

Oracle 白皮书
2010 年 9 月

Oracle Clusterware 11g 第 2 版

引言.....	1
Oracle 网络基础架构简介.....	2
使用 Oracle 网络基础架构简化网络安装.....	4
典型安装和高级安装.....	4
前提条件检查、Secure Shell 设置和 FixUp 脚本.....	4
Oracle ASM 中存储的 Oracle Clusterware 文件.....	5
使用 Oracle Clusterware 简化网络管理.....	6
网络命名服务和自动虚拟 IP 分配.....	7
基于策略和角色分离的集群管理.....	7
集群化（集群识别）命令.....	9
基于 EM 的资源和集群管理.....	9
提高可用性 — 高级调优.....	11
高级可用性.....	11
隔离的灵活性及第三方集群解决方案支持.....	11
冗余互连使用.....	12
Oracle Clusterware 11g 第 2 版中的无重启节点隔离.....	13
集群时间同步和网络管理.....	14
Cluster Health Monitor: 与 Oracle 网络基础架构相集成.....	14
使用 Oracle Clusterware 管理 Oracle RAC 数据库.....	15
使用 Oracle Clusterware 管理任何类型的应用程序.....	15
总结.....	16

引言

Oracle Clusterware 是一个可移植集群软件，它允许将独立的服务器整合到一起，这样它们就能像一个系统似地相互协作了。Oracle Clusterware 作为 Oracle Real Application Clusters (RAC) 所需的集群技术，首次与 Oracle Database 10g 第 1 版一起发布。Oracle Clusterware 是一个独立的集群基础架构，它与 Oracle RAC 完全集成，能够保护故障切换集群中的任何应用程序。

Oracle Clusterware 11g 第 2 版是 Oracle 集群解决方案发展过程的里程碑产品。Oracle Clusterware 与 Oracle 自动存储管理 (ASM) 相结合，已经成为 Oracle 网络基础架构软件。这两种技术的整合将网络应用程序管理提升至一个新的水平。

Oracle 网络基础架构引入了新的服务器池的概念，支持将网络划分为若干服务器群。“角色分离的管理”适用于集群、存储和数据库管理功能严格分离的组织。无论大小如何，集群识别命令以及基于 Enterprise Manager 的集群和资源管理都能够简化网络管理。Oracle ASM 的更多改进（例如，新的 ASM 集群文件系统、动态卷管理器）进一步完善了 Oracle 网络基础架构解决方案。

本文讨论的 Oracle Clusterware 11g 第 2 版新特性表明 Oracle Clusterware 不仅继续作为 Oracle RAC 的基础，并且已经发展为任何集群环境的基础。Oracle Clusterware 11g 第 2 版为任何应用程序提供了卓越的可用性和可伸缩性，免除了使用其他第三方集群解决方案的需要。

Oracle 网络基础架构简介

Oracle Clusterware 是一种将服务器场转换为集群的技术。集群是一组独立但相互连接的服务器，它们能够像一个系统似的协同工作。Oracle Clusterware 是该系统中支持协同工作的关键所在。

Oracle Clusterware 作为运行 Oracle Real Application Clusters (RAC) 所需的底层集群软件引入到 Oracle Database 10g 第 1 版中。作为 Oracle RAC 体系的一部分，Oracle Clusterware 不仅应用于 Oracle ASM 集群版，还紧密集成到 Oracle RAC 集群体系中。

Oracle Clusterware 作为一个全面的集群解决方案，其使用范围已不局限于 Oracle RAC。在这些环境中，Oracle Clusterware 作为故障切换集群解决方案为所有应用程序提供保护。

在这两种环境中，Oracle Clusterware 能够管理集群中的资源、进程和应用程序，同时维护节点成员身份并确保隔离。

通过 Oracle 网络基础架构，Oracle 将 Oracle ASM（成熟的 Oracle 数据库存储管理解决方案）和 Oracle Clusterware 集成到同一软件包中。Oracle 通过组合这两个集群环境中最强大的产品建立了通用的网络基础。

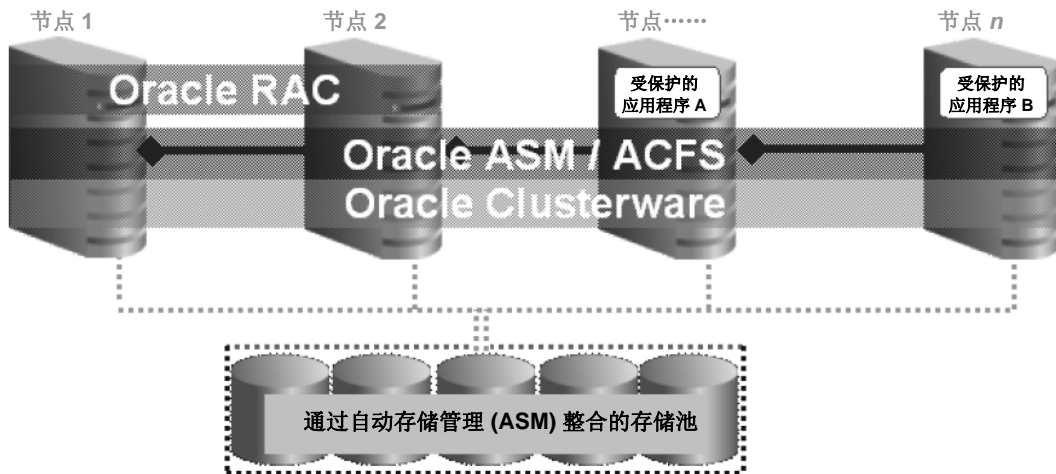


图 1 — Oracle Clusterware 配置概述

这两个产品的增强功能，如基于 Oracle ASM 的集群文件系统 (ACFS)、基于 Enterprise Manager 的用于管理整个集群的图形用户界面、以及服务器池和网络即插即用 (GPnP) 特性，能够对任何规模、类型或部署了任何数量应用程序的网格进行动态管理。

