

白皮书

使用 **Oracle** 数据库 11g 控制存储成本

作者: **Brian Babineau**
Bill Lundell

2008 年 2 月

目录

目录	i
执行概要	1
引言	2
应对信息增长	2
了解存储成本	3
资本费用	3
管理数据库存储	3
常被忽略的运作费用	3
制定信息生命周期管理战略	4
创建多层式 IT 基础架构	4
高效缩减数据	4
Oracle 数据库 11g 优化 ILM 存储战略	5
通过 Oracle 数据库分区利用分层式存储	5
通过压缩提高系统使用率	5
通过 Oracle 数据库 11g 节省成本	6
结论	7

所有商标名称都是它们各自公司的资产。Enterprise Strategy Group (ESG) 认为本出版物中信息的来源是可靠的，但 ESG 不对此作出担保。本出版物可能包含 ESG 的观点，并随时可能更改。本出版物的版权属于 Enterprise Strategy Group, Inc.。任何未经 Enterprise Strategy Group, Inc. 明确许可、以硬拷贝、电子形式或其他形式将此出版物的全部或部分复制或发布给无权获得本出版物的人的行为都是违反美国版权法的，将需要承担民事责任，并在适用时付诸刑事起诉。如果您有任何问题，请致电 (508) 482-0188 联系 ESG 客户关系部门。本 ESG 白皮书的编写获得了 Oracle 的协助与资金支持。

执行概要

ESG 的新近研究结果表明，一般的数据库预计每年增长 25%。这就需要更多资源以保存和管理公司数据。公司购置更多系统而不是更有效地利用现有的系统，由此产生的低效资源使用率（大多数组织认为可接受的的存储利用率为 50%）带来的是开支增加。

这种信息过快增长的一种解决之道是信息生命周期管理 (ILM)，可通过平衡基础架构成本与公司数据的可访问性和可用性要求，帮助组织保持信息价值与其底层 IT 基础架构的一致性。ILM 利用性能、可靠性和成本特性各异的多种存储层，有助于降低存储成本。时间较长、活跃性降低的数据可被移至更合适的存储层，并同样保持联机、可供访问。将 ILM 与压缩及消除数据重复等存储优化解决方案结合使用，可实现可计量的成本节省。

Oracle 数据库 11g 的 ILM 功能完善了分层的存储策略。ILM Assistant、分区和高级压缩等特性提供了一个解决方案，根据数据的存在时间和业务价值，将其迁移到最适宜的存储介质。Oracle 数据库 11g 在压缩方面的一个主要改善，在于其可与 INSERT 和 UPDATE 等常规数据操作一起使用。因此，数据压缩是持续进行的，而不是像上个案例中一样，只在批量负载操作过程中执行。Oracle 数据库 11g 还提供了性能优化，将压缩开销减至最低，达到有利于事务处理和数据仓储负载的最佳程度。

ESG 认为，Oracle 数据库的 ILM 功能可以节省 85% 的存储相关成本。客户还可以节省电力、冷却和地面空间方面的成本。很明显，在实现成功的 ILM 策略时，分层式存储和压缩是要考虑的重要因素。通过 Oracle 数据库 11g，用户可从自己的数据中获取最大的经济价值。

引言

数字内容的快速增长可追溯到多个因素。组织引入计算作为其成功的助力之一。不论是仅仅几个主要应用程序，还是完整的联网基础架构，从通信到业务文档等所有内容都转为数字形式。

当今的组织面临着过多关于大量数据保留策略的政府和行业法规，这也是数据呈指数级增长的原因之一。以前，不再使用的数据要么删除，要么保留 — 保留方式通常使其难以查找和检索。现在，可以永久保留这些信息，且必须能在合理的时限内为 IT 部门查找到。

当然，信息量的增加和保存时间的延长，必定意味着成本的增加。如果信息增长速度快、且大部分要保存，您就需要额外的存储和管理资源。但信息保留时间延长也有其优势。组织现在可以利用大量信息池来提升自己的竞争优势和扩大创收机会。

组织还必须考虑到信息保存时间的延长给 IT 部门带来的操作影响。增加新硬件将带来对昂贵的数据中心场地的需求，并增加电力和冷却资源的成本。

虽然在存储能力方面增加的采购成本难以避免，但现在组织可以提高这些投资的有效性。方法之一就是采用信息生命周期管理 (ILM) 策略，这样企业可以使用不同的存储层来根据业务价值管理信息。利用分层式的存储实现方式，ILM 解决方案有助于识别某项数据的相关业务价值，然后自动将该数据移至最合适的存储层。此外，ILM 解决方案还可通过压缩技术对这些存储层进行优化。通过压缩和存储分层的结合使用，组织的信息存储成本得以降低。

