Vastbase G100 V2.2

(Build 5)

技术白皮书

北京海量数据技术股份有限公司

【版权声明】

©2007-2022 北京海量数据技术股份有限公司 版权所有

本文档著作权归 **北京海量数据技术股份有限公司**（简称“海量数据”）所有，未经海量数据事先书面许可，任何主体不得以任何形式复制、 修改、抄袭、传播全部或部分本文档内容。

北京海量数据技术股份有限公司保留所有的权利。

【商标声明】

海量数据 VASTDATA 及其它海量数据产品和服务相关的商标均为 **北京海量数据技术股份有限公司** 及其关联公司所有。

本文档涉及的第三方主体的商标，依法由权利人所有。

【服务声明】

本文档意在向客户介绍海量数据全部或部分产品、服务的当时的整体概况，部分产品、服务的内容可能有所调整。您所购买的产品、服务的种类、服务标准等应由您与海量数据之间的商业合同约定，除非双方另有约定，否则，海量数据对本文档内容不做任何明示或模式的承诺或保证。

目录

1. 产品介绍

1.1. 产品定位

Vastbase G100 支持 SQL2003 标准语法，支持主备部署的高可用关系型数据库。

- 多种存储模式支持复合业务场景。
- NUMA 化内核结构支持高性能。
- 主备模式，CRC 校验支持高可用。

1.2. 应用场景

- 交易型应用

  大并发、大数据量、以联机事务处理为主的交易型应用，如电商、金融、O2O、电信 CRM/计费等，应用可按需选择不同的主备部署模式。

- 物联网数据

  在工业监控和远程控制、智慧城市的延展、智能家居、车联网等物联网场景下，传感监控设备多，采样率高，数据存储为追加模型，操作和分析并重的场景。

1.3. 技术特点

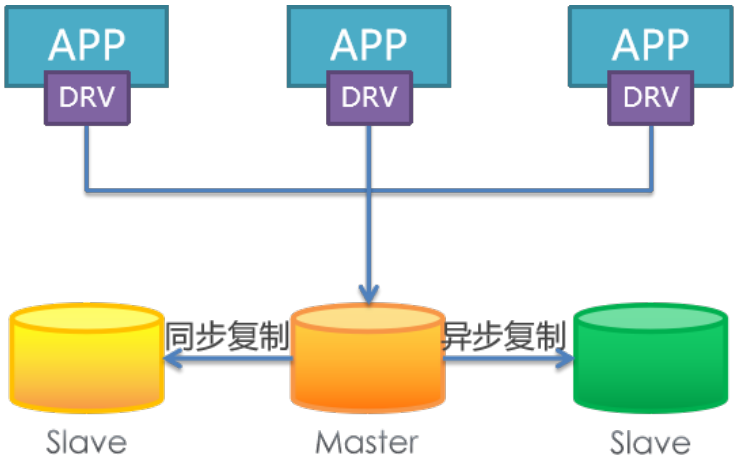
Vastbase G100 相比其他数据库主要有多存储模式，NUMA 化内核结构和高可用等产品特点。

- 多存储模式
  - 行存储，支持业务数据频繁更新场景。
  - 列存储，支持业务数据追加和分析场景。
  - 内存表，支持高吞吐，低时延，极高性能场景。
- NUMA 化内核结构
  - 关键数据结构分区，降低数据访问冲突。
  - 关键数据结构 NUMA 化，降低数据结构访问时延。
  - 关键业务线程绑核，避免核间线程漂移。
- 高可用
  - 支持主备同步，异步多种部署模式。
  - 数据页 CRC 校验，损坏数据页通过备机自动修复。
  - 备机并行恢复，10 秒内可升主提供服务。

1.4. 软件架构

Vastbase G100 是集中式数据库系统，在这样的系统架构中，业务数据存储在单个物理节点上，数据访问任务被推送到服务节点执行，通过服务器的高并发，实现对数据处理的快速响应。同时通过日志复制可以把数据复制到备机，提供数据的高可靠和读扩展。

Vastbase G100 逻辑架构如 0 所示。



2. 部署方案

Vastbase G100 主要支持单机部署和一主多备部署两种部署形态

2.1. 常用概念

- 单机

  单机指的是只有一个数据库实例。

- 双机

  双机指的是系统中存在主备数据库实例，主实例支持读写，备实例支持只读。

- 一主多备

  一主多备指的是在系统存在一个主机，多个备机。最多支持 8 个备机。

- 冷热备份

  冷备份： 是指备份就是一个简单的备份集，不可以提供服务。

  热备份： 是指备份集群可以对外提供服务 。

2.2. 部署形态汇总

单机和双机两种部署形态方案介绍请见表 1-1。

表1-1 Vastbase G100 部署形态汇总表

部署形态	技术方案	高可用	基础设置要求	场景特点	技术规格
单机	单机	无高可用能力	单机房	•对系统的可靠性和可用性无任何要求 •主要用于体验试用以及调测场景	•系统 RTO 和 RPO 不可控 •无实例级容灾能力，一旦出现实例故障，系统不可用 •一旦实例级数据丢失，则数据永久丢失，无法恢复
主备	主机+备机	抵御实例级故障	单机房	•节点间无网络延迟 •要求承受集群	•RPO=0 •实例故障 RTO<10s •无 AZ 级容灾能力

部署形态	技术方案	高可用	基础设置要求	场景特点	技术规格
				内实例级故障 •适用于对系统可靠性要求不高的场景	•推荐主备最大可用模式
一主多备	主机 +多个备机 Quorum	抵御实例级故障	单机房/跨机房	•节点间无网络延迟 •要求承受集群内实例级故障	•RPO=0 •实例故障 RTO<10s •推荐主备同步模式 •最少 2 个副本，最多 8 个副本

兼容列表

支持的 CPU 范围：
鲲鹏、飞腾、至强、申威、龙芯、兆芯、海光

支持的 OS 范围：
openEuler、麒麟 V10、UOS(openEuler 版)、CentOS 7.x/8.x(ARM 版本)

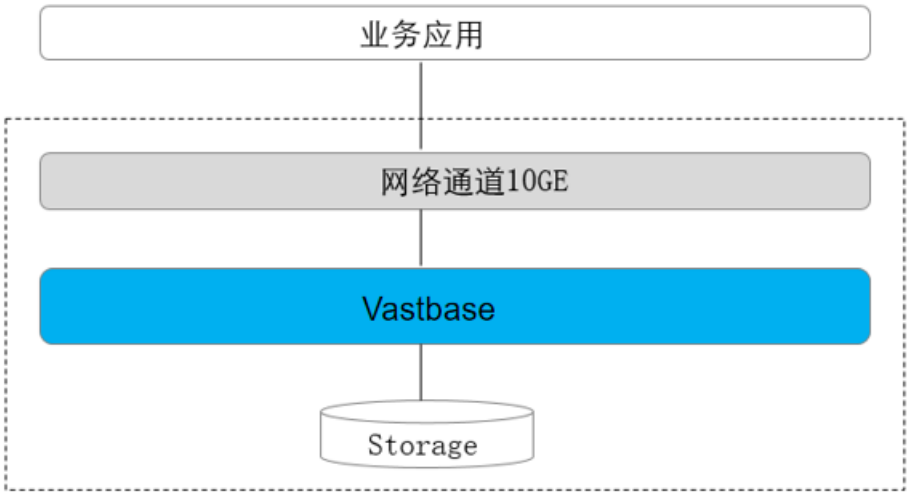
2.3. 部署方案介绍

整体部署方案可以分为三类：单机部署、一主一备部署、一主多备部署。具体在后续每个小节分解。

2.3.1. 单机部署

单机部署形态是一种非常特殊的部署形态，这种形态对于可靠性、可用性均无任何保证。由于只有一个数据副本，一旦发生数据损坏、丢失，只能通过物理备份恢复数据。这种部署形态，一般用于数据库体验用户，以及测试环境做语法功能调测等场景。不建议用于商业现网运行。