Oracle 网格基础架构不仅是 Oracle RAC 的基础，还作为任何集群环境的基础，为各种应用程序提供卓越的可用性和可伸缩性。

通过增加强健的存储管理解决方案，免除了使用任何第三方管理产品的需要，从而降低了网格环境的成本，如图 2 所示：

1. 使用 Oracle 网格基础架构作为整合平台将各种应用程序整合在基于集群的网格中，可以降低资本支出 (CAPEX)。
2. 使用 Oracle Clusterware 降低基于整合环境的数据中心成本，从而降低了运营成本 (OPEX)。此外，使用统一的管理界面管理较少服务器上运行的更多应用程序，使管理成本得以降低。

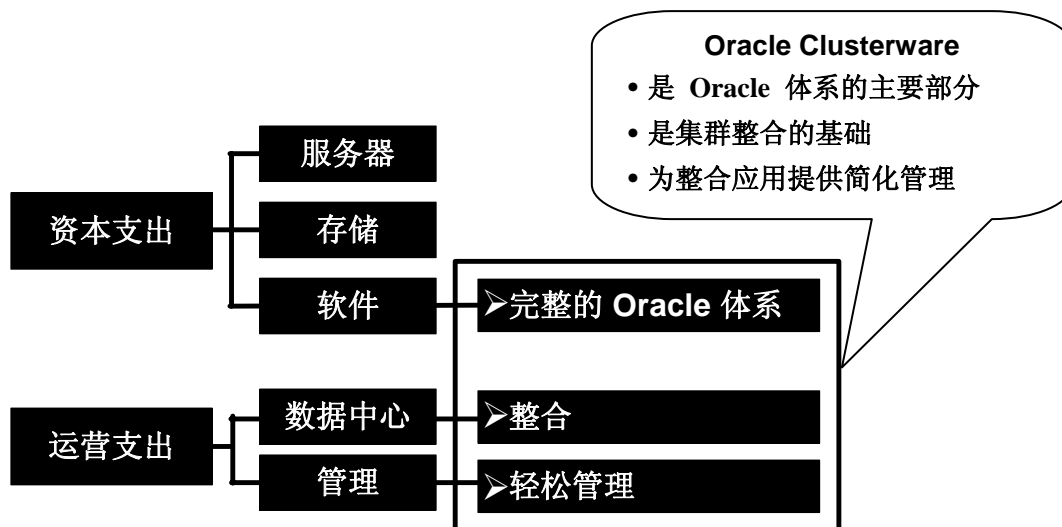


图 2：使用 Oracle Clusterware 节省成本

使用 Oracle 网格基础架构简化网格安装

通过以下方式简化和改进 Oracle 网格基础架构的安装，实现通用网格基础架构：

1. **验证** — 对使用 Oracle Universal Installer (OUI) 进行交互式安装期间的任何输入立即进行验证，以免以后出现故障。
2. **自动化** — 新的安装程序经过优化，能够自动完成特定步骤并使用典型配置。
3. 对于**大规模部署**，Oracle 网格基础架构提供新的“仅限于软件的安装”，将软件的部署与配置相分离。安装完软件之后，可以稍后进行配置。
4. 对于**单实例环境**，Oracle 网格基础架构可以“独立服务器”配置安装，提供 Oracle ASM 以及重启单实例数据库的功能（即 Oracle Restart）。

典型安装和高级安装

可以对 Oracle 网格基础架构进行“典型安装”或“高级安装”。典型安装提供了一种快速安装 Oracle 网格基础架构的方式，通过推荐的默认值简化安装，提供较少的自定义环境选项。如果需要较高的灵活性，可以选择高级安装，它针对较复杂的体系结构提供更多的自定义选项。

前提条件检查、Secure Shell 设置和 FixUp 脚本

要基于独立的服务器构建网格，每个服务器必须满足特定的前提条件，集群软件才能在这些服务器上正常运行，确保集群内的安全通信和稳定的集群操作。现在，专为 Oracle Database 11g 第 2 版而设计的 Oracle Universal Installer (OUI) 使用 Cluster Verification Utility (CVU) 为 Oracle 网格基础架构安装执行所有的前提条件检查。

如果因服务器配置未满足某些 Oracle 网格基础架构安装要求而导致前提条件检查失败，OUI 将列出这些要求并提供纠正指导。

OUI 检查远程节点的能力依赖正确的 Secure Shell (SSH) 设置。OUI 将检查能否通过 SSH 访问集群中的节点，只有在正确设置 SSH 之后才能继续进行安装。

此外，新的 Oracle Universal Installer 提供了在集群上自动配置 SSH 的选项。如果已经在计算机上（局部）设置了 SSH，OUI 将提供仅检查设置、检查和更正或从头设置等选择。

完成 SSH 的设置并执行所有检查之后，OUI 会列出检查失败的详细信息。对于一些检查（取决于检查的性质），新的 OUI 能够以 FixUp 脚本的方式加以修正，如图 3 所示。

