

Oracle Exadata 存储扩展 X9M-2

Oracle Exadata 数据库云平台经过集成设计，与其他平台相比，能够以更低的成本提供更高的性能和容量。它适用于所有类型的现代数据库负载，包括联机事务处理 (OLTP)、数据仓库 (DW)、内存分析、物联网 (IoT)、财务、游戏、合规性数据管理以及高效整合的混合负载。Oracle Exadata 存储扩展机架是为 Exadata 数据库云平台添加存储容量最简便、最快速、最可靠的方式。作为 Oracle 数据库云平台的自然延伸，Exadata 存储扩展机架可满足超大型关键业务数据库的需求。

集成设计，支持快速可靠部署

Oracle Exadata 数据库云平台是运行 Oracle 数据库的成本效益极佳、性能超强的平台。Exadata 易于部署，甚至适用于要求最严苛的关键任务系统，其数据库服务器、存储服务器和网络均由 Oracle 专家预先配置、预先调优和预先测试。该系统经过了广泛的端到端测试和验证，可确保所有组件（包括数据库软件、操作系统、虚拟机管理程序、驱动器、固件）均能无缝协同工作，消除性能瓶颈或单点故障。Exadata 存储扩展机架使用与 Exadata 数据库云平台相同的安全、高度可用的横向扩展架构构建，支持您根据需求添加存储，不必按峰值容量要求进行选型。

超强性能和超大容量

借助 Exadata 存储扩展 X9M 机架，客户可以为任何基于 RoCE 的 Exadata 数据库云平台增加 Exadata 存储容量和带宽。该存储扩展机架专为需要海量数据的数据库部署而设计，这些数据包括历史或归档数据、备份、文档、图像、XML 和 LOB 等。该扩展机架无需设置 LUN 或挂载点，因此配置起来十分简单。只需使用几个简单的命令，即可联机完成存储配置，并将存储添加至数据库。



Oracle Exadata 存储扩展 X9M-2

关键特性

- 为采用 RDMA over Covered Ethernet (RoCE) 网络的 Oracle Exadata 数据库云平台增加存储容量
- 100 Gb/秒 RoCE 网络
- 包含 4 至 19 台 Oracle Exadata 存储服务器
- 每机架未压缩的镜像可用容量高达 1663 TB
- 存储服务器多达 608 个 CPU 核专用于进行 SQL 处理
- 每个机架高达 28.5 TB 的持久内存加速
- 完全冗余，确保高可用性

通过弹性配置提供超强的系统可扩展性和扩容能力

Exadata 存储扩展 X9M 可为您提供比以往更高的扩展灵活性。Exadata 存储扩展 X9M 支持配置和购买四分之一机架（其中包含四台存储服务器），然后根据需求一次添加一台存储服务器，最多可达 19 台。由于可灵活添加 4 到 19 台存储服务器，因此这些配置可满足各种不同的应用需求。除了可将 Exadata 存储扩展 X9M 的配置进行升级外，Oracle 还继续使用构建块方法，通过集成的 RoCE 网络将 Exadata 存储扩展 X9M 与 Oracle 数据库云平台 X9M 连接，从而将您的系统轻松扩展至任意规模。Exadata 存储扩展 X9M 几乎能以任意组合与 Exadata 数据库云平台 X9M 弹性机架系统配合工作。通过 RoCE 线缆和内部交换机，您可以轻松地将多达 12 台 Oracle 数据库云平台 X9M 机架和 Exadata 存储扩展 X9M 机架连接在一起。一台 Exadata X9M-2 传统全机架配置¹，连同 11 个 Exadata 存储扩展柜 X9M 机架²可提供高达 47 PB (48,168 TB) 的磁盘裸容量和 7,136 个专用于 SQL 处理的 CPU 核。通过额外的 RoCE 交换机创建更大规模的配置。

在将新的 Exadata 存储扩展 X9M 机架连接基于 RoCE 的 Exadata 云平台后，系统的存储容量和性能均有提高。此时，系统可运行于单一系统映像模式，也可以通过逻辑分区整合多个数据库。使用 Exadata 数据库云平台和 Exadata 存储扩展机架可轻松实现横向扩展。自动存储管理 (ASM) 可以在系统联机状态下动态、自动地在各个 Exadata 存储服务器之间均衡分布数据，从而在各个机架之间均匀地分配 I/O 负载，充分利用所有硬件并将扩展机架轻松纳入整体配置之中。另外，还可以使用 I/O 资源管理器将 I/O 带宽按比例分配给系统中的各个数据库和用户以实现业务服务级别目标。

通过把数据密集型处理卸载到存储节点提供卓越性能

如今，数据量呈指数级增长，传统的存储阵列难以按所需的速率将数据从磁盘和闪存快速传输至数据库服务器以充分发挥 CPU 的处理能力。而配有多颗 CPU 的现代服务器每秒可以处理数十至数百 GB 的数据量，这种速度远远超出了传统架构存储阵列的存储机头和存储网络可以达到的水平。

Exadata 数据库云平台 X9M 的横向扩展架构不仅可提供高性能和高可扩展性，其采用的独有技术还可以将数据密集型 SQL 操作卸载到各个 Oracle Exadata 存储服务器。通过将 SQL 处理推送到 Exadata 存储服务器，从磁盘、闪存和持久内存中读取数据就可以立即在所有存储服务器上并行执行数据筛选和处理。存储系统只会将与查询直接相关的行和列发送至数据库服务器。

例如，如果某查询要查找三月份订单额超过 1000 美元的客户，那么 Exadata 系统的处理方式：将表扫描卸载至 Exadata 存储去执行，过滤掉金额小于 1000 美元的所有订单，过滤掉三月份以外的订单，仅提取相关的客户名称。最终传输至数据库服务器的数据量将减少数个量级。这将大幅加快查询执行速度、消除瓶颈和显著降低数据库服务器的 CPU 开销。