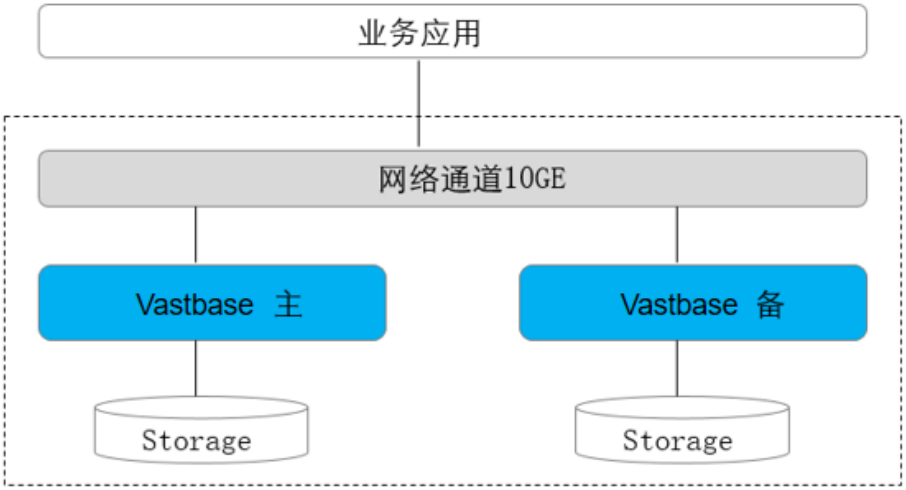
图1-2 单机部署形态图



2.3.2. 主备部署

主备模式相当于两个数据副本，主机和备机各一个数据副本，备机接受日志、做日志回放；

图1-3 主备部署形态图



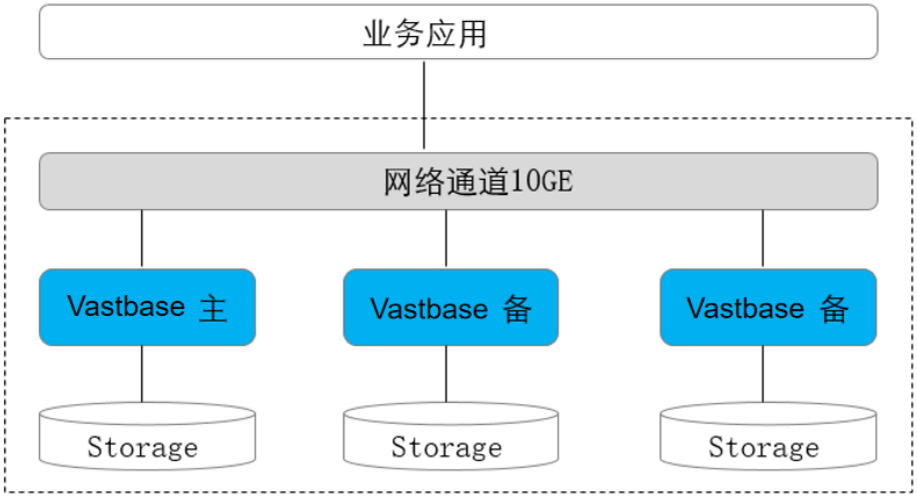
2.3.3. 一主多备部署

多副本的部署形态，提供了抵御实例级故障的能力。适用于需要抵御个别硬件故障的应用场景和要求机房级别容灾场景。

一般多副本部署时使用 1 主 2 备模式，总共 3 个副本，3 个副本的可靠性为 4 个 9，可以满足大多数应用的可靠性要求。

- 主备间 Quorum 复制，至少同步到一台备机，保证最大性能。
- 主备任意一个节点故障，不影响业务的进行。
- 数据有三份，任何一个节点故障，系统仍然有双份数据确保继续运行。任何一个备份都可以升主。
- 主备实例之间不可部署在同一台物理机上。

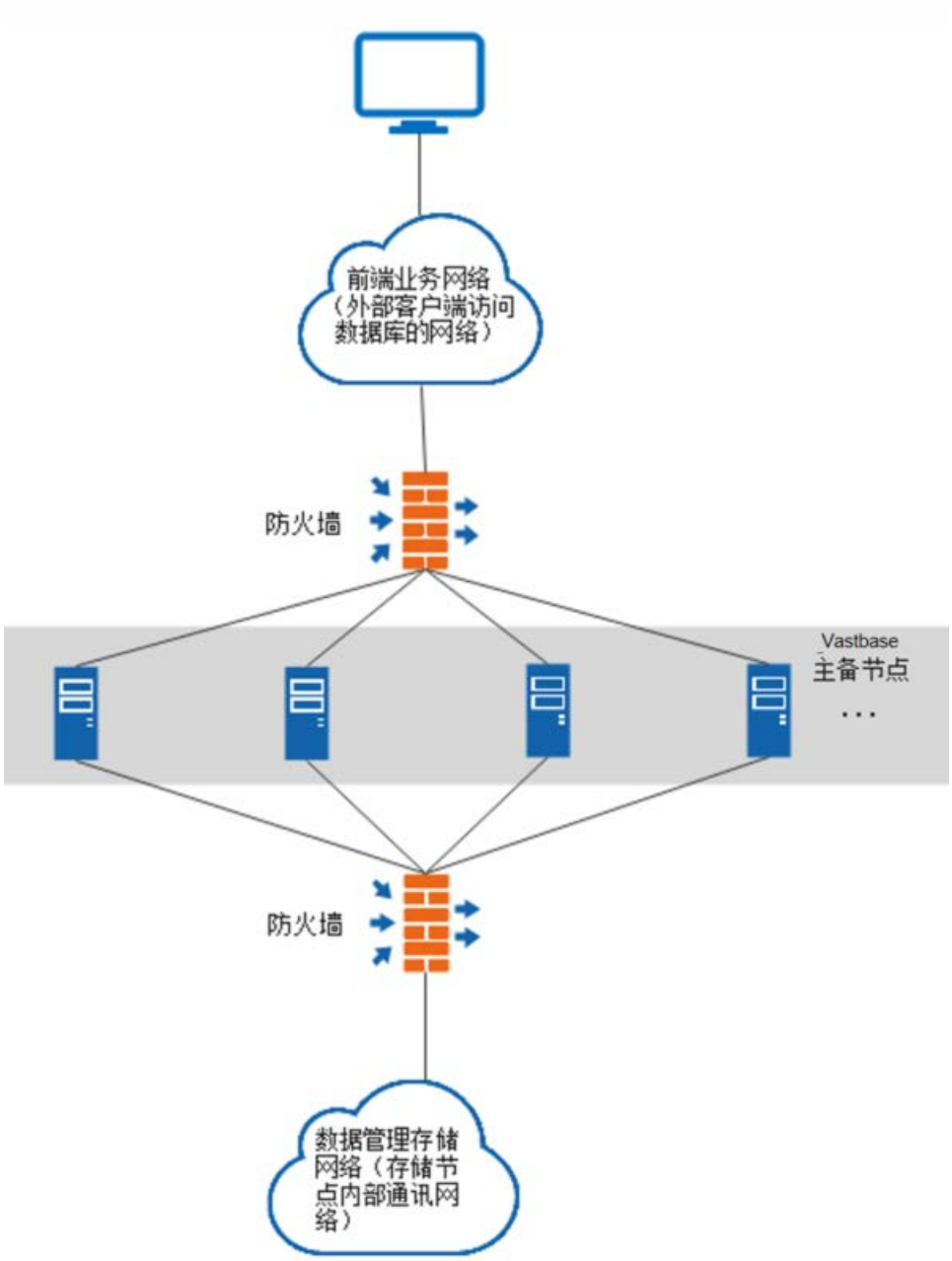
图1-4 一主多备部署形态图



3. 典型组网

为了保证整个应用数据的安全性，建议将 Vastbase G100 的典型组网划分为两个独立网络：前端业务网络和数据管理存储网络。

图1-5 典型组网



网络划分说明如表 1-2 所示。

表 1-2 网络划分

类型	描述
数据库管理存储网络	DBA 通过此网络调用 OM 脚本管理和维护 Vastbase G100 实例。同时，用于 Vastbase G100 主备通信组网。数据库管理存储网络也是应用执行系统监控的网络。
前端业务网络	外部客户端通过此网络访问 Vastbase G100 数据库。

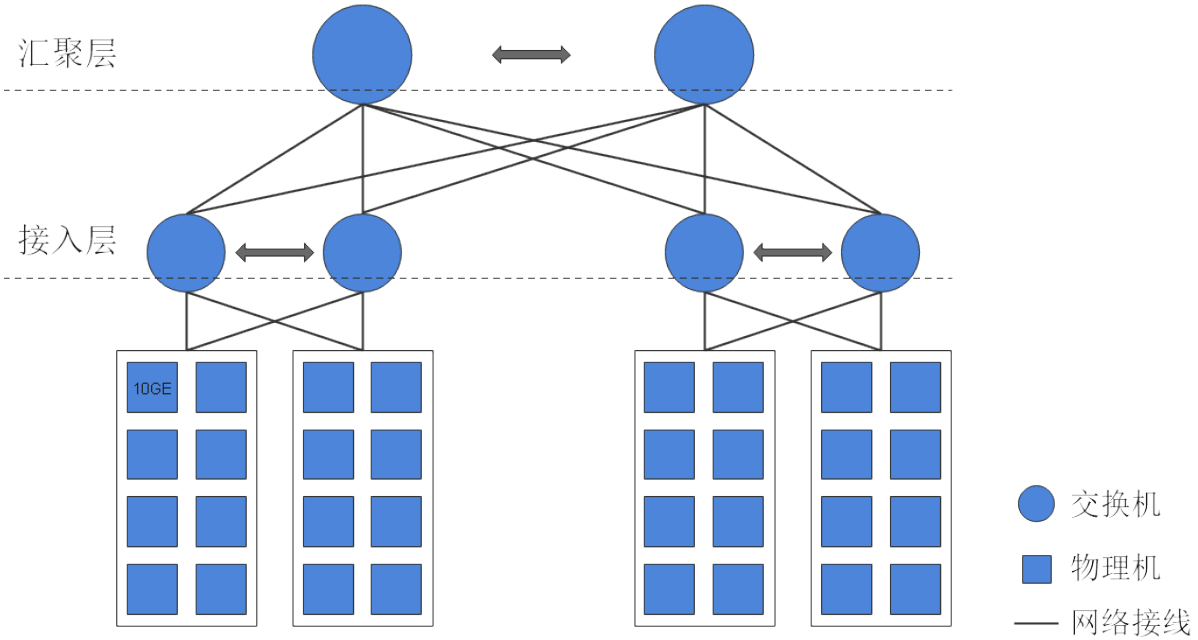
该典型组网有如下优点：

- 业务网络与数据库管理存储网络的隔离，有效保护了后端存储数据的安全。
- 业务网络和数据库管理存储网络的隔离，可以防止攻击者通过互联网试图对数据库服务器进行管理操作，增加了系统安全性。

网络独占性及 1:1 的带宽收敛比是 Vastbase G100 数据库网络性能的基本要求。因此，在生产系统中，对图 1-6 的后端存储网络，需满足独占性及至少 1:1 收敛比的要求。例如，图 1-6 中，其本质是 Fattree 组网方式。为实现收敛比 1：1，交换网络层级每提高一层，带宽增加一倍。图中每根加粗连接线代表 80GE 带宽，即 8 台物理机带宽上限之和。接入层每单台交换机下行带宽 160GE，上行带宽 160GE，收敛比 1:1；汇聚层每单台交换机接入带宽 320GE。