然后，必须在所有需要修正的节点上以 root 用户运行 FixUp 脚本。另外，还可以手动更改配置。但是并不推荐这样做，可以暂时忽略检查中的问题并稍后进行修正。

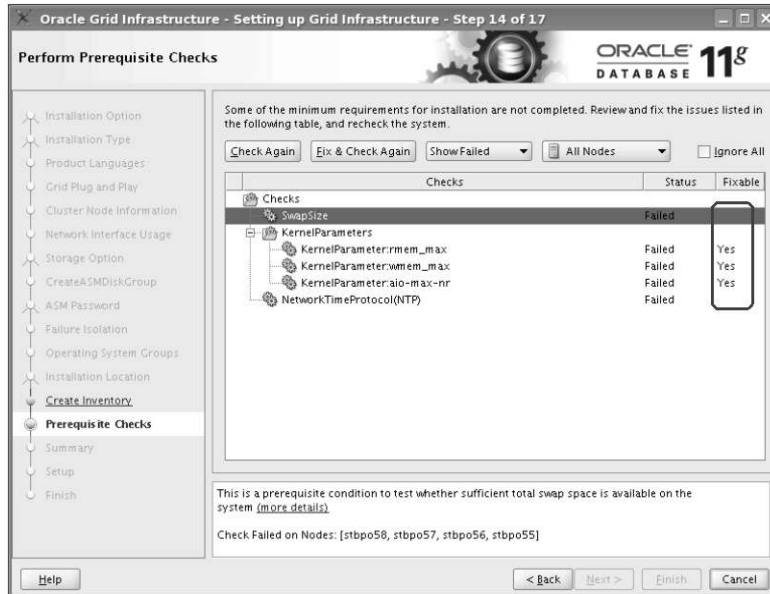


图 3: OUI 前提条件检查结果页面和 FixUp 脚本

Oracle ASM 中存储的 Oracle Clusterware 文件

Oracle 自动存储管理 (ASM) 现已与 Oracle Clusterware 一起集成到 Oracle 网络基础架构中。Oracle ASM 和 Oracle Database 11g 第 2 版相结合提供了完整的存储解决方案。该解决方案能够存储 Oracle Clusterware 文件，即 Oracle 集群注册表 (OCR) 和表决文件 (VF，也称为表决磁盘)，从而免除了使用第三方存储解决方案的需要。

这两种文件与其他数据库文件一起存储在 Oracle ASM 磁盘组中，因此能够利用 ASM 磁盘组的冗余配置。这表示常规冗余磁盘组将保存一个双向镜像 OCR。磁盘组中一个磁盘发生故障不会妨碍对 OCR 的访问。就高度冗余磁盘组（三向镜像）而言，两个独立磁盘出现故障也不会影响对 OCR 的访问。而采用外部冗余，Oracle 不提供任何保护。

为了防止物理磁盘故障，每个磁盘组只允许一个 OCR。Oracle 建议使用两个独立的磁盘组来保存 OCR，以免在发生 OCR 中断而不是磁盘故障时造成集群操作中断。

对表决文件的管理类似于 OCR。这些文件遵循 Oracle ASM 磁盘组的冗余配置，但是其管理方式不同于磁盘组中的常规 ASM 文件。每个表决磁盘放置在磁盘组的一个特定磁盘上。表决文件的磁盘及其在磁盘上的位置存储在 Oracle Clusterware 内部。

使用 Oracle Clusterware 简化网格管理

由于完善了 Oracle Clusterware 和 Oracle 网络基础架构的管理，而使得用户无需考虑网格规模，就可以轻松管理网格中的各种应用程序。对于 Oracle RAC 数据库管理而言，通过整合专门的管理工具，使管理网格中的 Oracle RAC 数据库变得更加容易。

这些管理方面的改善至少简化了以下某个方面的管理：

1. 管理任何规模的网格 — 集群中的服务器最多可达几百个
2. 使 Oracle Clusterware 不依赖第三方解决方案
3. 易于维护 — 减少（手动）管理任务的数量
4. 简化第三方应用程序的管理（仅使用一种管理工具）

为避免前期不必要的第三方应用程序配置步骤以及后期的重新配置步骤，大多数新的管理特性在初始安装期间就进行合理配置。

在 Oracle ASM 中管理 Oracle Clusterware 文件就是这些特性的一个示例。这些特性提供了在共享存储体系结构中存储 Oracle Clusterware 文件的简单方法，从而简化了安装，另外，这些文件不需要任何持续管理，因此带来了更多好处。

其他示例包括网格命名服务和自动虚拟 IP 分配以及 Oracle 集群时间同步服务，通常，在安装期间启动这些服务能够简化 Oracle 网格基础架构的安装。这些特性显示了它们在日常网格管理方面的真正优势。

网格命名服务和自动虚拟 IP 分配

对能够按需添加或删除节点的大型网格环境进行动态管理，需要动态的命名模式。像虚拟互联网协议 (VIP) 地址、域名服务 (DNS) 中的服务器名等静态命名在动态环境中会出现问题，因为集群中节点数量的任何变化都需要相应的更新。

新的网格命名服务 (GNS) 解决了这一问题。根据动态主机配置协议 (DHCP) 分配的 VIP 地址（这里的 DHCP 服务器不属于 Oracle 网格基础架构软件），现在可以通过 GNS 动态解析名称了。实施 GNS 和自动 VIP 分配意味着在集群中添加或删除服务器时，不再需要执行任何 VIP 和名称分配。就这点而言，集群实现了自我管理。

