



Oracle 白皮书  
2014 年 6 月

## Oracle Exadata 数据库云服务器和 Exadata 存储 服务器技术概述

引言 .....	2
Exadata 产品系列 .....	5
Exadata 工程化系统 .....	5
Exadata 数据库云服务器 .....	6
Exadata 存储服务器 .....	11
Exadata 存储扩展机架 .....	17
Exadata 数据库云服务器架构 .....	21
数据库服务器软件 .....	23
Exadata 存储服务器软件 .....	27
Exadata 智能扫描处理 .....	28
混合列压缩 .....	31
Exadata 智能闪存缓存特性 .....	32
Exadata 中的 I/O 资源管理 .....	33
Exadata 中的数据库网络资源管理 .....	35
Exadata 中的服务质量 (QoS) 管理 .....	35
Exadata 存储管理和数据保护 .....	36
总结 .....	40

## 引言

Oracle Exadata 数据库云服务器为运行 Oracle 数据库提供了一个性能和可用性最高的平台。Exadata 是一个现代化的架构，配有可横向扩展的行业标准数据库服务器、可横向扩展的智能存储服务器以及一个可连接所有服务器和存储的内部极速 InfiniBand 结构。Exadata 所用的独有软件算法在存储、基于 PCI 的闪存和 InfiniBand 网络中实现了数据库智能，因此，与其他平台相比，能以更低的成本实现更高的性能和容量。Exadata 可运行所有类型的数据库负载，包括联机事务处理 (OLTP)、数据仓库 (DW) 以及各种混合负载。Exadata 数据库云服务器可以简单快速地部署，能为您最重要的数据库提供全面的支持和保护，因此是整合数据库云的理想基础平台。

Exadata 数据库云服务器是一个易于部署的系统，包含了运行 Oracle 数据库所需的一切硬件。数据库服务器、存储服务器和网络均经过 Oracle 专家预先配置、调优和测试，其部署时间可比通常部署高性能系统所需时间节省数周或数月。广泛的端到端测试确保了所有组件能够无缝协同工作，并且系统中不存在可能影响整个系统的性能瓶颈或单点故障。

由于所有 Exadata 数据库云服务器均采用完全一致的配置方式，因此客户可从部署了 Exadata 数据库云服务器来运行任务关键应用程序的其他数千名用户获得丰富的经验。此外，客户所使用的机器与 Oracle 支持部门用于发现和解决问题的机器以及 Oracle 工程部门用于开发和测试 Oracle 数据库的机器也是完全一样的。由此可见，Exadata 平台不但为运行 Oracle 数据库进行了最彻底的测试和调优，还具备最高的可支持性。

Oracle Exadata 数据库云服务器运行标准的 Oracle 数据库。因此，如今使用 Oracle 数据库的任何应用程序都可无缝迁移至 Exadata 数据库云服务器平台，无需对应用程序进行任何更改。

Exadata 数据库云服务器采用一种可横向扩展的架构来承载数据库服务器和存储服务器。Exadata 配置会周密地平衡 CPU、I/O 和网络吞吐量，避免瓶颈问题。随着 Oracle Exadata 数据库云服务的增长，用户可以采用平衡的方式添加数据库 CPU、存储和网络，从而确保可扩展性而不造成瓶颈。横向扩展架构能适应任何规模的负载，允许从小规模配置无缝扩展到大规模的配置，同时还能避免性能瓶颈和单点故障。

Exadata 还采用了一项独特的技术，能够将数据密集型 SQL 操作分流到 Oracle Exadata 存储服务器。通过将 SQL 处理推送到 Exadata 存储服务器，当从磁盘中读取数据时就可以立即在所有存储服务器上并行执行数据过滤和处理。Exadata 存储分流降低了数据库服务器的 CPU 占用率，并大大减少了在存储服务器与数据库服务器之间移动的数据量。Exadata 智能闪存缓存通过加快 I/O 操作速度极大提高了 Oracle 数据库处理速度。闪存提供对数据库对象的智能缓存以避免物理 I/O 操作，此外，闪存还可以加速数据库日志操作。Exadata 存储提供一种高级压缩技术，即混合列压缩 (HCC)，此技术通常可提供 10 倍甚至更高层次的数据压缩，从而将有效数据传输速度提高一个数量级。Oracle Exadata 是世界上最安全的数据库云服务器。Exadata 存储基于卓越的 Oracle 数据库安全性功能而构建，能够查询完全加密的数据库，开销接近于零，并且查询速度可达到每秒数百 GB。Exadata 数据库云服务器的这些特性和许多其他特性共同构成了该产品杰出性能的基础。

借助 Exadata 存储扩展机架，可以提高 Exadata 数据库云服务器的 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，用于存储超出 Exadata 数据库云服务器容量的数据。在 Exadata 存储扩展机架中装备有标准 Exadata 存储服务器和支持基础架构，从而支持在 Exadata 数据库云服务器中轻松部署 Exadata 存储配置扩展。在使用 Exadata 存储扩展机架时，可获得和实现 Exadata 存储的所有优势和功能。

Exadata 数据库云服务器的设计使其既可与 Oracle Exalogic 中间件云服务器协同工作，也可独立于后者单独工作。Exalogic 中间件云服务器为运行 Oracle 融合中间件和 Oracle 融合管理软件提供了一个最佳运行平台。Exadata 和 Exalogic 的组合是一个全面的硬件和软件集成设计解决方案，可为包括 Oracle E-Business Suite、Siebel、PeopleSoft 应用程序在内的所有企业应用程序提供高性能。

Oracle SuperCluster 采用了 Exadata 存储技术，用来增强 Oracle 数据库的性能。SuperCluster 可用于托管 Oracle 融合中间件、Oracle 融合管理软件、通用应用程序以及 Oracle 数据库，这是一款基于 SPARC 服务器的高性能集成平台。它是一个工程化系统，旨在托管整个 Oracle 软件解决方案体系。除了 SuperCluster 中内置的 Exadata 存储服务器之外，该系统还可以使用 Exadata 存储扩展机架来增加容量和带宽。

## Exadata 产品系列

Exadata 产品系列的基础是 Oracle Exadata 数据库云服务器。Oracle Exadata 数据库云服务器是完全集成的完整数据库系统，它包含了用以快速、轻松部署任何具有最佳性能和最高可用性的企业数据库的所有组件。Exadata 存储服务器（Exadata 存储或 Exadata 单元）在数据库云服务器中用作 Oracle 数据库的存储。它所运行的 Exadata 存储服务器软件提供了独一无二、强大的 Exadata 技术，包括智能扫描、存储索引、智能闪存缓存、智能闪存日志、闪存缓存压缩、IO 资源管理器、网络资源管理和混合列压缩。借助 Exadata 存储扩展机架，可以快速简便地提高现有数据库云服务器或 SuperCluster 部署的 Exadata 存储容量和带宽。

### Exadata 工程化系统

Oracle Exadata 工程化系统的设计和构建旨在使其成为性能和可用性最高的 Oracle 数据库运行平台。用于运行 Oracle 数据库的传统自定义系统无法实现工程化系统可提供的性能和可用性水平。自定义数据库系统中使用的组件往往并不均衡并且配置不佳，这会造成瓶颈问题，从而降低整体系统性能。Exadata 系统采用集成设计并且经过端到端的优化，因而能够提供最佳性能。该工程化系统中的硬件组件、数据库软件和库、操作系统和设备驱动程序、固件、网络配置以及所有其他组件全部经过优化，能够协同工作。多年的调优经验使 Exadata 工程化系统能够提供运行 Oracle 数据库的最佳性能平台。

自定义数据库系统无法实现 Exadata 系统能够提供的可用性和正常运行时间。在自定义数据库系统中，复杂的组件间故障模式未经过测试，且各个组件也未集成为一个单元。而 Exadata 系统会处理所有可能的故障模式，包括节点故障、链路故障、存储故障和交换机故障。这意味着，该系统整体的正常运行时间更长，部署和运营风险更小。

Exadata 数据库云服务器是预先配置的系统，当天即可完成部署，从而在数据库部署过程中节省大量集成工作、成本和时间。Exadata 系统在交付时已经过组装、调试，随时可以运行。所有 Exadata 系统都相同，不存在特别的配置问题，因此所有 Exadata 用户都可获益于通用平台增强的可支持性。由于 Exadata 系统具有通用性，因此 Oracle 支持可以为数据库部署提供最佳支持。而自定义平台的构建和运营则需要顶尖人才。Exadata 让 IT 人员免于进行组件集成和测试工作，转而专注于企业业务需求。通过端到端的单一供应商支持以及对所有组件的端到端统一监视，Exadata 系统可以减少 IT 人员费用。

Exadata 系统可运行所有的现有 OLTP 和 DW 应用程序，可以现成地利用整整 30 年的 Oracle 数据库开发成果。在 Exadata 系统上部署应用程序数据库，不需要任何认证。Exadata 系统可以利用全部 Oracle 数据库生态系统 — IT 技能、人员、合作伙伴和技术。Exadata 系统甚至可以为最复杂的应用程序如 Oracle E-Business Suite、PeopleSoft、Siebel、SAP 以及内部构建的自定义应用程序透明地提供最佳的功能、性能和可用性。

### Exadata 数据库云服务器

Exadata 数据库云服务器有两种型号。Exadata 数据库云服务器 X4-2 可以从 2 个 24 核数据库服务器（最多 1 TB 内存）和 3 个 Exadata 存储服务器扩展到 8 个 24 核数据库服务器（最多 4 TB 内存）和 14 个 Exadata 存储服务器（全部在一个机架中）。Exadata 数据库云服务器 X4-8 包含 2 个 120 核数据库服务器（共有 4 TB 内存，可扩展到 12 TB）和 14 个 Exadata 存储服务器（在一个机架中）。X4-2 通过在一个机架中实现最大程度的可扩展性，提供到 Exadata 数据库云服务器系列的便捷入口点。X4-8 适用于需要大内存的大型部署或需要将多个数据库整合到一个系统中的情况。两个型号都运行 Oracle Database 11g 第 2 版和 Oracle Database 12c 数据库软件。



Exadata 数据库云服务 X4-2

### Exadata 数据库云服务 X4-2

Exadata 数据库云服务器 X4-2 有四种款型 — 全机架、半机架、四分之一机架和八分之一机架 — 具体取决于要部署的数据库的大小、性能和 I/O 要求。随着处理需求的增长，一种款型可联机扩展为另一种款型，确保提供顺畅的升级路径。此外，Exadata X4-2 可以轻松扩展到 18 机架网格，该网格具有 3,456 个 CPU 内核以及 12 PB 原始存储。所有 X4-2 数据库云服务器都包含以下组件：

- 行业标准数据库服务器，每个服务器预先配置有：运行频率为 2.7 GHz 的双插槽十二核 Intel® Xeon® E5-2697 v2 处理器，最大 512 GB 内存，四个 600 GB 10,000 RPM 磁盘，两个 40 Gb/秒 InfiniBand 端口，两个 10 Gb 以太网端口，四个 10/1 Gb 以太网端口，以及双冗余、可热插拔电源。数据库服务器上预先安装了 Oracle Linux 5 Update 9（含 Unbreakable Enterprise Kernel 2）和 Solaris 11 Update 1。可以在部署系统时为数据库云服务器选择需要的操作系统。
- Exadata 存储服务器，每台服务器预先配置有：运行频率为 2.6 GHz 的双插槽六核 Intel Xeon E5-2630 v2 处理器，96 GB 内存，3.2 TB Exadata 智能闪存缓存，12 个连接到一个存储控制器（512MB 带电池保护缓存）的磁盘，双端口 InfiniBand 连接，嵌入式 Integrated Lights Out Manager (ILOM) 以及双冗余、可热插拔电源。Exadata 存储服务器配备有 1.2 TB 高性能 10,000 RPM 磁盘或 4 TB 大容量 7,200 RPM 磁盘。在 Exadata 单元上预先安装了所有 Exadata 存储服务器软件。
- Sun 四倍数据速率 (QDR) InfiniBand 交换机和电缆，形成 40 Gb/秒 InfiniBand 结构以实现数据库服务器与 Exadata 存储服务器间通信以及 RAC 节点间通信。
- 以太网交换机，用于远程管理和监视数据库云服务器。
- 所有这些组件都封装到一个带配电装置 (PDU) 的定制 42U 机架中。

各种组件相互之间的数量配比经过精心选择，目的是为所有数据库应用程序提供最高的性能、高度可用的系统，以及 CPU 与 I/O 能力的最佳平衡。下表给出 Exadata 数据库云服务器 X4-2 每种款型的硬件组件数量。