主要优势

- 该系统经过预先配置、预先测试并且针对所有数据库应用进行了优化
- 每机架来自 SQL 的未压缩 I/O 带宽高达 1,140 GB/秒
- 可横向伸缩的集成设计存储架构，预先配置为能够在联机状态下轻松扩展系统容量和性能
- 易于升级，可满足任意规模应用的需求
- 使用随带的混合列压缩功能时单机架可存储 13 PB 以上用户数据
- 无需使用外部交换机即可将多达 12 台 Exadata 数据库云平台和 Exadata 存储扩展机架连接在一起以扩展配置。通过额外的 RoCE 交换机创建更大规模的配置

¹ 采用 8 台 Exadata X9M-2 数据库服务器和 14 台 Exadata X9M-2 大容量存储服务器的传统全机架配置

² 采用 19 台 Exadata X9M-2 大容量存储服务器的满配 Exadata 存储扩展机架

每台 Exadata 存储服务器 X9M-2 搭载两颗 Xeon® x86 处理器用于处理数据卸载。配备 19 台存储服务器的单个 Exadata 存储扩展柜 X9M 机架中共包含 608 个处理器核用于数据库服务器卸载。Exadata 存储服务器中的 CPU 不是要取代数据库 CPU，而是用于加速数据密集型负载处理，其工作原理类似于显卡加速图像密集型负载。

通过压缩优化存储使用和 I/O

Exadata 存储服务器提供了一种独有的压缩功能，称为**混合列压缩 (HCC)**，可显著减少大型数据库的存储占用。混合列压缩技术是在数据库表中组织数据的一种创新性的方法，它结合使用行方法与列方法来存储数据。这种混合方法既可获得列存储的压缩优势，又可避免纯列格式的性能劣势。

利用混合列压缩，Exadata 可对 Oracle 数据库实现极高水平的数据压缩并减少 I/O，从而大幅降低成本并显著提高性能，这对于分析负载尤为有效。存储节省取决于具体的数据类型，通常在 5 至 20 倍之间。平均存储节省达到了 10 倍，处于行业领先水平。对于传统系统而言，较高的数据压缩会以降低性能为代价，因为这会增加 CPU 的解压缩负担。而对于 Exadata 数据库云平台，它能够将解压缩卸载至 Exadata 存储的处理器中，而且因实现了高度压缩而减少了 I/O，因此在多数情况下，使用混合列压缩可加快分析负载的运行速度。

混合列压缩支持两种模式。**仓库压缩**模式适用于读取密集型负载，可大幅节省存储，同时增强分析性能。**归档压缩**模式可提供极高的压缩率，适用于极少访问但仍须保持联机的数据。此外，这类数据现在可以无缝存储在 XT 存储服务器中，以便进一步降低成本。

在 OLTP 系统中，混合列压缩可用于压缩较旧且不太活跃的数据，而较新、活跃度较高以及经常需要更新的数据则可以采用高级行压缩技术进行压缩。Oracle Database 18c 及更高版本能够联机更改各个表分区所采用的压缩类型（即使表上使用了全局索引），以确保随着数据的老化和活跃度降低在不同压缩类型之间进行无缝分层。

针对数据分析，Exadata 实施了一种用于加快报告生成和分析查询的特殊算法，即**Exadata 列闪存缓存**。列闪存缓存功能在 Exadata 闪存中实现了一种双格式架构，当频繁扫描的混合列压缩数据加载到闪存缓存时，它自动将这些数据转换为纯列格式。在闪存中的纯列数据的智能扫描速度更快，因为扫描只读取选定的列，从而减少了闪存 I/O 和存储服务器 CPU 占用。这在保持 OLTP 单行查找的卓越性能的同时加快了报告和分析查询的速度。

相关产品

- Oracle Exadata 数据库云平台 X9M-2
- Oracle Exadata 数据库云平台 X9M-8
- Oracle Exadata 存储服务器 X9M-2
- Oracle 数据库企业版 11g、12c、18c 和 19c
- 真正应用集群 (RAC)
- 分区
- 多租户
- Database In-Memory
- 高级压缩
- 高级安全
- 活动数据卫士 (ADG)
- GoldenGate
- 真正应用测试 (RAT)
- OLAP
- 高级分析
- 商务智能
- Enterprise Manager
- Oracle Linux
- Oracle Linux Virtualization

相关服务

Oracle 提供下列服务：

- 高级客户服务
- Oracle 标准系统支持服务
- Oracle 白金服务
- Oracle 咨询服务
- Oracle 大学课程

基于 RDMA 的突破性网络结构

Exadata X9M 版本采用同样的超高速云级网络结构，即从 Exadata X8M 开始引入的 RDMA over Converged Ethernet (RoCE)。RDMA（远程直接内存访问）可允许一台计算机在不占用操作系统或 CPU 资源的情况下直接从另一台计算机内存中访问数据，从而实现高带宽和低延迟。网卡直接读取/写入内存，无需额外的复制或缓冲过程，而且延迟非常低。RDMA 是 Exadata 高性能架构中的一个重要组成部分，并且在过去十年中得到了持续调优和增强，可支持多种 Exadata 专有技术，例如 Exafusion Direct-to-Wire 协议和 Smart Fusion Block Transfer。由于 RoCE API 基础架构与 InfiniBand 的基础架构相同，因此**可以在 RoCE 上使用所有现有的 Exadata 性能特性。**

Exadata X9M 版本采用双端口 PCIe Gen 4 网卡，能够支持 2 个 100Gb/秒双活 RoCE 网络，总吞吐量为 200Gb/秒。这使得这款全球运行速度最快的数据库云平台变得更快。在采用新的共享持久内存加速器的 Exadata X9M 上运行的实际数据库负载已经打破了 Exadata X8M 之前创下的 1600 万次读 IOPS 的基准测试记录，达到 **OLTP 2760 万次读 IOPS (8K IO)³**。