在安装期间（或作为安装后的维护任务）可以选择是否使用 GNS 和新的动态名称解析，通常选择启用。在激活 GNS 之前，必须将企业 DNS 中的一个域委托给 GNS 和 GNS IP（DNS 中唯一的静态项）。当客户需要访问委托域中的节点时，就可以透明地进入集群。

基于策略和角色分离的集群管理

使用 Oracle Grid Naming Service (GNS) 可以解决在普通动态环境中使用指定的固定名称的问题。即使使用 GNS，也需要动态分配应用程序，才能在集群中的特定服务器或（更一般地来讲）硬件资源上运行。Oracle Clusterware 中新的服务器池的概念解决了这一问题。

服务器池将集群划分成几部分。它们在集群中创建逻辑分区，通常将运行类似负载的服务器组合到一个服务器组中。因此服务器池能够按需动态分配容量，并在必要时确保隔离（可以将服务器池看作“集群中的一组专用服务器”）。

服务器池并不决定应用程序或资源在集群中的位置。策略来保证应用程序在适当数量的服务器上运行。应用程序必须运行在能满足其负载要求的最少数量的服务器上。因此，除其名称外还需要为服务器池定义三个属性：

1. **Min** — 指定服务器池中运行的最小服务器数。
2. **Max** — 指定服务器池中运行的最大服务器数。
3. **Imp** — 指定服务器池之间的相对重要性。在将服务器分配给服务器池或由于故障需要重新分配服务器池中的服务器时，会涉及到这个参数。

由于应用程序现在被定义为在服务器池而不是指定的服务器上运行，因此在所有服务器池上应用这些属性可以定义一个应用程序分配策略，将应用程序分配给集群中的硬件资源。

在这种情况下，仍然可以“按传统方式”管理集群 — 使用指定的服务器并将应用程序和资源分配给不同的独立服务器。这是为了保证向后兼容性。

此外，在网格中使用异构服务器时也需要使用指定的服务器。其前提通常是只采用异构服务器。如果在同一集群中使用计算资源（CPU /内存）完全不同的服务器，在将应用程序分配到服务器池时，就需要使用指定的服务器来反映这些区别。

当集群作为共享网格基础架构为企业内多个组织群体提供服务时，以服务器池为单位采用角色分离的管理就非常合适。

现在，在 IT 组织之间实施了严格职责分离的企业，将能够体现分离的效果，即使这些组将并发运行于同一共享网格基础架构上。角色分离的管理，通过限定每个组只能管理和操作分配给它们的（硬件）资源来确保不同组（组织）之间不会争夺资源。

在 Oracle Clusterware 11g 第 2 版中，角色分离的管理分为两种模式，如图 4 所示。其中，横向模式与 Oracle 之前的版本相同，在实施 Oracle RAC 体系时使用操作系统用户/组凭证。Oracle Clusterware 11g 第 2 版使用了专用的组分配。

“纵向模式”在 Oracle Clusterware 11g 第 2 版中首次推出，在共享同一网格基础架构的各个组织群体之间实现明确分离。这个新方法基于服务器池概念，通常指定一些操作系统 (OS) 用户来修改分配给其他操作系统用户的服务器池，以实现管理。

负责安装（也就是安装 Oracle Clusterware 软件）的操作系统用户和主用户（管理员或 root 用户，取决于 OS）也是集群的管理员。

在 Oracle Clusterware 11g 第 2 版中，使用内部访问控制列表 (ACL) 分配服务器池操作权限。以后的版本将计划集成目录服务，例如，Oracle Internet Directory。默认安装无法预见纵向职责分离。

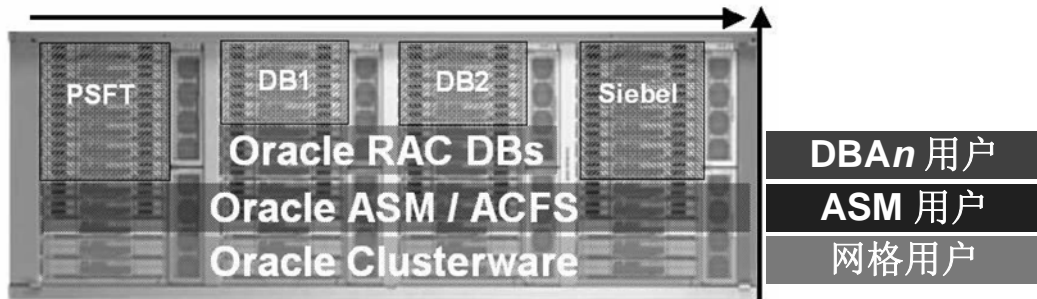


图 4：角色分离的管理

集群化（集群识别）命令

管理大型网络需要对远程节点执行管理任务（远程操作）。要管理具有数百个节点的网格，就必须同时调用集群中的所有节点操作或一组节点的操作。Oracle Clusterware 提供了基于集群识别命令行的命令，即集群化命令。例如，通过使用一条命令对集群中每个服务器进行查询就可以检查集群的状态。

基于 EM 的资源 and 集群管理

Oracle Enterprise Manager (EM) 扩展了集中式集群管理的理念并将其提升至新的水平。Oracle EM 不仅可以用来管理整个集群，还可以通过一个图形用户界面管理集群中的所有应用程序。

