

Oracle Exadata 与 Netezza TwinFin™ 的比较

自Netezza TwinFin 发布以来，不到一年的时间里已经有超过个客户开始使用这款设备，其中包括：

- 马萨诸塞州Blue Cross Blue Shield
- BlueKai
- Catalina 市场
- 康威物流公司 (Con-way Freight)
- DataLogix
- Epsilon (一家联营数据公司)
- interCLICK
- 美国洲际交易所 (IntercontinentalExchange)
- 日本医学数据中心有限公司(Japan Medical Data Center)
- Kelley Blue Book
- 马什菲尔德医疗集团 (Marshfield Clinic) 和马什菲尔德临床研究基金会 (Marshfield Clinic Research Foundation)
- MediaMath
- MetroPCS
- MicroAd
- MyLife.com
- 纽约泛欧交易所 (NYSE Euronext)
- 美国西北太平洋国家实验室 (Pacific Northwest National Laboratory)
- Premier公司
- 西夫韦公司 (Safeway)
- 尼尔森公司 (Nielsen Company)
- Wind Telecom



目录

页码

1	简介	4
2	联机事务处理 (OLTP) 和数据仓库	6
3	查询性能	8
4	操作简便性	17
5	价值	23
6	Exadata 与 TwinFin 结合使用	28
7	结论	29



1

简介

Netezza 专注于用于查询与分析大数据的技术。该公司的创新数据仓库设备在市场上引发了强烈震动。为了降低分析数据的运行和拥有成本，我们的许多客户已放弃使用 Oracle 数据仓库。现在，Oracle 向市场推出了 Exadata，这种机器似乎可以完成 TwinFin 所能执行的所有任务，而且也可以处理联机事务。Exadata 的优势在于简化了 Oracle 客户的联机事务处理 (OLTP) 系统，同时降低了这些系统的成本。它通过单一可伸缩的高性能 OLTP 环境，整合了多个系统，消除了重复性并降低了复杂性。然而，当客户尝试将其分析数据库整合到 Exadata 中时，这些优势会受到影响。Netezza 一直以来都为 Oracle 客户带来价值。时至今日，已有200多家组织机构将分析工作负荷从Oracle OLTP数据库转移至Netezza数据仓库设备，从而实现了两全其美的效果。

我们将 Exadata 和 TwinFin 作为数据仓库平台进行比较，且本着中立的态度和观点撰写本文，但为了确保可信度，我们征求了 Philip Howard (Bloor Research 研究室主任) 和 Curt Monash (Monash Research 总裁) 的意见。

要创新，就需要在思维和行动上都能突破常规，这样才能用新的方法去解决问题。Netezza 专注于客户对数据仓库的需求。TwinFin 为我们客户的仓库查询提供了出色的性能。TwinFin 也为客户带来了前所未有的简便性；在工作负荷不断动态变化的情况下，保持一致的服务级别所需的管理任务变得如此简单，任何对于 SQL和Linux的基础知识有所了解的人都可以进行相关的操作。TwinFin 的性能以及简便性降低了拥有和运行其数据仓库的成本。更为重要的是，我们的客户可部署以前曾认为超出其能力范围的分析应用程序，从而创造新的业务价值。

**“Netezza 是 Exadata 的灵感来源之一。Teradata 是 Exadata 的灵感来源之一。” Larry Ellison 于 2010 年 1 月 27 日表示
“我们要感谢他们促使我们不断努力，并让我们进军硬件业务。”**

“Netezza 是 Exadata 的灵感来源之一。Teradata 也是 Exadata 的灵感来源之一。” Larry Ellison 于 2010 年 1 月 27 日 “我们要感谢他们促使我们不断努力，并让我们进军硬件业务。”¹ 这番评论展现了 Larry Ellison 一贯的潇洒作风，但是他的评论中有一个严肃观点：只有最优秀的才会吸引 Oracle 的注意。Exadata 代表着 Oracle 的战略方向，调整其 OLTP 数据库管理系统，与 Sun 提供的大规模并行存储系统合作。Oracle 在推出 Exadata V2 时承诺，处理联机事务和分析查询均可实现极致性能。毋庸置疑，Oracle 在 OLTP 方面的确造诣深厚。但是与 OLTP 相比，数据仓库和分析在软硬件方面的需求全然不同。Exadata 的数据仓库凭据需要详细审查，尤其是对于简便性和价值而言。

本白皮书以回顾联机事务处理与数据仓库中的查询与分析处理之间的差异作为开篇。然后从查询性能、操作简便性和价值等几个方面讨论 Exadata 和 TwinFin。

我们仅要求读者像我们的客户和合作伙伴一样：将数据库管理系统工作原理的观念抛诸脑后，迎接全新的思维方式，准备以更少的努力（而不是更多）来实现更好的结果。

...准备好获得 事半功倍的经验。

Netezza 对 Exadata 计算机没有直接访问权限。我们很幸运地从许多评估了两种技术并最终选择 TwinFin 的公司那里获得了详细反馈。考虑到 Oracle 的规模及其对 Exadata 的重视程度，能够公开获得的有关 Exadata 的信息出乎意料的少。Oracle 引用的使用案例几乎没有为讨论提供什么信息，这一点本身也引起了不少行业追随者（例如《信息周刊》）的关注。² 本着开放精神，本文中提供的信息可供共享。但由于我们的错误而导致的任何错误结果都不是我们在有意进行误导。

¹ http://oracle.com.edgesuite.net/ivt/4000/8104/9238/12652/lobby_external_flash_clean_480x360/default.htm

² http://www.informationweek.com/news/business_intelligence/warehouses/showArticle.jhtml?articleID=225702836&cid=RSSfeed_IWK_News

2 联机事务处理 (OLTP) 和数据仓库

OLTP 系统执行许多简短事务。每个事务的范围都很小，通常只包含一条或几条记录，并且由于具备较高可预测性，因此通常会缓存数据。虽然 OLTP 系统可处理大量数据库查询，但是该技术重在对当前数据集执行写入操作（UPDATE、INSERT 和 DELETE）。这些系统通常用于处理特定的业务流程或功能，例如管理支票帐户的当前余额。其数据结构经常采用第三范式 (3NF)。OLTP 系统的事务类型比较稳定，其数据要求易于理解，因此辅助数据结构（如索引）可以在记录传输到内存进行处理之前，有效地在磁盘上查找记录。

相比之下，数据仓库系统的特点在于，主要是对当前和历史数据集进行频繁的数据库读取 (SELECT) 操作。与 OLTP 操作仅访问少量记录不同，数据仓库查询可能会扫描包含十亿行记录的表，并将其记录与来自多个其他表的记录进行联接。而且，无法预测企业用户将会请求对仓库中的数据执行哪些查询，从而降低了缓存和索引策略的价值。仓库中数据结构的选择包括从 3NF 到多维模型（如星型和雪花模式）。

每个系统中用于填充典型仓库的数据结构都反映了特定业务流程的需要。将数据加载到仓库之前，首先会进行清理、消除重复和整合性数据。

本文将数据仓库划分为第一代或第二代。尽管这种分类可能经不起深入推敲，但是它反映了我们有多少客户在讨论挖掘数据价值的发展之路。

第一代数据仓库通常需整晚加载。它们通过缓慢发展的基于 SQL 的报表和仪表板的稳定主体，向其企业提供信息。因为这些简单仓库在某种程度上类似于 OLTP 系统（其工作负荷和数据要求易于理解且比较稳定），所以组织通常采用用于 OLTP 数据库管理产品。

对于投入使用的产品：数据库管理员分析每个报表的数据要求并生成索引以加速数据检索。在仓库中的数据量超出事务系统中通常管理的数据量之前，OLTP 技术和技巧的沿用似乎比较成功。

进入本世纪后，企业和公共部门机构的数据量每年通常会增加 30-50%。在 OLTP 领域中大获成功的技术和实践经验已经越来越不适用于数据仓库，用于执行数据检索的索引就是这方面的一个例子。当数据库系统处理数据加载作业时，它还要忙于更新多个索引。如果数据量较大，此过程将变得非常缓慢，从而导致加载作业在分配的处理窗口超时运行。即使加班加点，技术团队仍然无法达到与企业商定的服务级别要求。由于业务部门需要等待报表和数据进行使用，因此生产效率会不断下降。

在OLTP领域中 大获成功的技术和实践经验已经越来越不 适用于数据仓库...