应对信息增长

数字化提高了信息的价值 — 数据更易于检索、共享和分析。组织试图利用这些优势以及从自己的信息中获取尽可能多的价值时，新的数据在不断产生。这些信息大部分（尤其是关键的业务数据）都是通过数据库来管理的，并且以良好的态势持续增长。根据 ESG 的研究¹，85% 的反馈者预计其主数据库实例年增长率为两位数。这一数据来源于预定机票或获取新客户联系信息等日常活动中的信息。

在数据库环境中，IT 部门通常会出于多种原因创建多个辅助实例，作为主数据集的完整备份或子集。辅助实例用于测试和开发环境、数据仓库和报表操作。这些实例的增加，进一步扩大了必须存储和管理的数据库信息的规模。

数据保护和灾难恢复流程也是导致数据库信息增长的原因。必须对主要数据进行备份，以防止因损坏或误删除导致信息丢失。在帮助提高数据可用性的同时，这些解决方案要求具备包括磁带或光系统等多种媒介格式的存储能力。

¹ 信息来源：ESG 研究报告：2007 Database Archiving Survey, 2007 年 12 月

IT 部门必须处理要求将数据保存到指定时间的法规、公司管理和法律要求。按照法规，组织必须保留特定的业务记录 — 其中很多都保存在数据库中。从法律的角度而言，组织必须保留可能的证据，并使其可供公司法律顾问在需要时查看。为达到这些要求，往往需要将数据库记录副本移至成本较低的磁盘系统或磁带介质中。经历过电子举证的组织中，约有 **60%** 不得不生成数据库记录。存储这些数据需要额外的资源，并且这些资源必须隔离以达到保存和安全目的。

了解存储成本

不论组织是新建辅助数据库实例来加载数据仓库，还是复制数据以用于诉讼支持目的，都必须增加其存储能力来应对数据库信息的增长。除了很多 IT 部门都关注的存储系统的购置成本和相关管理或保护软件之外，要了解数据保存的所有成本还必须考虑其他的一些开支。

资本费用

面临信息增长的情况，组织的应对方式是增加更多的同类存储能力。表面上，这似乎是最佳的方法，因为数据库管理员和存储管理员希望能避免所有的“意外”。但不考虑类型和业务价值，为大部分数据库信息使用性能及可用性相同的同样存储系统配置，这至少也是低效的。

需要使用第三方存储设备来备份信息，并且在某些情况下还要对数据进行归档以达到合规性要求。直到最近，组织还是依赖于磁带系统作为备份介质，并将介质分开保存以备在灾难恢复时使用。**由于数据库数据增长及其对组织的重要性，更多地使用了磁盘系统供备份和业务连续性操作之用。**可以在两个系统之间复制信息以达到可用性超高的配置，同时部署更为密集的存储系统以提高备份的可用性。信息增长时，需要使用以保存和保护信息的第三方系统也随之增多。

随着对能力方面的要求增加，组织还必须购置存储联网组件，包括主机总线适配器 (HBA) 及交换机。一些组织可能仍然认为服务器附带的存储或直接连接的存储就够用了，但联网存储由于具备集中式备份和管理的优势而成为更佳的选择。此外，进行采购决策时还应将备份和复制软件解决方案（尤其是根据存储能力进行许可的软件解决方案）纳入考虑因素。

管理数据库存储

虽然存储硬件成本确实是持续下降的，但 IT 部门仍然必须对所有底层基础架构进行管理。存储管理任务远远不止是系统配置和维护 — IT 部门必须在寻求优化存储系统性能的方法的同时，保证可用性，确保信息始终可供相应的业务用户访问。**ESG 研究²表明，一些组织存储预算的 30% 都用在人力成本上。**

常被忽略的运作费用

存储设备需要相应的电力供应，数据中心冷却系统也是一样。现在，电力支出对于数据中心运作的重要性日益显著。**虽然数据中心场地无疑价格昂贵，但是许多设施还没有足够的电力资源来运行所有硬件。**服务器和存储系统的密度也涉及到电力方面的问题，因为这些配置需要更多的冷却，消耗更多电力。

² ESG 的研究估计，不同的组织可能有所不同。信息来源：Enterprise Strategy Group, 2007 年

组织可能会满足于通过购置更多硬件来解决信息增长问题。但这一做法会因电力问题和数据中心场地不足而降低可行性。这一恶性循环看起来无法停息（在许多主要城市地区电力不足的情况下），而其结果则是引导组织实施能有助于自身更好地利用已有容量的解决方案。