共享持久内存加速

Exadata X8M 在闪存缓存前端引入了持久内存 (PMEM) 数据和提交加速器，可将访问远程存储数据的延迟降低若干个数量级。持久内存是一项现代芯片技术，可在 DRAM 与闪存之间添加一个极具性价比的大容量存储层。由于持久内存物理上存在于存储服务器的内存总线上，因此可直接以内存速度执行读取操作，其速度相比闪存大幅提升。与 DRAM 不同，写入操作是持久性的，在断电后仍然会保留。通过采用 RDMA 来远程访问持久内存，**Exadata 持久内存数据和提交加速器可绕过网络和 I/O 堆栈，消除成本极高的 CPU 中断和上下文切换，可将延迟减少 10 倍**，即从 200 μ s 缩短至不到 19 μ s。智能 Exadata 系统软件还可以确保跨存储服务器建立数据镜像，从而提供额外的容错能力。Exadata 在 Oracle 数据库与 Exadata 存储服务器之间实现了独有的端到端集成，可自动在数据库缓冲区缓存、持久内存和闪存缓存之间高效地缓存热度最高的数据块。将持久内存添加到存储层意味着此新缓存层的聚合性能可以由任何服务器上的任何数据库动态使用。与通用存储架构相比，这是一项显著的优势，可以避免跨数据库实例共享存储资源。

此外，智能 Exadata 系统软件特性还可以提高日志写入性能。日志写入延迟对于 OLTP 性能至关重要，更快的日志写入意味着更快的提交速度。相反，日志写入速度的任何下降都可能会导致数据库中断。借助基于 RoCE 的 Exadata，**Exadata 持久内存提交加速器**可以自动让数据库向持久内存发起单向 RDMA 日志写入。在 RDMA 和持久内存技术的支持下，无需确认即可执行日志写入操作，而且智能软件会将写入操作分布在多个服务器上，从而确保弹性。这提高了日志写入操作的性能。

³ 采用 10 台 Exadata X9M-2 数据库服务器和 12 台 Exadata X9M-2 极速闪存存储服务器的弹性配置

这个层还实现了自动化安全和管理流程。持久内存会在安装时自动完成配置，而不需要任何用户交互。硬件监视也经过预先配置。持久内存只能由使用数据库访问控制的数据库访问，从而确保数据的端到端安全性。在 Exadata X9M 中部署持久内存是一个非常简单的透明过程。



图 1. Intel® Optane™ 持久内存模块

极速闪存存储服务器：破纪录的 I/O 性能

Exadata **极速闪存 (EF) 存储服务器**是针对数据库进行了优化的全闪存 Exadata 数据库云平台的基础。每台 EF 存储服务器包含 8 个 6.4 TB Flash Accelerator F640v3 NVMe PCI 闪存驱动器，可提供 51.2 TB 的原始闪存容量，对于通常的数据库负载，其预期使用寿命可达到 8 年甚至更久。Exadata 将这些闪存设备直接置于高速 PCIe4 接口上（而不是在低速磁盘控制器的后方），可提供超强的性能。Exadata X9M 在闪存前端配有共享持久内存作为加速层，12 个 128 GB Intel® Optane™ 持久内存模块，可进一步提高性能。

Exadata X9M 结合使用了横向扩展存储、RDMA over Converged Ethernet 网络、数据库卸载、持久内存加速器和 PCIe 闪存，可提供极高的内存和闪存性能。当添加到 Exadata 数据库云平台 X9M-2 时，全配（机架）的 Exadata 存储扩展 X9M-2 配有 19 台极速闪存存储服务器，使用持久内存加速器，可实现高达**每秒 4370 万次的随机 8K 数据库读取 I/O 操作和 1160 万次的随机 8K 闪存写入 I/O 操作 (IOPS)**。

对于需要超高性能的数据仓库环境，Exadata X9M 极速闪存存储服务器能够以**每台服务器高达 75GB/秒的速度扫描**，扫描吞吐量增加高达 **1.39TB/秒⁴**⁵。

这些数据是在单机架 Exadata 系统中以标准 8K 数据库 I/O 规模运行 SQL 负载时测定的真实的端到端性能结果。Exadata 处理实际 Oracle 数据库负载的性能比传统存储阵列架构高若干数量级，也远非当今的全闪存存储阵列（其架构瓶颈会限制闪存吞吐量）可比。



图 2. 闪存加速器 PCIe 卡

⁴ 1425 GB/秒时测量的智能扫描速率

⁵ 采用 19 台 Exadata X9M-2 大容量存储服务器的满配 Exadata 存储扩展机架

大容量存储服务器：分层式磁盘闪存和持久内存以磁盘的成本提供共享内存性能

另一种 Exadata 存储选件是**大容量 (HC) 存储服务器**。该服务器包含 12 个 18 TB SAS 磁盘驱动器，磁盘原始总容量为 216 TB。它还包含 4 个闪存加速器 F640v3 NVMe PCIe 卡，可提供 25.6 TB 的闪存原始总容量。Exadata X9M 在闪存前端添加了共享持久内存加速层，12 个 128 GB Intel® Optane™ 持久内存模块，可进一步提高性能。使用智能软件部署 **Exadata Smart PMEM Cache**，只有热度最高的数据库块会自动缓存到此新层中。可通过 RDMA 直接从数据库访问，从而以极低的延迟提供极高的 I/O 速度。

HC 存储服务器中的闪存可直接用作闪存盘，不过一般都是将其配置在磁盘存储前、Exadata 持久内存数据加速器后的闪存缓存 (**Exadata 智能闪存缓存**) 来提供超强性能。Exadata 智能闪存缓存与 PMEM 缓存同步运行，可自动缓存频繁访问的数据，而将不常访问的数据保留在磁盘中。这既提供了闪存的高 I/O 速率和快速响应速度，还兼具磁盘的大容量和低成本优势。Exadata 作为唯一理解数据库负载的服务器，**知道何时不用缓存那些会对整体性能产生负面影响的数据**。例如，如果由备份或大表扫描引起的大规模写入 I/O 可能会中断更高优先级的 OLTP 或扫描操作，这些大规模 I/O 将绕开闪存缓存并直接进入磁盘。否则，Exadata 系统软件会缓存这些 I/O，以便于利用额外的备用闪存容量和 I/O 带宽来优化性能。除了自动缓存以外，管理员还可以使用 SQL 语句来确保将特定的表、索引或分区优先保存于闪存缓存中。

在实际的数据库负载处理中，Exadata 智能闪存缓存的命中率通常可以超过 95% 甚至达到 99%，产生的有效闪存容量可达到裸闪存容量的许多倍。例如，一个传统全配（机架）系统⁶ 提供的有效闪存容量往往接近于 900 TB 的可用磁盘容量。

