  - 索引加密：只支持对基表为加密表的索引进行加密。
  - 加密范围：支持对表和索引的数据文件中的数据进行加密，不支持对网络传输、xlog文件、pg_static系统表文件和core文件等其他介质中的数据进行加密。
  - 废弃数据：不对数据库清理机制产生的废弃数据进行加密。
  - 超长数据：由于加密表的数据页容量小于非加密表，如果非加密表中含长度接近8000字节的单条非toast数据，请谨慎该表转换为加密表，否则可能出现加密表可能无法存储超长数据的异常。解决异常的方案是通过ALTER .. SET (enable_tde=off)语法将加密表还原为非加密表，异常场景示例如下：
    - 异常示例1：将表转换为加密表后，执行VACUUM FULL tablename语法语法失败。
    - 异常示例2：将表转换为加密表后，UPDATE旧数据失败。
- 特性交互约束
  - 备份恢复：不支持细粒度备份恢复。
  - repair：调用repair函数修复加密表的数据页时，不对生成的临时文件中数据进行加密。

- 其他约束
  - 性能劣化：与非加密表相比，在加密表上进行DML操作时，性能会较小劣化。
  - 数据膨胀：与非加密表相比，加密表数据文件中存储了加密信息，存储空间缩小5%以内。

依赖关系

本特性依赖外部密钥管理服务提供密钥管理功能。

2.6.16 基于标签的强制访问控制

可获得性

本特性自503.2.0版本开始引入。

特性简介

基于标签的强制访问控制特性支持用户对主体和客体设置安全标签，并基于系统设定好的强制访问控制策略规则执行访问控制。通过给主体（用户或角色）和客体（表或表的列）设置合适的安全标签来控制用户/角色可以操作数据库的哪些表或表的列。

系统新增语法和系统表支持安全标签的创建删除和记录；然后在针对数据库表或表的列进行权限校验的位置，增加主体和客体安全标签的比较逻辑，通过比较安全标签级别和范围是否符合强制访问控制策略规则来决定校验是否通过，不通过则拒绝访问。

客户价值

基于标签的强制访问控制是由系统管理人员设置主体和客体的安全标签，系统会按照严格的强制访问控制策略进行访问控制，规则由系统决定，不能更改，从而对数据库中的敏感信息提供更严格的权限控制。

特性描述

首先用户根据业务需要创建由等级和范围组成的安全标签，并将安全标签分别应用到主体（用户或角色）和客体（表或表的列）上。然后当强制访问控制检查开关打开（enable_mac_check=on），用户执行DML操作时，系统会自动根据内置的基于安全标签的强制访问控制策略校验主体是否被允许访问客体，如果校验不通过，则访问失败。

- 安全标签由等级和范围两部分组成，两者中间用冒号分隔，形式如：等级类别:范围类别，其中等级类别有且仅由一个等级组成，范围类别可由多个范围组成，但至少需要有一个范围，例如“L1:G2,G41,G6-G27”。
- 等级分类中有1024个等级，命名为Li，其中1≤i≤1024，等级满足偏序关系（若i ≤ j，则Li ≤ Lj），例如等级L1小于等级L3。
- 范围分类中有1024个范围，命名为Gi，其中1≤i≤1024，范围之间无法比较大小，但可以进行集合运算，多个范围之间用逗号分隔，连字符表示区间，例如{G2-G5}表示{G2,G3,G4,G5}，集合{G1}是集合{G1,G6}的子集。
- 等级和范围的首字母L和G均为大写；L和G之后至少要有一个数字字符，且第一位非零，不允许出现其他非数字字符；{Gxxx-Gyyy}形式中数字yyy必须大于等于xxx。

- 不符合要求的等级和范围均为非法输入，系统会报错。

基于安全标签的强制访问控制策略规则由系统内置设定，用户不能更改：

- 插入（INSERT）策略：只有主体安全标记等级小于等于客体安全标记等级且主体安全标记范围是客体安全标记范围的子集时才允许插入数据。
- 查询（SELECT）策略：只有主体安全标记等级大于等于客体安全标记等级且主体安全标记范围是客体安全标记范围的超集时才允许查询数据。
- 修改（UPDATE）和删除（DELETE）策略：只有主体安全标记等级等于客体安全标记等级且主体安全标记范围等于客体安全标记范围时才允许修改和删除数据。
- 若客体未标记，则强制访问控制策略对该客体不生效。
- 若主体未标记，则不能访问任意带有标记的客体。