在 Oracle 网络基础架构的第 1 版中，只能使用 Oracle Enterprise Manager DatabaseControl 对集群及其所有资源（应用程序和数据库）进行整体管理。在新版本的 Oracle Enterprise Manager Grid Control 中，将继续沿用这一功能。

但是要使用 Oracle EM Database Control，至少要在集群中安装一个集群识别数据库。该数据库不需要在集群上运行并执行任何与资源相关的操作。但只有当 Oracle Enterprise Manager Database Control 使用的数据库处于联机状态时才能进行所有集群操作。

ORACLE Enterprise Manager 11g
Database Control

Cluster Database

Cluster: cluster7 >

Manage Resources

Page Refreshed Jun 4, 2009 11:27:29 AM PDT Refresh

Oracle Grid Infrastructure provides High availability framework to protect any application that is registered with Grid Infrastructure. You can Create, Administer and Monitor these Resources using this interface.

Resources 23 (10 4)
(Including Internal Oracle Resources)

Search Go Advanced Search

☒ Show Oracle Resources

Add Resource Add Application VIP

View Edit Delete Start Stop Relocate

Select All Select None Show All Details Hide All Details

Select	Details	Name	Cardinality	Current State	Target State	Running Hosts	Resource Type	Owner
<input type="checkbox"/>	Show	ApacheVIP	1	UP	UP	stbpo57	app appvip type	root
<input type="checkbox"/>	Show	MyApache	1	UP	UP	stbpo57	cluster_resource	root
<input type="checkbox"/>	Show	myclock	1	UP	UP	stbpo57	cluster_resource	oracle
<input type="checkbox"/>	Show	oraDATA.dg	Runs on all servers	UP	UP	stbpo55, stbpo56, stbpo57, stbpo58	ora.diskgroup type	oracle
<input type="checkbox"/>	Show	oraLISTENER.lsnr	Runs on all servers	UP	UP	stbpo55, stbpo56, stbpo57, stbpo58	ora.listener type	oracle
<input type="checkbox"/>	Show	oraLISTENER_SCAN1.lsnr	1	UP	UP	stbpo56	ora.scan_listener type	oracle
<input type="checkbox"/>	Show	oraLISTENER_SCAN2.lsnr	1	UP	UP	stbpo58	ora.scan_listener type	oracle
<input type="checkbox"/>	Show	oraLISTENER_SCAN3.lsnr	1	DOWN	DOWN	n/a	ora.scan_listener type	oracle
<input type="checkbox"/>	Show	ora.asm	Runs on all servers	UP	UP	stbpo55, stbpo56, stbpo57, stbpo58	ora.asm type	oracle
<input type="checkbox"/>	Show	ora.eons	Runs on all servers	UP	UP	stbpo55, stbpo56, stbpo57, stbpo58	ora.eons type	oracle
<input type="checkbox"/>	Show	ora.gsd	Runs on all servers	DOWN	DOWN	n/a	ora.gsd type	oracle
<input type="checkbox"/>	Show	ora.net1.network	Runs on all servers	UP	UP	stbpo55, stbpo56, stbpo57, stbpo58	ora.network type	root
<input type="checkbox"/>	Show	ora.oc4j	1	UP	UP	stbpo55	ora.oc4j type	oracle
<input type="checkbox"/>	Show	ora.ons	Runs on all servers	UP	UP	stbpo55, stbpo56, stbpo57, stbpo58	ora.ons type	oracle
<input type="checkbox"/>	Show	ora.racdb.db	Runs on server pool(s) ora.RACpool	DOWN	DOWN	n/a	ora.database type	oracle
<input type="checkbox"/>	Show	ora.registry.acfs	Runs on all servers	UP	UP	stbpo55, stbpo56, stbpo57, stbpo58	ora.registry.acfs type	root
<input type="checkbox"/>	Show	ora.scan1.vip	1	UP	UP	stbpo56	ora.scan_vip type	root
<input type="checkbox"/>	Show	ora.scan2.vip	1	UP	UP	stbpo58	ora.scan_vip type	root
<input type="checkbox"/>	Show	ora.scan3.vip	1	DOWN	DOWN	n/a	ora.scan_vip type	root
<input type="checkbox"/>	Show	ora.stbpo55.vip	1	UP	UP	stbpo55	ora.cluster_vip_net1 type	root
<input type="checkbox"/>	Show	ora.stbpo56.vip	1	UP	UP	stbpo56	ora.cluster_vip_net1 type	root
<input type="checkbox"/>	Show	ora.stbpo57.vip	1	UP	UP	stbpo57	ora.cluster_vip_net1 type	root
<input type="checkbox"/>	Show	ora.stbpo58.vip	1	UP	UP	stbpo58	ora.cluster_vip_net1 type	root

View Edit Delete Start Stop Relocate

图 5: Oracle Enterprise Manager 资源概述

提高可用性 — 高级调优

Oracle Clusterware 11g 第 2 版能够提供更高的可用性。例如，新的基于代理的监视系统监视所有资源。这些内存驻留代理允许使用更少的资源进行更频繁的检查。更频繁的检查意味着更快速的故障检测和更短的恢复时间。Oracle 监听器的平均故障检测时间已经从 5 分钟缩减到 30 秒，而检查间隔从 10 分钟缩减到 1 分钟。

此外，Oracle Clusterware 还降低了软件维护所需的计划停机时间。Oracle Clusterware 11g 第 2 版使用“Oracle 集群零停机补丁”、“异地升级”和“仅限于软件的安装”以最短停机时间执行所有软件维护操作。