组织要重新定义利用其数据所需的方式，本文将此开发称为第二代数据仓库。这些新仓库可方便地管理大型数据集，充当企业存储器。在执行询问时，这些数据仓库可以找回数年前记录的事件；这些长期记录的数据可增加预测分析应用程序的准确性。持续不断缓慢输入的方式正在替换整晚的批量加载，从而减小了从事件记录到以其他数以百万计的事件为环境的事件分析之间的延迟。除填充报表和仪表板的简单 SQL 外，仓库还处理线性回归、贝叶斯和其他高级分析的数学算法。如果发现一款高利润产品的销售量只在五个商店突然出现大幅提升，则可以让零售商了解具体发生的情况和原因。了解具体原因之后，零售商就可以采取策略在所有 150 个商店都开展类似的销售活动。用于构建仓库基础的计算系统必须能够在不影响常规报表和仪表盘的情况下应付此类突然激增的需求。企业用户需要可按照其选择的时间和方式，自由地利用其数据。他们对即时性的强烈需求使性能取决于管理员的优化工作的技术再无容身之地。

分析这些从 TB 到 PB 的数据量需要计算机能够在许多计算节点之间传播数据，每个节点都可与网格中所有其他对等节点并行分析较小数据集。能够在经常变化的工作负荷下维持快速且一致的查询性能的仓库可实现新类别的应用程序，从而形成有价值的见解并激励数据驱动决策制定的采用。

Oracle Exadata 的查询性能

在兼并 Sun 时，Oracle 已得出结论：Netezza 在行业中保持了十年的领先优势：当所有部分（软件和硬件）都针对其目标进行优化时，数据仓库系统可实现最高效率。Exadata 由通过快速网络连接的两个子系统形成：一个通过 InfiniBand 通信的智能存储子系统以及一个配有真正应用集群 (Real Application Clusters, RAC) 的 Oracle Database 11g V2。单机架系统包括由大规模并行处理 (MPP) 网格中的 14 个存储服务器（称为 Exadata 单元）组成的存储层，配有作为八个对称多处理节点的共享磁盘群集运行的 Oracle RAC 数据库。

在兼并 Sun 时，Oracle 已得出结论：Netezza 实现了（软件和硬件）针对其目标进行优化时，数据仓库系统可实现最高效率。

Oracle 将 Exadata 的存储层标记为智能层，因为该层可先处理 SQL 投影、限制和简单联接筛选，³然后再将生成的数据集放在网络上供 Oracle RAC 进行下游处理，此技术称为智能扫描。但是，是否在 Exadata 的 MPP 层中处理数据取决于一些特定因素，包括用于存储数据的表类型、数据最近是何时达到仓库的以及下面讨论的其他因素。

智能扫描不会处理使用称为索引组织表 (IOT) 的 Oracle 结构管理的数据。在 2010 年 2 月发布的《Oracle® Database Concepts 11g Release 2 (11.2)》(Oracle® Database 概念 11g 版本 2 (11.2)) 中的第 3-20 页上，文档建议“当数据的相关部分必须存储在一起或数据必须在物理上按特定顺序存储时，可使用索引组织表。此类型的表通常用于信息检索、空间和 OLAP 应用程序”。2010 年 2 月发布的《Oracle® Database Administrator's Guide 11g Release 2 (11.2)》(Oracle® Database 管理员指南 11g 版本 2 (11.2)) 中的第 19-53 页上写道：“索引组织表适用于特定程序索引结构的建模。例如，包含文本、图像和音频数据的基于内容的信息检索应用程序需要使用索引组织表进行有效建模的倒排索引”。Netezza 客户可以在 TwinFin 中运行这些第二代数据仓库应用程序。例如，对于为市场营销活动整理姓名和地址数据，以及为国家安全局在审查入境口岸数据时列出相关人员姓名进行的复杂文本分析，Levenshtein 编辑距离算法在这类应用程序中至关重要。遵循 Oracle 的建议并采用 IOT 管理文本、图像、音频和空间数据的组织会发现，由于 Exadata 无法应用大规模并行处理的强大功能来处理数据，其应用程序只能以次优方式执行。

第一代仓库（通常在 SMP 服务器上作为 OLTP 数据库运行）缺乏运行企业所需的所有查询的能力，而操作数据存储 (ODS) 则是架构师用于解决此缺点的一个结构。ODS 由多个操作系统定期更新，集成数据并可向应用程序同时提供事务和分析服务。运行 MPP 架构的第二代数据仓库使组织有机会将操作数据存储整合到仓库

³ 《A Technical Overview of the Sun Oracle Exadata Storage Server and Database Machine》(Sun Oracle Exadata 存储服务器和数据机器的技术概述) - 发表于 2009 年 10 月的 Oracle 白皮书。

中。Exadata 无法将智能扫描的功能应用于包含单个活动事务（插入、更新、删除）的数据块，一个Exadata 数据块可以包含许多记录。Oracle 的这一古怪的架构特点增加了客户将 Exadata 视为 ODS 的复杂性。

尽管 Oracle 在最新版本的数据库中提供了 511 个函数，但是其中只有 319 个函数能在 Exadata 的 MPP 存储层进行处理。智能扫描的制约性还会影响函数谓词 – 例如，Exadata 不会使用谓词 `months_between(d,sysdate) = 0` 或 `months_between(d,current_date) = 0` 或 `months_between(d,to_date('01-01-2010',' DD-MM-YYYY')) = 0` 来利用时间日期函数。⁴

Exadata 的存储层提供 Bloom 筛选器来实现一个大表与一个小表之间的简单联接，无法在 MPP 中处理其他更加复杂的对象。分析查询通常需要更加复杂的联接，但 Exadata 不支持这种联结。可以考虑这样一个简单示例：一个国际零售商需要深入了解位于英国商店的销售额的美元价值，此简单 SQL 查询需要三个表（销售、货币和商店）之间的联接。

```
选择 sum(sales_value * exchange_rate) us_dollar_sales
```

（从销售额、货币和商店）

其中 `sales.day = currency.day`

， `stores.country = 'UK'`

， `currency.country = 'USA'`。

Exadata 在 MPP 层中无法处理这三个表的联接，相反，它必须将计算所需要的全部数据通过网络转移至 Oracle RAC，这样做的效率非常低。

用户定义的函数 (UDF) 在第二代仓库中起核心作用。UDF 将查询功能扩展到 SQL 之外，并实现了嵌入式分析处理。Exadata 无法在 MPP 中处理 UDF，而是必须将处理职责推给 Oracle RAC（一种在设计和能力上都无法分析大型数据集的 OLTP 数据库）。

即使在简单报表中，Exadata 的存储层也无法处理不同的聚合，而这种聚合在分析查询中的使用十分普遍。

⁴ <http://antognini.ch/2010/08/exadata-storage-server-and-the-query-optimizer-%E2%80%93-part-4/>

