IBM Netezza 数据仓库设备架构

针对高性能数据仓储和分析的平台

IBM

目录

简介

任何企业的成功都依赖于及时拥有最佳的可用信息来制定合理的决策。做不到这一点会导致浪费商机、浪费时间和资源，甚至使组织处于风险之中。但是，发现关键信息来引导最优化行动，意味着要对数十亿个数据点和数 PB 数据进行分析，不管是预测结果、识别趋势还是通过大量不定性绘制出最佳行动路线。可以获得此类随需应变智能的公司能够更快速地作出反应，并制定出优于其他竞争对手的决策。

持续的分析创新可以为公司提供意想不到的智能，令企业的各个领域都获益匪浅。但是，当人们迫切地需要关键信息时，用于交付信息的平台应该是他们最不用花心思的地方。平台应像电灯开关一样简单、可靠且即时，能够处理超乎想象的数据量和工作负载，并不会因复杂性而停滞不前。平台的构建还必须考虑持久性，具备的技术基础能够在更多用户运行日益复杂的工作负载以及数据量持续增长的情况下仍维持性能，同时提供最低的总体拥有成本。

以简单的设备提供极限性能

IBM Netezza 数据仓库设备通过构建的平台以简单的一体化设备来获得行业领先的极高性价比，改变了数据仓库和分析的未来前景。这是高级分析领域内新的篇章，能够冲破一切阻碍、毫不妥协地迅速应对巨大的处理挑战。对于用户及其组织，这意味着即使在需求全方位迅速增长时，也能向所有需要的人提供最佳智能。

带有分析功能的 IBM Netezza 数据仓库设备具有革新的设计，所基于的准则使 IBM 能够提供市场中最佳性价比的服务产品。作为专为高速分析而设计的设备，其能力并非源自于最强大且最昂贵的组件，而是源于如何组装正确的组件并使其协同工作以最大化性能。大规模并行处理 (MPP) 将多核 CPU 与 IBM 独一无二的 FPGA Accelerated Streaming Technology (FAST™) 引

擎相结合，由此交付的性能是比之价格昂贵许多的系统所难以企及的。同时，作为易于使用的设备，该系统可以真正做到即开即用、性能卓越，无需建立索引或调优。设备简单性扩展至应用程序开发，支持快速创新，并能够将高性能分析带给最广泛的用户和流程。

本文介绍了 IBM 的 Asymmetric Massively Parallel Processing™ (AMPP™) 架构，并描述了系统如何编排查询和分析以实现其前所未有的速度。我们将看到 IBM Netezza 数据仓库设备软件和硬件如何相结合以最大限度地利用每个关键组件，以及专为查询大量数据的成千上万名用户而优化的系统是如何真正发挥效用的。这是独一无二的数据仓库和分析平台，具备无与伦比的性价比，随时准备满足当今的需求并应对未来的挑战。

架构原理

IBM Netezza 数据仓库设备在紧凑型系统中集成了数据库、处理和存储，专为分析处理而优化并且易于灵活拓展。该系统架构基于以下核心宗旨，这些宗旨已成为 IBM 在业内具有领先性价比的标志：

紧靠数据源进行处理

IBM Netezza 数据仓库设备架构基于基本的计算机科学原理：对大量数据集进行操作时，如果不是绝对必要，请勿移动数据。IBM 架构充分践行了这一原理，使用“现场可编程门阵列 (FPGA)”商品组件来尽早过滤出数据流中的无关数据，速度可快至数据可流出磁盘时便加以过滤。这一紧靠数据源的数据淘汰过程避免了 I/O 瓶颈，并使下游组件（例如 CPU、内存和网络）不必处理过多的数据，从而对系统性能带来显著的乘数效应。

