

基于 Oracle10g 的企业级信息集成解决方案的应用研究

Application and research of enterprise level Information Integration resolvable base on Oracle10g

王宇熙 (临安 浙江林学院信息工程学院 311300)
 曾 斌 (临安 浙江林学院现代教育技术中心 311300)
 曾 凯 (临安 浙江林学院信息工程学院 311300)
 姜小丽 (临安 杭州森嘉集团 311300)

摘要:本文介绍了基于 Oracle10g 的企业级信息集成解决方案。通过 Oracle10g 提供的功能强大的完整网络计算解决方案,使得当企业的业务优先级发生变化时,企业能够在一个分布式环境中随时随地根据需要访问信息,根据企业自身需要高效调整它们的资源,以适应不断变化的业务需求。

关键词:信息集成 Oracle10g 分布式 SQL Oracle Streams Oracle 透明网关

1 企业信息集成的需求

企业进行了合并或并购,或者选择了与某个业务合作伙伴交换数据,企业可通过下列方式,来最高效地利用它们的计算资源:

- (1) 在同类和异类数据库之间交换数据
- (2) 启用应用程序之间的通信
- (3) 与客户、合作伙伴和供应商交换信息
- (4) 在数据库之间复制数据

信息集成解决方案提供了一个数据集成视图,从而改善了决策过程。它允许多个应用程序协作运行,以提高运行效率。

2 基于 Oracle10g 的企业级信息集成解决方案

Oracle 数据库 10g 为了满足各种各样的需求,提供了各种信息集成解决方案,主要有:

- 整合 - 将所有数据移动到一个数据库,然后从一个中心位置进行管理。

- 联合 - 从表面看,数据集成在一个虚拟的数据库中,但实际仍然保留在当前的位置上。

- 共享 - 在多个数据库和应用程序数据存储库中维护多个信息副本。

2.1 整合

Oracle 数据库 10g 允许将各种异类信息整合到一个已证明具有可伸缩性的数据库中。Oracle 支持各种数据类型(包括音频、视频、XML、电子邮件、消息以及其他数据类型),几乎能在所有平台(从 Windows、Linux 到 Tru64)上运行。

虽然整合提供了形式最简单的信息集成,但是通常最难以实现。某些部门可能不希望整合它们的数据,或者在某些旧式系统中无法整合。而大多数企业都希望将它们的数据整合到较少的数据库中。数据库的总数减少将使企业管理较少的服务器和数据库,从而降低管理成本。由于应用程序只需部署一次(或者很少的几次),因此企业可以更快速地部署应用程序。最后,因为更快速的集中数据查询,所以企业能够更加快速地访问全球数据,并使整合成为数据仓储和商务

智能应用程序的理想策略。

2.2 联合

联合使信息保留在原始位置,信息在那里进行正常的维护和更新。从外表看,多个数据源集成到一个虚拟数据库中,隐藏了整合视图背后数据库的数量和类型。Oracle 联合数据库解决方案允许访问 Oracle 和非 Oracle 数据,以及结构化数据和非结构化数据,同时隐藏了应用程序数据的实际物理位置。因为不必移动任何信息,所以联合是信息集成的最快方式,达到了事半功倍的效果。联合还提供了对于不能整合的信息的支持,例如旧式应用程序或需要本地所有权的应用程序。应该使用联合来支持不经常访问信息的特别集成。Oracle 数据库 10g 支持多个数据源,以及访问这些数据源的多种方法。

(1) 使用 Oracle 分布式 SQL 访问数据。Oracle 分布式 SQL 对最终用户隐藏所有对象的物理位置,从而使一个包含 Oracle 数据库和非 Oracle 数据库的分布式数据库系统看起来就像一个 Oracle 数据库。企业可以利用此功能来使其所有数据库看起来就像一个数据库,从而降低了分布式系统的某种复杂程度。Oracle 使用数据库链接允许用户在一个数据库上访问远程数据库中的对象。本地用户可以访问到远程数据库的链接,而不必成为该远程数据库的用户。

