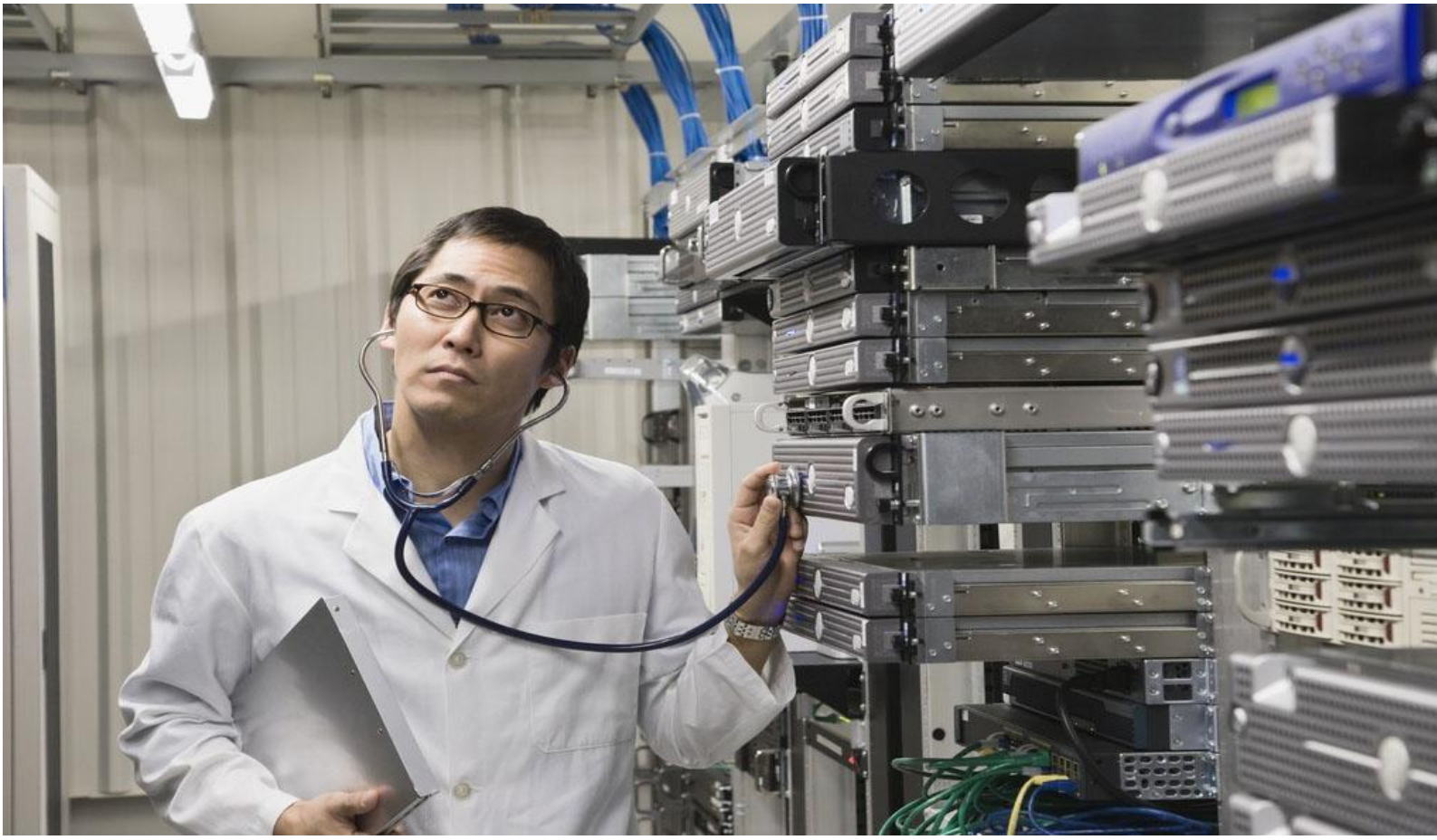
Oracle 系列电子书

Oracle RAC 最佳实践

由于最终用户习惯于获得瞬间响应时间，Oracle 为其产品提供持续可用性方面受到了前所未有的挑战，因此 RAC 应运而生。

- *Oracle RAC 几个常见的错误观点*
- *Oracle RAC 实施最佳实践*
- *Oracle RAC 技术最佳实践*
- *如何准确定义 RAC 数据库的职业角色*