对于测试系统，上述要求可以适当降低。

图1-6 数据库管理存储网络组网示例



4. 数据库核心技术

4.1. 面向应用开发的基本功能

4.1.1. 支持标准 SQL

Vastbase G100 数据库支持标准的 SQL。Vastbase G100 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。

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 Vastbase G100 程序的迁移代价。

4.1.2. 支持标准开发接口

提供业界标准的 ODBC 及 JDBC 接口，保证用户业务快速迁移至 Vastbase G100。

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

4.1.3. 事务支持

事务支持指的就是系统提供事务的能力，支持全局事务的 ACID，保证事务的原子性、一致性、隔离性和持久性。。

事务支持及数据一致性保证是绝大多数数据库的基本功能，只有支持了事务，才能满足事务化的应用需求。

- 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。

- 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。

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：

- 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。详细语法参见《Vastbase G100 开发者指南》。

4.1.4. 函数及存储过程支持

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。

存储过程是 SQL、PL/SQL 的组合。存储过程可以使执行商业规则的代码从应用程序中移动到数据库。从而，代码存储一次能够被多个程序使用。

- 允许客户模块化程序设计，对 SQL 语句集进行封装，调用方便。
- 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。
- 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。
- 为了处理 SQL 语句，存储过程进程分配一段内存区域来保存上下文联系。游标是指向上下文区域的句柄或指针。借助游标，存储过程可以控制上下文区域的变化。
- 支持 6 种异常信息级别方便客户对存储过程进行调试。

Vastbase G100 支持 SQL 标准中的函数及存储过程增强了存储过程的易用性。

存储过程具体的使用方式可以参考《Vastbase G100 开发者指南》。

4.1.5. PG 接口兼容

兼容 PSQL 客户端，兼容 PostgreSQL 标准接口。

4.1.6. 支持 SQL hint

支持 SQL hint 影响执行计划生成、提升 SQL 查询性能。

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，等多个手段来进行执行计划的调优，以提升查询的性能。

4.1.7. Copy 接口支持容错机制

Vastbase G100 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。

4.2. 高性能

4.2.1. CBO 优化器

Vastbase G100 优化器是典型的基于代价的优化（Cost-Based Optimization，简称 CBO)。在这种优化器模型下，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。

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。

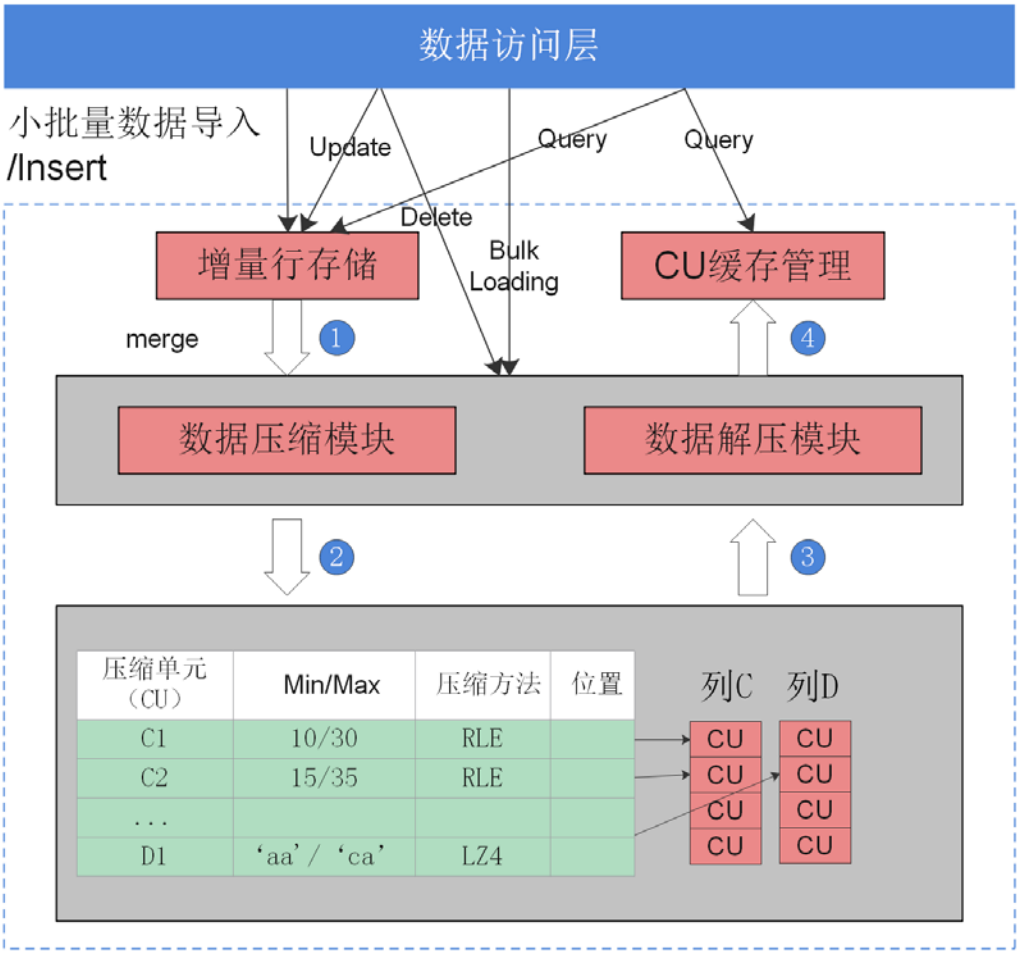
4.2.2. 行列混合存储

Vastbase G100 支持行存储和列存储两种存储模型，用户可以根据应用场景，建表的时候选择行存储还是列存储。

一般情况下，如果表的字段比较多（大宽表），查询中涉及到的列不很多的情况下，适合列存储。如果表的字段个数比较少，查询大部分字段，那么选择行存储比较好。

列存储方式如图 1-7 所示。

图1-7 列存储示意图



在大宽表、数据量比较大的场景中，查询经常关注某些列，行存储引擎查询性能比较差。例如气象局的场景，单表有 200~800 个列，查询经常访问 10 个列，在类似这样的场景下，向量化执行技术和列存储引擎可以极大的提升性能和减少存储空间。

行存表和列存表各有优劣，建议根据实际情况选择。

- 行存表

  默认创建表的类型。数据按行进行存储，即一行数据紧挨着存储。行存表支持完整的增删改查。适用于对数据需要经常更新的场景。

- 列存表

  数据按列进行存储，即一列所有数据紧挨着存储。单列查询 IO 小，比行存表占用更少的存储空间。适合数据批量插入、更新较少和以查询为主统计分析类的场景。列存表不适合点查询，insert 插入单条记录性能差。

行存表和列存表的选择原则如下：

- 更新频繁程度

数据如果频繁更新，选择行存表。

- 插入频繁程度

  频繁的少量插入，选择行存表。一次插入大批量数据，选择列存表。

- 表的列数

  表的列数很多，选择列存表。

- 查询的列数

  如果每次查询时，只涉及了表的少数（<50%总列数）几个列，选择列存表。

- 压缩率

  列存表比行存表压缩率高。但高压缩率会消耗更多的 CPU 资源。

4.2.3. 自适应压缩

当前主流数据库通常都会采用数据压缩技术。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。自适用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。

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。

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	√	√	√	√	√	√

具体场景如：支持类手机号字符串的大整数压缩、支持 numeric 类型的大整数压缩、支持对压缩算法进行不同压缩水平的调整。

4.2.4. 分区

在 Vastbase G100 系统中，数据分区是在一个实例内部对数据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。

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。
- 方便维护：如果分区表的某个分区出现故障，需要修复数据，只修复该分区即可。
- 均衡 I/O：可以把不同的分区映射到不同的磁盘以平衡 I/O，改善整个系统性能。

目前 Vastbase G100 数据库支持的分区表为范围分区表：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。

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。

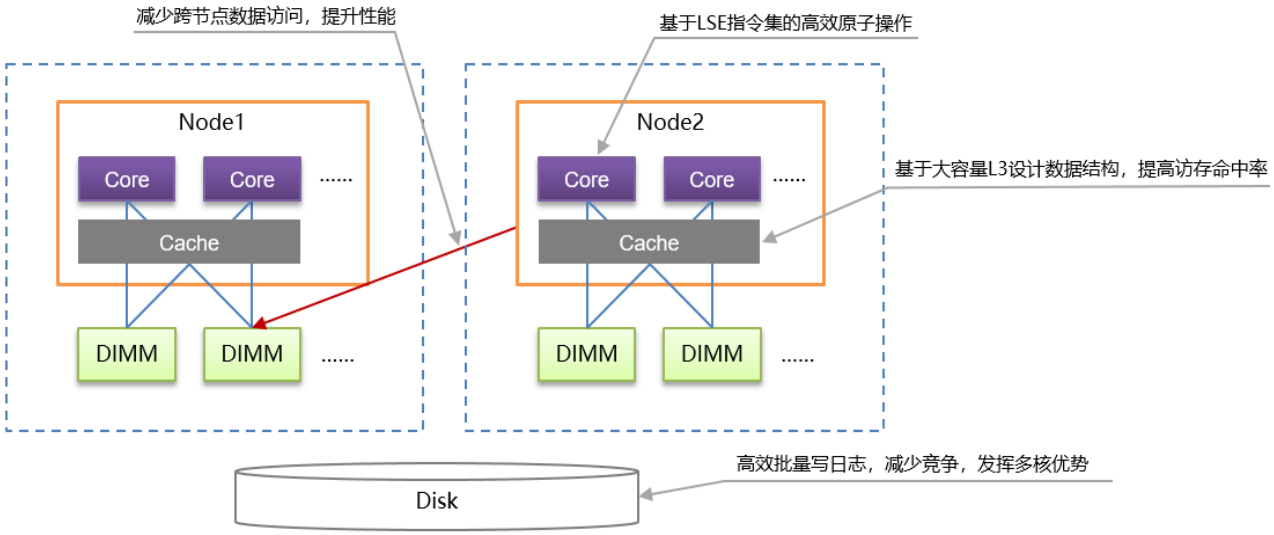
实际业务中，时间经常被作为查询对象的过滤条件。因此，用户可考虑选择时间列为分区键，键值范围可根据总数据量、一次查询数据量调整。