平衡的大规模并行架构

IBM Netezza 数据仓库设备架构结合了对称多处理 (SMP) 和 MPP 的最佳元素，创建专用于对数 PB 数据迅速运行分析的设备。每个架构组件（包括处理器、FPGA、内存和网络）都经过精心挑选和优化，能够以磁盘的物理特性所允许的速度快速为数据提供服务，同时最小化成本和能耗。IBM Netezza 数据仓库设备软件可编排这些组件，在数据流上以管道方式并行运作，从而最大化利用率，并从每个 MPP 节点提取最大的吞吐量。除原始性能外，这一平衡的架构可向并行执行的千余个处理流提供线性可扩展性，同时提供非常经济的总体拥有成本。

针对高级分析的平台

MPP 和紧靠数据源处理数据的原理同样适用于对大型数据集进行高级分析。IBM Netezza 数据仓库设备使复杂的非 SQL 算法轻松嵌入其 MPP 流的处理元素中，而不具备并行编程或网格编程通常所具有的错综复杂性。根据大量数据对“流上”的任何复杂因素运行分析这一能力避免了将数据移至单独的硬件所造成的延迟及相关成本。它还使性能出现数量级提升，由此使 IBM Netezza 数据仓库设备成为合并数据仓储和高级分析的理想平台。

设备简单性

IBM Netezza 数据仓库设备的架构可精简日常运作并使其实现自动化，从而使用户免于应对平台的底层复杂性。在与设备的任何其他设计方面进行权衡考量时，简单性原则始终是最重要的原则。不同于其他解决方案，它只是运行（迅速处理高要求的查询和混合的工作负载），无需其他系统所需的调优。甚至通常耗时的任务（例如，安装、升级和确保高可用性与业务连续性）也极大地简化，节省了宝贵的时间和资源。

加速创新和性能改进

IBM Netezza 数据仓库设备架构的主要目标之一是改善性价比，最终提供比竞争技术更快速的创新功能。虽然使用基于刀片的开放式组件使 IBM Netezza 数据仓库设备能够极快地采用技术增强，但是 FPGA 的加速效应、均衡的硬件配置和紧密耦合的智能软件相结合所提供的整体性能收益要远远大于这些单独的元素所提供的性能收益。事实上，IBM Netezza 数据仓库设备自推出以来，每隔两年即提供四倍多的性能提升（摩尔定律[1]的两倍），远高于其他著名的供应商。[2]

灵活的配置和极大的可扩展性

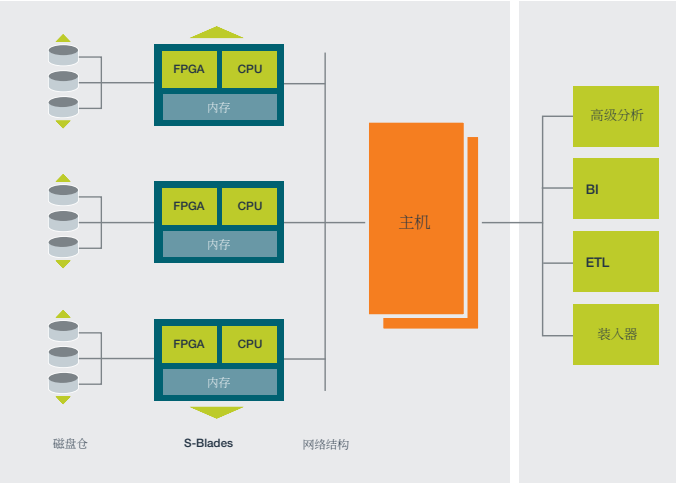
IBM Netezza 数据仓库设备能够以模块化方式将可查询用户数据从数百 GB 扩展至数十 PB。系统架构非常适合为不同的数据仓库和分析市场细分需求提供服务。使用基于刀片的开放式组件允许在配置中轻松修改磁盘-处理器-内存比率，以满足以性能为中心或者以存储为中心的需求。该架构还支持基于内存的系统，为任务关键型应用程序提供极快的实时分析。

以下页面仔细审视了 IBM 是如何将这些原理付诸实践的。

系统构建块

IBM Netezza 数据仓库设备的性能优势主要部分来自于其独一无二的 AMPP 架构，该架构将 SMP 前端与无共享 MPP 后端相结合以进行查询处理。该架构的每个组件均经过仔细挑选和集成，以造就均衡的综合性系统。每个处理元素都对多个数据流执行操作，尽早过滤出无关数据。千余个此类定制的 MPP 流协同工作，以分步处理工作负载。