此外，Exadata 智能闪存缓存还可以使用 Exadata 写回 (Write Back) 闪存缓存技术来缓存数据库块写入。写入缓存可消除大规模 OLTP 和批处理负载的磁盘瓶颈。使用 RoCE 网络连接到现有的 Exadata 数据库云平台时，一个全配 Exadata 存储扩展 X9M-2⁷ 的闪存写入容量高达**每秒 1160 万次 8K 闪存写入 IOPS**。Exadata 写入缓存具备透明性、持久性和完全冗余性，其性能相当于数十个由数千磁盘驱动器组成的企业磁盘阵列。

与其他基于闪存的解决方案相比，Exadata 的另一大优势是支持在 RAM、持久内存和磁盘之间自动进行数据分层。许多存储供应商已经开发了全闪存阵列，可实现比传统阵列更高的性能。这些全闪存式阵列在性能上有提升，而 Exadata 在磁盘与闪存之间实现的智能数据分层则具有强大的成本优势，这是因为闪存相对比较昂贵，其大小限制了可从中受益的总数据大小。此外，这些闪存阵列也无法从 Exadata 独有的数据库感知的存储优化技术中获益。某些闪存阵列提供的通用重复数据删除技术在虚拟化桌面基础设施环境中效果出众，但对数据库不起作用。

⁶ 采用 8 台 Exadata X9M-2 数据库服务器和 14 台 Exadata X9M-2 大容量存储服务器的传统全机架

⁷ 采用 19 台 Exadata X9M-2 大容量存储服务器的满配 Exadata 存储扩展机架

Exadata 不仅在容量上比一般全闪存式阵列大得多，其性能也更加出色。Exadata 采用经过优化的集成架构，采用 100 Gb/秒的 ROCE (RDMA over Converged Ethernet) 横向扩展的网络结构，配备极速 PCI 闪存，将数据密集型操作卸载至存储并且所有算法都专门针对数据库进行了优化，其吞吐量是全闪存式存储阵列所无法企及的。

扩展存储服务器容量：针对不常用数据的低成本 Exadata 存储

第三种 Exadata 存储选件是**扩展 (XT) 存储服务器**。每台 Exadata XT 存储服务器包含 12 个 18 TB SAS 磁盘驱动器，原始总磁盘容量为 216 TB。

此存储选件可将 Exadata 的运营和管理优势扩展到很少访问但又必须保持联机的数据。Exadata 扩展 (XT) 存储服务器具备以下特性：

- 高效 — XT 存储服务器和 HC 存储服务器一样提供很高的容量，并且具备混合列压缩功能
- 简单 — XT 存储服务器既能为 Exadata 增添容量，同时又能保持对应用和 SQL 透明，并能保持当前的工作模式
- 安全 — XT 服务器支持客户将用于联机数据的 Exadata 安全模式和加密功能扩展到不常用的数据
- 快速、可扩展 — 与其他针对不常用数据的存储解决方案不同，XT 存储服务器已经集成到 Exadata 结构中，因而支持快速访问，并且易于横向扩展
- 兼容 — XT 服务器只是另一种款式的 Exadata 存储服务器，客户可以将 XT 存储服务器添加到任何 Exadata
- 机架中成本低 — Exadata 存储软件许可证是可选的。

借助 Exadata 扩展 (XT) 存储服务器，企业可以一直使用值得信赖的并且持续经过验证的 Exadata 解决方案来满足其长期数据保留合规性需求，无需在多个平台上管理信息生命周期的运营风险和成本。在 Exadata 配置中结合使用极速闪存存储、大容量存储和扩展存储服务器，客户能够定义真正的信息生命周期管理策略。随着数据老化，数据可以在三层存储之间移动，确保将其存储在正确的介质上，以满足使用和保留要求。结合 Oracle 高级压缩的自动数据优化功能，客户可以定义相应的策略来自动执行这种数据移动和不同压缩级别变化。

极速备份与恢复性能

Exadata 存储扩展机架的一个重要优势是作为 Exadata 数据库云平台备份的目标。当对未压缩数据进行备份，将其写入 Exadata 存储扩展机架中的镜像磁盘时，能以高达每小时 48 TB 的速度创建完全数据库备份。对于增量备份，有效备份速度可提高至**数TB/小时**。当增量备份与混合列压缩结合使用时，有效备份速度可进一步提高至**数PB/小时**。

Exadata 存储扩展机架上的磁盘备份可以直接使用而不会影响性能，也不必进行还原操作。只有在备份到 Exadata 存储扩展系统时，才可以使用这种独特的备份功能。这是迄今为止备份和恢复 Oracle Exadata 数据库云平台的最快速、最简便的方法。

关键任务的高可用性

Exadata 数据库云平台经过精心设计，旨在提供极高水平的可用性。**该平台针对所有故障类型提供全方位防护**，从磁盘、服务器和网络之类的简单故障，到复杂的站点故障和人为错误，可防范所有故障。每个 Exadata 数据库云平台都采用**完全冗余的硬件配置**，包括冗余网络、冗余配电单元 (PDU)、冗余电源以及冗余数据库服务器和存储服务器。Oracle RAC 可防范数据库服务器故障。Oracle ASM 通过数据镜像防范磁盘或存储服务器故障。Oracle RMAN 可以极其快速高效地将数据备份到磁盘或磁带。Oracle 闪回技术支持回退数据库、表甚至行级的用户错误。使用 Oracle 数据卫士 (DG)，用户可以在高可用性架构 (MAA) 配置下部署第二个 Oracle Exadata 数据库云平台，在远程站点透明地维护数据库的实时副本，针对主数据库故障和站点灾难提供全面保护。

分析公司 IDC 认为，采用 MAA 配置的 Exadata 系统的**可用性达到了 99.999% 以上**，该系统被归为 IDC AL4 容错细分市场类别⁸。

Exadata 通过许多独特的方法来确保各种不同故障情况下的高可用性，这也体现了 Exadata 的软硬件深度集成的设计原则。**即时故障检测**就是这样一种独有功能。在非 Exadata 平台上，检测服务器故障需要停工等待很长的超时，从而导致应用长时间中断。基于 RoCE 的 Exadata 实现了基于 RDMA 的独有的**亚秒级节点死机检测**，几乎可完全消除应用中断的情况。