制定信息生命周期管理战略

创建多层式 IT 基础架构

可通过分层式体系结构部署 ILM，以帮助控制成本。当企业能更好地把握对更多信息的存储和管理之后，将会很快意识到购置更多同类型的硬件不是解决之道。信息生命周期管理 (ILM) 是一种战略，而非产品。它可帮助组织了解信息的价值，并将这一价值与相应的 IT 基础架构相统一。要达到这一目标，组织必须部署一个杂糅的 IT 基础架构 — 每一层有不同的成本、性能和可靠性量度，还有一种用于在这些不同层中识别和移动信息的方法。

IT 基础架构中应用分层式策略的最简单的方面是存储，因为组织可以从多种系统配置中选择部署。连接到存储系统的方法有多种，存储介质的种类也有多种 — 包括多种磁盘驱动器，提供不同的可用性和性能级别。此外，对存储系统电力利用的优化方式有了改进，可以在磁盘驱动器不被频繁访问时减慢其转速（这些系统有时称作 MAID，即大规模独立磁盘阵列）。有多种存储系统都支持在同一个设备中使用多种驱动器，这样组织无需购置和管理多个机箱就可以构建不同的层。这些配置选择有不同的价位和成本效益，组织可以根据不同的标准来进行投资。

配置组合可能不同，但也有一些供组织用于构建不同层的常规配置。例如，对于关键业务应用程序，通常使用更高性能的磁盘驱动器（成本也较高）。而对于备份或为满足合规性要求而保留信息的目的，则可使用应用程序响应时间较长、密度较高、价格较低的驱动器。组织可以将不通的驱动器类型部署在同一个系统中，也可分散到不同系统中。此外，使用频率较低的数据可以用 MAID 或其他节省电力的解决方案存储，有助于控制使用率。

大多数组织都部署单独的磁带存储，这类存储介质形式属于购置成本最低的一种。磁带不仅便于携带，而且符合灾难恢复和长期存档要求，但磁带上存储的数据不能即时访问。

信息的价值是随时变化的，因此如果组织要将所有的数据都保存在主要的高性能、价格昂贵许多的磁盘驱动器上，将需要花费惊人的成本。组织可以选择将信息转移至成本较低的磁盘驱动器，这样即保持了信息的可用性，又降低了成本。分层式存储有助于企业在信息访问和可用性要求与存储成本之间做出平衡。上述主要的考虑因素可以帮助购买模式一成不变（通常是价格较高的配置）的 IT 部门做出决策。

高效缩减数据

分层式 IT 基础架构只是通向信息生命周期管理的一个战术性步骤。随着信息 — 以前存储在数据库中的关系和结构化数据以及非结构化信息（文件、网页等）— 呈指数级增长，组织必须寻求更有效使用其 IT 资源的方法。对整个 IT 环境中的数据进行重复数据消除和应用压缩，可以对资本和运营预算产生极大的影响。

通过重复消除和压缩技术，可在不删除数据的情况下实现透明数据量缩减。消除重复数据确保只存储、备份或复制每个数据库记录和文件的一个副本，供灾难恢复之用。压缩可缩小数据库或文件的大小，进一步减少存储或管理的信息的数量。除了减少存储数据所需的磁盘空间之外，压缩技术还可提升磁盘的输入输出性能，从而提高内存使用效率，并实现网络带宽的最大化。对主数据库中的数据进行了压缩，可缩减复制以用于备份、灾难恢复和测试开发的数据量，从而节省更多存储成本。数据越少，运行基础架构所需的系统和人力就越少。

Oracle 数据库 11g 优化 ILM 存储战略

通过 Oracle 数据库分区利用分层式存储

ILM 战略也需要可以移动和管理信息的解决方案。在数据库管理方面，组织必须确保数据的访问和完整性。

组织需要能够自动化将信息迁移到最合适的存储层这一过程的软件。Oracle 数据库 11g 可简化数据库分区，使得用户可以根据定义的业务规则，将多种类型的数据存储在多个存储层中。通过将较旧数据移到价格较低、性能要求不高的磁盘驱动器上，客户的数据库存储成本得以降低。粒度级别的范围很广，管理员可根据需要的时间间隔进行灵活分区，并可通过虚拟列和组合分区，进一步细化其标准。通过 Oracle ILM Assistant，客户可创建 ILM 要求，以便将数据存储到最节省成本的存储层上。