	Exadata 数据库云服务器 X4-2 全机架	Exadata 数据库云服务器 X4-2 半机架	Exadata 数据库云服务 X4-2 四分之一机架	Exadata 数据库云服务 X4-2 八分之一机架 <sup>1</sup>
<b>数据库服务器</b>	8	4	2	2
• 用于数据库处理的 CPU 内核数	192	96	48	24
• 最大内存 (GB)	4,096	2,048	1,024	1,024
<b>Exadata 存储服务器</b>	14	7	3	3
• 用于 SQL 处理的 CPU 内核	168	84	36	18
• Exadata 智能闪存缓存 (TB)	44.8	22.4	9.6	4.8
• 用于数据库存储的磁盘数	168	84	36	18
<b>InfiniBand 交换机</b>	2	2	2	2
<sup>1</sup> 从物理上来讲，八分之一机架中的硬件与四分之一机架相同，但仅启用了一半的组件（数据库处理 CPU 内核、SQL 处理 CPU 内核、闪存和磁盘）。从八分之一机架升级到四分之一机架可通过软件激活硬件来完成。				

Exadata 数据库云服务器 X4-2 硬件

### Exadata 数据库云服务 X4-8

Exadata 数据库云服务器 X4-8 通过提供包含大型 SMP 数据库服务器的网格架构和 Exadata 存储网格，组合了出色的横向和纵向扩展架构。以前，一台大型 SMP 本身就需要全机架设备，难以进一步横向扩展。Exadata X4-8 使用两个基于 Intel 的超紧凑型 120 核 Sun 服务器，以建立高性能、高可用的数据库网格。每个服务器包括 2 TB 内存，40 Gb/秒 InfiniBand 以实现内部连接，以及 10 Gb 以太网以连接到数据中心。Exadata X4-8 可以轻松扩展到 18 机架网格，该网格具有 4,320 个 CPU 内核以及 12 PB 原始存储。新的 Exadata X4-8 可为所有业务应用程序提供极限性能，并支持大规模数据库整合。

Exadata 数据库云服务器 X4-8 以全机架配置的形式提供，运行 Oracle Database 11g 第 2 版或 12c，它包括以下技术。

- 两个行业标准数据库服务器，每个服务器预先配置有：运行频率为 2.80 GHz 的八插槽 15 核 Intel® Xeon® E7-8895 v2 处理器，2 TB 内存，七个 600 GB 10,000 RPM 磁盘，八个 40 Gb/秒 InfiniBand 端口，八个 10 Gb 以太网端口，十个 1 Gb 以太网端口，以及双冗余、可热插拔电源。数据库服务器上预先安装了含 Unbreakable Enterprise Kernel 的 Oracle Linux 5 Update 10。
- 14 个 Exadata 存储服务器，每个服务器预先配置有：运行频率为 2.6 GHz 的双插槽六核 Intel Xeon E5-2630 v2 处理器，96 GB 内存，3.2 TB Exadata 智能闪存缓存，12 个连接到一个存储控制器（512MB 带电池保护缓存）的磁盘（1.2 TB 高性能 10,000 RPM 磁盘或 4 TB 大容量 7,200 RPM 磁盘），双端口 InfiniBand 连接，嵌入式 Integrated Lights Out Manager (ILOM)，以及双冗余、可热插拔电源。在 Exadata 单元上预先安装了所有 Exadata 存储服务器软件。
- 两个 Sun 四倍数据速率 (QDR) InfiniBand 交换机和线缆，这些交换机和线缆形成了 40 Gb/秒 InfiniBand 结构以实现数据库服务器与 Exadata 存储服务器间通信以及 RAC 节点间通信。
- 以太网交换机，用于远程管理和监视数据库云服务器。
- 所有这些组件都封装到一个带配电装置 (PDU) 的定制 42U 机架中。

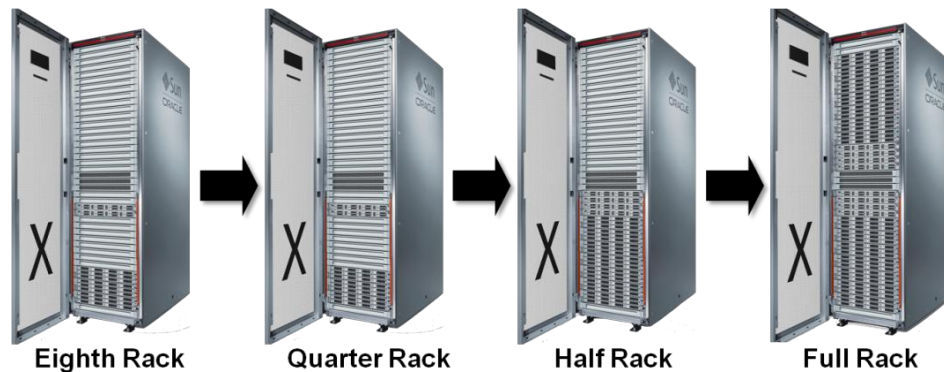
同样，各种组件相互之间的数量配比经过精心选择，目的是为所有数据库应用程序提供最高的性能、高度可用的系统，以及 CPU 与 I/O 能力的最佳平衡。

	Exadata 数据库云服务器 X4-8
<b>数据库服务器</b>	2
• 用于数据库处理的 CPU 内核数	240
• 最大内存 (GB)	12,288
<b>Exadata 存储服务器</b>	14
• 用于 SQL 处理的 CPU 内核	168
• Exadata 智能闪存缓存 (TB)	44.8
• 用于数据库存储的磁盘数	168
<b>InfiniBand 交换机</b>	2

Exadata 数据库云服务器 X4-8 硬件

### 数据库云服务器的升级能力

随着处理需求的增长，每种款型的 Exadata 数据库云服务器 X4-2 可以增加容量和能力，确保提供顺畅的升级路径。Oracle 人员可以轻松执行从八分之一机架到四分之一机架再到半机架再到全机架的现场联机升级。



数据库云服务 X4-2 升级

上述每种升级都可使用升级包。从物理上讲，八分之一机架中的硬件与四分之一机架相同，但仅启用了一半的主要组件（数据库处理 CPU 内核、SQL 处理 CPU 内核、闪存和磁盘）。从八分之一机架升级到四分之一机架可通过软件激活非活动硬件来完成。四分之一机架到半机架、半机架到全机架的升级包括让系统支持下一级更大型配置的必需组件，这些组件由 Oracle 支持负责进行安装。

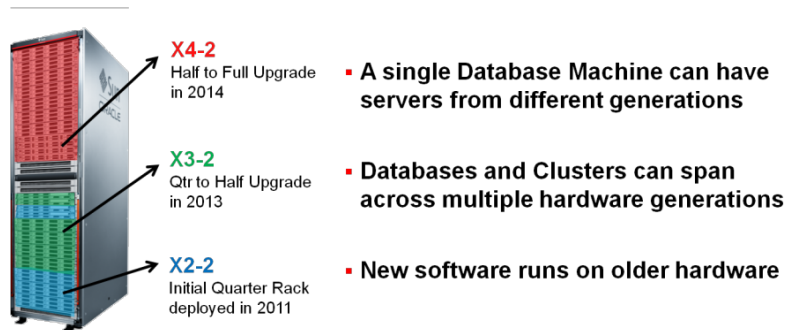
Exadata 数据库云服务器是极其强大的系统，但是仍然可使用构建块方法来让 Exadata 数据库云服务器扩展到几乎任意规模。我们可以使用系统中的 InfiniBand 结构连接多个 Exadata 数据库云服务器 X4-2 系统，以构成单个规模更大的系统映像配置。类似地，也可以连接多个 Exadata 数据库云服务器 X4-8 机架。因为所有 InfiniBand 基础架构（交换机和端口布线）在设计上都提供了这种扩展的可能，所以可以通过在机架之间连接 InfiniBand 线缆实现这种扩展。用户只需连接 InfiniBand 线缆并安装内部 InfiniBand 交换机即可将最多 18 个机架连接在一起。而通过额外的外部 InfiniBand 交换机则可以建立规模更大的配置。用户可以对 X4-2 全机架和半机架的任意组合进行连接。在两种情况下可以将四分之一机架与其他机架相连。两个四分之一机架彼此相连，或一个四分之一机架与全机架和半机架的一种配置相连。Exadata 数据库云服务器的这种固有的扩展功能可支持任何应用程序所需的最大规模的数据库。

此外，可以用相同的方式利用同样的 InfiniBand 结构将 Exalogic 中间件云服务器连接到 Exadata 数据库云服务器。在无需任何外部交换机的情况下，最多可以将十八个全机架 Exalogic 和 Exadata 系统连接在一起。



八个连接在一起的 Exadata 数据库云服务器 X4-2 机架形成一个系统

Exadata 数据库云服务器允许将新一代处理器和 Exadata 存储无缝部署到现有 Exadata 数据库云服务器中，因此能保护您的投资。同样，新版软件也与早前的 Exadata 数据库云服务器兼容。下面列举一个跨多代 Exadata 数据库云服务器升级和扩展的例子。



X2-2 四分之一机架升级到含 X3-2 服务器的半机架，然后升级到含 X4-2 服务器的全机架

## Exadata 存储服务器

Exadata 存储服务器运行 Exadata 存储服务器软件，该软件提供 Exadata 数据库云服务器所独有的强大软件技术，其中包括智能扫描、智能闪存缓存、智能闪存日志、IO 资源管理器、存储索引和混合列压缩。

Exadata 存储服务器（也称作 Exadata 单元）的硬件组件经过精心选择，可以满足高性能数据库处理的需要。同时，Exadata 软件经过优化，可以最充分地利用硬件组件和 Oracle 数据库。每个 Exadata 单元都为数据库提供了杰出的 I/O 性能和带宽。Exadata 存储服务器中的 CPU 内核专用于提供在 Exadata 存储中执行的一些特性，如智能扫描 SQL 处理。

Exadata 存储基于每个 Oracle 数据库的高度安全性功能而构建，能够查询完全加密的数据库，开销接近于零，并且查询速度可达到每秒数百 GB。这是因为它将解密处理从软件转移到了 Exadata 存储服务器硬件。Exadata 存储服务器中使用的 Oracle 软件和 Intel 处理器提供了高级加密标准 (AES) 支持，从而帮我们实现了这一目标。



Exadata 存储服务器 (Exadata 单元)

### Exadata 智能闪存缓存

Exadata 系统采用最新的 PCI 闪存技术，而非闪存磁盘。PCI 闪存将闪存直接置于高速 PCI 总线上（而非设置在低速磁盘控制器和管理器的后方），可提供超强的性能。每台 Exadata 存储服务器都包含 4 个 PCI 闪存卡，其闪存的总原始容量为 3.2 TB。每台全机架 Exadata 数据库云服务器 X4-2 或 X4-8 都包含 56 个 PCI 闪存卡，可提供 44.8 TB 的原始物理闪存容量。

Exadata 闪存可直接用作闪存磁盘，不过一般都是将其配置为磁盘前端的闪存缓存，因为缓存可为更多的数据提供闪存级性能，而闪存可直接存储的数据有一定限制。

Exadata 智能闪存缓存可自动将频繁访问的数据缓存到 PCI 闪存中，同时将不常访问的数据保留在磁盘驱动器中。这样，仅利用大容量、低成本的磁盘就可实现闪存级性能。Exadata 智能闪存缓存可识别数据库负载，知道何时应避免缓存数据库较少访问或太大不适于缓存的数据。例如，Exadata 可以判断 I/O 何时用于备份、何时用于表扫描、何时用于存储将很快被删除的临时结果。除了自动缓存以外，管理员还可以选择提供 SQL 指令来确保将特定的表、索引或分区始终保存于闪存中。表可以保留在闪存中，不必像传统存储那样必须移至不同的表空间、文件或 LUN。

Exadata 智能闪存缓存可自动将活跃的 IO 密集型数据移入闪存，将活跃度较低、IO 活动较少的数据保留在磁盘中，从而为较大的数据（通常可达到机器物理闪存容量的数倍）提供闪存级的 IO 速率、吞吐量和响应时间。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 90% 甚至达到 98%，而闪存容量却不到磁盘容量的十分之一。这种高命中率意味着 Exadata 智能闪存缓存可提供比物理闪存缓存大 10 倍以上的有效闪存容量。例如，全机架 Exadata 数据库云服务器 X4-2 的有效闪存容量高达 440 TB。

除了智能缓存带来的优势以外，Exadata 智能闪存缓存压缩还可以透明地压缩存入闪存缓存的用户数据，从而动态增加闪存缓存的容量。这样，用户可以在闪存中存放更多数据，进一步降低在磁盘驱动器上访问数据的需求。这种压缩和解压操作对于应用程序和数据库是完全透明的。Exadata 智能闪存缓存压缩通过硬件加速提供零性能开销的压缩和解压，即使在每秒运行数百万次 I/O 操作或每秒处理数百 GB 数据时也是如此。