高级可用性

现在，使用更高级的新型高可用性框架，Oracle Clusterware 11g 第 2 版能够为集群中的所有应用程序提供更高可用性。用来监视所有已部署资源的基于代理的监视系统只是其中的一个示例，另一个示例是扩展的相关性模型。

如今的应用程序不再基于唯一的组件或进程。Oracle Siebel 或 Oracle Peoplesoft、SAP 等复杂的应用程序都基于多个组件或进程。

要为整个应用程序提供高可用性就需要不断监视多个进程。使用更灵活的相关性模型，就可以在 Oracle Clusterware 11g 第 2 版中轻松地对这些和其他实际配置建模。

Oracle Clusterware 为应用程序开发人员提供集群 API，以便通过编程与 Oracle Clusterware 交互。这些 API 现已整合到代理开发框架中，通过它们可以更简单高效地开发 Oracle Clusterware 代理。此外，Oracle Clusterware 11g 第 2 版提供的默认脚本代理可以保护各种应用程序，就像在之前的版本中一样。

隔离的灵活性及第三方集群解决方案支持

传统上，Oracle Clusterware 使用 STONITH (Shoot The Other Node In The Head) 比较隔离算法来确保在集群完整性存在危险、“裂脑”等的情况下数据的完整性。对于 Oracle Clusterware，这意味着会通过一个本地进程从集群中强制删除一个或多个节点（隔离）。

除此传统的隔离方法之外，Oracle Clusterware 当前支持基于远程节点终止的隔离方法。这个方法的思想是，使用一种外部机制重启故障节点，而且即不需要 Oracle Clusterware 也不需该节点上运行的操作系统的协作。为了提供此功能，Oracle Clusterware 11g 第 2 版支持智能管理平台接口 (IPMI) 规范，这是一个标准的管理协议。

为了使用 IPMI，并且为了能够对集群中的一台服务器进行远程隔离，该服务器必须装备有主板管理控制器 (BMC)，BMC 支持局域网 (LAN) 上的 IPMI。一旦在集群的每台服务器中装备好该硬件，即可在 Oracle 网格基础架构的安装过程中，或者在安装之后使用 CRSCTL 执行安装后管理任务的过程中激活 IPMI。

Oracle Clusterware 继续在其下支持第三方集群解决方案。对于经过认证的解决方案（在 My Oracle Support/Certify 中可找到经过认证的解决方案），Oracle Clusterware 将以遵从第三方集群解决方案作出节点成员关系决定的方式来与第三方集群解决方案相集成。只有在一定时间内未作出决定的情况下，Oracle Clusterware 才会执行纠正性措施，使用上述隔离机制中的一种。

在 Oracle Clusterware 下维护第三方集群解决方案增加了集群体系的复杂度，使集群管理更加困难。因此，Oracle 建议避免在同一系统中使用多个集群解决方案。对于 Oracle RAC 环境，值得注意的是 Oracle Clusterware 是必需的，它提供了所有需要的功能。因此不再需要其他第三方解决方案。

冗余互连使用

在以前的版本中，为了在节点间利用冗余网络连接（用作冗余、专用、私有的信道即“互连连接”），需要绑定、中继、成组或类似的技术。如今，Oracle Clusterware 提供一个集成的解决方案来确保“冗余互连”。此功能从 Oracle Database 11g 第 2 版，补丁集一 (11.2.0.2) 开始提供。

该冗余互连使用特性并非直接作用于网络接口上。相反，它基于一种多监听端点体系结构，在该体系结构中，对每个专用网络分配一个高可用的虚拟 IP (HAIP) 地址（最多支持 4 个接口）。

默认情况下，Oracle Real Application Clusters (RAC) 软件使用所有的 HAIP 地址进行专用网络通信，以便在指定用于专用网络的接口组上提供负载均衡。如果一个专用互连接口出现故障或处于非通信状态，则 Oracle Clusterware 将相应的 HAIP 地址透明地转移给其余正常工作的接口中的一个。

从 Oracle Database 11g 第 2 版补丁集一 (11.2.0.2) 开始，Oracle RAC 数据库、Oracle 自动存储管理（集群 ASM），以及 Oracle Clusterware 组件如 CSS、OCR、CRS、CTSS 和 EVM 组件采用冗余互连特性，除此之外的 Oracle 软件及第三方软件不支持该特性。

Oracle Clusterware 11g 第 2 版中的无重启节点隔离

如前所述，Oracle Clusterware 使用 STONITH (Shoot The Other Node In The Head) 比较隔离算法来确保在集群完整性存在危险、“裂脑”等的情况下数据的完整性。对于 Oracle Clusterware，这意味着会通过一个本地进程从集群中强制删除一个或多个节点（隔离）。

在 Oracle Clusterware 11g 第 2 版，补丁集一 (11.2.0.2) 之前，通过“快速重启”相应的服务器来进行节点隔离。“快速重启”包括关闭和重启过程，该过程在执行关闭时既不等待任何 IO 操作完成也不等待文件系统进行同步。在 Oracle Clusterware 11g 第 2 版，补丁集一 (11.2.0.2) 中，为了尽可能避免这样的重启而对此机制进行了更改。

在 Oracle Clusterware 11g 第 2 版中已对此算法进行了改进，集群中一些为 Oracle RAC 所需要的子组件如果出现故障，不一定会引发对节点的立即隔离（重启）。相反，在集群中会尝试清除这一故障并重启故障子组件。只有在故障组件清除不成功时，才会重启节点以便强制进行清除。