特性增强

无。

特性约束

- 初始用户，具有SYSADMIN权限的用户或者继承了内置角色gs_role_seclabel权限的用户有权限创建、删除和应用安全标签。
- 只有初始用户才能应用或取消初始用户和具有persistence属性的用户的安全标签。
- 对主体和客体设置安全标签，主体支持用户和角色，客体支持普通表（pg_class中的relkind='r'）和普通表的列，不支持在系统表和系统表的列上应用安全标签。
- 打开强制访问控制开关后，基于安全标签强制访问控制与原有的自主访问控制（ACL权限和ANY权限）是“与”的关系，需要都满足才能访问成功。
- 出于防呆考虑，初始用户和系统管理员默认也受强制访问控制限制，但是可以随时更改或取消安全标签来使自己满足强制访问控制策略规则。
- 对于设置了安全标签的表，需要谨慎授予其他用户对该表的trigger权限，以免其他用户利用触发器绕过强制访问控制策略规则。

依赖关系

无。

2.6.17 敏感数据发现

可获得性

本特性自505.1.0版本开始引入。

特性简介

敏感数据发现功能提供函数gs_sensitive_data_discovery()和gs_sensitive_data_discovery_detail()，通过调用的不同函数，指定扫描对象和敏感数据分类器，得到对应扫描对象不同明细级别的敏感数据信息。

客户价值

敏感数据发现功能配合数据库内其他数据标记（如**动态数据脱敏机制**，**基于标签的强制访问控制**等）和保护特性（如**透明数据加密**，**行级访问控制**等）能够帮助客户：

1. 有效识别敏感信息和资产。
2. 提供更全面的数据保护能力。
3. 满足客户隐私数据保护以及符合监管规范的需求。

特性描述

本特性以函数调用的形式实现功能。分为两个部分：

1. 敏感数据发现函数框架的实现：包含内置函数的添加、参数校验、采样扫描、对采样数据应用分类器、扫描结果处理、异常处理、函数执行动作记录审计日志等。
2. 敏感数据分类器的具体算法实现：本次支持内置以下分类器（不区分大小写），分别为Email（电子邮件）/ PhoneNumber（电话号码）/ CreditCard（信用卡）/ ChineseName（中文姓名）/ EncryptedContent （加密数据）/ all（以上五种分类器全部选择）。

   具体分类原理及命中规则如下**表2-3**所示。

表 2-3 敏感数据分类器及分类器说明

分类器	分类器说明
电子邮件	匹配形如“example@example.com”、“example@example.co.uk”、“example@example.com.org”等的电子邮件地址。
电话号码（中国）	匹配手机号码，匹配手机号码，必须是（+国家代码）手机号格式，必须是11位数字，形如：+86 xxxxxxxxxxx，xxxxxxxxxxx，xxx xxxx xxxx，xxx-xxxx-xxxx。
信用卡	信用卡卡号位数需满足16位并且满足**Luhn（mod 10）**算法要求。 **说明** Luhn（mod 10）算法： 1. 从卡号最后一位数字开始，逆向将奇数位（1、3、5等）相加； 2. 从卡号最后一位数字开始，逆向将偶数位数字，先乘以2（如果乘积为两位数，则将其减去9），再求和； 3. 将奇数位总和加上偶数位总和，结果能被10整除则认为是合法的信用卡号，否则不是。

分类器	分类器说明
中文姓名	使用中国常见的姓氏进行判断。需要满足以下约束： ● 仅支持简体中文。 ● 对于某些少数民族姓名，可能存在漏报场景。 ● 仅对姓名长度为2-4的数据进行判断，超出此范围的数据不认为满足姓名规则。 ● 姓或名拆开为单独的列不会被识别成敏感。 ● 仅姓在前名在后的数据会被认为是符合规则的，姓在后名在前的数据也不会被识别为敏感。
加密数据	通过计算数据信息熵，并与设置的**阈值**相比较，信息熵高于阈值，则认为数据为敏感。该**阈值**保证数据库中所有加解密算法得到的密文均能识别为密文敏感数据。