4.2.5. SQL by pass

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，直接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。

4.2.6. 鲲鹏 NUMA 架构优化

图1-8 鲲鹏 NUMA 架构优化图



- ❖ Vastbase G100 根据鲲鹏处理器的多核 NUMA 架构特点，进行针对性一系列 NUMA 架构相关优化，一方面尽量减少跨核内存访问的时延问题，另一方面重分发挥鲲鹏多核算力优势，所提供的关键技术包括重做日志批插，热点数据 NUMA 分布，CLog 分区等，大幅提升 TP 系统的处理性能。
- ❖ Vastbase G100 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。
- ❖ Vastbase G100 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。

4.3. 高并发

在 OLTP 领域中，数据库需要处理大量的客户端连接。因此，高并发场景的处理能力是数据库的重要能力之一。

对于外部连接最简单的处理模式是 per-thread-per-connection 模式，即来一个用户连接产生一个线程。这个模式好处是架构上处理简单，但是高并发下，由于线程太多，线程切换和数据库轻量级锁区域的冲突过大导致性能急剧下降，使得系统性能（吞吐量）严重下降，无法满足用户性能的 SLA。

因此，需要通过线程资源池化复用的技术来解决该问题。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。

4.4. Oracle 兼容性说明

4.4.1. 内置函数

Vastbase G100 兼容的 Oracle 内置函数如下：

内置函数	说明
ADD_MONTHS(DATE,INTEGER)	返回 date 加上 integer 的值，返回值为 DATE 类型。
BITAND(NUMBER1,NUMBER2)	返回 number1 和 number2 按位与的值，返回值为 BIT 类型。
CONVERT(CHAR, DEST_CHAR_SET[, SOURCE_CHAR_SET])	将 source_char_set 字符集编码格式的输入字符串 char，转换为 dest_char_set 字符集编码格式，该函数只在服务端生效。
COSH(N)	返回参数 n 的双曲余弦值；
DECODE(EXPR,SEARCH1, RESULT1[[,SEARCH2, RESULT2],.....][, DEFAULT])	将 expr 表达式值依次与所有 search 表达式（search1，search2.....）进行比较，如果与 searchn 进行匹配，则返回 resultn，否则返回 default。如果没有 default，返回 null。
DUMP(TEXT)	
EMPTY_BLOB()	返回一个空 BLOB 类型。
EMPTY_CLOB()	返回一个空 CLOB 类型。
INSTR(源字符串，目标字符串，起始位置，匹配序号)	string2 的值要在 string1 中查找，是从 start_position 给出的数值（即：位置）开始在 string1 检索，检索第 nth_appearance（几）次出现 string2。
LAST_DAY(DATE)	返回 date 所在月的最后一天。
LENGTHB(CHAR)	返回 char 字符的字节长度，char 支持的类型为所有字符串类型（如 CHAR, VARCHAR2, NCHAR, NVARCHAR2 等）或可隐式转换为字符串的类型（如 integer 等）。
LISTAGG(MEASURE_EXPR[, 'DELIMITER'])	
LNNVL(CONDITION)	返回条件表达式 condition 的取反值，返回类型为 BOOLEAN。
LPAD(STRING, PADDED_LENGTH, [PAD_STRING])	lpad 函数从左边对字符串使用指定的字符进行填充.
LTRIM(C1,[,C2])	删除左边出现的字符串
MOD(N2, N1)	支持除数为 0 的情况，即当 n1 为 0 时，直接返回 n2。
MONTHS_BETWEEN(DATE1, DATE2)	返回 date1 与 date2 之间相差的月数，当 date2 早于 date1，返回值为负数。
NANVL(N2, N1)	当单精度或双精度浮点数 n2 的值为 NAN，则返回 n1，否则返回 n2。
NEW_TIME(DATE, TIMEZONE1, TIMEZONE2)	返回 timezone1 的 date 时间对应的 timezone2 的时间，返回值类型为 DATE。
NEXT_DAY(DATE, CHAR)	返回 date 后的第一个自然日（包括工作日、休息日和节假日），返回值为 DATE 类型。
NUMTODSINTERVAL(N, INTERVAL_UNIT)	根据 INTERVAL 数据类型单元 interval_unit（取值范围：'DAY'、'HOUR'、'MINUTE'、'SECOND'），将数字 n 转换为数据类型 INTERVAL DAY TO SECOND。

NLSSORT(CHAR[, NLSPARAM])	按照 nlsparam 指定的排序字符集对字符串 char 进行排序，默认使用 char 字符串字符集排序；
NVL(EXPR1, EXPR2)	返回 expr1, expr2 中第一个非空值。
NVL2(E1,E2,E3)	如果 e1 为 NULL，则函数返回 e3，若 e1 不为 null，则返回 e2。
REGEXP_COUNT(SOURCE_CHAR, PATTERN, POSITION, MATCH_PARAM)	返回指定字符串 source_char 中的指定位置 position 开始，匹配正则表达式模式 pattern 的次数。其中，match_param 参数会影响正则表达式匹配规则，比如 match_param='i'忽略大小写等。
REGEXP_INSTR(SOURCE_CHAR, PATTERN[, POSITION[, OCCURRENCE[, RETURN_OPT[, MATCH_PARAM[, SUBEXPR]]]]])	该函数拓展了 INSTR 函数的功能，允许使用正则表达式匹配，返回值类型为 INTEGER。 position 表示查找起始位置。 occurrence 表示查找 pattern 在 source_char 的第几次出现。 return_opt： 取值为 0 表示返回模式匹配的起始位置。 取值为 1 表示返回模式匹配的结束位置。 match_param 表示正则表达式模式匹配控制参数，如区分大小写等。 subexpr 表示正则表达式分组匹配的组号。
REGEXP_REPLACE(SOURCE_CHAR, PATTERN [, REPLACE_STRING [, POSITION [, OCCURRENCE [, MATCH_PARAMETER]]]])	搜索的字符串的正则表达式模式 REPLACE 函数的功能。
REGEXP_SUBSTR(SOURCE_CHAR, PATTERN[,POSITION[,OCCURRENCE[,MATCH_PARAM[,SUBEXPR]]]])	按正则表达式在 source_char 字符串中匹配子字符串。 source_char 为查找的输入字符串，支持所有字符串类型（如 CHAR, VARCHAR2, NCHAR, NVARCHAR2 等）或可隐式转换为字符串的类型（如 integer 等）。 pattern 为子字符串匹配的正则表达式。 position 为指定匹配的起始字符位置。 occurrence 为 pattern 在 source_char 出现的次数。 match_parameter 为正则表达式控制参数。 subexpr 为 pattern 的第几个子表达式，范围为 0~9。
ROUND(DATE, FMT)	按照 fmt 指定的日期格式对 date 进行四舍五入处理，返回值类型为 DATE。如果省略 fmt，则返回最近的一天。
ROUND(N,PRECISION)	返回 n 的四舍五入值，precision 为精度值。
RPAD(STRING, PADDED_LENGTH, [PAD_STRING])	从右边对字符串使用指定的字符进行填充.
RTRIM(C1,[,C2])	删除右边出现的字符串
SINH(N)	返回数字 n 的双曲正弦值，当 n 类型为 BINARY_FLOAT，返回类型 BINARY_DOUBLE，否则返回值类型为 NUMERIC。
SUBSTR(CHAR,POSITION[,SUBSTRING_LENGTH])	返回 char 字符串中第 position 个字符开始，长度为 substring_length 的子字符串。若不指定 substring_length，则截取到字符串结尾。

SUBSTRB(CHAR, POSITION[, SUBSTRING_LENGTH])	返回 char 字符串中第 position 个字节开始，长度为 substring_length 字节的子字符串。若不指定 substring_length，则截取到字符串结尾。
SYS_GUID()	返回 RAW 类型的全局唯一标识。
TANH(N)	返回参数 n 的双曲正切值；
TO_BLOB(CHAR)	将 char 字符串转换为 BLOB 类型，char 支持的类型为所有字符串类型（如 CHAR, VARCHAR2, NCHAR, NVARCHAR2 等）或可隐式转换为字符串的类型（如 integer 等）。
TO_CLOB(CHAR)	将 char 字符串转换为 CLOB 数据类型。
TO_CHAR(CHAR)	增加 char 支持的类型：char、character、varchar。
TO_DATE(CHAR[,FMT])	将 char 时间字符串按照 fmt 格式转换为 date 数据类型，char 支持的类型有 CHAR、VARCHAR2、NCHAR、 NVARCHAR2、TIMESTAMP。如果省略 fmt，则 char 必须采用 DATE 数据类型的默认格式。
TO_MULTI_BYTE(STR)	将单字节输入字符串转换为多字节字符串。
TO_NUMBER(EXPR)	将 expr 表达式值转换为 number 类型。
TO_TIMESTAMP(CHAR[,FMT])	将 char 时间字符串按照 fmt 格式转换为 timestamp 数据类型，char 支持的类型有 CHAR、VARCHAR2、NCHAR、NVARCHAR2、TIMESTAMP。如果省略 fmt，则 char 必须采用 TIMESTAMP 数据类型的默认格式。
TO_TIMESTAMP_TZ	将 char 时间字符串按照 fmt 格式转换为 timestamp WITH TIMEZONE 数据类型，char 支持的类型有 CHAR、VARCHAR2、NCHAR、 NVARCHAR2、TIMESTAMP。如果省略 fmt，则 char 必须采用 TIMESTAMP WITH TIMEZONE 数据类型的默认格式。
TO_SINGLE_BYTE(CHAR)	将多字节输入字符串转换为单字节字符串。
TRUNC(DATE[, FMT])	按照 fmt 指定的日期格式对 date 进行截断处理，返回值类型为 DATE。如果省略 fmt，则默认日期格式为'DDD'。