(2) 使用 XQuery 访问数据。Oracle 积极参与了 W3C XML Query Working Group 的 XQuery 开发工作,这是一个 XML 查询标准。XQuery 提供了搜索和操作 XML 数据的标准方式。XML 数据的某些特性不能使用关系表表示,例如元素的顺序。XQuery 是专门为访问 XML 数据而设计的,因此具有一些 SQL 中没有的功能。处理 XML 数据时,XQuery 可以降低应用程序的复杂性。除了查询 XML 数据,XQuery 还可以访问关系数据。在 Oracle 技术网站上 Oracle 提供了一个基于现有标准的 XQuery 实施。客户可以免费下载该实施,并使用它来访问 Oracle 数据库 10g 的内部或者外部数据。

(3) 将 Web 服务作为 SQL 数据源进行访问。Web 服务是当前最常用的在系统之间交换数据的方式。Web 服务几乎可以是任何类型的应用程序,不但可以向其他应用程序展示它所完成的操作,还可以为授权的应用程序或授权方执行该操作。

Oracle 数据库 10g 可以是一个 Web 服务使用者,也就是说,它可以调用来自 Java 类、PL/SQL 过程和触发器的外部 Web 服务。另外,Oracle 可以使一个 Web 服务看起来就像一个 SQL 行数据源。此特性使用户可以使用 SQL 查询该 Web 服务。然后您可以使用很多的 SQL 聚合函数来轻松地分析该 Web 服务提供的数据库。Oracle 数据库 10g 也可是一个 Web 服务提供程序。Oracle 将很多数据库操作都展示为 Web 服务。例如,您可以通过 Web 服务来查询数据库。您也可以使用 Web 服务执行消息传递操作,例如,使一个消息进入 Streams 队列或取消排队。有执行 DML 操作的 Web 服务,如更新、插入和删除。还有一些 Web 服务可以调用 PL/SQL 过程和函数,并将结果返回调用该 Web 服务的应用程序。

(4) 用 UltraSearch 查找信息。Oracle Ultra Search 提供了针对多个信息库的统一搜索和定位功能,其中包括 Oracle 数据库和非 Oracle 数据库、IMAP 邮件服务器、Web 服务器提供服务的 HTML 文档以及磁盘文件等等。您还可以搜索文档元数据和内容。UltraSearch 通过对被索引的数据源进行爬网,来为文档编写索引。用户可以对 Oracle 数据库以及支持 ODBC 标准的任何其他数据源进行爬网。UltraSearch 能够对企业的各种内容资产进行门户搜索,并充分发挥了 Oracle 的核心功能:平台伸缩性和可靠性。例如,假设您有一个应用程序,它提供库存位置、价格、新闻和信息。库存位置可能来自传统的数据库数据源,最新的价格可能来自 Web 服务,新闻和信息可能来自各种结构化和非结构化的数据源。UltraSearch 可以为这些数据源建立可搜索的索引,同时允许建立显示相关信息的查询,这些信息包括库存位置、价格以及到每个库存相关新闻和信息的链接。

2.3 共享

随着企业的发展壮大,能够共享多个数据库和应用程序中的信息就变得越来越重要了。通常情况下,企业会从很多技术共享信息中进行选择,每种信息针对一个特定的业务问题。虽然这些目标解决方案最初看起来比较容易使用和实施,但是一旦该企业的需求超出了它的能力之外,它们将无法提供任何帮助。开发人员必须及时实施多个工具来构建一个解决方案,且复杂程度按指数规律增加。

企业通常需要集成来自各种数据源的数据。当客户通过销售人员甚至合作伙伴进行在线订购时,企业需要共享 OLTP 更新、数据库事件和应用程序消息。此信息通常需要分配到各种目标,例如异类复制数据库、消息队列系统、数据仓库存放区、操作性数据存储器、其他应用程序以及备用数据库。

Oracle 提供了各种信息共享功能,其中包括 Oracle Streams、物化视图、可传输表空间和数据泵。使用这些功能可搜索到最经常访问的数据,从而缩短响应时间,消除对远程数据存储的依赖性。