特性增强

无。

特性约束

- 仅支持对扫描范围内可访问对象中的列应用分类器进行判断，无权限对象则跳过，函数执行结果中回显NOTICE提示存在无权限访问的对象。
- 仅支持对用户创建的模式、普通表、分区表、二级分区表进行扫描，不支持对snapshot、AI、系统schema等模式进行扫描，不支持对索引、物化视图等进行扫描。如果指定的扫描范围本身是规格范围外的对象，则会提示不支持；如果指定的扫描范围本身支持，但其中包含的某些对象不支持，则不会提示。
- 中文姓名分类器中只支持简体中文。对于少数民族姓名可能无法识别，会有漏报出现。
- 不建议在数据库业务繁忙阶段调用敏感数据发现函数、不建议多个客户端并发执行敏感数据发现函数、不建议在短时间内多次调用敏感数据发现函数，性能敏感场景下，可以开启IO管控和流控后再调用敏感数据发现函数。
- 由于敏感数据发现是在堆表层进行扫描，不会受到行级访问控制策略的影响。添加了行级访问控制的行数据也可以被正常扫描。

依赖关系

无。

2.7 负载管理

2.7.1 支持 I 层高时延逃生能力

可获得性

本特性自V500R002C10SPC500版本开始引入。

特性简介

I层异常会导致数据库SQL执行时延升高，进而导致内存或者线程池出现过载问题，针对此场景GaussDB提供自动逃生能力。

客户价值

当数据库由于I层异常导致SQL执行时延升高，会话堆积，内存或线程池过载无法对外提供服务时，能够快速实现逃生，短时间内恢复对外提供服务的能力。

特性描述

- 数据库内存出现过载问题时，快速kill会话并禁止新连接接入，待内存恢复正常状态后恢复对外服务能力。内存过载和恢复正常的内存阈值通过GUC参数resilience_memory_reject_percent设置，默认关闭该功能。
- 数据库线程池使用率过载时，快速kill会话并禁止新连接接入，待会话数降低到线程池可承受能力时恢复对外服务能力。线程池使用率过载和恢复正常的阈值通过GUC参数resilience_thread_reject_cond设置，默认关闭该功能。

特性增强

无。

特性约束

- 内存或者线程池过载触发逃生能力时，默认不对sysadmin或monitoradmin权限的用户的session做清理操作，若想对sysadmin或monitoradmin权限的用户的session做清理操作，请设置参数resilience_escape_user_permissions，具体请参考resilience_escape_user_permissions的详细描述和使用方法。
- 升级模式下，不触发该特性功能。

依赖关系

无。

2.7.2 并发场景支持抗过载逃生能力

可获得性

本特性自503.0.0版本开始引入。

特性简介

慢SQL导致的过载问题在现有的多个测试场景中经常出现，通常应用层在业务上需要保证对外提供的服务具备稳定可靠的SLA，每当出现慢SQL出现以后应用层会通过增加对数据库的连接请求数来确保SLA目标达成，因此对于数据库来说则是连接请求数增多导致的并发陡增问题，在现有的实现机制上由于连接数增多会导致CPU、内存等资源消耗增加，同时由于慢SQL无法执行完成导致执行slot被长时间占用，新的业务请求无法进入，最终导致业务吞吐量托底并且无有效回复手段。针对这一慢SQL入侵场景，本特性针对该场景下提升过载逃生的韧性能力，通过韧性的增强，能够在大并发场景下数据库服务端保持一定能力的稳定业务输出。

客户价值

当数据库由于慢SQL无法快速执行完成导致执行slot被长时间占用，新的业务请求无法进入，最终导致数据库无法对外提供服务时，能够提升过载逃生的韧性能力，通过韧性的增强，能够在大并发场景下数据库服务端保持一定能力的稳定业务输出。

特性描述

- 支持韧性检测能力：通过定义慢SQL的执行时间来实现慢SQL检测，一旦检测出慢SQL后则启动承受能力。
- 支持韧性承受能力：慢SQL入侵被认定以后，慢SQL在总量上受限于预先设定值，避免所有的执行slot都被慢SQL所占用，能够在整体上承受慢SQL的入侵。
- 支持韧性调整能力：慢SQL入侵被认定以后，在慢车道的管控态运行，系统对其IO资源使用加以限制和隔离降低对其他作业影响，确保系统整体KPI不全面恶化。
- 支持实时观测：慢SQL进入管控态后，可以通过视图查询慢SQL的具体运行状态。
- 支持灵活容错：对于资源充足或者偶发慢SQL需要具有一定包容性，零星异常SQL仍然有机会执行完毕。
- 支持灵活配置：对于预期执行时间基于规则设置，后续可自适应的选择合理的预期执行时间。