4.4.2. 高级函数包

- dbms_alert
- dbms_assert
- dbms_output
- dbms_pipe
- dbms_random
- dbms_utility
- utl_file

4.4.3. 系统视图

Vastbase G100 额外兼容的 Oracle 系统视图如下：

- user_tab_columns
- user_tables
- user_cons_columns
- user_constraints
- user_objects

- user_procedures
- user_source
- user_views
- user_ind_columns
- dba_segments

4.4.4. 数据类型

Vastbase G100 额外兼容的 Oracle 数据类型如下：

数据类型名称	数据类型
数字类型	NUMBER
变长字符串类型	VARCHAR2，NVARCHAR2 VARCHAR，NVARCHAR
定长字符串类型	CHAR，NCHAR
单精度精度二进制浮点类型	BINARY_FLOAT
双精度二进制浮点类型	BINARY_DOUBLE
二进制数据类型	RAW
二进制大对象类型	LONG，BLOB
字符大对象类型	CLOB
字节字符大对象类型	NCLOB
XML 数据类型	XMLType
日期时间类型	DATE
时间戳数据类型	TIMESTAMP， TIMESTAMP WITH TIME ZONE
本地时间戳数据类型	TIMESTAMP WITH LOCAL TIME ZONE

4.4.5. 隐式转换

Vastbase G100 额外兼容的 Oracle 隐式类型转换特性如下：

- 定长字符串类型 CHARACTER 与 NUMBER/NUMERIC，INT4，INT8，FLOAT4，FLOAT8 之间相互转换。
- 变长字符串类型 VARCHAR/VARCHAR2 与 NUMBER/NUMERIC，INT4，INT8，FLOAT4，FLOAT8 之间相互转换。
- 文本类型 TEXT 与 NUMBER/NUMERIC，INT2，INT4，INT8，FLOAT4，FLOAT8 之间相互转换。
- 短整形 INT2 转换为 CHARACTER，VARCHAR/VARCHAR2。
- 二进制大对象 BLOB 与二进制 RAW 之间相互转换。

4.4.6. 预定义参数

Vastbase G100 额外兼容的 Oracle 预定义参数如下：

预定义参数	说明
NLS_DATE_FORMAT	日期格式定义参数。
NLS_TIMESTAMP_FORMAT	不带时区的时间戳格式定义参数。
NLS_TIMESTAMP_TZ_FORMAT	带时区的时间戳格式定义参数。

4.4.7. 宏变量

Vastbase G100 额外兼容的 Oracle 宏变量如下：

- SYSDATE：获取当前系统时期。
- SYSTIMESTAMP：获取当前系统时间戳。
- DBTIMEZONE：获取当前数据库时区。
- SESSIONTIMEZONE：获取当前会话时区。
- ROWNUM：获取查询结果中的元组序号。
- ROWID：获取一条记录的一个相对唯一地址值。

4.4.8. 操作符

Vastbase G100 额外兼容的 Oracle 操作符如下：

- 求差集操作符：MINUS。
- 不等于操作符：^=。
- 时间 + 整数操作符：date + int。
- 时间 - 整数操作符：date – int。
- 时间 - 时间操作符：date – date。

4.4.9. SQL 与 PL/SQL 语法

Vastbase G100 兼容的 Oracle SQL 与 PL/SQL 语法如下：

- 支持 ORACLE 的 CREATE SEQUENCE 语法。
- 支持 ORACLE 的 CREATE/ALTER DATABASE 语法。
- 支持 ORACLE 的 CREATE/ALTER VIEW 语法。
- 支持 ORACLE 的 CREATE TABLE 语法。
- 支持 ORACLE 的 CREATE TABLESPACE 语法。
- 支持 ALTER TABLE ADD CONSTRAINT USING INDEX 语法。

- 支持外连接运算符（+）。
- 支持 CONNECT BY..... START WITH .....查询。
- 支持 ROWNUM 伪列。
- 支持 ROWID 伪列。
- 支持全局临时表。
- 支持 synonym。
- 支持字段表达式创建索引：CREATE INDEX ON COLUMN_EXPR。
- 支持 ALTER TABLE...MODIFY 修改表字段。
- 支持 VARCHAR、CHARACTER 数据类型指定长度单位。
- 支持 TYPE/NAME/VERSION/VALUE/INTERVAL 作为别名。
- 支持 ORACLE 的存储过程语法。
- 支持 ORACLE 的 DATE 数据类型。
- 支持 MERGE 操作语法：

```
MERGE [HINT] INTO table_name USING ({subquery | table_name | view_name}) alias
ON (condition) merge_update_clause merge_insert_clause;
```

- 支持时间间隔操作语法：

```
INTERVAL YEAR TO MONTH,INTERVAL DAY (l) TO SECOND (P);
```

- 支持存储过程游标语法：

```
CURSOR cursor_name [ parameter_list ] IS select_statement, TYPE type_name IS
REF CURSOR;
```

- 支持修改会话属性：

```
ALTER SESSION SET param_name = value;
```

- 支持匿名块。
- 支持存储过程跨模式访问。
- 支持子查询不指定别名。
- CREATE SEQUENCE 支持 NOCYCLE。
- CREATE/ALTER USER 语法中允许使用 IDENTIFIED BY 关键字替代 PASSWORD 关键字。
- UPDATE SET 语法中允许指定表名或别名修饰。
- ALTER TABLE 兼容 MODIFY NOT NULL 语法和 ENABLE 语法。
- 空字符串和 NULL 等价。
- 支持序列操作语法：sequence.CURRVAL，sequence.NEXTVAL。
- 删除表记录语法中 FROM 关键字。
- 支持数据类型 INTERVAL 与数字之间的运算操作：+、-、>、<、>=、<=、<>。
- 兼容分区表的 DML 操作：SELECT、INSERT、UPDATE、DELETE。
- 兼容 CREATE/ALTER MATERIALIZED VIEW 语法。
- 兼容 CREATE TYPE 语法。
- 兼容列约束的 enable/disable 语法。
- 兼容 FUNCTION 定义。

4.5. 高可用

4.5.1. 主备机

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，Vastbase G100 能够保证故障之前的数据无丢失，满足 ACID 特性。

主备环境可以支持主备和一主多备两种模式。主备模式下，备机需要重做日志，可以升主。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。主备主要用于一般可靠性的 OLTP 系统能够节省一定的存储资源。而一主多备提供更高的容灾能力，适合于更高可靠性要求的 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拉齐。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，拷贝的数据量比较大，耗时比较长，而增量 build 只拷贝差异文件，拷贝的数据量比较小，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。

4.5.2. 逻辑备份

Vastbase G100 提供逻辑备份能力，可以将用户表的数据以通用的 text 或者 csv 格式备份到本地磁盘文件，并在同构/异构数据库中恢复该用户表的数据。

4.5.3. 物理备份

Vastbase G100 提供物理备份能力，可以将整个实例的数据以数据库格式备份到本地磁盘文件中，并在同构数据库中恢复整个实例的数据。

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。

当前 Vastbase G100 支持全量备份与增量备份。除此之外还支持并行备份提高性能，对备份集进行压缩节省空间，制定备份保留策略等功能

4.6. 可维护性

4.6.1. 支持 WDR 诊断报告

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。

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，IO 状况，负载和 IO 是衡量 TP 系统最最要的特性。 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，IO 读写时间等。
Load Profile	从时间，IO，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，事务日质量，逻辑和物理读的量，读写 IO 次数、大小，登入登出次数，SQL、事务执行量，SQL P85、P90 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时间的发生次数，时间的等待时间。
Wait Classes	用于评估当前系统关键事件类型的性能。 主要包含数据内核在主要的等待事件的种类上的发布：STATUS、LWLOCK_EVENT、LOCK_EVENT、IO_EVENT。

诊断类别	描述
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，关键时间的详细性能诊断。 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	用于诊断后端作业性能的诊断。 包含页面操作，复制等后端操作的性能。
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	用于判断配置是够有变更。 包含当前所有配置参数的快照。

应用价值：

- 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，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。

4.6.2. 慢 SQL 诊断

慢 SQL 诊断分为两部分：实时慢 SQL 和历史慢 SQL。

- 实时慢 SQL 能根据用户提供的执行时间阈值，输出当前系统中正在执行的，且执行时间超过阈值的作业信息。
- 历史慢 SQL 能根据用户提供的执行时间阈值，记录所有超过阈值的执行完毕的作业信息。

历史慢 SQL 提供表和文件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 时间等。所有信息都是脱敏的。

应用价值：

- 实时的慢 SQL 提供给用户管理尚未执行完毕的作业的接口，用户可以手动结束异常的，消耗过多资源的作业。
- 历史慢 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。表和文件接口方便用户统计慢 SQL 指标，对接第三方平台。