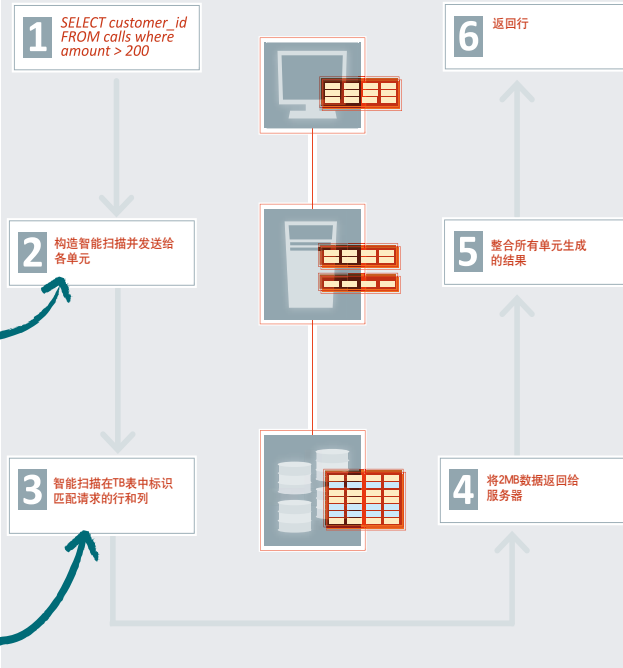
EXADATA 智能扫描处理

智能扫描的局限性

智能扫描并非十全十美，Exadata的MPP存储层无法处理下列情况：Oracle推荐用于管理文本、图像、音频和空间数据的索引组织表；包含活动事务（插入、更新、删除）的数据块；包含多个表的SQL连接和包含两个表的复杂连接；Oracle 511个数据库函数中的192个；用户定义的函数或不同的聚合（常见于简单报表中）。

MPP 利用不足

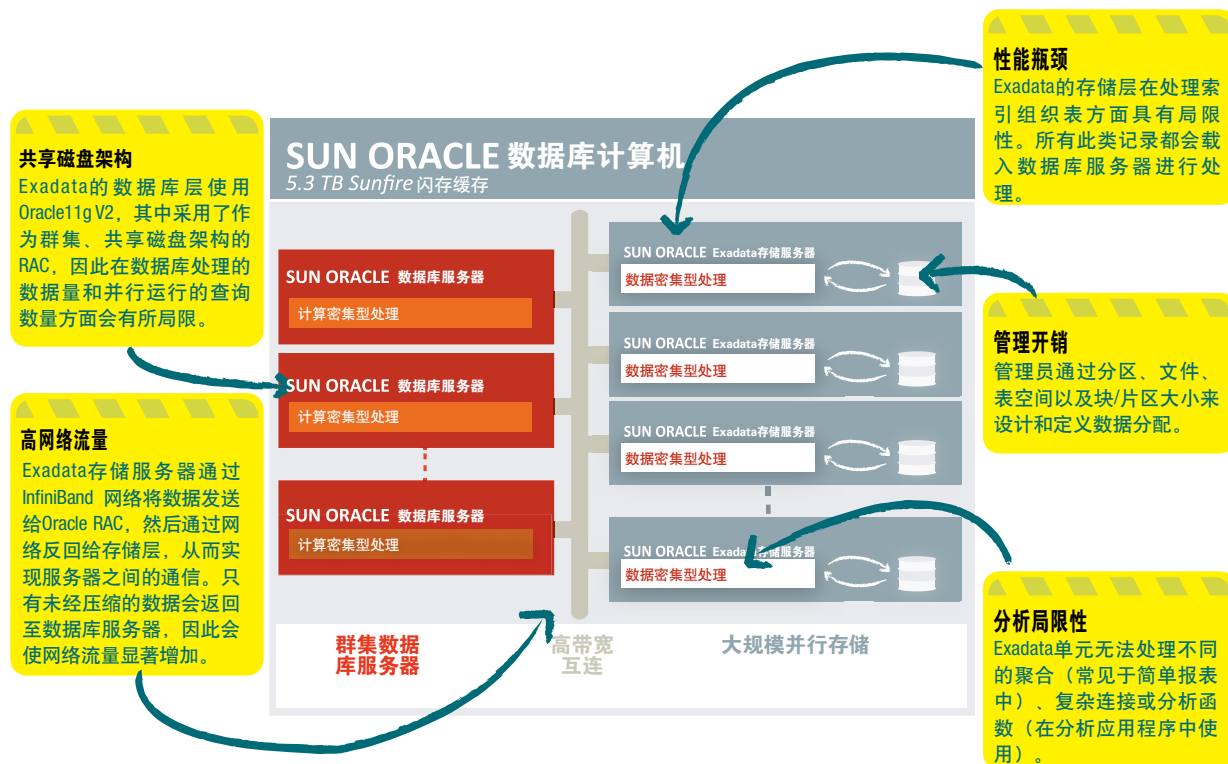
Exadata的工程设计没有充分利用MPP架构。数据库管理未能完全集成到存储层，这意味着要求硬件在MPP网格中处理的工作还远远不够。



Exadata 存储服务器无法与其他服务器通信，只能强制将所有通信都通过 InfiniBand 网络发送给 Oracle RAC，然后通过网络返回给存储层。此架构有益于联机事务处理，在这类处理中，可以通过将小型数据集从存储移动到数据库来满足每个事务（包含一个或几个记录）。对海量数据集的分析迫使 Exadata 在其存储与数据库层之间移动大量数据，从而会对查询性能产生消极影响。

Oracle 将其 40 Gb/秒交换机 InfiniBand 的使用定位为与 TwinFin 相比的一大优势，而实际上，Exadata 需要如此昂贵的网络是因为系统不平衡且效率低下。Exadata 存储服务器执行的工作过少，因此需将多于必要量的数据放置在网络上以移动到下游供 Oracle RAC 处理，导致后者需要执行的工作过多。

Exadata 存储层中的每个磁盘都由运行 Oracle RAC 的网格中的所有节点共享。此公共存储会形成以下风险：一个节点读取的页面正由其他节点进行更新。为了应付这种情况，Oracle 在节点之间强制执行协调。每个节



点都会检查其对等方的磁盘活动以防止发生冲突。Oracle 技术人员将此活动称为阻止 Ping 操作。在每个节点检查对等方的磁盘活动时消耗的计算周期，或一个空闲节点等待另一个节点完成操作时丢失的计算周期都没有得到利用。尽管对 OLTP 有所帮助，但是在为数据仓库专门设计的数据库中，这些周期不会浪费，而是可用于处理查询、挖掘数据和运行分析。

考虑下面这个示例：一个组织需要分析股票交易数据以识别股票市场的行业相关性。其算法计算斯皮尔曼等级相关系数（斯皮尔曼相关系数），通过评估描述参数间联系的难易程度来计算参数间的统计相关性。此分析可在以下方面形成有价值的见解：在一到十分钟的时间窗口内，特定股票是否会影响同一市场行业中其他股票的行为。数据集较大，较长历史的市场数据和每日交易，用于随市场的变化针对其历史记录进行分析。该分析需要在行业中的所有股票之间进行笛卡尔联接，同时对调查的股票计算成交量加权平均价格以及前一个收盘价回报率。运算结果会传递给斯皮尔曼等级相关系数函数，以计算该时间段内每个股票组合的总体协方差和标准偏差。在经过数小时的测试后，客户仍在等待 Oracle 10g 生成结果，后者的糟糕性能使得在最大时间窗口为 10 分钟的业务案例中，不能成功使用这种技术。

Exadata Oracle 11gR2 配有 MPP 存储系统和快速网络。但是，智能扫描无法处理笛卡尔联接（此分析过程的第一步），因此对于此计算没有任何价值。Exadata 会将所有记录放置在网络上，并将这些记录发送给 Oracle RAC。问题随之而来：Oracle RAC 的效率比 Oracle 10g 提高了多少？在处理算法时，Oracle 必须创建和管理临时数据集，并将这些数据集从内存写入存储设备。Exadata 的闪存缓存在这里可能会发挥一些作用，但是数据集的大小和算法的复杂性会强制将数据库进程写入磁盘。来自 Oracle RAC 的此数据流会通过网络返回（仍然被来自 MPP 存储层数据的数据所阻塞），并保持排队和未处理状态以等待满负荷工作的 Oracle RAC 的注意。Exadata 在服务器与共享磁盘之间的网络联接，以及 Oracle RAC（设计为 OLTP 数据库）都成为了整个处理过程的瓶颈。