TechTarget 中国



Oracle RAC（真正可用集群）

最佳实践

由于最终用户习惯于获得瞬间响应时间，Oracle 为其产品提供持续可用性方面受到了前所未有的挑战，因此 RAC 应运而生。——Don Burleson

随着最终用户习惯于获得瞬间响应时间，Oracle 为其产品提供持续可用性方面受到了前所未有的挑战。那些在 Redwood Shores(译者注：Oracle 公司总部所在地)的家伙们提供了一个重要的工具这就是 Oracle Real Application Clusters (RAC)，即真正可用集群。

什么是 RAC？简单来说就是一套允许单个数据库被多份 Oracle 程序同时访问的软件工具。如果一个服务器崩溃了，事务能够在最短的宕机时间内被重定向到其他存活的服务器上。

Oracle 的宣传称 RAC 是治愈多种疾病的良药，而 IT 厂家则会对这样的市场宣传产生误解，从而无法正确区别在高可用环境(HA)中使用 RAC 的成本和收益。

那么就让我们探索一些 Oracle RAC 最佳实践以及明了在使用这种基于集群技术的一些错误观点。在本次的 Oracle 技术电子书中，我们将会涉及以下内容：

- RAC 规划最佳实践
- RAC 实施最佳实践
- RAC 架构最佳实践
- 硬件架构和 RAC 性能
- RAC 备份和恢复最佳实践

- 性能调优最佳实践

Oracle RAC 几个常见的错误观点

Oracle RAC 是为了提供扩展性的

尽管 Oracle 公司希望你买小型刀片服务器然后使用他们的网格计算方案来获得“水平扩展”，但是实际上这并不是多数用户使用 RAC 的方法。注意：RAC 只是在超大型 IT 部门需要超过单个服务器提供极限的更多马力时的一种正统扩展方法。

作为 Oracle 最佳实践，要通过“垂直扩展”先进行单个服务器的扩容，先向上扩展再向外扩展。只有在你使单个服务器容量饱和之后再考虑“scale out”到多个服务器上。今天，单个服务器的内存和 CPU 马力比起前几年来说有了突飞猛进，因此比起往 RAC 环境中添加一个新服务器而言，增加单个机器的资源更加简单。在真实环境中，单个服务器能够处理每秒上千次的事务。只有世界上最大的那些 Oracle 数据库需要扩展到使用 RAC。

Oracle RAC 是独立的高可用解决方案

记住 RAC 只能保护你免于实例失效，这仅仅是众多可能引起非计划性中断的原因之一。为了真正的持续可用性，我们还必须部署多重镜像磁盘和冗余网络组

件。

为了每个 RAC 节点的可用性，需要多个冗余的主机总线适配器，多个网卡以及多个电源。就算只是在数据库实例产生了 failover，你也需要提供软件以允许多个主机总线适配器自动 failover，并且提供单个组件失效通知。

就像我们已经提到的，RAC 系统需要一个 cluster interconnect 来提供内存对内存(RAM-to-RAM)的数据块传输。Interconnect 必须非常快速，必须有高带宽和低延迟。Interconnects 包括：

- Dark fiber: Dense Wavelength Division Multiplexing (DWDM) technology
- Infiniband
- Myrinet

Cache fusion 上的瓶颈也是为什么 RAC 扩展或者说水平扩展有问题的另外一个原因。如果你的 cluster interconnect 无法处理这样的流量，那么额外的服务器将会降低整个系统的性能而不是提升它。解决这个问题的唯一办法就是改造应用以适应 RAC，或者采购更快的存储比如说固态硬盘。

Oracle RAC 保证了快速响应时间

事务响应时间总是重要的，但是它对于 RAC 数据库来说尤为重要。这是因为在连接阶段为了探测是否一个 RAC 节点或者机器已经失效而消耗了等待时间，因

此你必须保证新事务要小于 1 秒时间这样才能保证 2 秒的 failover 时间。

- **Oracle RAC 不需要灾难恢复组件**

除了在少数案例中使用了 DWDM 技术(也就是 dark fiber)，你仍然需要创建一份灾难恢复解决方案。因为 RAC 节点通常只间隔数英里，像飓风这样的自然灾害还是能够引起全局中断。因此 RAC 最佳实践中还是要包含一份地理上的快速失效接管解决方案，比如 Data Guard 或者更好一些，多路流复制。