闪存缓存压缩带来的收益因用户数据的可压缩性而异。处理未经压缩的表可实现最高的压缩率。索引的压缩率一般也非常不错。此外，在高级行和基本表压缩的基础上，Exadata 智能闪存缓存压缩仍然可以大幅扩充闪存缓存。尽管 OLTP 应用已经采用高级行压缩，其闪存缓存的总逻辑大小一般也可以再翻一倍。采用混合列压缩或 LOB 压缩的表经过了高度压缩，因而再进一步压缩的空间非常小。全机架 Exadata 数据库云服务器 X4-2 在启用闪存缓存压缩时，能提供高达 90 TB 的逻辑闪存缓存容量（未考虑数据库级压缩时）。全机架 Exadata 数据库云服务器 X4-8 也同样如此。

闪存性能通常会受到传统存储架构的约束和限制。而 Exadata 结合使用了横向扩展存储、InfiniBand 网络、数据库分流和 PCI 闪存，可提供超强的闪存性能。运行数据库负载时，单个全机架 Exadata 数据库云服务器 X4-2 实现了高达 100 GB/秒的数据扫描带宽速度，而每秒随机 8K 读取 I/O 操作 (IOPS) 最高可达 2,660,000 次。这样的性能比传统数据库架构高数个数量级。值得注意的是，这些数据是在单机架 Exadata 系统中以实际 IO 规模运行 SQL 负载时测定的真实的端到端性能结果，不是基于低级别 IO 工具进行的组件级测试。

使用写回闪存缓存特性，Exadata 智能闪存缓存还缓存数据库块写入。写入缓存可消除大规模 OLTP 和批处理负载的磁盘瓶颈。单个全机架 Exadata 数据库云服务器 X4-2 的闪存写容量的每秒 8K 写入 I/O 超过 1,960,000 次。Exadata 写入缓存透明、持久且完全冗余。Exadata 智能闪存缓存的 I/O 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。

为了进一步加快 OLTP 负载的处理速度，Exadata 智能闪存缓存还实施了一个专用于减少日志写入 I/O 延迟的特殊算法，即 Exadata 智能闪存日志。用户事务提交时间和关键更新执行时间对日志写入延迟非常敏感。智能闪存日志利用 Exadata 存储中的闪存缓存和 Exadata 磁盘控制器中的高速 RAM 内存来大幅减少日志写入延迟，避免其他闪存解决方案经常遇到的延迟高峰。Exadata 智能闪存日志算法是 Exadata 独有的特性。

Exadata 只采用闪存制造商设计的企业级闪存，具备极高的耐用性。因为 Exadata 专门针对关键任务负载而设计，所以不会采用在使用多年后会性能下降或发生意外故障的消费级闪存。对于大多数数据库负载，Exadata X4 中所采用的企业级闪存芯片的使用寿命可达到 10 年或更长时间。

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、闪存和磁盘之间自动进行数据分层。当数据库服务器直接使用第三方闪存卡或闪存磁盘时，闪存中的数据将只能由该服务器使用，因为本地闪存无法在多个服务器之间共享。这将导致无法使用 RAC 并且会将数据库部署规模限制为单一服务器，从而制约数据库的性能、可扩展性和整合能力。单一服务器中的任何组件故障（例如闪存卡）都会导致数据库访问丢失。本地闪存不支持 Exadata 所提供的智能闪存缓存和混合列压缩技术，并且会大幅增加管理复杂性。

实际经验表明，服务器本地闪存卡和闪存磁盘在遇到组件故障时会功能受损 (crippled)，从而导致数据库挂起、性能下降甚至系统崩溃。闪存产品无法避免的问题包括间歇性挂起、性能定期下降和供电期数据丢失，并且这些故障通常不会触发错误或警报来使闪存产品置于离线。更糟糕的是，这些问题有时会导致操作系统内部出现挂起，从而导致整个节点挂起或崩溃。Exadata 软件会自动检测并绕过性能较差或功能受损的闪存。检测到异常状况时，Exadata 会自动将 I/O 操作路由至备用存储服务器。

很多存储供应商已经认识到传统存储阵列架构本身限制了闪存性能的发挥，从而研发了新的全闪存阵列。这些全闪存阵列虽可提供比传统阵列更高的性能，但却舍弃了对磁盘和闪存

进行智能数据分层的成本优势。因此，受益于闪存的总数据大小将受到实际闪存大小的限制，而闪存的成本又非常高。Exadata 智能闪存缓存可自动将活跃的 IO 密集型数据置于闪存中，将活跃度较低、IO 活动较少的数据保留在成本较低的磁盘中，这通常可为物理闪存的十倍数据提供闪存级性能。数据库和闪存缓存压缩技术能进一步扩大 Exadata 闪存的容量。第三方闪存阵列不能从 Exadata 混合列压缩获益。

Exadata 不仅可提供比全闪存阵列高得多的容量，还可实现更高性能。全闪存存储阵列无法利用完全基于 InfiniBand 的扩展、快速 PCI 闪存、将数据密集型操作分流到存储及专门针对数据库进行了优化的算法来匹配 Exadata 经过集成、优化的架构所支持的吞吐量。

#### **Exadata 存储容量、性能、带宽和 IOPS**

每个 Oracle Exadata 存储服务器均配备 12 个 1.2 TB 10,000 RPM 高性能磁盘或 12 个 4 TB 7,200 RPM 大容量磁盘。基于高性能磁盘的 Exadata 存储服务器可提供高达 6 TB 的未压缩可用容量，以及高达 1.75 GB/秒的原始数据带宽。基于大容量磁盘的 Exadata 存储服务器可提供高达 20 TB 的未压缩可用容量，以及高达 1.5 GB/秒的原始数据带宽。当以压缩格式存储时，每个单元提供的用户数据量和数据带宽可显著增加。

下表说明数据库云服务器每种款型的存储容量。



	Exadata 数据库云服务器 X4-8 和 X4-2 全机架	Exadata 数据库云服务 X4-2 半机架	Exadata 数据库云服务器 X4-2 四分之一机架	Exadata 数据库云服务 X4-2 八分之一机架
Exadata 智能闪存缓存	44.8 TB	22.4 TB	9.6 TB	4.8 TB
原始磁盘容量				
• 大容量磁盘	672 TB	336 TB	144 TB	72 TB
• 高性能磁盘	200 TB	100 TB	43.2 TB	21.6 TB
可用容量	高达	高达	高达	高达
• 大容量磁盘	300 TB	150 TB	63 TB	30 TB
• 高性能磁盘	90 TB	45 TB	19 TB	9 TB
(未使用数据压缩)				

Exadata 数据库云服务器存储容量

注意：在计算原始磁盘容量时，1 TB = 1 万亿字节。实际的格式化容量要小一些。可用于数据库的可用容量是在考虑了镜像（ASM 常规冗余）并且保留一个空磁盘以自动应付磁盘故障之后计算得出的容量。

由于采用了 Exadata 智能闪存缓存，每个单元可提供极高的性能。Exadata 软件可以同时从闪存和磁盘进行扫描以实现最大带宽。闪存中的自动缓存使每个 Exadata 单元能够在访问未压缩数据时提供高达 5.4 GB/秒的带宽和 190,000 次数据库 8K IOPS。当数据以压缩格式存储时，可实现的用户数据容量、数据带宽和 IOPS 通常可提高多达 10 倍甚至更高。与 Oracle 数据库使用的传统存储设备相比，Exadata 存储实现了重大改进。

下表说明数据库云服务器每种款型的性能特点。

	Exadata 数据库云服务 X4-8 全机架	Exadata 数据库云服务 X4-2 全机架	Exadata 数据库云服务 X4-2 半机架	Exadata 数据库云服务器 X4-2 四分之一机架	Exadata 数据库云服务器 X4-2 八分之一机架
原始闪存数据带宽 (未使用数据压缩)	高达 100 GB/秒	高达 100 GB/秒	高达 50 GB/秒	高达 21.5 GB/秒	高达 10.7 GB/秒
数据库闪存读取 IOPS <sup>1</sup>	高达 2,660,000	高达 2,660,000	高达 1,330,000	高达 570,000	高达 285,000
数据库闪存写入 IOPS <sup>1</sup>	高达 1,960,000	高达 1,960,000	高达 980,000	高达 420,000	高达 210,000
原始磁盘数据带宽	高达	高达	高达	高达	高达
• 大容量磁盘	20 GB/秒	20 GB/秒	10 GB/秒	4.5 GB/秒	2.25 GB/秒
• 高性能磁盘	24 GB/秒	24 GB/秒	12 GB/秒	5.2 GB/秒	2.6 GB/秒
(未使用数据压缩)					
数据库磁盘 IOPS <sup>1</sup>	高达	高达	高达	高达	高达
• 大容量磁盘	32,000	32,000	16,000	7,000	3,500
• 高性能磁盘	50,000	50,000	25,000	10,800	5,400

Exadata 数据库云服务器 X4-2 和 X4-8 I/O 性能

## Exadata 存储扩展机架

我们精心设计的 Oracle Exadata 存储扩展机架 X4-2 能让您以最简便、最迅速、最可靠的方式为 Exadata 数据库云服务器或 SuperCluster 添加更多存储容量。作为 Exadata 数据库云服务的一个自然延伸，Exadata 存储扩展机架可用于满足最大的任务关键数据库的大数据需求。

Exadata 存储扩展机架专为需要海量数据的数据库部署而设计，所涉及的数据种类有：历史或存档数据、Exadata 数据库云服务器数据的备份和存档、文档、图像、文件和 XML 数据、LOB 以及其他大型非结构化数据。由于无需配置 LUN 或挂载点，扩展机架的配置极为简便。通过一些简单的命令，几分钟内就能配置好存储并将其添加到数据库中。

Exadata 数据库云服务器的性能优势来源于一种独特的技术，即 Oracle Exadata 存储服务器及其软件。通过将数据库处理工作推移到 Exadata 存储服务器中进行，所有磁盘能够以并行方式进行工作，减少了对数据库服务器 CPU 的占用，大幅减少了在存储和数据库服务器之间移动数据时对带宽的占用。Exadata 存储扩展机架由标准 Exadata 存储服务器和 InfiniBand 交换机组成，可以与您的 Exadata 数据库云服务器无缝集成。Exadata 存储扩展机架是一个高性能、大容量、高带宽、可横向扩展的存储解决方案，可为您的 Exadata 数据库云服务器部署带来高达 387 TB 的未压缩的镜像可用容量以及相应更高的 I/O 带宽。

Exadata 存储扩展机架现有三种型号可供使用，分别是：包含 18 台 Exadata 存储服务器的全机架配置；包含 9 台 Exadata 存储服务器的半机架配置；包含 4 台 Exadata 存储服务器的四分之一机架配置。对于任何应用程序都有合适的配置供您选择。随着处理需求的增长，一种型号可联机升级为另一种型号，确保提供顺畅的升级路径。这三种型号的扩展机架均装配有相同的 Exadata 存储服务器（1.2 TB 高性能磁盘或 4 TB 的大容量磁盘，以及 Exadata 智能闪存缓存），与 Exadata 数据库云服务器中使用的相同。

	Exadata 存储扩展机架 X4-2 全机架	Exadata 存储扩展机架 X4-2 半机架	Exadata 存储扩展机架 X4-2 四分之一机架
<b>Exadata 存储服务器</b>	18	9	4
• 用于 SQL 处理的 CPU 内核	216	108	48
• Exadata 智能闪存缓存 (TB)	57.6	28.8	12.8
• 用于数据库存储的磁盘数	216	108	48
<b>InfiniBand 交换机</b>	3	3	2

Exadata 存储扩展机架硬件

除了从小型 Exadata 存储扩展机架升级到大型机架外，Oracle 还继续使用一种构建块方法，用这种方法，您可以通过集成式 InfiniBand 结构将 Exadata 存储扩展机架连接到 Exadata 数据库云服务器，从而将您的系统轻松扩展至任意规模。Exadata 存储扩展全机架、半机架和四分之一机架几乎能以任意组合纳入到 Exadata 数据库云服务器的全机架、半机架和四分之一机架系统中。通过 InfiniBand 电缆，您可以轻松地将 18 个 Exadata 数据库云服务器机架和

Exadata 存储扩展机架连接在一起。具有一个 Exadata 数据库云服务器全机架 X4-2 和 17 个 Exadata 存储扩展全机架的 18 机架 X4-2 配置，其原始磁盘容量为 15,360 TB，具有 3,840 个专用于进行 SQL 处理的 CPU 内核。而通过额外的 InfiniBand 交换机甚至可以建立规模更大的配置。