由于信息移动是在数据库级别进行的，因此对于应用程序而言是透明的。因此，应用程序所有者或用户无需更改任何配置即可维护数据访问。所有数据移动都是在不断运行的情况下进行的。此外，组织还可保持所有数据库信息联机——并且成本得以降低。客户可使用任何工具进行数据查询，还可使用虚拟专用数据库 (VPD)，使这些操作仅针对用户有权查看的数据执行。

Oracle 数据库 11g 通过其 ILM Assistant 特性，实现了信息访问成本的降低，同时并未增加应用程序环境的复杂性。Oracle ILM Assistant 是用于引导管理员执行实现和管理 ILM 环境的工具。它可用于创建生命周期定义、提供关于数据移动、存档和删除时间的建议，并说明相关的存储要求和可能的成本降低。

通过压缩提高系统使用率

除了提供新功能允许客户利用分层式存储基础架构之外，Oracle 数据库 11g 还改善了之前更多是针对数据仓储环境的版本的桌面压缩功能。Oracle 数据库 11g 的高级压缩特性针对的是事务性应用程序，在更新和插入等数据操作过程中提供压缩功能。它使用一个专为相关数据设计的算法，消除数据库块和多列中的重复值。

压缩技术使用符号表，将重复值的一个实例存储为元数据。其他的副本将被替换为指向相应符号表条目的指针。数据块在压缩格式下可正常读取，不受任何影响。此外，与未压缩表相比，要检索同样多的数据，对压缩表的查询由于需要访问的数据块数量较少，因此查询性能显著提高。Oracle 的测试（结果公布在 2007 甲骨文全球大会上）表明，压缩使整表扫描操作的速度提高了 2.5 倍多，相应的磁盘输入输出数量减少了 3.5 倍多。该测试还显示写入操作的开销低到几乎可忽略（不到 3%），尤其是考虑到压缩带来的优势，这个开销更是微不足道。尽管这些数据并未得到 ESG 的证实，但客户仍可确信性能方面有所提升。读取压缩格式数据的另一个好处就是，由于 Oracle 数据库现在在同样大小的缓存内存中可容纳更多数据，因此内存使用效率有所提高。

Oracle 数据库 11g 的高级压缩特性还有助于控制非结构化信息的过快增长。ESG 预计组织中的非结构化信息占用了 70-80% 的存储，增加了对诸如 SecureFiles 之类的解决方案的需求。SecureFiles 是 Oracle 数据库 11g 引入的一项新技术，将非结构化信息存储在可使用标准算法压缩文件的数据库中。要进一步将非结构化信息的存储成本减至最低，客户可以使用 SecureFiles 的消除重复功能，来确保只存储每个文件的一个副本，其他的副本则用（指向唯一副本的）指针代替。除了降低存储成本之外，SecureFiles 的消除重复功能还可极大地提升写入和复制操作的性能。

Oracle 的高级压缩特性是无缝集成到 Oracle 的备份和恢复工具 Oracle 恢复管理器 (RMAN) 中的，并在将备份数据写入磁盘或磁带时对其进行压缩，因此有助于将数据库备份对存储基础架构的影响降至最小。恢复管理器和 Oracle 11g 高级压缩特性一般可带来 50-75% 的存储空间缩减，极大地节省了组织的 IT 预算，同时可提高完成备份的速度。