对于简单查询以外的所有查询， Exadata 都必须 将大型数据集从其存储层移动至其 数据库层，因此会使人们对其是否适合作为现代 数据仓库的平台产生疑问

对于数据仓库，Exadata 的 MPP 存储层主要限于执行谓词级别筛选（基于 SQL 语句的“where”子句的行级别限制，以及基于“select”子句的列级别预测），以及在扫描磁盘上存储的数据时为这些数据的解压缩提供 CPU 处理能力。此外，Exadata 的 MPP 层中的处理是条件性的，从而迫使管理员不得不去研究机器的复杂性。随后在运行 Oracle 11g RAC 的群集 SMP 头节点中对来自所有 Exadata 节点的结果执行所有其余复杂联接、聚合和筛选逻辑（繁重的仓库处理和数据分析）。Exadata 为希望将多个 OLTP 系统整合为单一平台的 CIO 提供了令人感兴趣的机会。除简单查询外，Exadata 都必须将大型数据集从其存储层移动至其数据库层，从而使人们对其是否适合作为现代数据仓库的平台产生疑问。

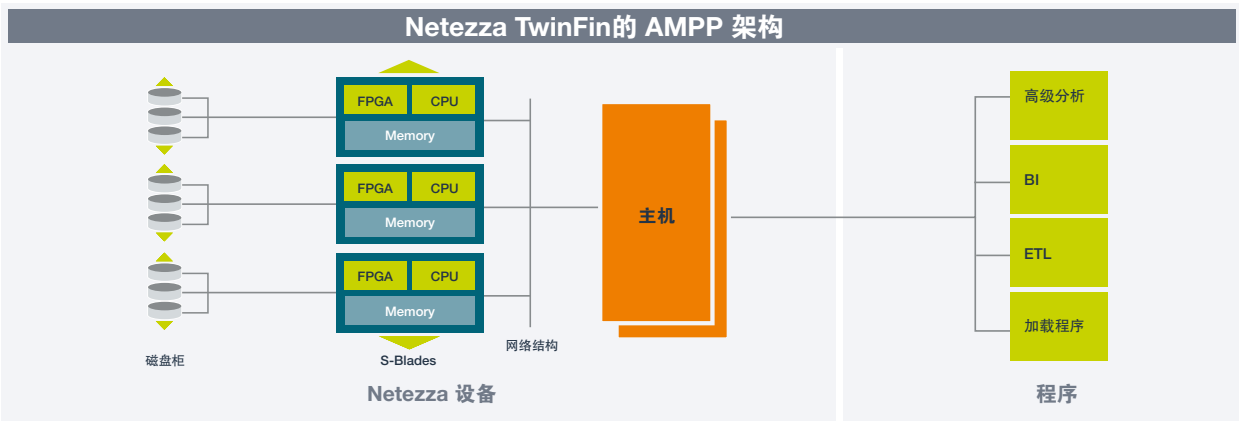
Netezza TwinFin 的查询性能

TwinFin 自始至终都设计为数据仓库平台。Netezza 使用非对称大规模并行处理 (AMPP) 架构，对称多处理主机⁵面向大规模并行处理节点组成的网格。TwinFin 利用此 MPP 网格处理繁重的数据仓库事物和数据分析。

TwinFin 网格中的节点称为 S-Blade（S刀片），这是一种包含多核中央处理单元 (CPU) 的独立服务器。每个 CPU 配有一个多引擎现场可编程门阵列 (FPGA) 和几千兆字节的随机存取内存。因为 CPU 具有其自己的内存，所以它们可只专注于数据分析，从不会被其他节点的活动所干扰（对于 Oracle RAC 中的阻止 Ping 操作时会发生这种情况）。

FPGA 是一种半导体芯片，此芯片配有大量内部可编程门，可用于实现几乎所有的逻辑功能，在管理流处理任务时特别有效。除了在 Netezza 产品的使用之外，FPGA 在数字信号处理、医疗成像和语音识别等这类应用中使用。Netezza 的工程师在我们设备的 FPGA 中内置了软件机器，以在数据到达 CPU 之前加快对数据的处理。在每个 Exadata 机架中，Oracle 专用 14 个八通道存储服务器所实现的功能少于 Netezza 通过我们刀片服务器中嵌入的 48 个 FPGA 实现的功能。每个 FPGA（仅仅 1” x 1” 平方硅）可以高效地完成工作，而且消耗的电量和产生的热量都非常低。

⁵ TwinFin 具有两个 SMP 主机以实现冗余，但是在任何时候都只有一个主机处于活动状态。



**Netezza 未采用具有已知
缺陷的旧系统，
也未使用更加智能的新存储层
进行平衡 TwinFin被设计为
用作数据仓库的优化平台**

Netezza 的 MPP 网格上的节点间通信在运行基于 IP 的自定义协议（完全利用总体截面带宽，并且甚至在持续、突发的网络流量条件下可消除拥塞）的网络结构上进行。网络经过优化，可扩展为一千多个节点，同时允许每个节点同时向所有其他节点发起大型数据传输。这些传输使数据仓库和高级分析的典型处理任务极其高效地执行。正如 SQL 语句可受益于 TwinFin 的 MPP 架构中的处理一样，处于高级分析中心的计算上复杂的算法也是如此。前几代的技术在物理上将应用程序处理与数据库处理分隔开来，这产生了低效和约束问题，因为大型数据集会从仓库传送到分析处理平台，然后再次返回。Netezza 将高级分析的繁重计算并入到其 MPP 网格中，在与数据物理上接近的每个 CPU 中运行算法，因而不需要移动数据并提升了性能。算法受益于在 Netezza 的 MPP 网格中的许多节点上运行，可从施加于可扩展性较低的群集系统的约束中解放出来。

TwinFin 展示了真正的 MPP 架构如何擅长计算斯皮尔曼等级相关系数 - 真正数据集上的真正工作负荷。客户将大量交易数据加载到 TwinFin 中，并将数据从实时市场持续不断地慢慢推送到仓库中。TwinFin 同时利用所有硬件和软件资源来执行查询的每个步骤。Netezza 的智能存储仅选择该市场行业所需的行，并且仅预测评估所需的列。联接结果会直接通过流传递给实现统计分析的代码，TwinFin 将这些代码下载到其 MPP 网格中的每个处理器，从而可并行运行复杂计算。来自 MPP 网格中每个节点的结果会通过网络返回给主机，以进行最终汇集并回呈给请求的应用程序。TwinFin 可在几分钟内完成分析（很好地满足了组织的 10 分钟时间窗口要求），随后只要市场处于开放状态以及只要组织希望评估动态市场中不同股票的行为，便可定期运行分析。除了金融服务行业，斯皮尔曼等级相关系数对于在计量经济、质量控制、生物信息和市场营销领域中工作的统计人员也十分有价值。

Netezza 未采用具有已知缺陷的旧系统，也未使用更加智能的新存储层进行平衡；TwinFin 设计为用于数据仓库的优化平台。TwinFin 可在其大规模并行网格中无条件地运行所有查询，提供卓越性能，并使程序员、管理员和用户可轻松工作。

Netezza在金融服务行业的客户使用精益化方法来分析管理其 **Oracle** 数据仓库所需的资源开销。他们在生成和维护索引、聚合、具体化视图和数据市场的过程中了解到，
其 IT 团队 90% 以上的工作注定是浪费的，或是没有增值的处理性工作