	Exadata 存储扩展机架 X4-2 全机架	Exadata 存储扩展机架 X4-2 半机架	Exadata 存储扩展机架 X4-2 四分之一机架
Exadata 智能闪存缓存	57.6 TB	28.8 TB	12.8 TB
原始磁盘容量			
• 大容量磁盘	864 TB	432 TB	192 TB
• 高性能磁盘	258 TB	129 TB	57 TB
可用容量	高达	高达	高达
• 大容量磁盘	387 TB	194 TB	85 TB
• 高性能磁盘	116 TB	58 TB	25 TB
(未使用数据压缩)			

Exadata 存储扩展机架容量

Exadata 存储扩展机架在大数据处理方面极具优势，这里以其作为 Exadata 数据库云服务器备份之目标为例来进行说明。在 Exadata 存储扩展机架中，如果对写入镜像磁盘的未压缩数据进行备份，整个数据库备份的创建速度可高达 27 TB/小时。如果执行数据库增量备份，备份速度则高达数百 TB/小时；如果对混合列压缩数据进行增量备份，备份速度高达数 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，也不必进行恢复。只有在备份到 Exadata 存储扩展机架时，才可以使用这种独特的备份功能。这是迄今为止最快速、最简便的 Oracle Exadata 数据库云服务器备份和恢复方法。

在将新的 Exadata 存储扩展机架连接到 Exadata 数据库云服务器后，系统的存储容量和性能将会同步提高。此时，系统可运行于单系统映像模式，还可进行逻辑划分以便整合多个数据库。Exadata 数据库云服务器和 Exadata 存储扩展机架组合方案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、自动地在各台 Exadata 存储服务器之间均衡分布数据。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

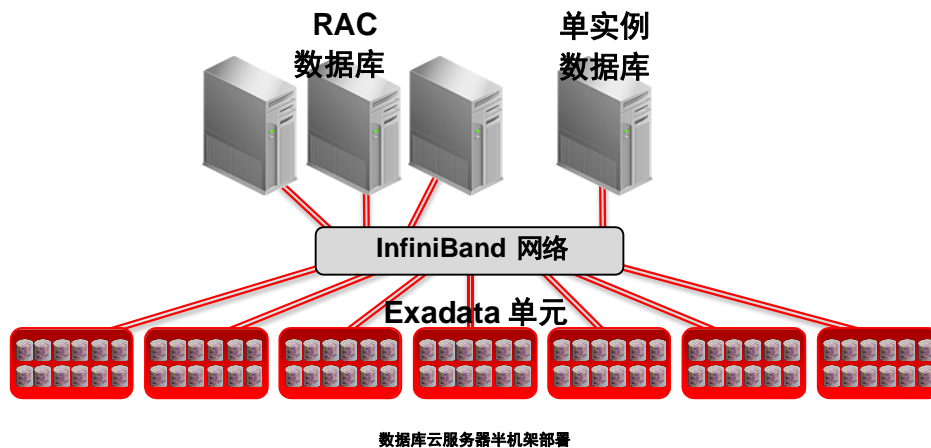
松纳入整体配置之中。另外，还可以使用 I/O 资源管理器将 I/O 带宽按比例分配给系统中的各个数据库和用户以实现业务服务级别目标。

	Exadata 存储扩展机架 X4-2 全机架	Exadata 存储扩展机架 X4-2 半机架	Exadata 存储扩展机架 X4-2 四分之一机架
原始闪存数据带宽 (未使用数据压缩)	高达 130 GB/秒	高达 65 GB/秒	高达 29 GB/秒
数据库闪存读取 IOPS <sup>1</sup>	高达 3,420,000	高达 1,710,000	高达 760,000
数据库闪存写入 IOPS <sup>1</sup>	高达 2,520,000	高达 1,260,000	高达 560,000
原始磁盘数据带宽 • 大容量磁盘 • 高性能磁盘 (未使用数据压缩)	高达 26 GB/秒 30 GB/秒	高达 13 GB/秒 15 GB/秒	高达 6 GB/秒 7 GB/秒
数据库磁盘 IOPS <sup>1</sup> • 大容量磁盘 • 高性能磁盘	高达 42,000 64,000	高达 21,000 32,000	高达 9,500 14,400
<sup>1</sup> 基于运行 SQL 的 8K IO 请求得出。			

Exadata 存储扩展机架 I/O 性能

## Exadata 数据库云服务器架构

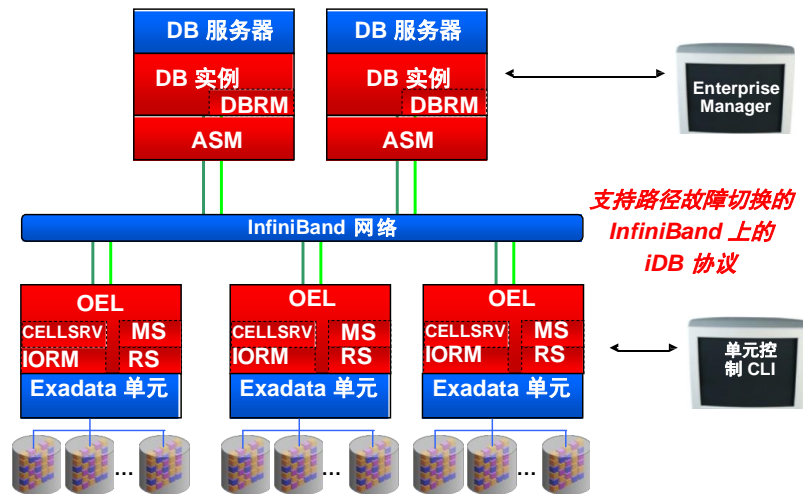
下图中是一个典型的数据库云服务器半机架部署的简化示意图。图中显示了两个 Oracle 数据库，其中一个是在跨三个数据库服务器部署的 Real Application Clusters (RAC) 数据库，另一个是在半机架配置中剩余的那个数据库服务器上部署的单实例数据库。（当然，为了在半机架上提供最佳性能和可用性，可将所有四个数据库服务器用于单个四节点 RAC 集群。）RAC 数据库可以用作生产数据库，而单实例数据库可以用于测试和开发。这两个数据库共享半机架配置中的七个 Exadata 单元，但是它们可以有各自独立的 Oracle 主目录以保持软件独立性。此配置的所有组件 — 数据库服务器、Exadata 单元、InfiniBand 交换机和其他支持硬件全部安置在数据库云服务器机架中。



数据库云服务器在服务器和存储之间使用最新的 InfiniBand 互连。每个数据库服务器和 Exadata 单元具有双端口四倍数据速率 (QDR) InfiniBand 连接，以实现高可用性。每个 InfiniBand 链路提供 40 Gb 带宽 — 比传统存储或服务器网络高出许多倍。此外，Oracle 的互连协议使用直接数据放置方法 (DMA — 直接内存访问)，将数据直接从线路移动到数据库缓冲区而无需进行额外的数据复制，从而确保极低的 CPU 开销。InfiniBand 网络兼具 LAN 网络的灵活性和 SAN 的效率。通过使用 InfiniBand 网络，Oracle 可确保网络不会成为性能瓶颈。该 InfiniBand 网络同时还为 Oracle 数据库 Real Application Cluster (RAC) 节点提供高性能集群互连。

Oracle Exadata 的架构设计旨在可横向扩展到任何级别的性能。为了实现更高的性能和更大的存储容量，可以在系统配置中添加更多的数据库服务器和 Exadata 单元，例如，可以从半机架升级到全机架。随着在系统配置中添加更多的 Exadata 单元，存储容量和 I/O 性能几乎呈线性提高。在 Exadata 配置中，从不进行也不需要单元到单元通信。

Exadata 解决方案的架构包括数据库服务器中的组件和 Exadata 单元中的组件。下图所示为四分之一机架配置的软件架构。



Exadata 软件架构

当使用 Exadata 时，大量 SQL 处理从数据库服务器分流到 Exadata 单元。除了为数据库提供传统的块服务之外，Exadata 还实现了从数据库实例到底层存储的功能传送。与传统存储相比，Exadata 存储的一个独到之处就是仅返回满足数据库查询的行和列，而不是整个被查询的表。Exadata 尽可能将 SQL 处理向数据（或磁盘）推进，并使所有磁盘并行操作。这可减少数据库服务器上的 CPU 使用，在数据库服务器和存储服务器之间移动数据时使用更少的带宽，并返回查询结果集而不是全部表。消减数据传输和数据库服务器负载可大大有益于传统上受限于带宽和 CPU 的数据仓储查询。免去数据传输对常常包括大量批处理和报表处理操作的联机事务处理 (OLTP) 系统也非常有益。

Exadata 对使用数据库的应用程序完全透明。数据库云服务器上运行的 Oracle Database 11g 第 2 版或 Oracle Database 12c 与传统系统上运行的完全相同 — 但在数据库云服务器上运行速度更快。当使用 Exadata 存储时，现有 SQL 语句（无论是即席语句还是打包或自定义应用程序中的语句）不受影响，不需要进行任何修改。该解决方案无需对应用程序进行任何

修改即可提供分流处理和带宽优势。Exadata 完全支持 Oracle 数据库的所有特性。Exadata 既适用于 Oracle 数据库的单实例部署，又适用于 Real Application Cluster 部署。无论采用还是不采用 Exadata，对 Oracle Data Guard、Oracle Recovery Manager (RMAN)、Oracle GoldenGate 等功能和其他数据库工具的管理都是相同的。用户和数据库管理员仍可利用他们目前熟悉的同样的工具和知识，因为一切就像运行传统的非 Exadata 存储一样。

由于数据库云服务器上存在与传统系统上相同的 Oracle 数据库和功能，因此管理数据库云服务器的 IT 人员必须拥有同一软件的类似知识，以便在数据库云服务器上对其进行管理。在管理数据库云服务器时，拥有 Oracle 数据库管理、备份及恢复、RAC 和 OEL 经验是十分重要的。

### 数据库服务器软件

Oracle 数据库软件已得到显著增强，以便利用 Exadata 存储。Exadata 软件在数据库服务器和 Exadata 单元之间进行了最佳分割。数据库服务器和 Exadata 存储服务器软件使用 iDB（智能数据库协议）进行通信。iDB 在数据库内核中实现，它将数据库操作透明地映射到 Exadata 增强操作。除了数据库提供的传统数据块传送之外，iDB 还实施功能传送架构。iDB 用于将 SQL 操作向下传送到 Exadata 单元中执行，并将查询结果集返回到数据库内核。Exadata 单元仅返回满足 SQL 查询的行和列，而不是返回数据库块。与现有 I/O 协议一样，iDB 也可以直接从磁盘读取以及向其写入字节范围，因此当无法进行分流处理时，Exadata 会和传统存储设备一样为 Oracle 数据库提供服务。但是，在可以进行分流处理时，数据库内核中的智能能够将诸如表扫描等操作向下传递到 Exadata 存储服务器上执行，因此只有请求的数据会返回到数据库服务器。

iDB 基于开放标准可靠数据报套接字 (RDS) 协议构建，并在 InfiniBand 上运行。ZDP（零损失、零复制数据报协议）— RDS 的一种零复制实现，用于消除不必要的块复制。在数据库服务器和 Exadata 单元上可以使用多个网络接口。这是一种极快的低延迟协议，可最大程度减少为 I/O 操作提供服务所需的数据副本数。

Oracle 自动存储管理 (ASM) 用作 Exadata 的文件系统和卷管理器。ASM 可虚拟化存储资源，并提供 Exadata 的高级卷管理和文件系统功能。在各个可用 Exadata 单元和磁盘上均匀地条带化数据库文件可在所有存储硬件之间产生均衡的 I/O 负载。ASM 执行非侵入性资源分配和重新分配的能力是在 Exadata 环境中实现共享网络存储功能的关键因素。ASM



提供的磁盘镜像与可热插拔 Exadata 磁盘一起，共同确保数据库可以容许单独磁盘驱动器出现故障。在单元之间进行数据镜像以确保单元故障不会导致数据丢失或阻碍数据的可访问性。这种大规模并行架构可提供无限可伸缩性和高可用性。

Oracle 数据库的 Database Resource Manager (DBRM) 特性已得到增强，以便用于 Exadata 环境。DBRM 允许用户定义和管理数据库内和数据库间的 I/O 带宽和 CPU、撤销、并行度、活动会话以及其他可管理的资源。这允许在数据库之间共享存储，同时不必担心一个数据库会独占 I/O 带宽并影响共享存储的其他数据库的性能。为用户组分配某一百分比的可用 I/O 带宽，而 DBRM 可确保实现这些目标。这通过数据库针对关联的数据库和用户组标记 I/O 来实现。这通过整个 I/O 系统为数据库提供完整的 I/O 优先级视图。数据库内用户组 I/O 分配在数据库服务器上定义和管理。数据库间 I/O 分配在 Exadata 单元中的软件内定义，并由 I/O Resource Manager (IORM) 管理。Exadata 单元软件确保在数据库之内和之间管理和正确分配数据库间 I/O 资源。总的来说，DBRM 确保每个数据库接收其指定的 I/O 资源量，并确保满足用户定义的 SLA。