⁸ 年全球 AL4 服务器市场份额：容错系统成为数字化转型平台，IDC，Paul Magurani Peter Rutten，2020 年 7 月

磁盘和闪存设备有时会因内部故障扇区恢复、内部固件重启或耗损均衡而出现 I/O 操作延迟很长的情况。这些延迟很长的 I/O 操作可能会导致任务关键型 OLTP 数据库出现停滞。借助 Exadata **独有的 I/O 延迟限制** 功能，当读取 I/O 延迟大大超过预期时，Oracle Exadata 系统软件可自动将读取 I/O 操作重定向到数据的 ASM 镜像副本。同样，它可自动将高延迟的写入 I/O 操作重定向到正常运行的闪存设备，从而消除写入操作期间的异常。Exadata 系统软件利用机器学习技术来预测濒临故障的组件，并采取主动措施稳妥地停用此类组件。如果磁盘确实发生故障，ASM 对磁盘上保存的数据执行重新平衡操作，应用可无中断地持续访问数据库。Exadata 支持在线更换磁盘、风扇、电源和 PCIe 闪存卡，从而避免停机。Exadata 系统软件进一步改进了重新平衡，当在存储服务器之间移动数据时保留闪存缓存填充信息和存储索引以保持一致的应用性能。在极少数情况下，当网络子系统内发生延迟尖峰时，Exadata 将数据库服务器发出的 I/O 重定向到另一台存储服务器。

Exadata 自动监视 CPU、内存、I/O、子系统、文件系统和网络。这种自动化利用了机器学习技术与我们通过数千次重大实际部署积累下来的深厚经验。例如，Exadata 可以检测出因不当使用系统资源而对数据库性能产生负面影响的情况，自动识别出这是由哪个进程造成的，进而发出告警，对此用户无需进行任何人工干预。

由于具备行业领先的可用性，领先企业纷纷部署 Exadata 数据库云平台来支持其至关重要的应用，包括银行间资金转账、在线证券交易、实时呼叫追踪以及基于 Web 的零售。Exadata 关键任务的高可用性功能不仅适用于 OLTP 负载，也适用于数据仓库和分析负载。

最高服务级别

Oracle 为 Exadata 产品系列提供了全面的支持服务，包括：24x7 硬件服务、系统监视、软件安装和配置以及其他标准服务和定制服务。

此外还有专为 Oracle 集成系统提供的特别有用的 **Oracle 白金服务**。白金服务提供故障监视服务，具有更快的响应速度，并且可以在需要时将问题快速上报至开发部门。根据白金服务协议，Oracle 支持工程师将远程执行软件维护和修补。白金服务为集成系统内所有软硬件（包括 Oracle 数据库）提供比以往所享有的更高的支持级别。对于 Exadata 客户来说，白金服务无需另外付费。

IT 敏捷性

Exadata 为数据库的运行提供了一个完备的系统，包含存储、服务器和内部网络。传统数据库系统的管理工作通常由多个管理团队分担执行，如数据库团队、存储团队和系统管理团队，各自负责不同的组件。相比之下，**Exadata 系统的管理工作通常仅需一个统一的数据库云平台管理 (DMA) 团队即可完成**。数据库云平台管理员全面掌控 Exadata 数据库云平台中的所有资源，包括存储资源。数据库云平台管理员可以部署新数据库和实施配置更改，不再需要多个组件管理团队协同作业 — 这些团队通常都有繁重的工作并且工作优先级各不相同。这样，数据库云平台管理员可以将工作重点放在与应用和业务相关的增强任务上，而不是与各组件团队沟通协调，或者调优和诊断低级配置问题。

显著降低成本

由于 Exadata 数据库云平台可提供超强性能、大存储容量和独有的压缩功能，因此，原本需要超大型传统硬件系统的负载现在可以在小得多的 Exadata 系统上运行。在选型方面，与传统系统相比，Exadata 系统规模通常要小 2-4 倍。

Exadata 可为大型数据集提供巨大的 RAM、闪存和磁盘空间。一个全机架 Exadata 存储扩展⁹ 上的裸磁盘存储可达到 3.9 PB，而裸闪存存储可高达 972 TB。此外，混合列压缩通常可将存储和内存容量提升 10 倍。通过在磁盘、闪存和内存层之间智能地移动活跃数据，Exadata 不但极大地提高了性能，而且还将成本降至极低。

Exadata 可将支持多种负载的众多数据库整合到单一云平台中，这是一项**独有的能力**。高端 OLTP、分析、批处理、报告和备份操作都能以绝佳的性能同时在一个或多个数据库中运行。**Exadata 具备超强性能和超大容量，支持用户将大量数据库和负载整合到一个 Exadata 平台中**。将数据库整合到 Exadata 中不仅可以降低系统硬件成本、软件成本，还能大幅降低持续运营成本。

统一的 Exadata 数据库云平台配置有助于显著节省成本。**Exadata 不仅实现了技术标准化，而且还实现了集成、测试、安全性、增强、调优和支持的标准化**。与传统系统相比，客户部署 Exadata 系统要快得多，所需工作量也少得多。低级别的调优、集成和维护任务得以减少甚至完全消除。每家 Exadata 用户采用的配置都有数千家其他用户和 Oracle 内部团队在使用，由于众多用户采用同样的配置，用户遇到问题的可能性大为减少，有了问题也能轻松快捷地解决，从而能够降低运营成本和停机成本。在发生问题时，客户只需与 Oracle 一家供应商打交道，因为整个系统（硬件、固件、操作系统、虚拟机管理程序和数据库层）都由 Oracle 拥有和提供支持。这种一家供应商支持模式可加快问题解决速度并减少停机时间，从而进一步增加收益。

⁹ 采用 19 台 Exadata X9M-2 大容量存储服务器的满配 Exadata 存储扩展机架

Exadata 的业务效益

除了超强性能、高可用性、高安全性以及跨本地部署和云部署的部署灵活性等运营优势外，Exadata 还能直接创造业务效益。

由于大大缩短了系统配置、调优和测试所需的时间，**Exadata 可加快新业务应用的市场投放速度**。此外，部署时间也从数月缩短至数日，并大幅降低了上线后遇到意外系统级问题的风险。在部署新应用时，不可预测的应用使用模式往往会造成性能问题。Exadata 具有极高的 I/O、网络和计算吞吐量，可轻松承受不可预测的峰值负载，不会减慢任务关键型负载的响应速度。Exadata 总体上加快了应用部署速度并降低了风险，从而让企业加快了创新步伐。