在 Oracle Clusterware 11g 第 2 版，补丁集一 (11.2.0.2) 中还进行了进一步的改进以便 Clusterware 可以尝试不重启节点而防止“裂脑”，这实现了客户要求保留节点和避免重启的要求，因为节点上运行着不受 Oracle Clusterware 管理的应用程序，若非如此，这些应用程序会因节点重启而被强制关闭。

使用新的算法，当决定从集群中删除一个节点时，Oracle Clusterware 首先会尝试关闭作为删除对象的节点计算机上的所有资源。特别地，终止 IO 进程，确保完全停止这些进程之后再进行下一步操作。如果因某种原因无法停止其中某些资源，或者无法终止所有 IO 进程，Oracle Clusterware 仍然会执行重启或使用 IPMI 从集群中强制删除该节点。

如果可以停止所有的资源，并且可以终止所有的 IO 进程，Oracle Clusterware 将会在相应的节点上关闭自身，但是会在其系统停止后重启。该重启由 Oracle Clusterware 11g 第 2 版中引入的 Oracle High Availability Services Daemon 发起。

集群时间同步和网络管理

集群节点间的时间同步至关重要。虽然集群中各服务器间的时间偏差并不一定会导致不稳定，但是时间不同步可以使集群的整体管理难以进行。其中一个原因是时间戳是用本地节点时间写入的。如果集群中时间偏差很大，日志分析可能会受到严重的影响。

一般会在数据中心使用一个中央时间服务器（为 NTP 所访问）来对服务器时间进行同步以防集群节点间出现时间偏差。这是避免突然调整各个节点时间的最佳方法，如果时间调整过于突然，可能会导致从集群中删除节点。

为使 Oracle 网格基础架构不受外部资源（故障）的影响，可以使用新引入的 Oracle Cluster Time Synchronization Service Daemon (OCTSSD) 作为备用方法在一个集群的各个服务器间进行时间同步。

Oracle CTSS 后台进程常常已安装并且总是在运行之中，但是对它的配置是根据系统中的实际情况来进行。如果系统中安装了 NTP，则 CTSS 以观察器模式启动，不进行时间同步。只有在集群的任何服务器上都没有 NTP 时，CTSS 才会以活动模式启动，以一台服务器作为基准服务器进行集群中的时间同步。

另外，在网络管理方面增加了更多的灵活性。在技术上，主要的增强是 Oracle Clusterware 管理下的一个网络资源。作为一个本地集群资源，它不断地监视着每台服务器上的网络。在发现网络中断时，Oracle Clusterware 管理下的相关资源（如 VIP）会得到通知，如有必要，会故障切换到另一个节点。

Oracle Clusterware 对集群中的每个子网维持一个网络资源。多子网支持是一个新的特性，该特性有助于在网格基础架构中整合应用程序和数据库。多子网支持特性支持在集群中使用不同的子网来独立地访问应用程序和数据库，这些子网就好像是数据库和应用客户端双方的独立环境。

Cluster Health Monitor: 与 Oracle 网格基础架构相集成

Cluster Health Monitor (CHM) 以前称为 Instantaneous Problem Detector for Clusters 或 IPD/OS，用于检测和分析与操作系统 (OS) 和集群资源相关的性能下降和故障，以便为运行 Oracle Clusterware 和 Oracle RAC 的集群中发生的许多问题提供更有力的说明。

该工具持续跟踪每个节点、进程和设备级的 OS 资源消耗情况。它收集和分析整个集群范围内的数据。在实时模式中，达到阈值后，即刻向操作者显示警报。为了进行根本原因分析，可以重放历史数据以便了解出现故障时发生的情况。

使用 Oracle Clusterware 管理 Oracle RAC 数据库

Oracle RAC 与 Oracle Clusterware 紧密集成。这意味着，只要使用 Oracle RAC 数据库就需要 Oracle Clusterware。在 Oracle 的体系中，没有任何其他集群软件可以代替 Oracle Clusterware，也不需要任何其他集群软件和 Oracle RAC 数据库一起运行。

类似 SRVCTL 这样的工具专用于管理集群中的 Oracle 预配置资源，因而可以简化 Oracle RAC 数据库的日常管理任务，这类工具证明了这两种产品的强大集成。

例如，从较低层次讲，在隔离和实例删除方面，Oracle RAC 被视为集群中的一个特殊的应用程序。Oracle Clusterware 11g 第 2 版与 Oracle RAC 数据库系统相集成的其他例证还有：内部代理，这些代理提供更多与 Oracle ASM 和 Oracle RAC 实例相关的状态信息、Clusterware 管理的服务以及单一客户端访问名称 (SCAN)（这使得客户端可以使用单一集群别名连接到集群中任何数据库）。

使用 Oracle Clusterware 管理任何类型的应用程序

所有类型的应用程序和进程都可作为集群中的资源由 Oracle Clusterware 来监视和管理。这些资源可以作为集群中每个节点上的本地资源而运行，从而提供基础架构组件。此外，还可将它们定义为集群资源，这种情况下，它们将以故障切换方式来管理应用程序。

非 Oracle 应用程序必须使用 CRSCTL 工具来管理。而 Oracle Enterprise Manager (EM) 数据库控制可用于管理集群中所有类型的资源。