Oracle 数据库文件系统 (DBFS) 由 Oracle Database 11g 第 2 版引入，可以提供对 Oracle 数据库中存储文件的高性能文件系统访问。DBFS 基于 SecureFiles 而构建，SecureFiles 特性由 Oracle Database 11g 第 1 版引入，可以提供高性能 SQL 以及对 Oracle 数据库中存储文件的编程访问。当与在数据库中存储文件的应用程序以及有关这些文件的元数据一同使用时，DBFS 和 SecureFiles 非常有用，特别是，DBFS 支持基于文件的工具和实用程序，允许它们对数据库中存储的文件进行操作。

DBFS 经过优化，可用于 Exadata 数据库云服务器环境：与高端存储阵列相比，它提供非常高的文件吞吐量，并允许在 Exadata 系统上直接实施数据仓储 ETL（提取、转换和加载）暂存处理。DBFS 还为在 Exadata 系统上部署的所有数据库之间可以共享的应用程序文件提供空间。而其他供应商的数据仓储产品为了执行这些功能需要客户购买额外付费的独立存储阵列来进行部署和管理。使用 DBFS，Exadata 客户可以在简化其数据仓储环境的同时提升性能。

Oracle 数据库的两个特性 — Oracle Database Quality of Service (QoS) Management 和 QoS Management Memory Guard，允许系统管理员直接管理 Oracle Exadata 数据库云服务器上托管的应用程序服务级别。使用基于策略的架构，QoS 管理将准确的运行时性能与资源量度

关联，使用专家系统分析此数据以识别瓶颈，并提出资源调整建议以便在动态负载情况下满足和保持性能目标。如果没有足够的可用资源，则 QoS 将通过牺牲不太关键的目标来保持更为业务关键的目标。QoS Management Memory Guard 与 Cluster Health Monitor 协作，可以检测出由于内存过量使用而存在故障风险的节点。作为应对办法，它自动阻止新连接以便保留现有负载，并在有了足够的可用内存之后恢复连接。

#### **Enterprise Manager 为 Exadata 数据库云服务器提供支持**

Oracle Enterprise Manager Cloud Control 12c 全盘管理 Exadata 数据库云服务器，提供对整个工程化系统从监视到管理和后续维护的全面生命周期管理。

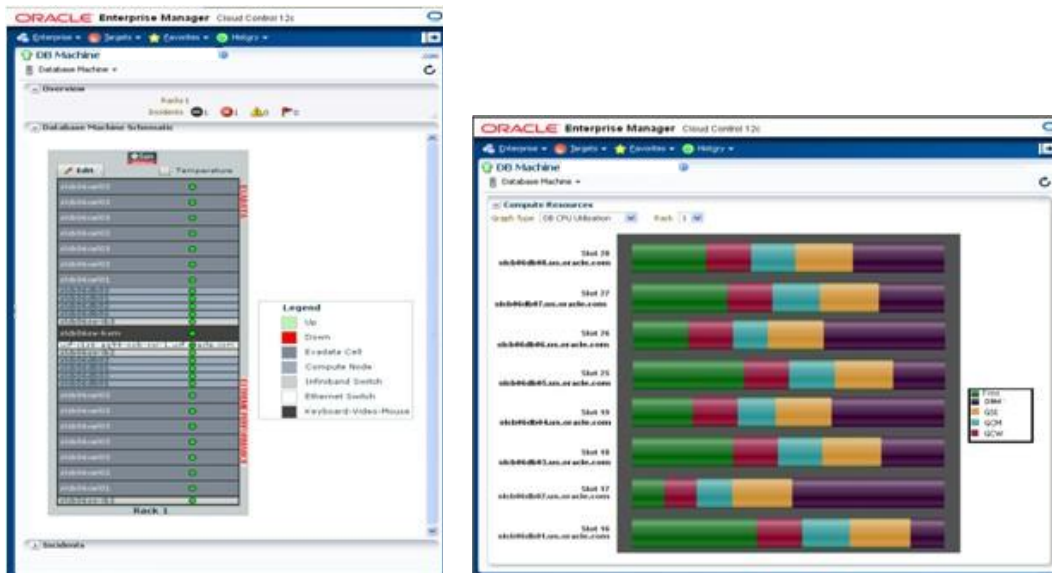
#### **集成式系统监视**

Oracle Enterprise Manager 提供全面的监视和通知，使管理员能够主动检测和响应 Oracle Exadata 数据库云服务器及其软硬件组件中存在的问题。管理员可以轻松调整这些监视设置，以满足其数据中心环境的需求。在获得警告通知后，管理员可以轻松查看警告历史记录以及问题组件的相关性能指标，例如 InfiniBand 端口的网络性能或 Exadata 存储单元的磁盘活动情况，从而确定问题的根本原因。通过直接连接 Exadata 硬件组件，Oracle Enterprise Manager 可以向管理员发出硬件相关故障的警报，并通过与 Oracle Automatic Service Requests (ASR) 集成来自动记录服务请求，以便 Oracle 支持人员立即查看。

原先在传统系统中需要数据库管理员、系统管理员和存储管理员共同协作才能诊断出来的问题，现在由于实现了对整个 Exadata 的集成式系统监视而在几分钟内便可诊断出来。

#### **多合一管理**

Oracle Enterprise Manager 为 Oracle Exadata 硬件和软件提供一个统一的视图，在该视图中，您可以查看所有组件的运行状况和性能，如数据库服务器、InfiniBand 交换机、Exadata 存储单元、Oracle 数据库、ASM 等组件。



使用 Enterprise Manager Cloud Control 12c 监视 Exadata

Oracle 数据库不经任何更改即可透明地在 Oracle Exadata 数据库云服务器上运行。但有时 DBA 需要从数据库下钻到存储系统，以发现和诊断性能瓶颈或硬件故障。Enterprise Manager 对 Exadata 硬件和软件的集成视图使 DBA 能够从数据库性能页面无缝导航到相关 Exadata 存储服务器以便隔离问题，无论这些问题是由于硬件组件造成的，还是由于同一存储子系统上运行的其他数据库造成的。SQL 监视功能能够实时分析 SQL 执行的性能，该功能是 Exadata 感知的，能够确定被分流到 Exadata 存储服务器的执行计划的计划运行情况，从而使 DBA 能够了解 SQL 语句的效率。

Enterprise Manager 根据特定受管组件的运行状况和性能特点来提供 Exadata 管理功能。例如，除了监视 InfiniBand 网络的性能外，管理员还可以在 Enterprise Manager 检测到端口降级时更改端口的设置。在 Exadata 存储单元上，如果管理员发现某个特定数据库占用了过多的 I/O 资源，从而影响到同一组存储单元上的其他数据库的性能，他们可以配置和激活 Enterprise Manager 中的 I/O 资源管理器计划。

## 整合规划

随着企业日益寻求将分散的数据库整合到 Oracle Exadata 基础架构中，管理员可以使用 Oracle Enterprise Manager 中的 Consolidation Planner 来确定不同 Exadata 配置下的最佳整合策略。利用实际硬件配置和 Enterprise Manager 中存储的服务器负载历史记录，Consolidation Planner 对源系统的负载进行分析，并计算整合计划在目标 Exadata 系统上的预期资源使用。Consolidation Planner 配备有一个丰富的硬件配置库，可以指导管理员为甚至是虚拟 Exadata 服务器（从 X4-2 到 X4-8 的不同型号）定义整合方案。现在，企业可以就其 Exadata 的确切配置制定更加明智和最佳的决策以适合其数据库整合需求。

对于 Oracle Exadata 数据库云服务器，管理经过集成设计，可与系统软硬件协同工作，这种集成设计不仅带来了卓越的高性能和高可用性，还简化了管理和整合工作。

## Exadata 存储服务器软件

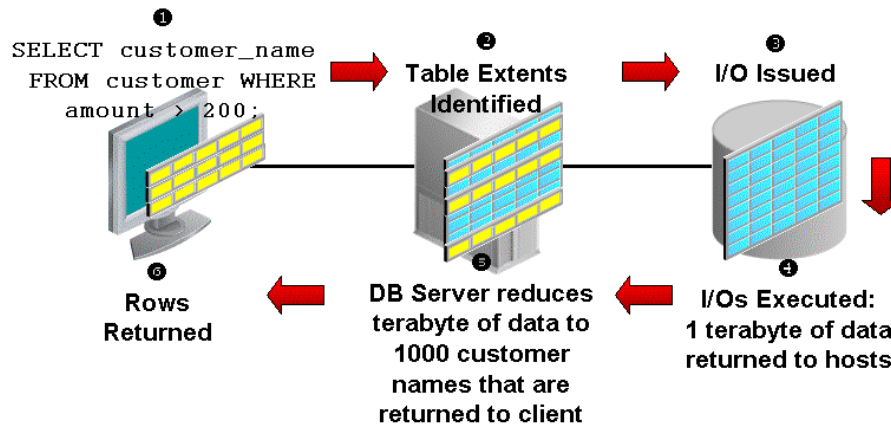
与任何存储设备一样，Exadata 存储服务器是一台具有 CPU、内存、总线、磁盘、NIC 以及服务器中常见的其他组件的计算机。它同样运行操作系统 (OS)，在 Exadata 环境中该操作系统是 Oracle Linux。驻留在 Exadata 单元中的 Exadata 存储服务器软件运行于含有 Unbreakable Enterprise Kernel 的 OL 之下。可以在受限模式下访问 OL 以管理 Exadata 单元。

CELLSRV (Cell Services) 是单元中运行的 Exadata 软件的主要组件，它提供大部分 Exadata 存储服务。CELLSRV 是多线程软件，它基于 iDB 协议与数据库服务器上的数据库实例通信，并为数据库提供块服务。它提供高级 SQL 分流功能，在无法进行 SQL 分流处理时提供 Oracle 块服务，并且执行 DBRM I/O 资源管理功能以测量发出 I/O 的不同数据库和用户组的 I/O 带宽。

单元中运行的 Oracle 软件的其他两个组件是 Management Server (MS) 和 Restart Server (RS)。MS 是管理和查询 Exadata 单元状态的主要接口。它与 Exadata 单元命令行界面 (CLI) 和 EM Exadata 插件协同工作，并且提供独立的 Exadata 单元管理和配置。例如，从单元中发出 CLI 命令以配置存储、查询 I/O 统计信息以及重新启动单元。此外，还提供了分布式 CLI，从而可以将命令发送到多个单元以简化对多个单元的管理。Restart Server (RS) 确保 Exadata 软件和服务持续运行。它用于更新 Exadata 软件。它还确存储服务启动和运行，并在需要时重新启动服务。

## Exadata 智能扫描处理

使用非 iDB 感知的传统存储时，所有数据库智能都存在于服务器上的数据库软件中。下面以表扫描为例，说明在这样的架构中是如何执行 SQL 处理的。



传统数据库 I/O 和 SQL 处理模型

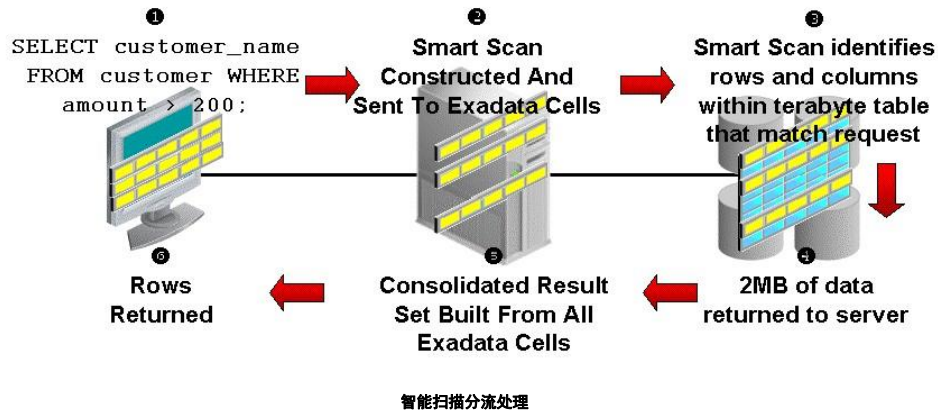
- ❶ 客户端发出一条 SELECT 语句，该语句带有一个谓词以筛选并仅返回想要的行。
- ❷ 数据库内核将此请求映射到包含被扫描表的文件和盘区。
- ❸ 数据库内核发出 I/O 以读取块。
- ❹ 将被查询表的所有块读取到内存。
- ❺ 然后对原始块进行 SQL 处理，搜索满足谓词的行。
- ❻ 最后将这些行返回到客户端。