Oracle RAC 实施最佳实践

RAC 数据库通常都遵循着跟任何 Oracle 数据库一样的最佳实践方案，只是其中有一些是 Oracle RAC 系统特有的。首先，很重要的一点是如何尽量缩小 RAC 节点间的物理距离而又能保证它们之间互相独立，这样才能避免所有节点同时失效的情况。

在一个繁忙的 RAC 数据库中，server interconnect(服务器内联)的速度对于系统响应速度有决定性作用。最佳实践推荐我们使用尽可能快的 interconnect，通常是比如 dark fiber 这样的光纤解决方案。

有些客户将 RAC 节点分布在相邻的独立建筑中，得益于超快的 dark fiber

interconnect 技术，可以使用“Extended RAC”架构来使 RAC 节点的距离大到 100 英里。这使您同时拥有了高可用性和灾难恢复能力。

不过 Dark fiber 非常昂贵，为了降低成本，大多数客户采取了将 RAC 和灾难恢复方案比如多路流复制结合在一起的最佳实践。

RAC 的要点是保证最终用户在一个服务器失效后自动重新连接到还存活的其它服务器上。这是通过应用服务器层面或者是 Oracle Transparent Application Failover (TAF)功能实现的。但是不论选择哪种方式，你都必须等待大概 3 秒，之后才会认为此服务器已失效而重新尝试连接到另外的服务器。

接下来，让我们更进一步探讨一下特定 RAC 技术的最佳实践。

● Oracle RAC interconnect 最佳实践

RAC 是多实例共享同一数据库的方法，共享数据块通过高速 interconnect 在节点之间传输，这称为 cache fusion。为了保证性能，关键之处在于密切关注 interconnect 层面并且记住以下几点：

RAC 喜欢较小的 block size，interconnect 必须拥有足够快速的网络硬件，RAC 负载均衡对性能至关重要。

● Oracle RAC 节点负载均衡最佳实践

我不太同意 Oracle 提出的负载均衡基于最小负载的实现方法，因为这增加了额外的 cache fusion。在实际环境中，相似业务的最终用户都将请求发送到同一 RAC 节点上。如果我们的 RAC 系统有不同类型的最终用户，我们会希望将负载均衡到不同的数据区域去。举例来说，客户处理可能在节点 1 上，订单处理在节点 2 上，而产品处理则在节点 3 上。将 RAC 最终用户通过数据需求来分组可以保证 cache fusion 负载降到最小。

RAC 喜欢较小的 block size，interconnect 必须拥有足够快速的网络硬件，RAC 负载均衡对性能至关重要。

● Oracle RAC 磁盘存储管理最佳实践

为了实施 RAC 系统，你应该使用共享存储设备因为很多服务器都必须同时存取磁盘。一个单实例数据库可以使用 Direct Attached Storage (DAS)这是一种连接到单一服务器上的一组廉价磁盘，而 RAC 则必须使用 Storage Area Network (SAN)，这是更昂贵更复杂的通常使用光纤通道连接到多个服务器的磁盘阵列。这

需要一组独立的硬件，从主机总线适配器连接到 SAN 上。因此 DBA 具有数据存储层面的完整知识就显得很重要。

● Oracle RAC 块大小最佳实践

最佳实践是在 RAC 环境中使用小的 2K block size 以在 cache fusion 时最小化“baggage”传输。因为 block size 是工作的单位，block size 越小，就能够通过更小的负载传输越高粒度的数据。如果你有较长的数据行(大于 2K)，则需要转而使用 4K 的 block size。

实施 RAC 集群仅仅是开始，持续监控 RAC 集群的健康状况在造成最终用户困扰之前就及时定位解决问题也是至关重要的。

● Oracle RAC 监控最佳实践

为了保证 RAC 节点永远不会碰到全局问题(译者注：所有节点都失效)，正确的监控架构都必须的。RAC 数据库很少在没有任何报警的情况下就失效。如果 DBA 懂得监控正确的指标，他就能够创建一套预警系统，能够及时发现问题并通知他，让他在实例崩溃之前就修复问题。