Oracle Exadata 的操作简便性

Oracle 依赖于受过培训、经验丰富且技能熟练的数据库管理员来设计数据库，以提供查询性能。设计技术是数据的物理分区，这表示可以将查询限制到特定相关数据已知的分区。分区是从 Oracle 的 OLTP 根继承下来的。在第二代仓库中，管理员无法知道明天或下周将运行哪些查询和分析。Exadata 包括 Oracle 的自动存储管理可在所有可用磁盘间自动执行分区的分拆。但是，管理团队仍必须创建分区、为跨实例共享的存储配置和管理磁盘组、选择和实现 2 通道镜像或 3 通道镜像以及配置分配单元大小。此外，Exadata 配置需要管理员创建和管理表空间、索引空间、临时空间、日志和范围。正如行业分析师 Curt Monash 所指出的：“不倚重于复杂分区来实现良好性能的系统也许表现得更好。”⁶

⁶ Curt Monash?www.dbms2.com/2009/09/21/notes-on-the-oracle-database-11g-release-2-white-paper/

如前一部分所述，无论 Exadata 是否在其 MPP 存储层中有条件地处理数据；此复杂性都要求管理员（甚至是使用过以前版本的 Oracle 的管理员）参加一系列培训课程。Oracle 建议“除了 Exadata 知识之外，确保您的 DBA 具有最新 (11gR2) 信息。”⁷ 获取最新知识需要每个管理员最少参加四次课程，共 18 天。刚接触 Oracle RAC 的管理员需要另外学习 5 天，以了解进行管理的复杂之处。

一个 Netezza 在金融服务行业的客户使用精益化⁸ 方法来分析管理其 Oracle 数据仓库所需的资源开销。他们在生成和维护索引、聚合、具体化视图和数据市场的过程中了解到，其 IT 团队 90% 以上的工作注定是浪费的，或是不会增值的处理。这种浪费的成本会转化为不必要的硬件和软件许可证成本、数以 TB 计的浪费的存储、延长的开发和数据加载周期、数据长时间不可用、陈旧数据、性能低下的加载和查询以及过多的管理费用。

Exadata 几乎没有采取什么行动来简化 Oracle 数据仓库的管理。管理员必须管理多个服务器层，每一层都具有要维护的操作系统映像、固件、文件系统和软件。Oracle 建议 DBA 应预计在管理 11g（Exadata 中的数据库版本）方面花费的时间⁹ 比在旧的 10g 部署上花费的时间要减少 26%。如果这在实践中得到了证实，Exadata 将客户浪费在无价值的管理上的时间减少了四分之一，则表示 Oracle 朝正确的方向迈出了一步。Netezza 的设备设计不会浪费客户的时间。“DBA 团队只需备份环境并管理设备的高级安全模型，仅此而已。他们无需执行任何其他工作（例如，在处理 Netezza 时不再需要编制索引的概念）。”¹⁰

⁷ http://www.oracle.com/global/hu/events/20100126/exadata_migration_budapest_robert_pastijn.pdf

⁸ “精益化”源自制造业，是使用 Six Sigma 的工具和技术分析浪费的资源开销，并以消除不会使产品或服务增值的活动为目标。

⁹ <http://www.dbms2.com/2009/09/21/notes-on-the-oracle-database-11g-release-2-white-paper/>

¹⁰ 将 Oracle 用于 OLTP 并将 Netezza 用于数据仓库的客户，引自链接的 Exadata 与 Netezza 论坛：http://www.linkedin.com/groupAnswers?viewQuestionAndAnswers=&gid=2602952&discussionID=11385070&sik=1275353329699&trk=ug_qa_q&goback=ana_2602952_1275353329699_3_1

企业用户不仅要求快速完成其查询，还期望获得一致的性能；一个昨天需要五秒钟完成但今天三分钟才完成的报表可能会创建一个票证，要求 IT 支持人员进行响应。数据仓库不可避免地要满足不断变化的动态工作负荷的需求。通过批处理作业或缓慢馈送从 OLTP 系统到达的数据会进行加载，管理任务（如备份和还原，以及在业务和仪表板看不到的后台运行的整理）会持续进行更新。同时，计算密集型应用程序（如预测哪些索赔或交易是欺诈或不合规矩的应用程序）会在仓库基础结构上形成突发的繁重负荷。向企业提供一致的性能对仓库形成了两个要求：一致的查询性能和有效的工作负荷管理（这可简化向所有需要进行的工作分配可用计算能力的过程，通常基于与企业达成一致的优先级）。

管理 Oracle 中的工作负荷需要管理员了解多个优化参数，其中许多参数高度依赖于其他参数。Exadata 中的情况更加复杂，某些参数必须为网格中的每个处理器设置为相同值。管理员不是自上而下控制系统，而是必须自下而上地工作 - 采用实验方法更改各个参数设置，然后观察生产环境中的效果。Oracle 的工作负荷管理理念在联机事务处理方面获得了成功，在这种情况下，每个应用程序置于数据库和服务器上的负荷可以在开发环境中进行配置，参数设置的效果进行了注明且易于理解，并可进行增量调整。在实现稳定系统后，便可将参数设置投入生产。

第二代仓库中没有“正常负荷”；企业用户将可在外部世界中发生各种事件的同时，对大型数据集随意执行计算需求比较高的分析。当其他负荷消失时，分析师不必在这段空闲时间内等待。仓库的动态特性暴露了 Oracle 针对工作负荷管理的优化方案的脆弱性。在负荷时增时减的情况下为具有不同应用程序和数据要求的大批用户实现和维护一致的性能，这是一项复杂的任务。即使是训练有素且具备丰富 Oracle 经验的管理人员团队也很难维护一个具有以下特点的数据仓库：能够向定期报表和加载作业提供一致的性能，同时为分析活动中出现的非预期但重要的激增需求提供冗余。

“我们与Netezza一起证明概念的方式是：他们发给我们一个盒子，我们将其放入数据中心并接入我们的网络。在 24 小时以内，就可以启动并且证明我没有夸大其词，一切就是这么简单。”

Oracle RAC 是一种复杂的技术，其调优参数晦涩难懂。Oracle 建议即使是在其以前版本数据库方面经验丰富的管理员也要接受一段时间的培训，由此可以肯定 Exadata 的复杂性。

Netezza TwinFin 的操作简便性

Netezza 的客户乐意在公开信息中确认我们的设备易于安装和使用。“我们与Netezza一起证明概念的方式是：他们发给我们一个盒子，我们将其放入我们的数据中心并接入网络。”他说道，“在 24 小时以内，我们便已启动并运行。我没有夸大其词，一切就是这么简单。”¹¹这段评论来自 Joseph Essas，他是 eHarmony, Inc.（一家已使用 Oracle 的数据库和 RAC 软件的公司）的技术副总裁。

减少时间以提高生产效率是个好的开始；Netezza 的理念是使数据仓库的所有阶段都简单易行。客户面临的首要任务是加载数据。TwinFin 自动执行数据分布。得自概念证明项目的经验是：客户使用自动分布将其数据加载到 Netezza，运行其查询并将结果与其高度优化的 Oracle 环境进行比较。除最简单的查询外，自动分配的优异性能足以使 TwinFin 胜过 Oracle。客户之后可能会对所有查询进行分析，从而确定可以通过按不同主键重新分配数据来加速的查询。TwinFin 可以简化此任务。

¹¹ <http://www.itworld.com/software/61575/eharmony-finds-data-warehouse-match-netezza>

对于简单方法，需要了解以下几点

- 无需群集互连（GES 和 GCS）监视/调优
- 无需特定于 RAC 的知识/调优（具有 RAC 经验的 DBA 不再那么重要）
- 无需数据库空间/表空间大小调整和配置
- 无需重做/物理日志大小调整和配置
- 无需日记/逻辑日志大小调整和配置
- 无需为表进行页/块大小调整和配置
- 无需为表进行范围大小调整和配置
- 无需临时空间分配和监视
- 无需集成操作系统内核建议
- 无需维持操作系统建议的修补级别
- 无需配置主机/网络/存储的 JAD 会话
- 无需查询（例如 first_rows）和优化器（例如 optimizer_index_cost_adj）提示
- 无需统计包（统计信息、缓存命中、等待事件监视）
- 无需内存调优（SGA、块缓冲区等）
- 无需计划/创建/维护索引
- 简单分区策略：哈希或轮循