4.6.3. 支持一键式收集诊断信息

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：

- 通过操作系统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

4.7. 数据库安全

4.7.1. 自主访问控制

管理用户对数据库的访问控制权限，涵盖数据库系统权限和对象权限。

支持基于角色的访问控制机制，将角色和权限关联起来，通过将权限赋予给对应的角色，再将角色授予给用户，可实现用户访问控制权限管理。其中登录访问控制通过用户标识和认证技术来共同实现，而对象访问控制则基于用户在对象上的权限，通过对象权限检查实现对象访问控制。用户在为相关的数据库用户分配完成任务所需要的最小权限从而将数据库使用风险降到最低。

支持三权分立权限访问控制模型，数据库角色可分为系统管理员、安全管理员和审计管理员。其中安全管理员负责创建和管理用户，系统管理员负责授予和撤销用户权限，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。客户可通过设置参数来选择是否开启三权分立控制模型。

4.7.2. 强制访问控制

数据安全标签是根据客体和主体的敏感标记对客体访问实行限制的一种方法。Vastbase G100 数据库安全增强选件利用敏感标记和策略来实现强制访问控制。

通过对涉及敏感数据的行进行敏感数据标签，并配置读写策略，用户可实现对存在敏感数据的记录根据用户权限级别进行访问限制上的区分。此配置必须由具备“数据库安全员”身份的用户进行修改，可实现对包括数据库管理员在内的非授权用户的数据访问权限进行控制，有效防止敏感数据泄露。

Vastbase G100 数据库数据安全标签可以是表级、行级或者是列级，甚至是混合策略。

4.7.3. 控制权和访问权分离

针对系统管理员用户，实现表对象的控制权和访问权分离，提高普通用户数据安全性，限制管理员对象访问权限。

该特性适用于如下场景，即对于有多个业务部门的企业，各部门间使用不同的数据库用户进行业务操作，同时存在同级别的数据库维护部门使用数据库管理员进行运维操作，业务部门希望在未经授权的情况下，管理员用户只能对各部门的数据进行控制操作（DROP、ALTER、TRUNCATE），但是不能进行访问操作（INSERT、DELETE、UPDATE、SELECT、COPY）。即针对管理员用户，表对象的控制权和访问权分离，提高用户数据的安全性。

系统管理员可以在创建用户时指定 INDEPENDENT 属性，表示该用户为私有用户。针对该用户的对象，数据库管理员（包含初始用户和其他管理员用户）在未经其授权前，只能进行控制操作（DROP、ALTER、TRUNCATE），无权进行 INSERT、DELETE、SELECT、UPDATE、COPY、GRANT、REVOKE、ALTER　OWNER 操作。

4.7.4. 数据库加密认证

采用基于 RFC5802 机制的口令加密认证方法。

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

创建用户所设置的口令被加密存储在系统表中。整个认证过程中口令加密存储和传输，通过计算相应的 hash 值并与服务端存储的值比较来进行正确性校验。

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口令的正确性。

4.7.5. 数据库审计

审计日志记录用户对数据库的启停、连接、DDL、DML、DCL 等操作。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。

用户可以通过参数配置对哪些语句或操作记录审计日志。

审计日志记录事件的时间、类型、执行结果、用户名、数据库、连接信息、数据库对象、数据库实例名称和端口号以及详细信息。支持按起止时间段查询审计日志，并根据记录的字段进行筛选。

数据库安全管理员可以利用这些日志信息，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。

4.7.6. 网络通信安全特性

支持通过 SSL 加密客户端和服务器之间的通信数据，保证客户的客户端与服务器通信安全。

采用 TLS 1.2 协议标准，并使用安全强度较高的加密算法套件，支持的加密算法套件如表 1-4 所示。

表1-4 加密算法套件

IANA 编码	IANA 套件名
0x00,0x9F	TLS_DHE_RSA_WITH_AES_256_GCM_SHA384
0x00,0x9E	TLS_DHE_RSA_WITH_AES_128_GCM_SHA256
0x00,0xA3	TLS_DHE_DSS_WITH_AES_256_GCM_SHA384
0x00,0xA2	TLS_DHE_DSS_WITH_AES_128_GCM_SHA256
0x00,0x6B	TLS_DHE_RSA_WITH_AES_256_CBC_SHA256
0x00,0x67	TLS_DHE_RSA_WITH_AES_128_CBC_SHA256
0x00,0x6A	TLS_DHE_DSS_WITH_AES_256_CBC_SHA256
0x00,0x40	TLS_DHE_DSS_WITH_AES_128_CBC_SHA256
0xC0,0x9F	TLS_DHE_RSA_WITH_AES_256_CCM
0xC0,0x9E	TLS_DHE_RSA_WITH_AES_128_CCM
0x00,0x39	TLS_DHE_RSA_WITH_AES_256_CBC_SHA
0x00,0x33	TLS_DHE_RSA_WITH_AES_128_CBC_SHA
0x00,0x38	TLS_DHE_DSS_WITH_AES_256_CBC_SHA
0x00,0x32	TLS_DHE_DSS_WITH_AES_128_CBC_SHA

4.7.7. 透明存储加密

为了防止数据文件被窃取后泄露用户数据，Vastbase 安全增强选项提供了全库透明加密功能。数据将在写入磁盘时被自动加密，从磁盘中读取时被自动解密，整个过程对用户以及上层应用是透明的。由于数据读写时需要进行额外的加解密操作，故会对数据库性能造成一定影响。

5. 技术指标

技术指标	最大值
数据库容量	受限于操作系统与硬件
单表大小	32TB
单行数据大小	1GB
每条记录单个字段的大小	1GB
单表记录数	2^{48}
单表列数	250~1600（随字段类型不同会有变化）
单表中的索引个数	无限制
复合索引包含列数	32
数据库名长度	64
对象名长度（除数据库名以外的其他对象名）	64
单表约束个数	无限制
并发连接数	10000
分区表的分区个数	32768
分区表的单个分区大小	32TB
分区表的单个分区记录数	2^{55}
LOB 最大容量	1GB-8203 字节
SQL 文本最大长度	1048576 字节

6. 术语表

表1-5 术语表

术语	解释
A – E	
ACID	在可靠数据库管理系统（DBMS）中，事务（transaction）所应该具有的四个特性：原子性（Atomicity）、一致性（Consistency）、隔离性（Isolation）、和持久性（Durability）。
Bgwriter	数据库启动时创建的一个后台写线程，此线程用于将数据库中脏页面写入到持久性设备（例如磁盘）中。
bit	比特。计算机处理的最小的信息单位。比特用来表示二进制数字 1 或 0，或者一种逻辑条件真或假。在物理上，比特表示一个电路上高或低的电压点或者磁盘上的磁化单程或其它。一个单独的比特位所传达的信息很少有意义的。然而，一个 8 位组却构成了一个字节，可用于表示如一个英文字母，十进制数字，或其它字符等多种类型的信息。
Bloom Filter	布隆过滤器。由 Howard Bloom 在 1970 年提出的二进制向量数据结构，它具有很好的空间和时间效率，被用来检测一个元素是不是集合中的一个成员，这种检测只会对在集合内的数据错判，而不会对不是集合内的数据进行错判，这样每个检测请求返回有“在集合内（可能错误）”和“不在集合内（绝对不在集合内）”两种情况，可见 Bloom filter 是牺牲了正确率换取时间和空间。
CIDR	Classless Inter-Domain Routing，无类域间路由 IP 编址方案。CIDR 摒弃传统的基于类（A 类：8，B 类：16，C 类：24）的地址分配方式，允许使用任意长度的地址前缀，有效提高地址空间的利用率。CIDR 表示方法：IP 地址/网络 ID 的位数。比如 192.168.23.35/21，其中“21”表示前面地址中的前 21 位代表网络部分，其余位代表主机部分。
CLI	Command-line Interface，命令行界面。应用程序和用户交互的一种方式，完全基于文本输入和输出。命令通过键盘或类似装置输入，由程序编译并执行。结果是以文本或图形的方式呈现在终端界面。
CU	Compression Unit，压缩单元。列存表的最小存储单位。
core 文件	当程序出现内存越界、断言失败或者访问非法内存时，操作系统会中止进程，并将当前内存状态导出到 core 文件中，以便进一步分析。 core 文件包含内存转储，支持全二进制和指定端口格式。core 文件名称由字符

术语	解释
	串 core 以及操作系统进程 ID 组成。 core 文件不依赖于任何平台。
Core Dump	通常在程序异常终止时，核心转储（Core Dump）、内存转储或系统转储用于记录特定时间计算机程序工作内存的状态。实际上，其它关键程序的状态经常在同一时间进行转储，例如处理器寄存器，包括程序指标和栈指针、内存管理信息、其它处理器和操作系统标记及信息。Core Dump 经常用于辅助诊断和纠错计算机程序问题。
DBA	Database Administrator，数据库管理员。指导或执行所有和维护数据库环境相关的操作。
DBLINK	DBLINK 是定义一个数据库到另一个数据库路径的对象，通过它可以查询远程数据库对象。
DBMS	Database Management System，数据库管理系统。数据库管理系统是为了访问数据库中的信息而使用的一个管理系统软件。它包含一组程序使用户可以进入、管理、查询数据库中数据。基于真实数据的位置，可以分为内存数据库管理系统和磁盘数据库管理系统。
DCL	Data Control Language，数据控制语言。
DDL	Data Definition Language，数据定义语言。
DML	Data Manipulation Language，数据操纵语言。
备份	备份件或者备份过程。指复制并归档计算机数据，当发生数据丢失事件时，可以用该复制并归档的数据来恢复原始数据。
备份和恢复	保护数据库防止由于媒介失效或人为错误造成的数据丢失过程中涉及的一组概念、过程及策略。
备机	Vastbase G100 双机方案中的一个节点，用于作为主机的备份，在主机异常时，备机会切换到主机状态，以确保能正常提供数据服务。
崩溃	崩溃（或系统崩溃）指计算机或程序（例如软件应用程序或操作系统）异常终止的事件。出现错误后，通常会自动退出。有时出现恶意程序冻结或挂起直到崩溃上报服务记录崩溃的详细信息。对于操作系统内核关键部分的程序，整个计算机可能瘫痪（可能造成致命的系统错误）。
编码	编码是指用代码来表示各组数据资料，使其成为可利用计算机进行处理和分析的信息。用预先规定的方法将文字、数字或其它对象编成数码，或将信息、数据转换成规定的电脉冲信号。