Exadata 具备超强的性能以及庞大的内存和闪存容量，可大幅加快用户响应速度，从而提高员工效率和客户满意度。**用户将有更多的时间来进行有益的工作，而不是将时间浪费在等待系统响应上。**

Exadata 的超强性能不仅可以提高业务效率，还可以**帮助业务用户做出更明智的决策、发现增长机遇和降低成本**。用户可以实时分析数据、探索各种可能性以及通过快速迭代找到更优秀的解决方案。Exadata 可助力实现：

- 实时业务数据分析
- 更快的财务关账
- 更完善的规划和预算
- 更有效、更快速的预测

总结

Exadata 提供了一个完全集成的数据库平台，采用新的硬件技术和**独有的软件**来确保实现超强性能、高可用性和高安全性。此外，Exadata 还有助于降低成本、简化管理和增强可支持性，从而能够大大提高业务敏捷性和效率。凭借诸多优势，Exadata 当之无愧地成为了运行 Oracle 数据库新的全球标准 — 无论是本地部署还是云端部署。

Exadata 存储扩展 X9M 的关键容量和性能指标

指标	四分之一机架 高容量	四分之一机架 极速闪存	单台服务器 高容量 (HC)	单台服务器 极速闪存 (EF)	单台服务器 扩展 (XT)	最高配置 高容量	最高配置 极速闪存
存储服务器数量	4	4	1	1	1	19	19
闪存指标							
最高 SQL 闪存带宽 ²	180 GB/秒	300 GB/s	45 GB/秒	75 GB/秒	0 GB/秒	855 GB/秒	1425 GB/秒
最高 SQL 闪存读取 IOPS ³	9,200,000	9,200,000	2,300,000	2,300,000	0	43,700,000	43,700,000
最高 SQL 闪存写入 IOPS ⁴	2,456,000	2,456,000	614,000	614,000	0	11,666,000	11,666,000
PCI 闪存容量 (裸)	102.4 TB	204.8 TB	25.6 TB	51.2 TB	0.0 TB	486.4 TB	972.8 TB
磁盘指标							
最高 SQL 磁盘带宽 ²	7.2 GB/秒	不适用	1.8 GB/秒	不适用	不适用	34 GB/秒	不适用
最高 SQL 磁盘 IOPS ³	10,000	不适用	2,600	不适用	不适用	49,000	不适用
数据容量 (裸) ⁵	864 TB	204 TB	216 TB	51 TB	216 TB	4104 TB	973 TB
综合指标							
数据容量 (可用) — 常规冗余 ⁶	327.1 TB	75 TB	82 TB	19 TB	82 TB	1664 TB	382.5 TB
数据容量 (可用) — 高冗余 ⁶	256.6 TB	59.0 TB	64.1 TB	14.7 TB	64.1 TB	1218.7 TB	280.2 TB
最高数据加载速度 ⁷	10 TB/小时	10 TB/小时	2.5 TB/小时	2.5 TB/小时	2.5 TB/小时	47.5 TB/小时	47.5 TB/小时
<p>实际系统性能随应用而不同。</p> <p>¹ EF = 极速闪存; HC = 高容量</p> <p>² 该带宽是在不采用数据库压缩的情况下运行 SQL 时得到的峰值物理扫描带宽。在采用数据库压缩的情况下, 有效数据带宽将高于此值。</p> <p>³ 基于运行 SQL 的 8K IO 请求得出。请注意, IO 大小对闪存 IOPS 的影响很大。其他供应商报出的 IOPS 可能基于与数据库无关的、较小的 IO 大小。</p> <p>⁴ 基于运行 SQL 的 8K IO 请求得出。执行 ASM 镜像之后在存储服务器上测得的闪存写入 I/O, 这通常会发出多个存储 IO 来保持冗余性。</p> <p>⁵ 裸容量以标准磁盘驱动器术语来计量, 1 GB = 10 亿字节。测量可用容量时使用的是常规二进制空间术语, 1 TB = 1024 * 1024 * 1024 * 1024 字节。</p> <p>⁶ 该容量是在采用镜像 (ASM 常规冗余) 并保留足够空间 (四分之一机架和半机架上一个磁盘, 全机架两个磁盘) 以便在常规冗余情况下磁盘出现故障时重建镜像保护这种配置下的实际可用空间。</p> <p>⁷ 有效闪存容量大于裸闪存容量, 该数值考虑了 Exadata 智能闪存缓存算法实现的高闪存命中率以及底层磁盘存储的大小。它是指通常可以存储在 Exadata 上并且以闪存内存的速度访问的数据文件的大小。</p> <p>⁸ 加载速度通常只受限于数据库服务器 CPU, 而不会受限于 IO。速度因加载方式、索引、数据类型、压缩和分区而异。</p>							

Exadata 存储扩展 X9M 硬件

四分之一机架

4 台 Exadata 存储服务器 X9M-2:

- 128 个 CPU 核，用于 SQL 处理
- 48 个 128 GB 持久内存
- 对于高容量四分之一机架：48 个 18 TB 高容量驱动器和 16 个 6.4 TB NVMe PCI 闪存卡；
- 对于极速闪存四分之一机架：32 个 6.4 TB NVMe PCI 闪存驱动器
- 3 台 36 端口 100 Gb/秒 RoCE 交换机

其他硬件组件：

- 42U 机架
- 以太网交换机，用于提供管理连接
- 2 个冗余配电单元 (PDU)

随附的备件包内含：

- 1 个 6.4 TB NVMe PCI 闪存卡和 1 个 18 TB 高容量磁盘，或者
- 1 个 6.4 TB NVMe PCI 闪存驱动器

Exadata 存储扩展 X9M 的连接与升级

连接 Exadata 数据库云服务器

通过 RoCE 网络连接达 12 个任意组合的 Exadata 数据库云平台 X9M 机架或 Exadata 存储扩展 X9M 机架。

通过外部 RoCE 交换机创建更大规模的配置。所连接的机架可以是 X9M 或 X8M 代硬件。

升级能力

搭建初始四分之一机架之后，可以根据需要一次添加一个或多个 HC、EF、XT 或 HC、EF 与 XT 存储服务器的组合，最多可扩展至最高配置（19 台存储服务器）。