提交给 TwinFin 的所有查询都会在大规模并行网格中自动处理，无需数据库管理员干预。查询和分析通过主机进入 TwinFin，在主机中，优化器、编译器和计划程序将查询和分析分解为许多不同的部分或片段，并将这些指令分发给处理节点（或 S-Blade）组成的 MPP 网格，所有这些节点（或 S-Blade）随后针对其本地管理的数据切片同时处理其工作负荷。

到达 TwinFin 的每个 S-Blade 的片段开始将压缩的数据从磁盘读入到内存中。FPGA 随后从内存缓冲区读取数据，并利用其压缩引擎进行解压缩，立即将磁盘中的每个块转换为等效于 FPGA 中的 4-8 个数据块。Netezza 的工程设计可以加速所有数据仓库中最慢的组件 – 磁盘。接下来，在 FPGA 中，数据会流入项目引擎，该引擎根据要处理的 SQL 查询的 SELECT 子句中指定的参数，筛选出列。只有满足 SELECT 子句的记录才会进一步向下游传递给限制引擎，该引擎根据 WHERE 子句中指定的限制，阻止那些处理查询不需要的行通过门。可见性引擎以流速度维护 ACID（原子性、一致性、隔离性和持久性）遵从性。所有这些工作（对不需要的列和行的持续调整）都在大小只有一平方英寸的节能高效的 FPGA 中完成。如果 TwinFin 无需移动数据，它便不会这么做。

FPGA 的预处理完成，它仅将生成的经过整理的记录集流回 S-Blade 内存，其中，CPU 会执行诸如排序、联接和聚合等更高级别的数据库操作（与 MPP 网格中的所有其他 CPU 并行执行这些操作）。CPU 还可能为高级分析处理应用嵌入在片段代码中的复杂算法。CPU 最后从整个数据流中汇集所有中间结果并为片段生成一个结果，按照片段代码的指示通过网络结构将该结果发送给其他 S-Blade 或主机。当 JOIN 所需的数据未在节点上并置时，TwinFin 的节点间网络结构会在数据库完成限制和投影之后，在处理周期后期高效且简单地重新分布。某些高度复杂的算法需要在节点间进行通信才能计算其答案。TwinFin 利用一种消息传递接口来传达中间结果并生成最终结果。

此外，因为原始压缩数据块仍保留在内存中，所以，在以后需要类似数据的查询中，可以通过 TwinFin 的表缓存（一种无需任何 DBA 培训或参与的自动机制）可以自动重用这些数据块。

在迁移到 Netezza 后 仅仅过了三个月客户便表示， 他的团队交付的分析应用程序比他们以前 使用 Oracle 三年中交付的应用程序还要多

因为 TwinFin 对所有任务全面应用并行机制，所以其工作负荷管理系统在控制可供各个及所有作业使用的设备计算资源方面发挥着关键作用。在 Netezza 的设备架构中，一个软件组件控制所有系统资源：处理器；磁盘；内存；网络。这种简洁性是 TwinFin 工作负荷管理系统的基础。TwinFin 的工作负荷管理系统使管理员可以方便地根据与企业达成一致的优先级将计算资源分配给用户和组，并为多个团体维护一致的响应时间。

TwinFin 消除了浪费时间的数据库调优工作。Netezza 设备有能力制定自己的智能决策，而无需任何调优，也几乎不需要系统管理。对于不断变化的动态工作负荷维护一致的性能只需要少量管理任务，具有 Linux 和 SQL 经验的任何人都能轻松胜任。管理员只需将 TwinFin 的资源分配给用户团体中的各个组，并将控制移交给工作负荷管理系统。从周而复始的数据库管理工作中解放出来之后，技术人员可以与企业一起研究创造价值的新数据利用方式。在迁移到 Netezza 后仅仅过了三个月，客户便表示，其团队交付的分析应用程序比他们以前使用 Oracle 三年中交付的应用程序还要多。处理分析应用程序的位置与管理数据的位置相近，以及利用处理 SQL 时所用的同一 MPP 平台，这表示组织真正有机会显著增加从数据获取的价值。

Exadata 的价值

对于大量数据和通常为计算密集型的复杂工作负荷，运行数据仓库所需的硬件和软件本身就代表着庞大的支出。许多使用 OLTP 数据库构建其第一代仓库的组织的经验是，培训和雇用专业技术人员的成本不仅超过了资产成本，而且年复一年地重复发生。这些成本使数据仓库项目不堪重负，而且极大减损了为组织创造的价值。正如同时作为 Netezza 和 Oracle 的金融服务客户执行的浪费分析所强调的那样，将 Oracle 用于数据仓库是劳动密集型的。Netezza 客户证实，技术要求苛刻的低级管理任务是毫无必要的；这样看来，操作 Oracle 数据库要求管理员将主要精力用于关注和料理底层技术是没有道理的，而 Netezza 客户则将这些时间用于通过利用其数据创造价值。

Exadata 的新存储层增加了管理员进行调优和管理的又一层复杂性。Exadata 是一款最新产品，仅有少量投入生产的数据仓库在使用该技术，因此对拥有成本进行预测可能还为时尚早。然而，客户应该预计到从 Exadata 获得一致的高性能将在培训、数据库设计和管理方面产生高额成本。

尽管添加新存储层可有条件地减少 Oracle 数据库的磁盘吞吐量瓶颈，但是 Exadata 的工程设计更大程度上只是为适应大规模并行处理而进行了调整，而不是完全利用该架构。Oracle 未能将数据管理完全集成到 Exadata 的存储层，这意味着要求硬件在其 MPP 网格中完成的工作太少。这样会提高购买 Exadata 的成本，客户需要为硬件付费，而这种硬件永远也不会被它的软件完全利用。这些成本会在仓库的使用寿命期间不断积累。客户需为其数据中心的未充分利用的空间付钱，这些空间如果用于容纳更加高效的计算机系统，则会带来更多的回报。

评估系统		Netezza	Oracle
项目		TwinFin 12	Exadata v2 (SAS)
性能和架构	MPP	<ul style="list-style-type: none"> · 真实 MPP · 针对数据仓库和分析进行优化 	<ul style="list-style-type: none"> · 混合 – 并行存储节点和 SMP 群集头节点 · 针对事务处理 (OLTP) 进行优化
	硬件架构	<ul style="list-style-type: none"> · 完全处理 S-Blade (1 个 CPU 核心 + 1 个 FPGA 核心/1 个磁盘驱动器) · 主要用于用户/应用程序接口的 SMP 主机节点 · 独立的刀片到刀片重新分配 	<ul style="list-style-type: none"> · 智能存储 (1 个 CPU 核心/1.5 个磁盘驱动器) · 运行 Oracle 11g RAC 的 SMP 群集节点 · InfiniBand (Exadata 节点到 SMP 群集) · 头节点参与所有数据重新分配
	数据流	<ul style="list-style-type: none"> · S-Blade 上的 FPGA 性能协助 – 解压缩、谓词筛选、行级别安全增强 · 95% 以上的工作在 S-Blade 上完成 	<ul style="list-style-type: none"> · Exadata 节点主要用于解压缩和谓词筛选。 · 大多数 DW 和分析工作在 SMP 头节点中完成
	数据库内分析	<ul style="list-style-type: none"> · 完全参与的 MPP 平台进行分析 · 用户定义的函数、聚合及表 · 语言支持: C/C++、Java、Python、R、Fortran · 范式支持: SQL、Matrix、Grid、Hadoop · 内置集合包含 50 多个主要分析 (完全并行) · 开源: 对 GNU 科学计算库和 CRAN 库的支持 · 集成开发环境: 带有向导的 Eclipse 和 R GUI 	<ul style="list-style-type: none"> · 仅限于头节点群集的分析处理 · 用户定义的函数和聚合 · 语言支持: C/C++、Java · 范式支持: SQL、Matrix (次要) · 基本分析函数
	扩展	<ul style="list-style-type: none"> · 线性性能和数据大小可扩展性 · 功能全面的企业级工作负荷管理及其他功能 	<ul style="list-style-type: none"> · 非线性性能和数据大小扩展 – 头节点群集的性能和 I/O 瓶颈
简便性	设备系统管理和集成	<ul style="list-style-type: none"> · 无需调优, 无需编制索引, 无需分区 · 为提供最佳性价比而开发的平衡系统 	<ul style="list-style-type: none"> · 性能严重依赖于调优 · 性能依赖于分区和索引的设计和管理