术语	解释
编码技术	呈现计算机软硬件识别的特定字符集数据的技术。
表	表是由行与列组合成的。每一列被当作是一个字段。每个字段中的值代表一种类型的数据。例如，一个表可能有 3 个字段：姓名、城市和国家。这个表就会有 3 列：一列代表姓名，一列代表城市，一列代表国家。表中的每一行包含 3 个字段的内容，姓名字段包含姓名，城市字段包含城市，国家字段包含国家。
表空间	包含表、索引、大对象、长数据等数据的逻辑存储结构。表空间在物理数据和逻辑数据间提供了抽象的一层，为所有的数据库对象分配存储空间。表空间创建好后，创建数据库对象时可以指定该对象所属的表空间。
并发控制	在多用户环境下同时执行多个事务并保证数据完整性的一个 DBMS 服务。并发控制是 Vastbase G100 提供的一种多线程管理机制，用来保证多线程环境下在数据库中执行的操作是安全的和一致的。
查询	向数据库发出的信息请求，包含更新、修改、查询或删除信息的请求。
查询操作符	Query Operator，也称为查询迭代算子（Iterator）或查询节点（Query Tree Node）。一个查询的执行可以分解为一个或多个查询操作符，是构成一个查询执行的最基本单位。常见的查询操作符包括表扫描（Scan），表关联（Join），表聚集（Aggregation）等。
持久性	数据库事务的 ACID 特性之一。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。
存储过程	存储过程（StoredProcedure）是在大型数据库系统中，一组为了完成特定功能的 SQL 语句集，经编译后存储在数据库中，用户通过指定存储过程的名称并设置参数（如果该存储过程带有参数）来执行它。
操作系统	操作系统 OS（operating system）由引导程序加载到计算中，对计算机中其它程序进行管理。其它程序叫做应用或应用程序。
大对象	大对象（Blob）在数据库中指使用二进制方式存储的数据。它通常可以用于存储视频、音频和图像等多媒体数据。
段	数据库中，一段指包含一个或多个区域的数据库中的一部分。区域是数据库的最小范围，由单元调用块组成。一个或多个段组成一个表空间。
F – J	
Failover	指当某个节点出现故障时，自动切换到备节点上的过程。反之，从备节点上切换回来的过程称为 Failback。
FDW	Foreign Data Wrapper，外部数据封装器。是 Postgres 提供的一个 SQL 接口，

术语	解释
	用于访问远程数据存储中的大数据对象，使 DBA 可以整合来自不相关数据源的数据，将它们存入数据库中的一个公共模型。
Freeze	在事务 ID 耗尽时由 AutoVacuum Worker 进程自动执行的操作。Vastbase G100 会把事务 ID 记在行头，在一个事务取得一行时，通过比较行头的事务 ID 和事务本身的 ID 判断这行是否可见，而事务 ID 是一个无符号整数，如果事务 ID 耗尽，事务 ID 会跨过整数的界限重新计算，此时原先可见的行就会变成不可见的行，为了避免这个问题，Freeze 操作会将行头的事务标记为一个特殊的事务 ID，标记了这个特殊的事务 ID 的行将对所有事务可见，以此避免事务 ID 耗尽产生的问题。
GDB	GNU 工程调试器，可以监控其它程序运行时的内部情况，或者其它程序要崩溃时发生了什么。GDB 支持如下四种主要操作（使 PDK 功能更加强大），辅助查找缺陷。 •启动程序，指定可能影响行为的任何因素。 •特定条件下，停止程序。 •程序停止时，检查发生了什么。 •修改程序内容，尝试纠正一个缺陷并继续下一个。
GIN 索引	Generalized Inverted Index，通用倒排索引。作用为处理索引项为组合值的情况，查询时需要通过索引搜索出出现在组合值中的特定元素值。
GNU	GNU 计划，又称革奴计划，是由 RichardStallman 在 1983 年 9 月 27 日公开发起的。它的目标是创建一套完全自由的操作系统。GNU 是“GNU's NotUnix”的递归缩写。Stallman 宣布 GNU 应当发音为 Guh-NOO 以避免与 new 这个单词混淆（注：Gnu 在英文中原意为非洲牛羚，发音与 new 相同）。Unix 是一种广泛使用的商业操作系统的名称。技术上讲，GNU 类似 Unix。但是 GNU 却给了用户自由。
gsql	Vastbase G100 交互终端。通过 gsql 能够以交互的方式输入查询，下发查询到 Vastbase G100，然后查看查询结果。或者，也可以从文件中输入。此外，gsql 还提供许多元命令和各种类似 shell 命令，协助脚本编写及自动化各种任务。
GUC	Grand Unified Configuration，数据库运行参数。配置这些参数可以影响数据库系统的行为。
HA	高可用性（HighAvailability），通过尽量缩短因日常维护操作（计划）和突发的系统崩溃（非计划）所导致的停机时间，以提高系统和应用的可用性。
HBA	host-based authentication，主机认证。主机鉴权允许主机鉴权部分或全部系统

术语	解释
	用户。适用于系统所有用户或者使用 Match 指令的子集。该类型鉴权对于管理计算集群以及其它完全同质设备非常有用。总之，服务器上的三个文件以及客户端上的一个文件必须修改，为主机鉴权做准备。
服务器	为客户端提供服务的软硬件的组合。单独使用时，指运行服务器操作系统的计算机，也可以指提供服务的软件或者专用硬件。
隔离性	数据库事务的 ACID 特性之一。它是指一个事务内部的操作及使用的数据对其它并发事务是隔离的，并发执行的各个事务之间不能互相干扰。
关系型数据库	创建在关系模型基础上的数据库。关系型数据库借助于集合代数等数学概念和方法来处理数据库中的数据。
归档线程	数据库打开归档功能时启动的一个线程，此线程用于将数据库日志归档到指定的路径。
故障接管	功能对等的系统部件对于故障部件的自动替换过程。系统部件包含处理器、服务器、网络、数据库等。
环境变量	定义进程操作环境某一方面的变量。例如，环境变量可以为主目录，命令搜索路径，使用终端或当前时区。
检查点	将数据库内存中某一时刻的数据存到磁盘的机制。Vastbase G100 定期将已提交的事务数据和未提交的事务数据存到磁盘，这些数据用来和 Redo 日志一起在数据库重启和崩溃时恢复数据库。
加密	用于传输数据的功能。通过该功能，可以隐藏信息内容，防止非法使用。
节点	将构成 Vastbase G100 集群环境的各台服务器（物理机或虚拟机）称为集群节点，简称节点。
纠错	系统自动识别软件和数据流上的错误并自动修正错误的能力，提升系统的稳定性和可靠性。
进程	在单个计算机上执行程序的实例。一个进程由一个或多个线程组成。其它进程不能接入某个进程已占用的线程。
基于时间点恢复	PITR（Point-In-TIme Recovery），基于时间点恢复是 Vastbase G100 备份恢复的一个特性，是指在备份数据和 WAL 日志正常的情况下，数据可以恢复到指定时间点。
记录	在关系型数据库中，每一条记录对应表中的每一行数据。
K – O	

术语	解释
逻辑复制	数据库主备或两个集群间的数据同步方式。区别于通过物理日志回放方式的物理复制，逻辑复制在两个集群间传输逻辑日志或通过逻辑日志对应的 SQL 语句实现数据同步。
逻辑日志	数据库修改的日志记录，可直接对应为 SQL 语句，一般为行级记录。区别于物理日志，物理日志是记录物理页面修改的日志。
逻辑解码	逻辑解码是一种通过对 xlog 日志的反解实现将数据库表的所有持久更改抽取到一种清晰、易于理解的格式的处理过程。
逻辑复制槽	在逻辑复制的环境下，逻辑复制槽用以防止 Xlog 被系统或 Vaccum 回收。Vastbase G100 中用于记录逻辑解码位置的对象，提供创建、删除、读取、推进等多个 SQL 接口函数。
MVCC	Multi-Version Concurrency Control，多版本并发控制。数据库并发控制协议的一种，它的基本算法是一个元组可以有多个版本，不同的查询可以工作在不同的版本上。一个基本的好处是读和写可以不冲突。
NameNode	NameNode 是 Hadoop 系统中的一个中心服务器，负责管理文件系统的名称空间(namespace)以及客户端对文件的访问。
OM	Operations Management，运维管理模块。提供集群日常运维、配置管理的管理接口、工具。
客户端	连接或请求其它计算机或程序服务的计算机或程序。
空闲空间管理	管理表内空闲空间的机制，通过记录每个表内空闲空间信息，并建立易于查找的数据结构，可以加速对空闲空间进行的操作（例如 INSERT）。
垃圾元组	是指使用 DELETE 和 UPDATE 语句删除的元组，Vastbase G100 在删除元组时只是打个删除标记，由 Vacuum 线程清理这种垃圾元组。
列	字段的等效概念。在数据库中，表由一列或多列组成。
逻辑节点	一个物理节点上可以安装多个逻辑节点。一个逻辑节点是一个数据库实例。
模式	数据库对象集，包括逻辑结构，例如表、视图、序、存储过程、同义名、索引、集群及数据库链接。
模式文件	用于决定数据库结构的 SQL 文件。
P – T	
Page	Vastbase G100 数据库关系对象结构中行存的最小内存单元。一个 Page 大小为默认为 8KB。