此升级包含的硬件组件包括：

- 用于连接所有组件的 RoCE、以太网线缆和适配器
- 对于每台额外的 HC 存储服务器：12 个 18 TB 高容量驱动器、4 个 6.4 TB NVMe PCI 闪存卡和 12 个 128 GB 持久内存模块；或者
- 对于每台额外的 EF 存储服务器：8 个 6.4 TB NVMe PCI 闪存卡和 12 个 128 GB 持久内存模块；或者
- 对于每台额外的 XT 存储服务器：12 个 18 TB 高容量驱动器

升级支持服务：

- 硬件安装和软件配置

Exadata 存储扩展 X9M 环境规格

指标	四分之一机架	最高配置	单台服务器
高度		78.74 英寸 (2000 毫米)	3.42 英寸 (86.9 毫米)
宽度		23.66 英寸 (601 毫米)	17.52 英寸 (445.0 毫米)
深度		47.13 英寸 (1197 毫米)	29.88 英寸 (759.0 毫米)
噪音 (工作)	9.2 B	9.5 B	8.0 B
采用大容量磁盘的环境			
重量	919.7 磅 (417.2 千克)	2069.2 磅 (938.6 千克)	76.7 磅 (34.8 千克)
最大功耗	20.1 kW (20.5 kVA)	17.2 kW (17.6 kVA)	0.8 kW (0.9 kVA)
常规功耗 ¹	3.2 kW (3.3 kVA)	12.1 kW (12.3 kVA)	0.6 kW (0.6 kVA)
最大负荷下的散热率	15,590 BTU/小时	58,839 BTU/小时	2,883 BTU/小时
	16,448 kJ/小时	62,075 kJ/小时	3,042 kJ/小时
常规负荷下的散热率	10,913 BTU/小时	41,187 BTU/小时	2,018 BTU/小时
	11,513 kJ/小时	43,453 kJ/小时	2,129 kJ/小时
最大负荷下的气流 ²	722 CFM	2724 CFM	133 CFM
常规负荷下的气流 ²	505 CFM	1907 CFM	93 CFM
采用极速闪存磁盘的环境			
重量	855.3 磅 (388.0 千克)	1763.3 磅 (799.8 千克)	60.6 磅 (27.5 千克)
最大功耗	4.6 kW (4.7 kVA)	17.4 kW (17.7 kVA)	0.9 kW (0.9 kVA)
常规功耗 ¹	3.2 kW (3.3 kVA)	12.2 kW (12.4 kVA)	0.6 kW (0.6 kVA)
最大负荷下的散热率	15,686 BTU/小时	59,293 BTU/小时	2,907 BTU/小时
	16,548 kJ/小时	62,554 kJ/小时	3,067 kJ/小时
常规负荷下的散热率	10,980 BTU/小时	41,505 BTU/小时	2,035 BTU/小时
	11,584 kJ/小时	43,788 kJ/小时	2,147 kJ/小时
最大负荷下的气流 ²	726 CFM	2745 CFM	135 CFM
常规负荷下的气流 ²	508 CFM	1922 CFM	94 CFM
工作温度/湿度: 5 °C 至 32 °C (41 °F 至 89.6 °F), 相对湿度 10% 至 90%, 无冷凝			
工作海拔: 最高 3048 米, 900 米以上每上升 300 米最高环境温度下降 1 °C			
¹ 常规功耗随应用负载而不同。			
² 气流方向必须从前往后。			

Exadata 存储扩展 X9M 相关法规和认证

法规 ^{1, 2, 3}	产品安全:	UL/CSA 60950-1、EN 60950-1、IEC 60950-1 CB Scheme (不同国家/地区之间存在差异)、UL/CSA 62368-1、EN 62368-1、IEC 62368-1 CB Scheme (不同国家/地区之间存在差异)
	EMC	
	辐射:	FCC CFR 47 第 15 部分、ICES-003、EN55032、KN32、EN61000-3-11、EN61000-3-12
	抗干扰性:	EN55024、KN35
认证 ^{2, 3}	北美 (NRTL)、CE (欧盟)、International CB Scheme、HSE Exemption (印度)、BSMI (中国台湾)、CCC (中华人民共和国)、EAC (EAEU 包括俄罗斯)、KC (韩国)、RCM (澳大利亚)、VCCI (日本)、UKCA (英国)	
欧盟指令 ³	2014/35/EU 低电压指令、2014/30/EU EMC 指令、2011/65/EU RoHS 指令、2012/19/EU WEEE 指令	

¹ 本文中引用的所有标准和认证均为最新官方版本。如需了解更多详细信息，请联系您的销售代表。

² 其他国家和地区的法规/认证也可能适用。

³ 有些情况下，仅在机柜级别的系统达到了法规和认证合规性（如适用）。

Exadata 存储扩展 X9M 支持服务

组件：

- 硬件保修：保修 1 年，正常营业时间（周一至周五早 8 点至晚 5 点）4 小时内通过 Web/电话进行回复，2 个工作日内到达现场/调换部件
- Oracle 标准系统支持服务包括 Oracle Linux 和 Solaris 支持服务以及 24x7 的 2 小时内现场硬件维修响应（维修地点需处于维修中心附近）
- Oracle 标准操作系统支持服务
- Oracle 客户数据与设备保留服务
- 系统安装服务
- 软件配置服务
- Oracle 白金服务
- 关键业务系统服务
- Oracle Exadata 启动包
- 系统升级支持服务，包括硬件安装和软件配置
- Oracle 自动服务请求 (ASR)