特性增强

无。

特性约束

- 仅对非sysadmin或monitoradmin权限的用户执行的select类型的语句进行慢SQL管控。
- 仅在线程池开启模式下生效。
- 资源管控仅在use_workload_manager=on时生效。

依赖关系

无。

2.7.3 资源管控

可获得性

本特性自503.1.0版本开始引入。

特性简介

资源管控作为后续企业级应用场景的关键诉求，在MetaERP等场景下有着重要的作用。当进程或者某一单一线程占用大量的资源时，需要资源管控能力来限制单一进程/线程使用的资源，避免出现单一进程/线程占用大量资源导致其他进程/线程资源不足的情况。

GaussDB目前已经识别的可管控的资源有：CPU、内存、I/O、最大并发数、磁盘空间等。

客户价值

补充数据库资源管理能力，针对不同用户限制不同的资源使用，避免出现某一用户占用全部资源导致其他用户申请不到资源的场景。

特性描述

- 支持进程级CPU管控和调整。
- 支持资源池粒度的CPU，用户和资源池可绑定，间接管控用户的CPU资源使用。
- 支持资源池级CPU使用的资源监控能力。
- 支持session/线程粒度的CPU管控能力。
- 支持资源池粒度的最大连接数的限制能力。
- SMP计划支持预占stream线程执行。
- 支持实例级别、资源池级别、session级别、SQL级别的内存管控和监控能力。
- 支持资源池级别、作业级别的I/O管控和监控能力。
- 支持资源池粒度的最大并发数管控能力。
- 支持进程级别的最大并发数管控能力。

特性增强

无。

特性约束

- 资源管控仅在use_workload_manager=on和enable_control_group=on时生效。
- 需要初始化control group文件系统。

依赖关系

无。

2.7.4 SQL 限流能力

可获得性

本特性自505.0.0版本开始引入。

特性简介

在数据库系统中，时常会出现某类SQL执行异常，占用较多系统资源，或者出现某类SQL因异常或业务需求并发激增，影响其他业务执行，甚至整个数据库系统无法响应其他业务请求的情况。为了解决该问题，GaussDB实现了SQL限流的能力，可以从多维度限制某类SQL执行的并发数。

客户价值

当数据库由于某类SQL并发数突增或者长时间执行，导致其他的业务请求无法执行，最终导致数据库无法对外提供服务时，通过本特性能够限制异常SQL的并发数，让正常的业务得到保障，提升系统韧性。

特性描述

本特性可以实现多种规则的SQL限流能力，限制某类SQL或实例的最大并发数，包括：

1. 根据Unique SQL ID进行限流：在明确某条SQL为慢SQL或者占用资源较高的SQL时，可以通过Unique SQL ID对该SQL进行限流，避免业务大量执行此SQL而影响其他业务；
2. 根据SQL类型及关键字进行限流：在明确某类SQL请求可能会随业务量增长而增长的时候，使用SQL类型和关键字对此类SQL进行限流；
3. 由于某些业务高峰是可以预知的，在仅希望在业务高峰时段对SQL请求进行限制的情况下，可以设置SQL限流的生效时间，避免限流规则常驻系统；
4. 当只希望限流规则作用于业务库，而对系统库的SQL不做控制，可以按不同库的维度进行限流；
5. 除了对某类SQL进行限流，本特性还提供实例级别的限流能力；
6. 对于限流规则，提供查询统计的能力，可以查询所有的限流规则，并根据规则列表对限流规则进行管理。此外，还提供查询限流规则限制访问的SQL次数。

特性增强

无

特性约束

- 当前不支持在备机创建限流规则，但主机创建的限流规则会自动同步到备机生效，待备机日志回放完成后生效。
- 对于Unique SQL ID限流，需要设置GUC参数enable_resource_track = on，instr_unique_sql_count > 0。
- 对于关键字限流，按照并发度排序，并发度越低优先级越高。关键字不区分大小写，支持模糊匹配。
- 数据库名称区分大小写。删除某个数据库，再创建同名的数据库，会导致所有已录入的针对这个数据库的限流规则失效。
- 基于资源的实例级最大活跃并发数限流，按照并发度排序，并发度越低优先级越高，资源利用率的采集存在10s左右的时间差。
- 对于限流周期结束的规则，会在下次限流规则触发时将is_valid标记为false，后续不再检查。
- 对于SQL限流次数的统计没有落盘，是数据库从启动到当前的累计次数。
- 游标、存储过程中的SQL语句不会被限流。
- 管理员用户执行的SQL语句不会被限流。