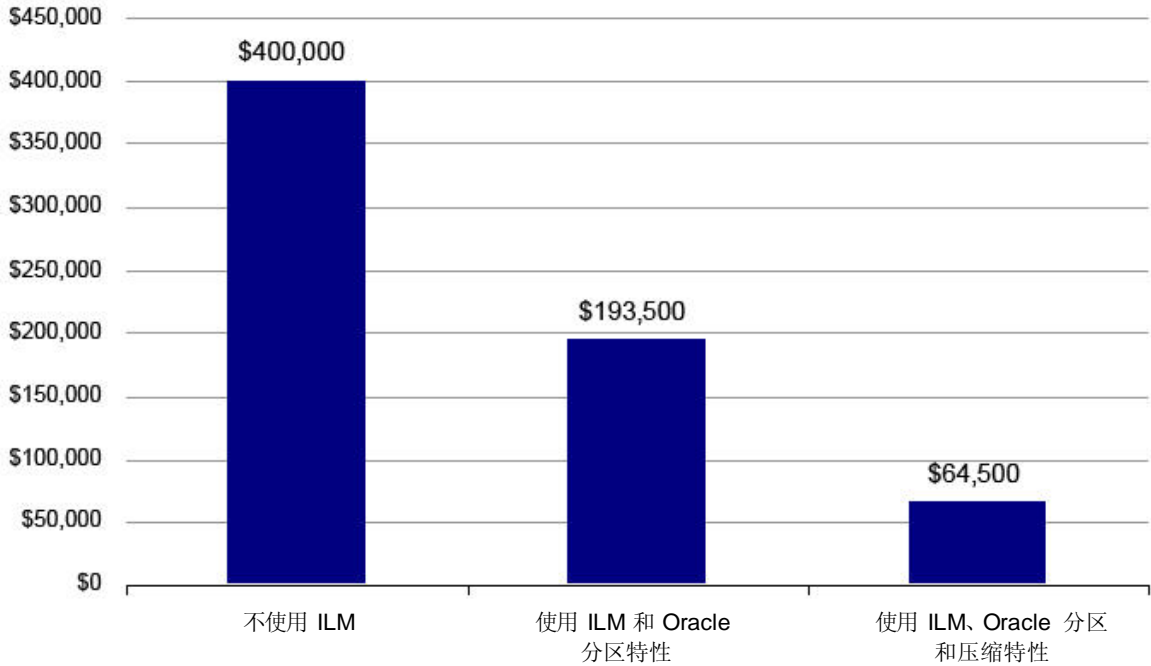
通过 Oracle 数据库 11g 节省成本

如果将使用 Oracle 数据库 11g 的 ILM 策略前后的总存储成本进行对比，则 ILM 特性更经济的效果就更为突出了。在本案例中，将使用 10 TB 的存储容量。ESG 市场分析表明，每 GB 的平均成本 — 包括购置和管理成本（电力、冷却和场地费用是可列入考虑的附加成本，但本例中不使用）— 折合美元为：40 美元用于主存储，25 美元用于辅助存储，12 美元用于密集型的存档存储。客户可以利用他们各自的存储成本量度来进行这类分析。主存储指高性能的光纤通道驱动器，辅助存储指的是性能稍低、容量稍大的光纤通道驱动器，存档存储则指的是大容量、较性能的 ATA 驱动器。

在这种情况下，一种做法就是不管访问活动级别，将所有数据库记录都在主存储上保留一年时间。应用前面提到的每 GB 主存储的价格，数据库存储的总成本为 400,000 美元。如果您的数据库支持大量执行频繁数据插入、刷新或删除的应用程序，这一选择比较有用。

降低存储成本的另一个方法就是假设 10 TB 数据中有 10% 是活动的，35% 活动性较低，而其余 55% 则是历史数据。主存储上将预留 1 TB 要作频繁更改的数据。组织处理其余 90% 的数据的方法可能各不相同，但通常活动性较低的数据和历史数据都会被移至辅助驱动器保留 6 个月，然后再被移至更为密集的驱动器保留 6 个月（以至数年，如果需要保存较长时间的话）。这样，系统中活动性较低的数据需要的空间为 3.5 TB，而不活动的数据或历史数据则需要 5.5 TB 空间。如果以相应的每 GB 成本价格来计算，主存储、辅助存储和历史存储的成本分别为 40,000、87,500 和 66,000 美元，总成本为 193,500 美元。这意味着一年期内节约的成本为 206,500 美元，比起将所有数据存储在主存储层上降低了 50%。图 1 描述了这些节省。

图 1. ORACLE ILM 对存储成本的影响 (1 年期)



基于 10 TB 的存储，所有每 GB 存储成本都是根据 ESG 研究的估计数据计算得出，根据各个组织的情况，结果可能有所不同。
信息来源: Enterprise Strategy Group, 2007 年

如果客户选择使用 Oracle 的高级压缩功能，可以节省更多成本。无论使用何种底层硬件，这一方法都可缩减至少 3 倍的数据存储量。10 TB 所需的维护费用从原来的 400,000 美元降至仅 60,000 多美元，减少了 335,000 多美元，存储系统费用的缩减率高达近 85%。

结论

组织必须认识到现有数据库存储管理方法（即，添加更多同类的硬件来应对数据增长，处理重复数据和非压缩信息）的低效。通过分层式存储、将数据移至不同存储层以及减少要管理的数据量等简单步骤，可极大地缩减资本及运营费用。组织至少应该评估一下如 Oracle 数据库 11g 等有助于节省这些成本同时无需应用程序更改的解决方案。



20 Asylum Street
Milford, MA 01757
电话: 508-482-0188
传真: 508-482-0218
www.enterprisestrategygroup.com