(1) Oracle Streams。Oracle Streams 是一个开放的信息共享解决方案。每个元素都支持行业标准语言和各类标准。Streams 支持从 Oracle 系统到非 Oracle 系统的捕获和应用。可以通过透明网关或者通用连接将更改应用到非 Oracle 系统。Streams 还包括一个这样的 API,它允许非 Oracle 数据源轻松提交或接收更改记录,从而使异类数据可双向移动。通过指定流操作的元素配置,可以满足特定的用户需求,如消息队列或数据复制。a) 消息队列:Oracle Streams 高级队列提供了一个用于处理事件的统一框架。它使得用户应用程序能够将消息排入到存放区、传播到订阅存放区、通知用户应用程序消息可以使用,以及将消息从目标位置取消排队。它支持消息排队系统的所有标准功能,包括多用户队列、发布和订阅、基于内容的路由、Internet 传播和转换。另外,Oracle 的消息网关功能还将 Oracle 数据库应用程序与其他消息队列系统(如 WebSphere MQ [以前称为 MQ Series] 和 Tibco)进行了集成。与传统的队列不同,Oracle Streams 存放区可保存各类消息,因此可以将各类消息排队到一个存放区内。Streams 还支持用户应用程序通知功能,此功能结合了压入模式的近乎实时的优点与弹出模式的可伸缩和可管理的优点。b) 数据复制:复制为当今很多企业面临的伸缩性、可用性和性能问题提供了一个解决方案。Oracle Streams 确定哪些信息是相关的,然后与需要该信息的用户进行共享。

Oracle Streams 复制自动从源数据库捕获更改,将其传播到一个或多个远程数据库,然后将这些更改应用到每个目标数据库。Oracle Streams 既可以复制对数据所作的更改(DML 更改),也可以复制对表结构所作的更改(DDL 更改)。数据库管理员可以选择哪些更

改要传播到每个目标数据库,并且指定如何路由这些消息。这种定向的网络路径使得复杂配置(其中包括 N 路和 hub-and-spoke(集中星型))成为可能,并实现了将更改一次传播到某个站点(用于以后扇出到其他目标)的联网配置。如果对多个位置中的相同数据进行了更改,Oracle Streams 会自动检测到冲突,并根据需要解决冲突。

(2) 物化视图。很多企业的一个共同需求是定期将产品目录传播到各个地区办事处,从而使销售人员从客户位置下达订单。为了满足这些需求,Oracle 提供了一种称为物化视图复制的复制类型。物化视图包含了一个表在某个时间点的完整或部分副本,可对这些副本进行增量维护。物化视图可以是只读的,也可以使用冲突解决方案进行更新。因为支持断开计算以及简单的海量部署,所以物化视图特别适用于移动计算应用程序。物化视图也经常用于数据仓库环境,通过计算和存储聚合数据(如汇总和平均值数据)来改善性能。

(3) 混合配置。客户可以充分利用 Oracle Streams 以及 Oracle 透明网关和 Oracle 物化视图的强大功能,来创建表面上跨越多个市场的配置,从而启用新类别的应用程序。另外,所有部署以及相关的元数据都是可兼容的。

(4) 与 Oracle 可传输表空间共享信息。可传输表空间利用了批量信息共享的机制。可以抽取一个表空间,进行复制或移动,然后将其插入另一个数据库中。抽取和插入操作速度很快,与表空间的大小无关。还可以以只读模式在多个数据库上同时装载一个表空间,这实际上是允许两个数据库共享信息,并且按并行方式单独处理数据。

Oracle 数据库 10g 的新增功能支持跨平台传输。现在数据库可以处理表空间中的数据,并可进行必要的格式更改(例如不同的字节顺序)以适应不同平台的不同标准。此版本还允许用户重命名表空间,如果将表空间传输到一个已经存在同名表空间的数据库,此项功能很有用。

(5) 与 Oracle 数据泵共享信息。Oracle 数据泵通过高速并行移动 Oracle 数据库内容的批量数据和元数据,来支持批量信息共享。数据泵可使 DBA 控制并行机制(其中包括使并行机制加速或减速,或者停止或重

新启动导出/导入的各类调节器),从而改善了输出和输入性能。新的网络模式允许用户直接通过网络在数据库之间传输数据,不产生任何的中间泵文件。

数据泵使用户能够移动或者复制表空间或者数据库的子集,比可传输的表空间具有更细的粒度。用户可以指定特定的传输对象,并可重新映射模式、数据文件和表空间。数据泵还能以交互模式运行,这种模式允许用户向转储文件中添加附加文件。另外,因为数据泵与 Oracle Streams 兼容,所以用户可以将对象导出到一个新的数据库,并使用 streams 来使得这些对象与生产数据库保持同步。