Exadata 包含许多硬件来补偿 Oracle RAC 作为一款分析型数据仓库的糟糕性能。机器的购买成本以及在数据中心的运行成本都十分高昂，对 Exadata 的单个机架进行供电和冷却所消耗的电能比单个 TwinFin 机架高出了将近 60%。

在成本损耗价值的同时，一个基本问题是，Exadata 是否可帮助客户创造价值。第一代数据仓库让客户了解到信息和数据正在发挥着越来越重要的作用，而通过本文前面讨论的第二代仓库的高级分析和其他功能可使这些数据释放出更大潜力。迄今为止，配备传统存储的 Oracle RAC 尚未证明可成功实现这一功能。Exadata 的存储层无法处理复杂联接、不同的聚合以及分析函数。难以想像两种各自无法进行高性能深入分析海量数据集的技术，通过快速网络连接并安装在同一个机架中，便可实现这一壮举。

Netezza TwinFin 的价值

Netezza 工程师将数据管理和分析深入集成到不共享任何内容的大规模并行网格中。这种创新的一个成果是为我们的客户带来了简便性，这直接使得拥有和操作数据仓库的成本比使用面向 OLTP 的数据库产品（如 Oracle 的产品）可能花费的成本要低得多。

对数据仓库的要求已不仅仅是处理简单 SQL，完全利用数据需要仓库能够运行空间分析、预测模型、调查图和其他分析应用程序。举例而言，一个组织需要评估灾难（如火灾和洪灾）对大型投保财产组合所产生的风险，因此使用 Oracle Spatial 和 Oracle 10g 构建了一个应用程序，用于确定易发生灾害地区的投保资产。分析超过 400 万的地点需要空间联接和处理计算密集型空间算法。评估火灾风险需要对 10,000 个多边形进行 430 亿次计算，而评估洪灾风险需要对 7,000 个多边形进行 300 亿次计算。Oracle 完成每次评估需要一个多小时。大风突然改变方向和强度；洪水猛涨。自然灾害的动态行为意味着风险概况在一个小时内可能会发生巨大变化。公司需要更高的敏捷性的工具，一旦业务需要，便可自由地分析情况。与 Oracle 10g 或 Exadata 不同，Netezza 在其大规模并行处理网格中运行空间分析的空间联接和计算密集型算法。Netezza 可在 1.6 秒内完成火灾风险评估，在 4.4 秒内完成洪灾风险评估，从而使公司可迅速从其数据获取完整价值。

TwinFin 从一开始就设计为适用于大规模并行处理 SQL 和高级分析应用程序，Netezza 将组织从专属语言中解放出来。合作伙伴和客户可以将现有应用程序移植到 TwinFin，也可以选择采用其选择的语言（包括 C++、Java、Python、R 和 Fortran）开发新的分析应用程序。使用 C 的组织可以利用 GNU 科学计算库中作为免费软件提供的 1,000 多个分析函数。¹²使用 R 的组织还可以使用 R 综合归档网络 (CRAN) 中公开提供的 2,000 多个程序包。¹³此外，Netezza 的合作伙伴和客户还可以选择使用 MapReduce/Hadoop，例如作为可高度扩展的吸收机制，以便在将面向公众的 Web 应用程序和 Web 日志生成的庞大数据集加载到 TwinFin 中进行分析之前，对这些数据集进行预处理。

¹² www.gnu.org/software/gsl/

¹³ <http://cran.r-project.org/>

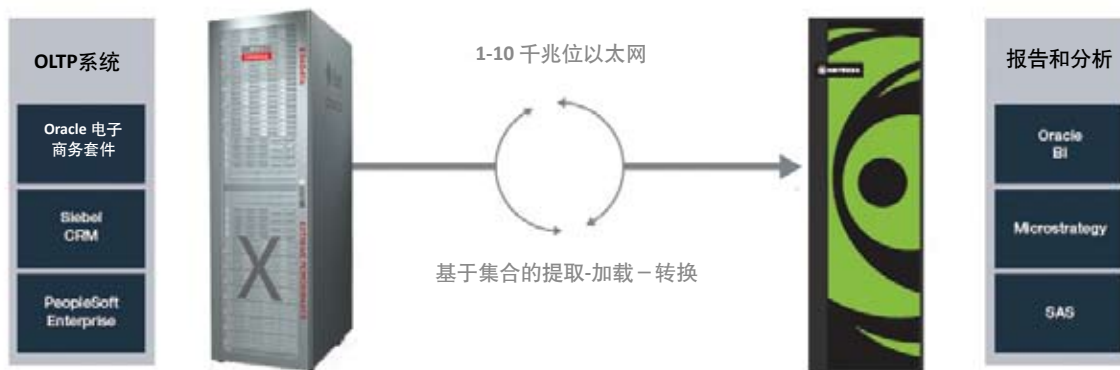
6

将 Exadata 与 TwinFin 结合使用

第 2 节讨论了联机事务处理与数据仓库之间的差异，OLTP 的优化、数据缓存策略和性能调优对于数据仓库几乎不起什么作用。由于其数据库是针对 OLTP 而设计，因此 Exadata 使 CIO 有机会将多个不同的事务系统（各自具有其自己的硬件和 Oracle 数据库）整合到单个可管理的环境中。

这种整合可能提供一些机会，可以在从源系统填充仓库的提取-转换-加载 (ETL) 阶段中降低成本并提高性能。因为他们向其第一代仓库中增加了更多来源，所以某些组织会无意中创建复杂的 ETL 网络。许多 Oracle 客户通过将 Oracle 的数据库用于 PL/SQL 以及该数据库系统专有的其他语法和接口，来构建自己的 ETL 应用程序。尽管 PL/SQL 基于行的处理的速度可能已足以转换较小数据量（在第一代仓库中较为典型），但是第二代仓库中普遍存在的大型数据集需要基于集合的处理所提供的较高性能。Oracle 的 GoldenGate 软件以及来自专业 ETL 供应商的其他产品可提供此功能。虽然不在本文所述范围之内，但是 Netezza Migrator™ 和 Netezza Data Virtualizer™ 也可以帮助 ETL 重建项目获得成功并提高查询性能。

Exadata（运行多个 Oracle 应用程序）将数据传递给 TwinFin 进行分析



Netezza 已在数据仓库方面成为 Oracle 的主要替代者，将数据仓库和数据集市从 Oracle 迁移到 Netezza 可创造新的机会而不是风险。多数 Netezza 客户已经这样做了，其中许多是通过与具有成功迁移的良好记录的系统集成公司合作来完成这一工作的。

Exadata 是从 Oracle 的 OLTP 平台演化而来的。Oracle 的数据库管理系统针对 OLTP 而设计，与数据仓库相比，OLTP 中的数据量相对较小。OLTP 系统的数据库活动可以在投入生产之前进行评估，管理员有时间设计、测试和优化每个事务的数据检索。数据仓库必须立即处理企业需要对其数据进行的任何查询，需要管理员介入调解的技术不适用于这项任务。将此技术应用于事务处理之外的角色会对负责管理和操作数据仓库的人员和流程施加巨大压力。