DBA 必须监控集群，共享磁盘，ASM(或者 OCFS)，数据库实例，监听，和更多的深层次指标，比如缓存一致性，interconnect 延迟，磁盘时间等等一系列事情。

虽然高成本的性能监控工具比如 Oracle Grid Cntrol 能够帮助初学者进行初步的 RAC 监控，但是一个 RAC DBA 还是应该具有编程技巧，使用查询数据字典，dbms——scheduler 以及邮件告警机制来创建属于自己的 RAC 监控架构。

如何准确定义 RAC 数据库的职业角色

Oracle RAC 数据库最佳实践的讨论先划上句号，让我们把注意力放在如何最好的定义 RAC 数据库的职业角色上。

● Oracle RAC 人员编制

RAC 数据库最佳实践是雇佣经验丰富的 RAC DBA 来管理集群，避免招聘那些仅仅有培训经历而无实际工作经验的人。

认识到人力资源成本是 Oracle 管理中的最昂贵部分是很重要的，这数十年来，硬件开销已经极大下降但是人力成本还是与以往一样。

要注意到拥有 RAC 技术的 Oracle 专家比普通的 DBA 更值钱。最近的 Oracle 从业人员工资调查指出，DBA 平均年薪是 97000 美元，RAC 专家通常可以拿到每年 140000 美元。而那些管理着价值上百万美元的 RAC 数据库的专家则通常年薪高达 250000 美元。

可惜，培养自己的 RAC DBA 并不是简单的事儿，培训教程非常昂贵，并且又没有能够代替真实环境经验的课程。培训自己的 DBA RAC 技术会使他更有市场竞争力，花费了数万美元来培训自己的 DBA 然后他跳槽去了更好的地方这样的情况并不鲜见。

● Oracle RAC 工作角色

系统管理员(SA，负责管理数据库和磁盘)和 RAC DBA(负责管理 RAC 数据库)之间永远都有冲突。同时也有网络管理员这个职位的明确定义，这是专门管理 RAC 数据库环境的集群心跳内联和节点间数据传输的职位。

如果 DBA 负责保障 RAC 数据库性能，那么他应该被被给予服务器和磁盘存储子系统的 root 权限，但是，并不是所有 DBA 都拥有管理复杂服务器和 SAN 环境的计算机知识，因此看个案而定是否给予此权限。

● Oracle RAC 培训

不能正确地培训 SA，DBA 和网络管理员就会导致系统的非计划停机，SAN 环境，比如 EMC，Tagmastore 和 NetApp 都有复杂的架构，这需要定期的培训课程。

磁盘配置也同样具有挑战性，RAC 只有在使用了特定的磁盘配置(比如 ASM，OCFS，RAW 或者第三方集群文件系统)时才能正常运转。这些工具也同样需要培训。

网络管理员也必须接受培训，来知道如何设置 cluster interconnect，也包括在诸如 Infiniband 和 DWDM 上配置 interconnects。

这些 RAC 相关职位中，DBA 有最大的学习曲线。他们必须要明白如何配置和管理所有复杂的 RAC 组件，包括 clusterware 和文件系统存储。

结论

总体来说，虽然 RAC 提供了持续的可用性，它也不是什么神话。有很多工作需要做才可以保证 RAC 数据库使用可用。每一个 RAC 数据库都有其特性，但是也有一些大家都知道的风险和陷阱。使用 Oracle RAC 最佳实践是保证成功的必修课。

我们的编辑团队

您若有何意见与建议，欢迎与我们的编辑联系。

诚挚感谢以下人员热情参与 TechTarget 中国《Oracle 系列电子书》的内容编辑工作！

诚邀更多的数据库专业人士加入我们的内容建设团队！

Donald Burleson

Donald Burleson 是全球闻名的 Oracle 专家，主攻领域为 Oracle 性能调优。曾出版过三十余本 Oracle 书籍，目前从事 Oracle 咨询相关工作。

张乐奕

TechTarget 中国特邀技术顾问。国内知名 Oracle 专家，曾就职于甲骨文公司，担任 Oracle 顾问角色，拥有多年 Oracle 数据库管理经验，并获得 Oracle ACE 称号。