对于大量查询来说，谓词常常会筛选掉读取的大多数行。但是，表中的所有块都需要读取、通过存储网络传输并复制到内存。读取到内存的行远多于完成请求的 SQL 操作所需的行。这将产生大量数据传输，这些传输会占用带宽并影响应用程序吞吐量和响应速度。

在数据库体系的存储层中集成数据库功能之后，可以更高效地执行查询和其他数据库操作。尽可能地靠近硬件实现数据库功能（对于 Exadata 来说是在磁盘一级），可以显著加快数据库操作速度并提高系统吞吐量。

使用 Exadata 存储，可以更高效地处理数据库操作。在 Exadata 存储中可以处理执行表扫描的查询，并且仅将所需的数据子集返回到数据库服务器。Exadata 存储单元可以执行行筛选、列筛选和一些联接处理（以及其他一些功能）。在这种情况下，仅将相关和所需的数据返回到数据库服务器。

下图说明使用 Exadata 存储时是如何进行表扫描操作的。



① 客户端发出一条 SELECT 语句，该语句带有一个谓词以筛选并仅返回想要的行。  
 ② 数据库内核确定 Exadata 存储可用，于是构造一个 iDB 命令来代表发出的 SQL 命令，然后将这个 iDB 命令发送到 Exadata 存储。  
 ③ Exadata 软件的 CELLSRV 组件扫描数据块以找出那些满足所发出的 SQL 命令的行和列。  
 ④ 仅将满足谓词的行和请求的列读取到内存。  
 ⑤ 数据库内核整合来自所有 Exadata 单元的结果集。  
 ⑥ 最后，将这些行返回到客户端。

智能扫描对于应用程序来说是透明，并且不需要进行任何应用程序或 SQL 更改。从 SQL EXPLAIN PLAN 中可以看出何时使用了 Exadata 智能扫描。返回的数据是完全一致的、事务性的，并且严格遵循 Oracle 数据库一致读取功能和行为。如果某单元在智能扫描期间停工，则未完成的智能扫描部分透明地路由到另一个单元来完成。智能扫描正确地处理 Oracle 数据库的复杂内部机制，其中包括：未提交的数据和锁定行、链接行、压缩表、国家语言处理、日期算术、正则表达式搜索、物化视图以及分区表。

Oracle 数据库和 Exadata 服务器协作执行各种 SQL 语句。将 SQL 处理从数据库服务器中移出可释放服务器 CPU 周期，并消除大量带宽使用以便更好地服务于其他请求。SQL 操作运行速度更快，并且更多的操作可以同时运行，因为 I/O 带宽争用减少。下面我们来看看因使用 Exadata 而受益的各种 SQL 操作。

### 智能扫描谓词筛选

Exadata 支持表扫描谓词筛选。仅将请求的行返回到数据库服务器，而不是表中的所有行。例如，当发出以下 SQL 时，仅将员工聘用日期晚于指定日期的行从 Exadata 发送到数据库实例。

```
SELECT * FROM employee_table WHERE hire_date > '1-Jan-2003';
```

这种仅将相关行返回到服务器的功能大大提高了数据库性能。对于更为复杂的查询来说，也存在这种性能提高，因此同样有益于复杂查询，包括带有子查询的查询。

### 智能扫描列筛选

Exadata 为表扫描提供了列筛选，也称为列投影。仅将请求的列返回到数据库服务器，而不是表中的所有列。例如，当发出以下 SQL 时，仅将 employee\_name 和 employee\_number 列从 Exadata 返回到数据库内核。

```
SELECT employee_name, employee_number FROM employee_table;
```

对于具有许多列或列中包含 LOB（大型对象）的表来说，可以节省大量 I/O 带宽。当配合使用时，谓词和列筛选可显著提高性能并减少 I/O 带宽使用。此外，列筛选也适用于索引，从而使查询速度更快。

### 智能扫描联接处理

Exadata 在大型表和小型查找表之间执行联接，这是使用星型模式的数据仓库的一个常见情形。这使用布隆筛选器实施，布隆筛选器是一种非常高效的概率方法，用于确定行是否为所需结果集的成员。

### 加密表空间和加密列的智能扫描处理

Exadata 存储支持加密表空间 (TSE) 和加密列 (TDE) 的智能扫描分流处理。这样可提高访问企业中最机密数据时的性能。

### 存储索引

存储索引是 Exadata 存储中提供的非常强大的功能，该功能有助于避免 I/O 操作。Exadata 存储服务器软件在 Exadata 单元中创建和维护存储索引（即有关数据库对象的元数据）。存储索引跟踪该单元上存储的表的各列的最小值和最大值。如果查询指定了 WHERE 语句，则在执行任何 I/O 之前，Exadata 软件将查看存储索引，以便通过比较列值与存储索

引中维护的最小值和最大值来确定具有指定列值的行是否存在于单元中。如果列值在最小值和最大值范围之外，则避免为该查询执行扫描 I/O。许多 SQL 操作的运行速度将显著提高，因为大量 I/O 操作自动由几个查找操作取代。为了最大程度减少运营开销，Exadata 存储服务器软件透明、自动地创建和维护存储索引。

### 数据挖掘模型评价的分流

数据挖掘模型评价分流到 Exadata 存储。这使得在数据库云服务器上的数据仓库部署成为更好、性能更佳的数据分析平台。所有数据挖掘评价函数（例如 prediction\_probability）都分流到 Exadata 进行处理。这不仅加快数据仓库分析速度，而且还减少数据库服务器 CPU 使用以及数据库服务器与 Exadata 存储之间的 I/O 负载。

### 其他 Exadata 智能扫描处理

分流到 Exadata 存储的另外两个数据库操作是增量数据库备份和表空间创建。使用 Exadata，增量数据库备份的速度和效率显著提高。当使用 Exadata 存储时，数据库中更改跟踪的粒度更为细化。使用 Exadata，在各个 Oracle 块级别跟踪更改，而不是在大的块组级别。这样，备份使用的 I/O 带宽更少，而备份运行速度更快。

使用 Exadata，还可更高效地执行文件创建操作。例如，当发出 Create Tablespace 命令时，不是立即执行操作，在服务器内存中格式化新表空间的每个块并将它们写入存储，而是向 Exadata 存储发送一个 iDB 命令，指示其创建表空间并格式化块。主机内存使用因此而减少，并且创建和格式化表空间块的相关 I/O 得以分流。这些操作节省了 I/O 带宽，这意味着更多的带宽可用于其他业务关键工作。

### 混合列压缩

数据压缩可显著减少大型数据库占用的存储空间。Exadata 提供了一种非常高级的压缩功能，称为混合列压缩 (HCC)。混合列压缩可实现最高级别的数据压缩，为企业节省大量成本并且显著提高性能（由于 I/O 减少）。平均节省 10 倍至 15 倍的存储空间，具体取决于使用 HCC 的方式。由于平均节省 10 倍的存储空间，因此 IT 管理人员在几年内很少需要并且通常不需要购买新存储。例如，如果实现 10 倍的存储空间节省，则 100 TB 数据库仅使用 10 TB 的物理存储。现在有 90 TB 的存储空间可用，IT 组织可以推迟很长一段时间再购买存储。



HCC 是一种在数据库块中组织数据的新方法。顾名思义，此技术混合使用行方法和列方法来存储数据。这种混合方法兼具两种技术的优势，实现了列存储的压缩优势，同时避免了纯列格式的性能不足之处。这里使用一种称为压缩单元的逻辑结构来存储一组经过混合列压缩的行。当加载数据时，从该组行中分离出列值，进行排序并组织在一起，然后进行压缩。一组行的列数据经过压缩后，将纳入压缩单元中。

在 Exadata 存储中提供对 HCC 数据的智能扫描处理，并执行列投影和筛选。可直接对经过混合列压缩的数据运行查询，不需要解压缩数据。不需要对满足查询谓词所需的数据进行解压缩，只对内存中要返回给客户端的列和行进行解压缩。解压缩过程在 Exadata 单元上发生，以便最大程度提高性能并从数据库服务器分流处理。假设混合列压缩表的压缩率通常为 10 倍，则与未压缩的数据相比，这可有效地使 I/O 速率提高 10 倍。

### Exadata 智能闪存缓存特性

Oracle 在 Oracle Exadata 存储服务器中直接实现了智能闪存缓存。Exadata 智能闪存缓存将频繁访问的数据保存在非常快速的闪存存储中，而将大部分数据保存在非常经济合算的磁盘存储中。该工作无需用户干涉自动完成。Oracle 闪存缓存之所以智能，是因为它知道何时避免缓存那些永不会重用或不适于缓存的数据。Oracle 数据库和 Exadata 存储允许用户在数据库表、索引和数据段级提供指令，以确保将特定数据驻留在闪存中。表可以保留在闪存中，而不必像传统存储与传统闪存磁盘那样，将表移动至不同的表空间、文件或 LUN。

Exadata 智能闪存缓存软件实现了对数据库读写的自动缓存。在全机架 X4 数据库云服务器中，该软件可提供高达 266 万次的 SQL 闪存 8K 读 IOPS。使用写回闪存缓存特性，Exadata 软件甚至可为要求最严苛的 OLTP 数据库带来超高性能。在全机架 X4 数据库云服务器中，写缓存技术可提供高达 196 万次的 SQL 闪存 8K 写 IOPS。此外，Exadata 智能闪存缓存在 Exadata 存储服务器重新启动时仍保持工作状态，且无需任何预热时间。

Exadata 智能闪存缓存还可用于降低日志写入 I/O 延迟，从而消除可能由于数据库日志操作而导致的性能瓶颈问题。用户事务提交时间对日志写入延迟非常敏感。另外，许多性能关键的数据库算法，如空间管理和索引划分也对日志写入延迟非常敏感。如今，Exadata 存储使用磁盘控制器中带电池保护的 DRAM 缓存来加快日志写入速度。写入到磁盘控制器缓存的速度通常很快，但在磁盘 IO 密集时该速度可能减慢。智能闪存日志功能利用 Exadata 存储中的闪存内存来加快日志写入速度。

闪存内存的平均写入延迟很低，但是其写入速度偶尔会不同寻常地很慢，比平均速度慢一或两个数量级。Exadata 智能日志的想法是，同时向闪存内存和磁盘控制器缓存执行重做写入操作，只要这两个操作中有一个操作完成，就结束该次写入。这确实让 Exadata 兼具磁盘和闪存两者之最佳优势。智能闪存日志功能不仅提高了用户事务响应速度，而且还因其加快了性能关键的数据库算法的速度，从而提高了 IO 密集型负载的整体数据库吞吐量。

数据库对 Exadata 智能闪存缓存的写入和日志记录都是以透明的方式进行的。数据库和 Exadata 存储软件可处理所有崩溃和恢复场景，除数据库恢复所需的常规操作外，无需任何额外或特别的管理员干预。从 DBA 的角度来看，系统的行为方式完全透明，DBA 不需要关心闪存用作数据文件或重做数据的临时存储这一事实。在行为上唯一的不同是数据库写入延迟持续较低。

### Exadata 中的 I/O 资源管理

使用传统存储，由于无法确定使用存储子系统中的 I/O 带宽的不同作业和用户的工作的优先级，阻碍了共享存储网格的建立。当多个数据库共享存储子系统时也会发生这种情况。使用 Exadata 存储时，Exadata 存储的 DBRM 和 I/O 资源管理功能可以防止一类工作或一个数据库独占磁盘资源和带宽，并确保满足用户定义的 SLA。DBRM 能够对数据库之间以及不同的用户与各类工作之间使用的 I/O 带宽进行协调并确定优先级。通过将数据库与存储环境紧密集成，Exadata 可以知晓工作类型和使用的 I/O 带宽量。因此，用户可以让 Exadata 系统识别各种类型的负载，为这些负载指定优先级，并确保最关键的负载得到优先处理。

在数据仓储或混合负载环境中，您可能希望确保为数据库中的不同用户和任务分配相对适量的 I/O 资源。例如，您可能希望为系统上的交互式用户分配 70% 的 I/O 资源，为批处理报表作业分配 30% 的 I/O 资源。这可以使用 Exadata 存储的 DBRM 和 I/O 资源管理功能很简单地实现。

Exadata 管理员可以创建一个资源计划，指定应如何确定 I/O 请求的优先级。这可通过将不同类型的工作放入各个服务组（称为用户组）中来实现。用户组可以由多个属性定义，这些属性包括用户名、客户端程序名、函数或查询已运行的时间长度。