主要特性和功能

Exadata 和数据库软件特性 — 分析

- 自动并行处理数据扫描并将数据扫描卸载至存储（独有特性）
- 在存储中基于“where”子句筛选行（独有特性）
- 在存储中基于所选列筛选行（独有特性）
- 对 JSON 和 XML 分析查询进行存储卸载（独有特性）
- 在存储中基于与其他表的联接筛选行（独有特性）
- 混合列压缩（独有特性）
- 存储索引数据忽略（独有特性）
- 按用户、查询、服务、数据库等管理 I/O 资源（独有特性）
- 在闪存缓存中自动转换为列格式（独有特性）
- 对表扫描进行智能闪存缓存（独有特性）
- 对索引快速全扫描进行存储卸载（独有特性）
- 对加密数据扫描进行存储卸载并确保 FIPS 合规性（独有特性）
- 针对 LOB 和 CLOB 的存储卸载（独有特性）
- 针对最小值/最大值运算的存储卸载（独有特性）
- 针对数据挖掘的存储卸载（独有特性）

- 如果存储 CPU 繁忙，则反向卸载到数据库服务器
- 自动列化数据（独有特性）
- 当数据加载至闪存缓存时，自动将其转换为内存中格式（独有特性）

Exadata 和数据库软件特性 — OLTP

- 数据库感知式 PCI 闪存（独有特性）
- Exadata 智能闪存缓存（独有特性）
- Exadata 智能闪存日志（独有特性）
- 智能写回闪存缓存（独有特性）
- 按集群、负载、数据库或用户确定 I/O 优先级以确保服务质量（独有特性）
- Exafusion Direct-to-Wire 协议（独有特性）
- 数据库智能网络资源管理（独有特性）
- Exachk 全体系验证（独有特性）
- 全体系安全性扫描（独有特性）
- 数据库范围的安全性（独有特性）
- 通过单元到单元 (Cell-to-Cell) 重新平衡保留闪存缓存和存储索引（独有特性）
- 全体系安全擦除（独有特性）
- 即时创建数据文件（独有特性）
- 智能融合块传输（独有特性）
- 控制每个数据库的闪存缓存大小（独有特性）
- In-Memory OLTP 加速（独有特性）
- Undo 块远程 RDMA 读取（独有特性）
- 每个容器数据库通过多租户选项支持 4000 个可插拔数据库（独有特性）

Exadata 和数据库软件特性 — 高可用性

- 即时检测节点或单元故障（独有特性）
- 内存中容错（独有特性）
- 亚秒级故障切换停滞磁盘或闪存上的 I/O（独有特性）
- 将备份卸载到存储服务器（独有特性）
- Exadata 数据验证（扩展式 H.A.R.D.）（独有特性）
- 优先恢复关键数据库文件（独有特性）
- 通过读取其他存储服务器来自动修复损坏的磁盘数据（独有特性）
- 避免在预计存在故障的磁盘上执行读取 I/O 操作（独有特性）
- 限制使用、关闭再打开暂时性能低下的驱动器（独有特性）
- 当镜像存储服务器发生故障时提供关闭防护（独有特性）
- 检测并禁用不可靠的网络链路（独有特性）
- 重新平衡时保留存储索引（独有特性）

可管理性特性

- 具有升级预置优化的嵌入式 Oracle Integrated Lights Out Manager (ILOM)
- Oracle Enterprise Manager Exadata 插件
- 活动 AWR 包含存储统计信息，以支持端到端监视（独有特性）
- 针对以太网连接的 IPv6 支持
- 按需扩容
- 单元 (Cell) 软件透明重启
- 闪存和磁盘生命周期管理警报
- 自动磁盘清理和修复
- Oracle Linux Virtualization 的可信分区
- 自动创建 VLAN
- Oracle Exadata Deployment Assistant
- 单独的管理交换机和连接

- 从远程服务器执行 Exacli 命令行管理
- 通过 Cellcli 命令行管理存储服务器
- DCLI 分布式命令行自动化工具
- 自动服务请求和补丁管理程序 (patchmgr) 支持：
 - 数据库服务器
 - 存储服务器
 - 配电单元 (PDU)
 - Cisco RoCE 和管理交换机

Oracle 数据库软件（需单独购买）：

- 对于数据库服务器：Oracle Database 11g R2 企业版、Oracle Database 12c 企业版 R1 和 R2、Oracle Database 18c 企业版 R1 以及 Oracle Database 19c 企业版。Oracle 数据库选件，如 Oracle 真正应用集群 (RAC)、Oracle 分区、Oracle 多租户、Oracle 活动数据卫士 (ADG)。关于所支持的特性，请参阅具体版本的相应文档。
- 对于存储服务器：Oracle Exadata 系统软件。允许将许可从一个系统转移至另一个系统，或转移至新系统。

Oracle 软件（随机附带）：

- 对于数据库服务器：采用 Unbreakable Enterprise Kernel 5 的 Oracle Linux 7 Update 7。零丢失零复制数据报协议 (ZDP) RoCEv2 协议，用于 Exadata 存储服务器和 Oracle 数据库之间的通信。该协议基于可靠数据报套接字 (RDS) OpenFabrics 企业发行版 (OFED)

联系我们

请致电 **400-699-8888** 或访问 oracle.com/cn。中国地区的用户请访问 oracle.com/cn/corporate/contact/，查找您当地 Oracle 办事处的电话号码。

 blogs.oracle.com

 facebook.com/oracle

 twitter.com/oracle

版权所有 © 2021, Oracle 和/或其关联公司保留所有权利。本文档仅供参考，内容如有更改，恕不另行通知。本文档不保证没有错误，也不受其他任何口头表达或法律暗示的担保或条件的约束，包括对特定用途的适销性或适用性的暗示担保和条件。我们特别声明拒绝承担与本文档有关的任何责任，本文档不直接或间接形成任何契约义务。未经预先书面许可，不允许以任何形式或任何方式（电子或机械的）、出于任何目的复制或传播本文档。

Oracle 和 Java 是 Oracle 和/或其关联公司的注册商标。其他名称可能是其各自所有者的商标。

Intel 与 Intel Xeon 是 Intel Corporation 的商标或注册商标。所有 SPARC 商标均为 SPARC International, Inc. 的商标或注册商标，需经许可方可使用。AMD、Opteron、AMD 徽标以及 AMD Opteron 徽标是 Advanced Micro Devices 的商标或注册商标。UNIX 是 The Open Group 的注册商标。0321