依赖关系

无

2.8 AI 能力

2.8.1 AI4DB: 数据库自治运维

2.8.1.1 数据库指标采集、预测与异常检测

可获得性

本特性自V500R002C00版本开始引入。

特性简介

本特性是GaussDB Kernel集成的、可以用于数据库指标采集、预测以及异常监控与诊断的AI工具，是DBMind套间中的一个组件。当前通过兼容Prometheus平台来采集数据库系统的指标，提供Prometheus exporter用于采集和加工数据库监控指标。通过监控指标时序数据，可以用来预测未来负载走向，诊断问题，同时还可以进行异常检测等。

客户价值

- 极大简化运维人员工作，释放大量劳动力，为公司节省成本。
- 用户可以通过指标采集、监控和预测功能提前感知问题，从而防止数据库发生意外，导致更大的损失。

特性描述

Prometheus是业内非常流行的开源监控系统，同时本身也是一款时序数据库。Prometheus的采集端称之为exporter，用来收集被监控模块的指标项。为了与Prometheus平台完成对接，DBMind分别实现了两款exporter，分别是用来采集数据库指标的openGauss-exporter，以及对采集到的指标进行二次加工的reprocessing-exporter。

本特性支持对采集到的指标进行预测，用户可通过修改配置文件来指定需要进行预测的关键系统指标（KPI），进而便于用户发现指标的走势，及时进行对应的运维操作。如预测内存使用率可以发现内存泄漏、预测磁盘使用情况可以在合适的时候扩容。基于AI的异常检测算法，可以发现指标的走势波动，进而促使用户及时地发现问题。

特性增强

在V500R002C10版本中，进行了大幅度改进，兼容Prometheus 平台，实现三个exporter 用于对接Prometheus。

特性约束

- 数据库状态正常，并且用户已将数据目录写入环境变量。
- Python 版本要求3.7及以上。
- 配置Prometheus 监控平台，并启动本服务，以便监控数据可被收集。

依赖关系

Prometheus

2.8.1.2 慢 SQL 根因分析

可获得性

本特性自V500R002C00开始引入。

特性简介

慢SQL一直是数据运维中的痛点问题，如何有效诊断慢SQL根因是当前一大难题，工具结合数据库自身特点融合了现网DBA慢SQL诊断经验，该工具可以支持慢SQL根因15+，能同时按照可能性大小输出多个根因并提供针对性的建议。

客户价值

为客户提供快速可靠的慢SQL发现及根因分析功能，极大简化了运维人员的工作。

特性描述

基于Prometheus数据采集方案，收集慢SQL根因分析需要的数据，包括系统资源信息（cpu usage、memory usage、IO）、负载信息（QPS）、大进程信息（包括外部大进程和数据库定时任务）、慢SQL文本信息、慢SQL开始执行时间和结束执行时间、慢SQL执行计划，临时文件信息等信息，而后，本功能根据AI算法计算最匹配的慢SQL根因，并给出对应的建议和置信度。

特性增强

无

特性约束

- 数据库状态正常、客户端能够正常连接。
- 具备Python3.7+的环境。
- 其中慢SQL的信息通过WDR报告获取，数据库WDR报告中会标记SQL是否是慢SQL，其相关GUC参数track_stmt_stat_level默认打开，否则需要用户手动打开，一般设置为track_stmt_stat_level=’off, L0’，更高级别对性能会有一定的影响。数据采集部分由Prometheus方案实现，故需要用户配置Prometheus数据采集平台，本功能只专注于算法并从Prometheus中获取指标的序列信息。

依赖关系

无

2.8.1.3 索引推荐

可获得性

本特性自V500R002C00开始引入。

特性简介

本功能是一个覆盖多种任务级别和使用场景的数据库智能索引推荐工具，其具备单Query索引推荐功能、虚拟索引功能、workload级别索引推荐功能，可以为用户提供可靠的索引建议。