(6) 用 Oracle Streams 移动批量数据。Oracle 数据库 10g Streams 可以移动任何操作系统文件、BFILE 和数据库,包括那些由表空间组成的数据库。此特性允许用户仅使用一个命令即可创建副本。您要做的只是指定要复制的表空间,和一个指向该副本数据库的 dB 链接。Streams 将自动创建指定表空间的副本,然后将此副本移动到目标数据库,并使此副本与生产数据源同步。用户还可以将需要的双向复制设置指定为该操作的一部分。副本被实例化之后,用户即可轻松修改该配置,例如,可以添加数据转换。

(7) 用集成技术在网格环境中供应数据。本文讲述的很多信息集成技术也可用于在网格环境中供应数据。用户可以使用粒度更细的 Oracle Streams 数据供应功能,通过一次提供一个记录来供应信息。还可以使用类似于分布式 SQL 的功能使数据在被需要之前一直保持在原位置,从而实现按需供应。批量数据移动工具(如可传输表空间和数据泵)是在网格中供应大量数据的高效方式。

假设一个生产数据库需要执行某些分析。在网格环境中,用户只需选择向生产 RAC 数据库添加节点,即可提供执行分析所需的附加 CPU 容量。然而,在现实中这种方式或许不可行。可能存在一些部门限制,阻止针对该生产数据库运行分析。或许您没有运行 RAC,或许在相应的网络和 SAN 上没有更多可分配的适合节点,但是在另一个网络或 SAN 上拥有节点。使用 Oracle 数据库 10g 的数据供应功能,可以传输包含要分析信息的表空间、在另一个系统上实例化该数据的 Streams 副本,以及在该系统上执行分析。

此功能对于仅使用一个命令即可移植到网格的操

作很有帮助,中间不会出现宕机时间。跨平台传输表空间允许在不同的平台和操作系统之间进行数据移植。当移动数据并在网格上设置新的数据库时,现有的数据库需保持联机状态。移动完成之后,针对数据库进行的更改将被捕获,并且应用于目标中。当复制完成并运行后,只需将应用程序指向网格中的新数据库,移植即可完成。某些情况下,共享和集成信息的最好方式就是在需要该信息的时候进行远程访问。正如本文前面所介绍的那样,Oracle 内置了分布式 SQL、远程过程调用和分布式事务。

用户可以查询其他数据库以获取需要的信息、必要时更新这些信息,以及访问其他服务器的程序逻辑。可以通过 Oracle 通用连接,或者使用透明网关来访问其他数据库管理系统中的数据。可以将 Oracle 数据和非 Oracle 数据进行合并,还可以通过网关使用 Oracle Streams,实现与非 Oracle 系统共享信息。

3 结论

快速变化的业务环境促使企业形成一个分布式的环境。综合在这样的情况下各个企业出现的信息集成需求时,Oracle 提供了一个功能强大的完整解决方案,在一个安全可靠的环境中满足每个需求。分布式 SQL 用于在 Oracle 数据库之间进行透明通信。通用连接和透明网关将此功能扩展到非 Oracle 系统。Oracle Streams 高级队列将通信功能从数据库扩展到了应用程序。另外,Oracle Streams 的复制功能可明显改善这些应用程序的可用性和可伸缩性。

参考文献

- 1 王海亮、王海凤、张立民等,精通 Oracle 10g 备份与恢复,中国水利水电出版社,2005。
- 2 Rick Greenwald、Robert Stackowiak、Jonathan Stern, Oracle Essentials, 3e: Oracle Database 10g, O'Reilly, 2004。
- 3 美 Ian Abramson Michael S. Abbey Michael Corey, 孙杨[译], Oracle Database 10g 基础教程,新华书店总店北京发行所,2004。
- 4 王海亮,精通 Oracle 10g PL/SQL 编程,中国水利水电出版社,2004。
- 5 王海亮等编,精通 ORACLE10G 系统管理,中国水利水电出版社,2005。