“Netezza 是其长期产品路线图与我们自己的路线图

完全一致的第一个数据库产品。

我们将此称为我们的按需数据库。”

– Steve Hirsch (纽约泛欧交易所的首席数据官)

Oracle 告诉客户 Exadata 在架构方面与 Netezza 颇为相似，但是要略胜一筹，因为 TwinFin 不支持每种数据类型或 SQL 标准，并且不支持数据挖掘或高并发性。Netezza 的客户对此并不同意：“这是其长期产品路线图与我们自己的路线图完全一致的第一个数据库产品。我们将此称为我们的按需数据库。”¹⁴ Steve Hirsch (纽约泛欧交易所的首席数据官) 说道。

¹⁴ www.netezza.com/customers/nyse-euronext-video.aspx

考虑到其不同的工作负荷特点，少数客户尝试在同一个基础结构上运行 OLTP 和数据仓库系统；这样做需要不断进行调整和优化。技术人员被置于一种左右为难的境地：要么接受 OLTP 和数据仓库的折中性能，要么徒劳地尝试不断重新配置数据库，以满足不同工作负荷相互冲突的需求。组织将继续在不同平台上运行其 OLTP 和仓库系统，每种平台各自针对工作负荷的需求进行专门配置。计划将 Oracle Exadata 用于 OLTP 的组织¹⁵ 可以将 Netezza TwinFin 用于数据仓库，将两者相结合以同时获得两方面的最佳性能。

唯一真正重要的是您自己的数据仓库 – 在您的数据中心运行于您的数据之上的您的应用程序。现场概念证明 (PoC) 使 IT 部门有机会彻底研究一项技术，了解使用 TwinFin 可以如何帮助其业务同行从数据中提取更大价值。充分利用此机会需要使用应用于其他项目的相同准则来管理 POC。Curt Monash 在其博客 “Best practices for analytic DBMS POCs” (分析 DBMS POC 的最佳实践) 中提供了明智的建议，¹⁶ 其中包括使独立顾问参与进来，以指导项目获得成功的结果。对于希望了解其仓库在 TwinFin 上如何执行的公司，Netezza 提供了其使用测试 (免费且无风险)。若要预订，请访问 www.netezza.com/testdrive。

Netezza TwinFin：享受非一般的使用体验。

了解更多信息

Netezza 十分确定您会喜欢 Netezza TwinFin 解决方案，我们邀请您使用您的数据在您的公司中尝试该解决方案 (完全免费)。获取使用测试十分容易。

在此处了解更多信息：www.netezza.com/testdrive

¹⁵ 请参阅 Curt Monash 的 DBMS2 博客 www.dbms2.com/2010/01/22/oracle-database-hardware-strategy/，了解他关于 Exadata 之类的技术用作整合企业的许多 Oracle 数据库的平台而不是运行一些繁重数据库管理任务的讨论。

¹⁶ <http://www.dbms2.com/2010/06/14/best-practices-analytic-database-poc/>

传播本文

请将此电子书与您的朋友和同事进行分享。

- [发送本文](#)
- 发布链接: www.netezza.com/eBooks/Exadata-TwinFin-Compared.pdf

访问我们

参加Netezza的路演活动，了解机构如何利用 TwinFin™ 平台实现高性能数据仓库、商务智能和高级分享 – 所有这些功能都包含在一个易于使用的设备中。 [了解更多信息。](#)

向我们提供反馈

您对此电子书提出的想法和论点有何评价？让我们了解您的喜好以及可能希望进一步讨论的内容。

- 与 Netezza 社区成员交流: www.enzeecomunity.com/groups/twinfin-talk
- Twitter 上有关本文的评论: [#twinfin](https://twitter.com/twinfin)

联系我们

访问我们的博客: www.enzeecomunity.com/blogs/nzblog

访问 Netezza 网站: www.netezza.com

访问 Netezza 社区网站: www.enzeecommmunity.com

将您的意见发送给我们: www.netezza.com/company/contact_form.aspx



关于作者

Phil Francisco (Netezza 产品管理和产品市场营销副总裁)

Phil Francisco 带来了在技术开发和全球技术市场营销领域方面的超过20年的经验。作为 Netezza 的产品管理和产品市场营销副总裁，他制定新的业务和产品策略、指导产品组合并推动产品市场营销计划。在加入 Netezza 之前，Francisco 曾是 PhotonEx (针对核心电信网络提供商的 40 Gb/s 光传输系统的领先开发商) 的市场营销副总裁。在为 PhotonEx 工作之前，Francisco 曾担任朗讯科技光纤网络事业部的产品市场营销副总裁，在那里，他与一些世界最大的电信运营商合作，规划和实施光网络解决方案。Francisco 先生拥有一项先进光网络架构的专利。他在宾夕法尼亚大学的摩尔电气工程学院，以优异成绩获得了电气工程的学士学位以及计算机科学的学士学位。他在斯坦福大学获得了电气工程的硕士学位，并在杜克大学的福库商学院完成了高级管理课程。阅读 Phil 的博客: www.enzeecomunity.com/blogs/nzblog.

关于 Netezza

Netezza – 先驱者, 领导者, 成熟企业

随着数据仓库设备的发明, Netezza 为淹没在数据中并努力提高处理速度和能力以分析和理解数据的全部含义的公司, 彻底改革和简化了分析方式。

现在, Netezza 是数据仓库设备领先者, 它将存储、处理、数据库和分析合并到单个系统中, 以其他方案三分之一的成本, 提供 10-100 倍的性能。我们再一次颠覆了性价比的标准。可以将 Netezza 视为一辆具有经济型轿车的价格和效率的法拉利。

凭借数百个客户 (包括 Nationwide、Neiman Marcus、Orange UK、The Sherwin-Williams Company、Virgin Media 等) 以及全世界的办事处, Netezza (NYSE: NZ) 已证明是成熟的解决方案, 可解决数据仓库和分析不断上升的成本和复杂性问题。我们令人印象深刻的全球合作伙伴体系 (包括 Ab Initio、Business Objects、Cognos、EMC、IBM、Informatica、Microsoft、MicroStrategy、SAS 和其他公司, 以及全世界范围内不断增加的系统集成商、分销商和开发商合作伙伴) 意味着我们的客户确信, Netezza 将很好地适用于他们所使用的任何现有基础结构。

我们推出了世界上第一个数据仓库设备, 并勇于挑战现状。我们创建了一个新的细分市场, 并建立了整个行业的发展规划。现在, 在客户的需求驱动下, 我们正在开发的解决方案能够解决更大、更复杂的企业范围挑战, 包括全球联盟、数据集成、传统系统集成、业务连续性、高级分析和遵从性。凭借良好的财务状况、盈利的商业模式、对成长的关注以及非常热情的“enzee”团体, 您可以确信, Netezza 将始终是数据分析领域强大的合作伙伴, 伴随您以及您不断增加的数据和需求共同成长。



Netezza Corporation
26 Forest Street
Marlborough, MA 01752

电话: +1 508 382 8200
传真: +1 508 382 8300

www.netezza.com

关于 Netezza Corporation:

Netezza (NYSE: NZ) 是数据仓库设备领域的全球领先者, 其设备可在企业规模扩展的情况下极度简化高性能实施过程。Netezza 的技术使组织能够以极快的速度处理捕获的大量数据, 从而为当前数据密集型行业 (包括数字媒体、能源、金融服务、政府、卫生和生命科学、零售和通信) 带来显著的竞争优势和运营优势。Netezza 的公司总部位于马萨诸塞州马尔伯勒, 并在北美、欧洲和亚太地区设有办事处。

有关 Netezza 的更多信息, 请访问 www.netezza.com。