客户价值

为客户提供快速可靠的索引推荐功能，极大简化了运维人员的工作。

特性描述

单query索引推荐功能支持用户在数据库中直接进行操作，本功能基于查询语句的语义信息和数据库的统计信息，对用户输入的单条查询语句生成推荐的索引；虚拟索引功能支持用户在数据库中直接进行操作，本功能将模拟真实索引的建立，避免真实索引创建所需的时间和空间开销，用户基于虚拟索引，可通过优化器评估该索引对指定查询语句的代价影响；对于workload级别的索引推荐，用户可通过运行数据库外的脚本使用此功能，本功能将包含有多条DML语句的workload作为输入，最终生成一批可对整体workload的执行表现进行优化的索引。

特性增强

无

特性约束

- 数据库状态正常、客户端能够正常连接。
- 当前执行用户下安装有gsql工具，该工具路径已被加入到PATH环境变量中。
- 具备Python3.7+的环境。

依赖关系

无

2.8.1.4 参数调优与诊断

可获得性

本特性自V500R002C00版本开始引入

特性简介

本功能是一款数据库集成的参数调优工具，通过结合深度强化学习和全局搜索算法等AI技术，实现在无需人工干预的情况下，获取最佳数据库参数配置。本功能不强制与数据库环境部署到一起，支持独立部署，脱离数据库安装环境独立运行。

客户价值

该工具可以在任意场景下，快速给出当前负载的调参配置，减少DBA的人工干预，提升运维效果，满足客户期望。

特性描述

调优程序包含三种运行模式，分别是:

- recommend: 通过用户指定的用户名等信息登录到数据库环境中，获取当前正在运行的workload特征信息，根据上述特征信息生成参数推荐报告。报告当前数据库中不合理的参数配置和潜在风险等；输出根据当前正在运行的workload行为和特征；输出推荐的参数配置。该模式是秒级的，不涉及数据库的重启操作，其他模式可能需要反复重启数据库。
- train: 通过用户提供的benchmark信息，不断地进行参数修改和benchmark的执行。通过反复的迭代过程，训练强化学习模型，以便用户在后面通过tune模式加载该模型进行调优。
- tune: 使用优化算法进行数据库参数的调优，当前支持两大类算法，一种是深度强化学习，另一种是全局搜索算法（全局优化算法）。深度强化学习模式要求先运行train模式，生成训练后的调优模型，而使用全局搜索算法则不需要提前进行训练，可以直接进行搜索调优。

特性增强

无

特性约束

- 数据库状态正常、客户端能够正常连接、且要求数据库内导入数据，以便调优程序可以执行benchmark测试调优效果。
- 使用本工具需要指定登录到数据库的用户身份，要求该登录到数据库上的用户具有足够的权限，以便可以获得充足的数据库状态信息。
- 使用登录到数据库宿主机上的Linux用户，需要将$**GAUSSHOME/bin**添加到PATH环境变量中，即能够直接运行gsql、gs_guc、gs_ctl等数据库运维工具。
- Python版本建议为Python3.7及以上，且运行环境中已经安装相应依赖，并能够正常启动调优程序。您可以独立安装一个python3.6+的环境，无需设置到全局环境变量中。不建议使用root用户权限安装本工具，如果以root身份安装完本工具，使用其他用户身份运行本工具时，需要确保配置文件有读取权限。
- 本工具支持以三种模式运行，其中tune和train模式要求用户配置好benchmark运行环境，并导入数据，本工具将会通过迭代运行benchmark来判断修改后的参数是否有性能提升。
- recommend模式建议在数据库正在执行workload的过程中执行，以便获得更准确的实时workload信息。
- 本工具默认带有TPC-C、TPC-H、TPC-DS以及sysbench的benchmark运行脚本样例，如果用户使用上述benchmark对数据库系统进行压力测试，则可以对上述配置文件进行适度修改或配置。如果需要适配用户自己的业务场景，需要您参照benchmark目录中的template.py文件编写驱动您自定义benchmark的脚本文件。

依赖关系

无

2.8.1.5 慢 SQL 发现

可获得性

本特性自V500R002C00版本开始引入。

特性简介

本功能是一个SQL语句执行时间预测工具，通过模板化方法，实现在不获取SQL语句执行计划的前提下，依据语句逻辑相似度与历史执行记录，预测SQL语句的执行时间。

客户价值

- 工具不需要用户提供SQL执行计划，对数据库性能不会有任何影响。
- 不同于业内其他算法只局限于OLTP或其他场景，本工具场景更加广泛。

特性描述

SQLdiag着眼于数据库的历史SQL语句，通过对历史SQL语句的执行表现进行总结归纳，将之再用于推断新的未知业务上。由于短时间内数据库SQL语句执行时长不会有太大的差距，SQLdiag可以从历史数据中检测出与已执行SQL语句相似的语句结果集，并基于SQL向量化技术和模板化方法预测SQL语句执行时长。