通过 Oracle Clusterware 11g 第 2 版，对应用程序进行注册以便在 Oracle Clusterware 中加以管理变得异常轻松。为此，只需完成以下四个步骤（包括集群中的故障切换测试）：

1. 创建一个应用程序特定操作脚本或单独的代理
2. 创建一个应用程序 VIP 以便访问该应用程序
3. 在 Oracle Clusterware 中配置和注册该应用程序
4. 检查该应用程序的启动/停止，验证成功后完成注册

Oracle Enterprise Manager 支持上述每个步骤以便轻松进行集中的部署。例如，通过图 6 所示的对话框能够将资源添加到集群中。其中包括操作脚本的创建和资源的注册。

通过其他对话框，可以创建应用程序 VIP 或在集群中的服务器间重新分配资源。大多数集群相关的管理操作即使在分配给 Oracle Enterprise Manager Database Control 的数据库关闭后也可以执行。

ORACLE Enterprise Manager 11g
Database Control

Cluster Database

Add Resource

General Parameters Advanced Settings Dependencies

* Name: MyApache

* Resource Type: cluster_resource (View) (Add)

Description: Failover resource for the Apache WebServer

☒ Start the resource after creation

Placement

The following parameters define where the resource would be placed.

Placement: ☐ Anywhere in the Cluster where this resource instance can be started
☒ Restrict or Favor the placement of resource to the Server Pools: AppsPool
☐ Restrict or Favor the placement of resource to the Hosts

Placement Policy: ☒ Restricted
☐ Favored

Cardinality: ☒ Specify number: 1
☐ Set to size of Server Pool(s) on which the resource is running

Degree: 1

Active Placement: ☐ Re-evaluate resource's placement during addition or restart of a cluster node

Action Program

Action Program defines the way to start, stop and check the status of a resource. Action Program could be an executable (Agent File) and / or a script (Action Script) that the Oracle Clusterware can invoke. Action Program should accept 'start', 'stop' or 'check' as argument to perform respective operations. User can implement all these operations using Agent File alone or Action Script alone or using a combination of both (some operations in Agent File and some in Action Script). If both implement the same operation, Agent File operation would override the Action Script operation.

Action Program: Use Action Script

Action Script Name: /myshared/scripts/myapache.scr (Create New Action Script)

☐ Overwrite if already exists (on any node of the cluster)

General Parameters Advanced Settings Dependencies

图 6: 使用 Oracle Enterprise Manager 向集群添加资源

总结

人们需要能够管理大型网格，无论其中部署了多少个应用程序，这日益成为一种挑战。Oracle 网格基础架构克服了这类挑战。使用服务器池、网格即插即用以及集成的标准管理工具，Oracle Clusterware 如今可管理网格中所有的应用程序，就好像这些程序运行于同一个系统之上，从而提高了可用性、可伸缩性和灵活性。

甲骨文（中国）软件系统有限公司

北京总部

地址：北京市朝阳区建国门外大街1号，国贸大厦2座2208室
邮编：100004
电话：(86.10) 6535-6688
传真：(86.10) 6505-7505

北京上地6号办公室

地址：北京市海淀区上地信息产业基地，上地西路8号，上地六号大厦D座702室
邮编：100085
电话：(86.10) 8278-7300
传真：(86.10) 8278-7373

上海分公司

地址：上海市卢湾区湖滨路222号，企业天地商业中心1号楼16层
邮编：200021
电话：(86.21) 2302-3000
传真：(86.21) 6340-6055

广州分公司

地址：广州市天河北路233号，中信广场53楼5301&5308室
邮编：510613
电话：(86.20) 8513-2000
传真：(86.20) 3877-1026

成都分公司

地址：成都市人民南路二段18号，四川川信大厦20层A&D座
邮编：610016
电话：(86.28) 8619-7200
传真：(86.28) 8619-9573

大连分公司

地址：大连软件园东路23号，大连软件园国际信息服务中心2号楼五层502号A区
邮编：116023
电话：(86.411) 8465-6000
传真：(86.411) 8465-6499

济南分公司

地址：济南市涿源大街150号，中信广场11层1113单元
邮编：250011
电话：(86.531) 8518-1122
传真：(86.531) 8518-1133

甲骨文软件研究开发中心（北京）有限公司

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：100094
电话：(86.10) 8278-6000
传真：(86.10) 8282-6455

甲骨文研究开发中心（深圳）有限公司

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

沈阳分公司

地址：沈阳市沈河区青年大街219号，华新国际大厦17层D单元
邮编：110016
电话：(86.24) 2396 1175
传真：(86.24) 2396 1033

南京分公司

地址：南京市玄武区洪武北路55号，置地广场19层1911室
邮编：210028
电话：(86.25) 8476-5228
传真：(86.25) 8476-5226

杭州分公司

地址：杭州市西湖区杭大路15号，嘉华国际商务中心702室
邮编：310007
电话：(86.571) 8717-5300
传真：(86.571) 8717-5299

西安分公司

地址：西安市高新区科技二路72号，零壹广场主楼1401室
邮编：710075
电话：(86.29) 8833-9800
传真：(86.29) 8833-9829

福州分公司

地址：福州市五四路158号，环球广场1601室
邮编：350003
电话：(86.591) 8801-0338
传真：(86.591) 8801-0330

重庆分公司

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) 6370-8898
传真：(86.23) 6370-8700

深圳分公司

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

甲骨文亚洲研发中心（上海）

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：86-21-6095 2555



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© 2010 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。