术语	解释
PostgreSQL	PostgreSQL 是一个开源的关系数据库管理系统(DBMS)，由全球志愿者团队开发。PostgreSQL 不受任何公司或个体所控制，源代码免费使用。
Postmaster	数据库服务启动时启动的一个线程。用于监听来自集群其它节点或客户端的连接请求。 主机上监听到备机连接请求，并接受后，就会创建一个 WAL Sender 线程，用于处理与备机的交互。
RHEL	Red Hat Enterprise Linux，红帽企业 Linux。
REDO 日志	记录对数据库进行操作的日志，这些日志包含重新执行这些操作所需要的信息。当数据库故障时，可以利用 REDO 日志将数据库恢复到故障前的状态。
SCTP	Stream Control Transmission Protocol，流控制传输协议。是 IETF 于 2000 年新定义的一个传输层协议。是提供基于不可靠传输业务的协议之上的可靠的数据报传输协议。SCTP 的设计用于通过 IP 网传输 SCN 窄带信令消息。
Savepoint	保存点。是一种在关系数据库管理系统中实现子事务（也称为嵌套事务）的方法。在一个长事务中，可以把操作过程分成几部分，前面部分执行成功后，可以建一个保存点，若后面的执行失败，则回滚到这个保存点即可，无需回滚整个事务。保存点对于在数据库应用程序中实现复杂错误恢复很有用。如果在多语句事务中发生错误，则应用程序可能能够从错误中恢复（通过回滚到保存点）而无需中止整个事务。
Session	数据库系统在接收到应用程序的连接请求时，为该连接创建的一个任务。它被 Session Manager 管理，完成一些初始化任务，执行用户的所有操作。
SLES	SUSE Linux Enterprise Server，由 SUSE 提供的企业级 Linux 操作系统。
SMP	Symmetric Multi-Processing，对称多处理技术，是指在一台计算机上汇集了一组处理器（多 CPU），各 CPU 之间共享内存子系统以及总线结构。操作系统必须支持多任务和多线程处理，以使得 SMP 系统发挥高效的性能。数据库领域的 SMP 并行技术，一般指利用多线程技术实现查询的并行执行，以充分利用 CPU 资源，从而提升查询性能。
SQL	Structure Query Language，结构化查询语言。数据库的标准查询语言。它可以分为数据定义语言（DDL），数据操纵语言（DML）和数据控制语言（DCL）。
SSL	Secure Socket Layer，安全套接层。SSL 是 Netscape 公司率先采用的网络安全协议。它是在传输通信协议（TCP/IP）上实现的一种安全协议，采用公开密钥技术。SSL 广泛支持各种类型的网络，同时提供三种基本的安全服务，它们都使用公开密钥技术。SSL 支持服务通过网络进行通信而不损害安全性。它在客户端

术语	解释
	和服务器之间创建一个安全连接。然后通过该连接安全地发送任意数据量。
收敛比	交换机下行带宽与上行带宽的比值。收敛比越高，流量收敛程度越大，丢包越严重。
TCP	Transmission Control Protocol，传输控制协议。用于将数据信息分解成信息包，使之经过 IP 协议发送；并对利用 IP 协议接收来的信息包进行校验并将其重新装配成完整的信息。TCP 是面向连接的可靠协议，能够确保信息的无误发送。
trace	一种特殊的日志记录方法，用来记录程序执行的信息。程序员使用该信息进行纠错。另外，根据 trace 日志中信息的类型和内容，有经验的系统管理员或技术支持人员以及软件监控工具诊断软件常见问题。
强一致性	任何查询不会瞬时的看到一个分布式事务的中间状态。
全备份	备份整个数据库。
全量同步	Vastbase G100 双机方案中的一种数据同步机制，是指把主机中的所有数据同步给备机。
日志文件	计算机记录自身活动的记录。
事务	数据库管理系统执行过程中的一个逻辑单位，由一个有限的数据库操作序列构成，事务必须满足 ACID 原则。
数据	事实或指令的一种表达形式，适用于人为或自动的通信、解释或处理。数据包含常量、变量、阵列和字符串。
数据分区	数据分区是指在一个数据库实例内部，将表按照划分为多个数据互不重叠的部分（Partition）。具体的分区方式可以有：范围分区（Range），它根据元组中指定字段的取值所处的范围映射到目标存储位置。
数据库	数据库是存储在一起的相关数据的集合，这些数据可以被访问，管理以及更新。同一视图中，数据库可以根据存储内容类型分为以下几类：数目类、全文本类、数字类及图像类。
数据库实例	一个数据库实例是一个 Vastbase G100 进程以及它控制的数据库文件。Vastbase G100 在一个物理节点上安装多个数据库实例。一个数据库实例也被称为一个逻辑节点。
数据库双机	Vastbase G100 提供的高可靠性双机方案。在此方案中，每个 Vastbase G100 逻辑节点标识为主机或备机。在同一时间内，只有一个 Vastbase G100 被标识为主机。双机初次建立时，主机会对每个备机数据做全量同步，然后做增量同

术语	解释
	步。双机建立之后的运行过程中，主机能接受数据读和写的操作请求，备机只做日志同步。
数据库文件	保存用户数据和数据库系统内部数据的二进制文件。
数据字典	数据字典是一系列只读的表，用来提供数据库的信息。这些信息包括：数据库设计信息、存储过程信息、用户权限、用户统计数据、数据库进程信息、数据库增长统计数据和数据库性能统计数据。
死锁	为使用同一资源而产生的无法解决的争用状态。
索引	数据库索引，是数据库管理系统中一个排序的数据结构，以协助快速查询、更新数据库表中数据。
统计信息	数据库使用统计信息估算查询代价，以查找代价最小的执行计划，统计信息一般是数据库自动收集的，包括表级信息（元组数、页面数等）和列级信息（列的值域分布直方图）。
停用词	在信息检索中，为节省存储空间和提高搜索效率，在处理自然语言数据（或文本）之前或之后会自动过滤掉某些字或词，这些字或词即被称为 Stop Words（停用词）。
U – Z	
Vacuum	数据库定期启动的清理垃圾元组的线程，根据配置参数可以同时启动多个。
verbose	verbose 选项指定显示在屏幕上的处理信息。
WAL	Write-Ahead Logging，预写日志系统。实现事务日志的标准方法，是指对数据文件（表和索引的载体）持久化修改之前必须先持久化相应的日志。
WAL Receiver	数据库复制时备机创建的一个线程的名称。此线程用于从主机接收数据、命令，并反馈确认信息至主机。一个备机只有一个 WAL Receiver 线程。
WAL Sender	数据库复制过程中，主机接受到备机的连接请求后创建的一个线程的名称。此线程用于发送命令、数据到备机，并从备机接收信息。一个主机可能会有多个 WAL Sender 线程，每一个 WAL Sender 线程对应一个备机的一个连接请求。
WAL Writer	数据库启动时创建的一个写 Redo 日志的线程，用于将内存中的日志写入到持久性设备（如：磁盘）。
Xlog	表示事务日志，一个逻辑节点中只有一个，不允许创建多个 Xlog 文件。
xDR	详单。用户面和信令面详单的统称，包括 CDR 和 UFDR、TDR 和 SDR。

术语	解释
物理节点	一个物理机器称为一个物理节点。
系统表	存储数据库元信息的表，元信息包括数据库中的用户表、索引、列、函数和数据类型等。
下推	Vastbase G100 是分布式数据库，其可以利用多 DN 分布式并行执行查询计划，即将 CN 中的查询计划下发到各 DN 中并行执行。这种行为称为下推。与将数据抽取到 CN 上执行查询的方式相比，下推可以大幅提升查询性能。
压缩	数据压缩，信源编码，或比特率降低涉及使用相比原来较少比特的编码信息。压缩可以是有损或无损。无损压缩通过识别和消除统计冗余降低比特位。无损压缩中没有信息丢失。有损压缩识别并删除次要信息，减少了比特位。减少数据文件大小的方法被普遍称为数据压缩，尽管其正式名称为源编码（数据源的编码，然后将其存储或传输）。
一致性	数据库事务的 ACID 特性之一。在事务开始之前和事务结束以后，数据库的完整性约束没有被破坏。
元数据	用来定义数据的数据。主要是描述数据自身信息，包含源、大小、格式或其它数据特征。数据库字段中，元数据用于理解以及诠释数据仓库的内容。
原子性	数据库事务的 ACID 特性之一。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。事务在执行过程中发生错误，会被回滚到事务开始前的状态，就像这个事务从来没有执行过一样。
脏页面	已经被修改且未写入持久性设备的页面。
增量备份	基于上次有效备份之后对文件修改的备份。
增量同步	Vastbase G100 双机方案中的一种数据同步机制，是指把主机中数据增量同步给备机，即只同步主备间有差异的数据。
主机	Vastbase G100 数据库双机系统中接受数据读写操作的节点，和所有备机一起协同工作。在同一时间内，双机系统中只有一个节点被标识为主机。
主题词	在标引和检索中用以表达文献主题的规范化的词或词组。
转储文件	转储文件是一种特定类型的 trace 文件。转储文件为响应事件过程中一次性输出的诊断数据，trace 文件指诊断数据的连续输出。
最小恢复点	最小恢复点是 Vastbase G100 提供的数据一致性保障手段之一。最小恢复点特性可以在 Vastbase G100 启动时检查出 WAL 日志和持久化到磁盘的数据的不一致性，并提示用户进行处理。

电话：010-82838118
地址：北京市海淀区学院路 30 号科大天工大厦 B 座 6 层
官网：www.vastdata.com.cn