特性增强

无

特性约束

- 需要保证用户提供的历史日志及待预测负载的格式符合要求，可以使用数据库GUC参数开启收集，也可以通过监控工具采集。
- 为保证预测准确率，用户提供的历史语句日志应尽可能全面并具有代表性。
- 按照要求配置python环境。

依赖关系

无

2.8.2 DB4AI: 数据库驱动 AI

可获得性

本特性自V500R002C00版本开始引入。

特性简介

DB4AI是指利用数据库的能力驱动AI任务，实现数据存储、技术栈的同构。通过在数据库内集成AI算法，令GaussDB具备数据库原生AI计算引擎、模型管理、AI算子、AI原生执行计划的能力，为用户提供普惠AI技术。不同于传统的AI建模流程，DB4AI“一站式”建模可以解决数据在各平台的反复流转问题，同时简化开发流程，并可通过数据库规划出最优执行路径，让开发者更专注于具体业务和模型的调优上，具备同类产品不具备的易用性与性能优势。

客户价值

- 通过本功能，用户无需手动编写AI模型代码，直接通过开箱即用的SQL语句即可执行机器学习模型的训练和预测，学习和使用成本极低；
- 避免数据碎片化存储和反复搬迁导致的额外开销；
- 更高的执行效率，本功能的AI模型训练效率极高，相比用户自行手动训练模型有数倍性能收益；
- 更严密的安全防护，从而避免训练AI模型导致数据泄露。

特性描述

数据库的原生DB4AI能力，通过引入原生AI算子，简化操作流程，充分利用数据库优化器、执行器的优化与执行能力，获得高性能的数据库内模型训练能力。更简化的模型训练与预测流程、更高的性能表现，让开发者在更短时间内能更专注于模型的调优与数据分析上，而避免了碎片化的技术栈与冗余的代码实现。

特性增强

V500R002C10版本中支持更多算法。

特性约束

数据库状态正常

依赖关系

无

2.8.3 ABO 优化器

2.8.3.1 智能基数估计

可获得性

本特性自503.0.0版本开始引入。

特性简介

智能基数估计利用库内轻量级算法进行多列数据分布建模，并且提供多列等值基数估计的能力。在数据分布倾斜并且列之间相关性强的数据场景下能够提供更准确的估计结果，从而给优化器提供准确的代价参考，提高计划生成准确率，提高数据库查询执行效率。

客户价值

用户可以通过创建智能统计信息改善多列统计的准确率，从而提升查询优化性能。

特性描述

智能基数估计首先利用数据库内数据样本进行数据分布建模，并且将模型压缩存储在数据库中。优化器在执行计划生成阶段触发智能估计，实现对代价更精确的估计，并且生成更优的计划。

特性增强

本版本增加了包含等值和范围条件复合查询的支持。

特性约束

- 数据库运行状态良好，无资源紧张状况。
- 仅支持FLOAT8、Double Precision、FlOAT4、REAL、INT16、BIGINT、INTEGER、VARCHAR、CHARACTER VARYING、CHAR、CHARACTER、NUMERIC、TIMESTAMP以及TIMESTAMP WITH TIMEZONE数据类型。
- 本特性支持不超过64列的查询基数估计，由于同时受到ANALYZE语法对于列数的约束，因此当前仅支持32列基数估计模型创建。
- 为了保证系统性能，模型创建只利用一定量的数据样本（最多1024*1024），如果数据过于稀疏，估计结果可能不准确。
- 为了能够充分利用有限的内存进行模型访问加速，建议创建AI统计列数量不超过100个，否则可能会触发内存替换，可以通过参数进行调整。
- 如果出现过长的变长字符串类型数据，可能会影响基数估计模型创建和估计的性能。