定义这些用户组之后，用户可以设置一个层次结构，规定哪个用户组享有 I/O 资源的优先权，以及为每个用户组分配多少 I/O 资源量。这种层次结构确定的 I/O 资源优先级可以同时应用于数据库内操作（即在数据库内发生的操作）和数据库间操作（即在不同数据库间发生的操作）。

多个数据库之间共享 Exadata 存储时，您还可以确定为每个数据库分配的 I/O 资源的优先级，从而防止一个数据库独占磁盘资源和带宽，以确保满足用户定义的 SLA。例如，可能有两个数据库共享 Exadata 存储。业务目标要求，对于组织而言，每个数据库都有其相应的价值和重要性。最后决定，数据库 A 应得到 33% 的总可用 I/O 资源，数据库 B 应得到 67% 的总 I/O 资源。为了确保为每个数据库中的不同用户和任务分配相对适量的 I/O 资源，定义了不同的用户组。

- 对数据库 A 定义了两个用户组
  - 为交互式营销活动保留 60% 的 I/O 资源
  - 为批处理营销活动保留 40% 的 I/O 资源
- 对数据库 B 定义了三个用户组
  - 为交互式销售活动保留 60% 的 I/O 资源
  - 为批处理销售活动保留 30% 的 I/O 资源
  - 为大客户销售活动保留 10% 的 I/O 资源

这些用户组分配是相对于为每个数据库分配的总 I/O 资源而言的。

对于客户而言，将多个数据库整合到一个 Exadata 数据库云服务器上是一个可节省成本的解决方案。在 Exadata 存储服务器软件 11.2.2.3 及更高版本中，可以使用 Exadata I/O Resource Manager (IORM) 来对数据库云服务器上运行的不同数据库启用或禁用闪存使用。这让客户能够保留闪存以便用于性能最关键的数据库。在 Exadata 存储服务器软件 11.2.3.1 及更高版本中，除了上述基于百分比的策略之外，还可以通过基于共享的策略向共享 Exadata 存储的不同数据库分配资源。在将大量数据库整合到 Exadata 系统时，通过这种机制可轻松地管理资源。

实际上，Exadata I/O Resource Manager 解决了传统存储技术未能解决的一个难题：能够对共享存储子系统的多个数据库和用户的工作进行平衡和确定优先级，藉此建立共享网格存储环境。Exadata I/O 资源管理可确保针对共享 Exadata 存储的多个数据库满足用户定义的 SLA。这确保了每个数据库或用户均可获得适量份额的磁盘带宽以满足业务目标。

## Exadata 中的数据库网络资源管理

Exadata 还实现了独一无二的数据库网络资源管理，确保报告、批处理和备份等网络密集型负载不会影响对响应时间敏感的交互式负载。RAC 缓存融合通信和日志文件写入等对延迟敏感的网络操作会自动移至服务器和存储网卡以及 InfiniBand 网关交换机中的消息队列的头部，从而绕至任何对延迟不敏感的消息之前。延迟关键消息甚至会抢先于已经通过网络发送了一部分的非延迟关键消息，从而即使存在大型网络 DMA 操作也能确保快速响应。

## Exadata 中的服务质量 (QoS) 管理

Oracle Exadata QoS Management 是基于策略的自动化产品，用于监视整个系统的负载请求。它对应用程序之间共享的资源进行管理，并调整系统配置以保持应用程序运行于业务需要的性能级别。它可从容应对系统配置和需求的变化，从而避免应用程序性能级别的额外波动。

Oracle Exadata QoS Management 监视目标系统上的每个工作请求的执行。从工作请求要求连接到数据库以使用数据库服务时，它便开始跟踪该工作请求。完成工作请求所需的时间，即响应时间（也称为端到端响应时间或往返时间）是指从发起数据请求到完成数据请求的时间。通过准确地测量响应时间的两个部分（使用资源花费的时间以及等待使用资源花费的时间），QoS Management 可以快速发现系统中的瓶颈。然后，它给出重新分配资源的建议以解除瓶颈，从而保持或恢复服务级别。此时会提醒系统管理员需要进行这种重新分配，管理员只需在 QoS Management 信息板上单击按钮即可完成这一操作。还提供了有关此操作对整个集群预期性能影响的完整详细信息。最后，保存所有操作和策略更改的审计日志以及历史系统性能图。

Oracle Exadata QoS Management 管理系统中的资源以便：

- 当提供的资源足以满足需求时，即使负载更改，也会满足应用程序的业务级性能要求；

- 当提供的资源不足以满足需求时，Oracle Exadata QoS Management 通过牺牲不太关键的业务性能要求来设法满足更为关键的业务性能要求；
- 当负载状况严重超出容量时，资源仍然可用。

#### 使用 Oracle Exadata QoS Management 的优势

在一般企业中，如果应用程序的响应时间未在可接受的级别内，则为了解决该问题可能要花费很长的时间。通常，管理员首先会询问以下问题：“我们是否正确配置了系统？通过更改参数是否可以解决该问题？我们是否需要更多的硬件？”遗憾的是，很难准确地回答这些问题；结果通常是进行几个小时徒劳无益令人沮丧的试验。

Oracle Exadata QoS Management 具有以下优势：

- 缩短了管理 Oracle Real Application Clusters (Oracle RAC) 资源的系统管理员的工作时间，降低了对他们的专业技能要求
- 有助于减少执行中断的次数
- 缩短解决可限制或降低应用程序性能的问题所需的时间
- 当负载变化时为系统提供稳定性
- 使添加或删除服务器可以对应用程序透明地进行
- 降低因服务器故障而导致的对系统的影响
- 有助于确保满足服务级别协议 (SLA)
- 实现更有效的硬件资源共享
- 保护现有负载免受过量使用内存而引起的服务器故障的影响
- Exadata 存储虚拟化
- Exadata 提供一组丰富的、完善的强大存储管理虚拟化功能，这些功能可利用 Oracle 数据库、Exadata 软件和 Exadata 硬件的优势。

#### Exadata 存储管理和数据保护

如前所述，Exadata 单元是一个运行 Oracle Linux (OL) 以及 Oracle Exadata 存储服务器软件的服务器。首次启动时，该单元与任何其他计算机一样进行引导，直到变为 Exadata 存储服务模式。前两个磁盘驱动器具有小型逻辑单元编号 (LUN) 分片（称为系统区域），大小约为 30 GB，保留用于 OL 操作系统、Exadata 软件和配置元数据。系统区域包含 Oracle Database 自动诊断信息库 (ADR) 数据和其他有关 Exadata 单元的元数据。管理员不必管理

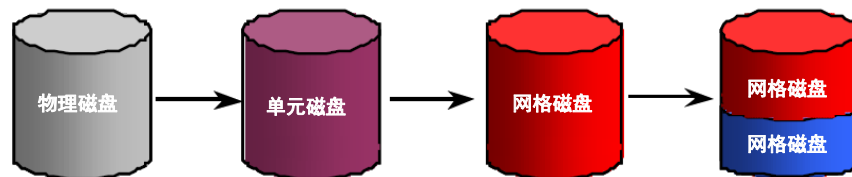
系统区域 LUN，因为它是自动创建的。其内容自动跨物理磁盘进行镜像，以防范驱动器故障并允许热插拔磁盘。这两个磁盘驱动器的剩余部分可用于存储用户数据。

### Exadata 用户存储虚拟化

使用自动存储管理 (ASM) 来管理 Exadata 单元中的存储。ASM 卷管理、条带化和数据保护服务使其成为卷管理的最佳选择。ASM 提供了针对驱动器和单元故障的数据保护、最佳性能以及极其灵活的配置和重新配置选择。

单元磁盘是物理磁盘（除去系统区域 LUN（如果存在））的虚拟表示，并且是管理员在 Exadata 单元中管理的主要磁盘对象之一。一个单元磁盘由一个 LUN 表示，这个 LUN 是在发现物理磁盘时由 Exadata 软件自动创建和管理的。

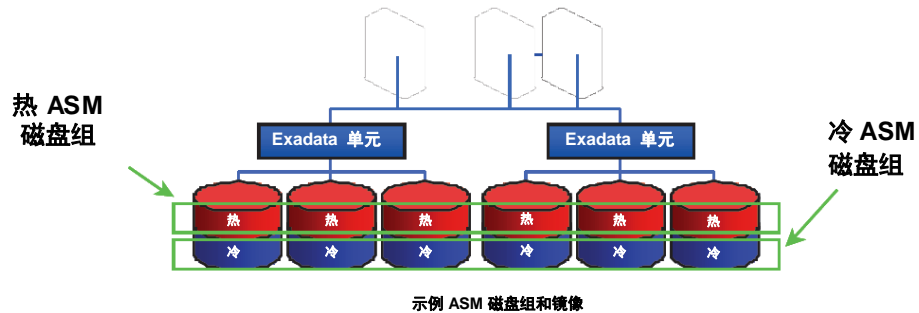
单元磁盘可以进一步虚拟化为一个或多个网格磁盘。网格磁盘是分配给 ASM 作为 ASM 磁盘的磁盘实体，ASM 对它们进行管理以便为数据库存储用户数据。最简单的情况是单个网格磁盘占据整个单元磁盘。但是，还可以将一个单元磁盘分成多个网格磁盘分片。通过在一个单元磁盘上放置多个网格磁盘，管理员可以将该存储分隔到多个具有不同性能或可用性要求的池中。可以通过网格磁盘分片来分配单元磁盘的“热”、“暖”和“冷”区域，或分隔数据库共享的 Exadata 磁盘。例如，可以对某个单元磁盘进行分区，以便一个网格磁盘位于物理磁盘的较高性能部分并配置为三重镜像，而另一个网格磁盘位于磁盘的较低性能部分并用于归档或备份数据，没有任何镜像。可以使用网格磁盘功能来实施信息生命周期管理 (ILM) 策略。



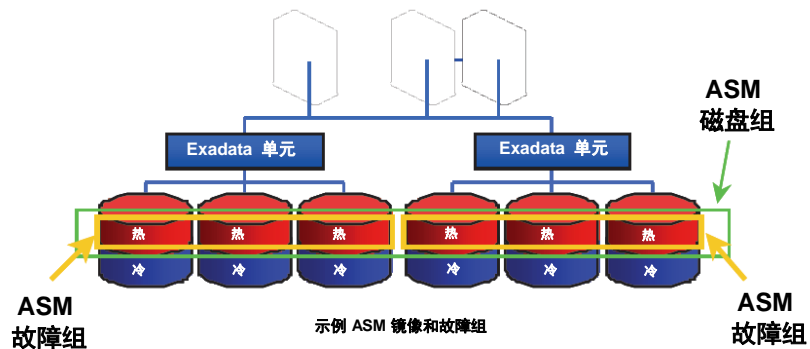
网格磁盘虚拟化

以下示例说明了单元磁盘和网格磁盘在更复杂的 Exadata 存储网格中的关系。

配置单元磁盘和网格磁盘之后，跨 Exadata 配置定义 ASM 磁盘组。定义了两个 ASM 磁盘组：一个跨“热”网格磁盘，另一个跨“冷”网格磁盘。所有的“热”网格磁盘都放置在一个 ASM 磁盘组中，所有的“冷”网格磁盘都放置在另一个磁盘组中。将数据加载到数据库时，ASM 将在这些磁盘组中均匀地分配数据和 I/O。可以对这些磁盘组启用 ASM 镜像以便在磁盘故障时提供保护，可以对两个磁盘组或其中一个磁盘组启用该功能，也可以都不启用。可以对每个磁盘组单独开启或关闭镜像。



最后，为了在某个 Exadata 单元整个出现故障时提供保护，定义了 ASM 故障组。故障组可确保将镜像的 ASM 盘区放置在不同的 Exadata 单元上。



使用 Exadata 和 ASM:

- Exadata 软件自动配置单元磁盘（创建 LUN）。
- 多个网络磁盘可以共存于物理磁盘上，以根据数据库应用程序的需要实现相应的性能，或通过 Exadata 构建 ILM 策略。
- ASM 自动跨 Exadata 磁盘和单元进行数据库数据条带化，以确保均衡的 I/O 负载和最佳的性能。
- ASM 动态添加和删除功能实现非侵入性单元和磁盘分配、取消分配以及重新分配。
- ASM 镜像和 Exadata 单元的热插拔功能在磁盘故障时提供透明的数据保护和可访问性。

- ASM 提供双重或三重镜像，以根据数据的重要性进行相应的保护。
- Exadata 自动创建 ASM 故障组，以在单元故障时提供透明的数据保护和可访问性。
- 在 ASM 中自动提供可用网格磁盘，并且可以非常容易地对其进行配置。无需像常规存储那样创建挂载点或 LUN。所有这些工作将自动完成。