依赖关系

依赖于数据库内的多列统计信息创建语法和数据采样算法。

2.8.3.2 自适应计划选择

可获得性

本特性自503.0.0版本开始引入。

特性简介

本特性通过触发基于基表条件选择率的计划选择，以及对于使用了部分索引和offset的查询提供缓存多计划管理和自适应选择。典型场景下能够提升数倍查询吞吐。

客户价值

通过本功能，用户可以通过维护多个缓存计划实现适应不同的查询参数，从而提升查询执行性能。

特性描述

自适应计划选择作用于使用通用缓存计划进行计划执行的场景。通过使用范围线性扩张进行缓存计划探索，通过范围覆盖匹配进行计划选择。自适应计划选择弥补了传统单一缓存计划无法根据查询条件参数进行变化带来的性能问题，并且避免了频繁调用查询优化。

特性增强

本版本特性增加了在GPC开启的场景下对于gplan/aplan/cplan切换的能力。

特性约束

- 数据库运行正常。
- 用户成功登录数据库。
- 用户创建数据库，数据表并导入数据。

依赖关系

依赖于数据库内的计划缓存功能。

2.8.3.3 自适应代价估计

可获得性

本特性自505.1.0版本开始引入。

特性简介

自适应代价估计功能基于均匀混合模型（UMM），以及代价参数模型，提供代价估计的能力。利用负载监控线程监控模型准确度，实现快速高效的负载管理和模型增量更新，保证估计准确率；利用实时高效的查询谓词特征识别最优的基数估计策略；用于解决现网场景中数据和执行环境变化场景下，代价估计失真从而导致计划不优的问题。

客户价值

通过本功能可以解决如下问题：

- 现网中数据复杂多变，静态统计信息无法支撑准确代价评估和执行计划质量问题。
- 当前的基数估计方法无法支撑连接算子选择和连接顺序选择问题。
- 在业务程序运行时，复杂多变的负载会引起执行计划不准确的问题。从而覆盖更多客户业务场景，发挥更大价值。

特性描述

本特性主要是利用真实查询反馈来纠正已有的代价模型和基数估计模型。

1. 反馈基数估计：利用UMM模型和哈希匹配建模历史负载和选择率之间的映射关系，用于解决join算子和scan算子由于统计偏差导致的基数估计误差大问题，解决在此场景下查询计划质量提升，降低查询时延。
2. 反馈代价矫正：利用查询负载算子反馈及时获得对于算子执行时间感知，发现误差较大时，针对代价参数进行拟合调整，从而获得更准确的代价评估，针对由于代价模型不合理导致的慢SQL，提高查询效率。

特性增强

无

特性约束

- 反馈基数估计
  a. 目前只支持基表扫描算子（SeqScan、IndexScan、IndexOnlyScan、BitmapHeapScan）和连接算子（NestLoop、MergeJoin、HashJoin）以及Sort、Hash、Material算子的基数估计（支持范围不包括针对分区表的操作算子，当前功能不支持分区表），且不能含有不支持的约束条件。目前支持的约束条件包括：
     - 简单的连接条件（col1::type =/<>/>/< col2::type，支持类型转换）。
     - 数值范围条件（col1::type =/>/< const，BETWEEN处理为大于和小于两个条件，const类型支持OID、NUMERIC、INT、FLOAT、TIMESTAMP）。
     - 简单的字符串比较条件（col1::type =/<>/LIKE/NOT LIKE const，const类型支持BPCHAR，VARCHAR，TEXT）。

     在优化器做基数估计时，若一个算子及其子算子的类型和条件在支持范围内，则会使用反馈模型做基数估计；否则会使用默认方法。但即使上层算子不在支持范围内，其子算子仍可能使用反馈模型做基数估计 。
  b. 收集到新SQL语句信息后会对后续执行计划产生可能劣化的跳变，只有收集到足够跳变计划信息的优化器才可以逐渐收敛生成最优/较优计划。迭代次数的上限是查询的可能路径数量，实际迭代次数远小于此上限。在JOB数据集的一百多条查询中，95%的查询能够在2轮迭代以内收敛到稳定计划，剩余查询也在最多7轮迭代后收敛。
  c. 依赖参数log_min_duration_statement >= 0生效，且不收集调用存储过程或者函数内部执行语句的查询反馈信息。
- 反馈代价矫正

  代价矫正的结果只用于SeqScan、IndexScan、IndexOnlyScan、BitmapHeapScan、HashJoin、NestLoop和MergeJoin算子的代价估计。在优化器做代价估计时，若一个算子及其子算子的类型和条件在反馈基数支持范围内，则会使用反馈模型的基数和矫正代价计算代价；否则会使用默认方法。

依赖关系

依赖于数据库内cplan/aplan计划生成发挥作用。