### 迁移到 Exadata 存储

有多种技术可以将数据迁移到数据库云服务器。可以使用 Oracle Recovery Manager (RMAN) 进行迁移，用这个工具从传统存储进行备份并将数据恢复到 Exadata。还可以使用 Oracle Data Guard 简化迁移。为此，首先创建一个基于 Exadata 存储的备用数据库。备用数据库可以使用 Exadata 存储，生产数据库可以使用传统存储。通过执行快速转换，只需几秒钟即可将备用数据库转换为生产数据库。这提供了内置安全网，因为，如果发生无法预料的问题，您可以非常稳妥地撤销迁移。还可以使用可传输表空间和数据泵迁移到 Exadata。任何用于在 Oracle 数据库之间移动数据的技术都可以用于 Exadata。

### Exadata 中的数据保护

Exadata 在设计时采用了客户惯于对 Oracle 产品期待的同样的高可用性 (HA) 标准。在 Exadata 环境中，所有数据库特性和工具与其在传统非 Exadata 存储环境中一样有效。用户和数据库管理员将使用熟悉的工具，能够利用他们现有的 Oracle 数据库知识和过程。Exadata 架构消除了所有单点故障。在 Exadata 中纳入了镜像、故障隔离、驱动器和单元故障防护等熟悉的特性，以确保持续的数据可用性和数据保护。下面介绍 Exadata 服务器中其他确保高可用性的特性。有关在 Exadata 数据库云服务器上使用这些技术和最佳实践的更多信息，请参见以下位置中的 Exadata 最高可用性架构最佳实践白皮书：<http://www.oracle.com/technetwork/database/features/availability/exadata-maa-best-practices-155385.html>。

### Exadata 中内置的硬件辅助恢复数据 (HARD)

Oracle 硬件辅助恢复数据 (HARD) 计划是一个旨在防止发生数据损坏的全面的计划。尽管数据损坏很少发生，但是一旦发生，可能对数据库产生灾难性影响，进而波及到业务。Exadata 中嵌入了增强的 HARD 功能，以提供更高级别的数据保护和端到端的数据验证。Exadata 对存储在其中的数据执行大量验证，包括校验和、块位置、魔数、头尾检查、对齐



错误等。在 Exadata 中实施这些数据验证算法可防止将损坏的数据写入永久性存储中。此外，无需手动步骤即可提供这些检查和保护，而将 HARD 用于常规存储时，则需要手动步骤。

#### **Data Guard**

Oracle Data Guard 是 Oracle 数据库的软件特性，用于创建、维护和监视一个或多个备用数据库，以便在发生故障、灾难、错误和损坏时保护您的数据库。Data Guard 无需修改即可与 Exadata 协同工作，并可用于生产数据库和备用数据库。通过将 Active Data Guard 与 Exadata 存储一起使用，可以将查询和报表从生产数据库分流到极速备用数据库，并确保生产数据库上的关键工作不受影响，而同时仍提供灾难保护。

#### **闪回**

Exadata 利用 Oracle 闪回技术提供一组特性来查看并及时恢复数据。闪回特性在 Exadata 环境中和非 Exadata 环境中的使用相同。使用闪回特性，可以查询历史数据、执行更改分析、执行自助修复，以便在数据库联机时从逻辑损坏中恢复。实际上，Exadata 中内置的 Oracle 闪回特性让用户拥有类似快照的功能，让用户能够将数据库恢复到发生错误之前的某一时间。

#### **Recovery Manager (RMAN) 和 Oracle Secure Backup (OSB)**

Exadata 与 Oracle Recovery Manager (RMAN) 协同工作，以实现高效的 Oracle 数据库备份和恢复。所有现有 RMAN 脚本无需更改即可用于 Exadata 环境。RMAN 旨在与服务器密切协作，在备份和恢复期间提供块级损坏检测。RMAN 在备份过程中通过文件多路复用和备份集压缩来优化性能和空间使用，并与 Oracle Secure Backup (OSB) 和第三方介质管理产品集成以进行磁带备份。

## **总结**

如今，企业日益需要利用统一的数据库平台以便将所有应用程序部署和整合到同一个通用基础架构上。无论是 OLTP、DW 还是混合负载，通用基础架构都可提供数据中心所需的效率和可重用性，并提供实际的内部网格计算。为不同应用程序构建或使用自定义的专用系统是一种浪费，成本也非常高昂。企业日益需要处理更多的数据，同时也意识到其 IT 预算捉襟见肘。对 IT 软硬件总拥有成本 (TCO) 的审视结果通常使人们趋向于选择一个通用的高性能基础架构来部署所有应用程序。通过将基于 Exadata 的数据库云服务器纳入到 IT 基础架构中，公司可以：

- 提高数据库性能，能够在相同的时间内做更多的事情。
- 通过将部署整合到一个通用基础架构上，采用可伸缩、渐进的步骤应对变化和发展。
- 提供任务关键数据可用性和保护。

# 甲骨文（中国）软件系统有限公司

## 北京远洋光华中心办公室

地址：北京市朝阳区景华南街5号远洋光华中心C座21层  
邮编：100020  
电话：(86.10) 6535-6688  
传真：(86.10) 6515-1015

## 北京汉威办公室

地址：北京市朝阳区光华路7号汉威大厦10层1003-1005单元  
邮编：100004  
电话：(86.10) 6535-6688  
传真：(86.10) 6561-3235

## 北京甲骨文大厦

地址：北京市海淀区中关村软件园24号楼甲骨文大厦  
邮编：100193  
电话：(86.10) 6106-6000  
传真：(86.10) 6106-5000

## 北京国际软件大厦办公室

地址：北京市海淀区中关村软件园9号楼国际软件大厦二区308单元  
邮编：100193  
电话：(86.10) 8279-8400  
传真：(86.10) 8279-8686

## 北京孵化器办公室

地址：北京市海淀区中关村软件园孵化器2号楼A座一层  
邮编：100193  
电话：(86.10) 8278-6000  
传真：(86.10) 8282-6401

## 上海名人商业大厦办公室

地址：上海市黄浦区天津路155号名人商业大厦12层  
邮编：200001  
电话：(86.21) 2302-3000  
传真：(86.21) 6340-6055

## 上海腾飞浦汇大厦办公室

地址：上海市黄浦区福州路318号腾飞浦汇大厦508-509室  
邮编：200001  
电话：(86.21) 2302-3000  
传真：(86.21) 6391-2366

## 上海创智天地10号楼办公室

地址：上海市杨浦区淞沪路290号创智天地10号楼512-516单元  
邮编：200433  
电话：(86.21) 6095-2500  
传真：(86.21) 6107-5108

## 上海创智天地11号楼办公室

地址：上海市杨浦区淞沪路303号创智天地科教广场3期11号楼7楼  
邮编：200433  
电话：(86.21) 6072-6200  
传真：(86.21) 6082-1960

## 上海新思大厦办公室

地址：上海市漕河泾开发区宜山路926号新思大厦11层  
邮编：200233  
电话：(86.21) 6057-9100  
传真：(86.21) 6083-5350

## 广州国际金融广场办公室

地址：广州市天河区珠江新城华夏路8号合景国际金融广场18楼  
邮编：510623  
电话：(86.20) 8513-2000  
传真：(86.20) 8513-2380

## 成都中海国际中心办公室

地址：成都市高新区交子大道177号中海国际中心7楼B座02-06单元  
邮编：610041  
电话：(86.28) 8530-8600  
传真：(86.28) 8530-8699

## 深圳飞亚达科技大厦办公室

地址：深圳市南山区高新南一道飞亚达科技大厦16层  
邮编：518057  
电话：(86.755) 8396-5000  
传真：(86.591) 8601-3837

## 深圳德赛科技大厦办公室

地址：深圳市南山区高新南一道德赛科技大厦8层0801-0803单元  
邮编：518057  
电话：(86.755) 8660-7100  
传真：(86.755) 2167-1299

## 大连办公室

地址：大连软件园东路23号大连软件园15号楼502  
邮编：116023  
电话：(86.411) 8465-6000  
传真：(86.755) 8465-6499

## 苏州办公室

地址：苏州工业园区星湖街328号苏州国际科技园5期11幢1001室  
邮编：215123  
电话：(86.512) 8666-5000  
传真：(86.512) 8187-7838

## 沈阳办公室

地址：沈阳市和平区青年大街390号皇朝万鑫国际大厦A座39层3901&3911室  
邮编：110003  
电话：(86.24) 8393-8700  
传真：(86.24) 2353-0585

## 济南办公室

地址：济南市历源大街150号中信广场11层1113单元  
邮编：250011  
电话：(86.531) 6861-1900  
传真：(86.531) 8518-1133

## 南京办公室

地址：南京市玄武区洪武北路55号置地广场19层1911室  
邮编：210018  
电话：(86.25) 8579-7500  
传真：(86.25) 8476-5226

## 西安办公室

地址：西安市高新区科技二路72号西安软件园零壹广场主楼1401室  
邮编：710075  
电话：(86.29) 8834-3400  
传真：(86.25) 8833-9829

#### 重庆办公室

地址：重庆市渝中区邹容路68号大都会商厦1611室  
邮编：400010  
电话：(86.23) 6037-5600  
传真：(86.23) 6370-8700

#### 杭州办公室

地址：杭州市西湖区杭大路15号嘉华国际商务中心810&811室  
邮编：310007  
电话：(86.571) 8168-3600  
传真：(86.571) 8717-5299

#### 福州办公室

地址：福州市五四路158号环球广场1601室  
邮编：350003  
电话：(86.591) 8621-5050  
传真：(86.591) 8801-0330

#### 南昌办公室

地址：江西省南昌市西湖区沿江中大道258号  
皇冠商务广场10楼1009室  
邮编：330025  
电话：(86.791) 8612-1000  
传真：(86.791) 8657-7693

#### 呼和浩特办公室

地址：内蒙古自治区呼和浩特市新城区迎宾北路7号  
大唐金座19层北侧1902-1904室  
邮编：010051  
电话：(86.471) 3941-600  
传真：(86.471) 5100-535

#### 郑州办公室

地址：河南省郑州市中原区中原中路220号  
裕达国际贸易中心A座2015室  
邮编：450007  
电话：(86.371) 6755-9500  
传真：(86.371) 6797-2085

#### 武汉办公室

地址：武汉市江岸区中山大道1628号  
武汉天地企业中心5号大厦23层2301单元  
邮编：430010  
电话：(86.27) 8221-2168  
传真：(86.27) 8221-2168

#### 长沙办公室

地址：长沙市芙蓉区韶山北路159号通程国际大酒店1311-1313室  
邮编：410011  
电话：(86.731) 8977-4100  
传真：(86.731) 8425-9601

#### 石家庄办公室

地址：石家庄市中山东路303号石家庄世贸广场酒店14层1402室  
邮编：050011  
电话：(86.311) 6670-8080  
传真：(86.311) 8667-0618

#### 昆明办公室

地址：昆明市三市街六号柏联广场写字楼11层1103A室  
邮编：650021  
电话：(86.871) 6402-4600  
传真：(86.871) 6361-4946

#### 合肥办公室

地址：安徽省合肥市蜀山区政务新区怀宁路1639号平安大厦18层1801室  
邮编：230022  
电话：(86.551) 6595-8200  
传真：(86.551) 6371-3182

#### 广西办公室

地址：广西省南宁市青秀区民族大道136-2号华润大厦B座2302室  
邮编：530028  
电话：(86.771) 391-8400  
传真：(86.771) 577-5500



Oracle Exadata 数据库云服务器和

Exadata 存储服务器技术概述

2014 年 6 月

作者: Gurmeet Goindi

参与编著: Juan Loaiza, Tim Shetler,  
Ashish Ray

公司网址: <http://www.oracle.com> (英文)

中文网址: <http://www.oracle.com/cn> (简体中文)

销售中心: 800-810-0161

售后服务热线: 800-810-0366

培训服务热线: 800-810-9931

欢迎访问:

<http://www.oracle.com> (英文)

<http://www.oracle.com/cn> (简体中文)

版权© 2014 归 Oracle 公司所有。未经允许, 不得以任何形式和手段复制和使用。

本文的宗旨只是提供相关信息, 其内容如有变动, 恕不另行通知。Oracle 公司对本文内容的准确性不提供任何保证, 也不做任何口头或法律形式的其他保证或条件, 包括关于适销性或符合特定用途的所有默示保证和条件。本公司特别声明对本文档不承担任何义务, 而且本文档也不能构成任何直接或间接的合同责任。未经 Oracle 公司事先书面许可, 严禁将此文档为了任何目的, 以任何形式或手段(无论是电子的还是机械的)进行复制或传播。

Oracle 是 Oracle 公司和/或其分公司的注册商标。其他名字均可能是各相应公司的商标。