AMPP 架构



让我们来仔细查看一下该设备的关键构建块：

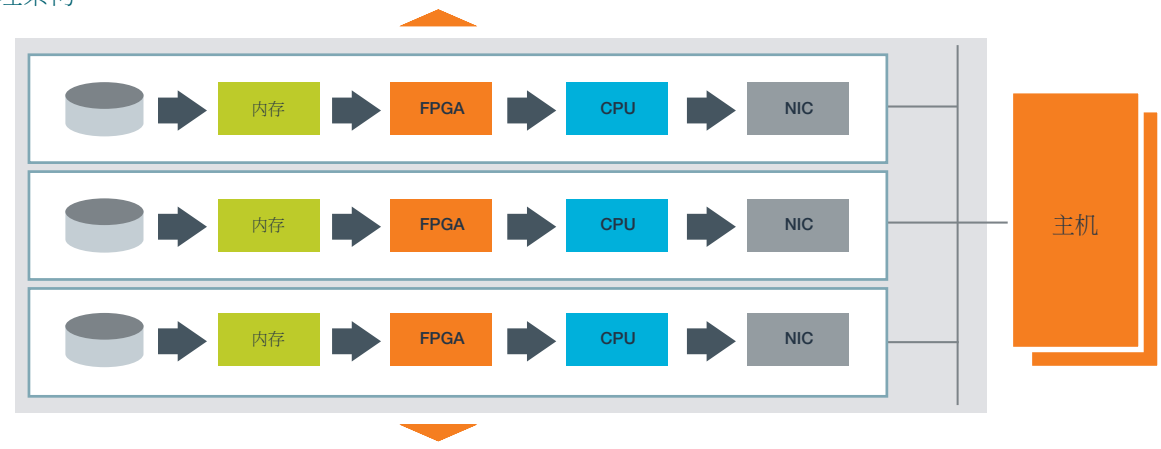
- IBM 主机： SMP 主机是高性能的运行 Linux 的 IBM 服务器，这些服务器在主动/被动配置中进行设置以实现高可用性。主动主机为外部工具和应用程序提供标准化的接口。它将 SQL 查询编译为可执行的代码段（称为片段），创建经优化的查询计划，并将这些片段分发到 MPP 节点以供执行。
- 片段刀片 (S-Blade)： S-Blade 是智能的处理节点，组成了该设备的加速 MPP 引擎。每个 S-Blade 都是独立的服务器，包含强大的多核 CPU、多引擎 FPGA 和千兆字节的 RAM，所有这些都经过负载均衡并且可并行工作以交付最优异的性能。这些 CPU 内核设计有充足的动态余量，以便针对高级分析应用程序的大量数据运行复杂的算法。
- 磁盘仓： 磁盘仓包含高密度、高性能的 IBM 存储磁盘，这些磁盘受到 RAID 保护。每个磁盘都包含数据库表中的一部分数据。磁盘仓通过高速互连连接至 S-Blade，高速互连使 IBM 中的所有磁盘能够以尽可能最高的速率同时使数据流入 S-Blade。
- 网络结构： 所有系统组件都通过高速网络结构连接。IBM 运行基于 IP 的定制协议，该协议充分利用光纤网涵盖各区段的总带宽，并可以消除拥堵，即使在发生持久的突发性网络流量的情况下也是如此。该网络可优化为扩展至超过 1000 个节点，同时允许每个节点启动对所有其他节点同时进行的大量数据传输。

注： 所有系统组件都为冗余组件。虽然主机为主动/被动型，但是该设备中的所有其他组件均可热插拔。用户数据已完全制作镜像，实现超过 99.99% 的可用性。

在 S-Blade 内爆发极限性能

一个片段处理器（众多处理器之一）：一般化商品组件和 IBM Netezza 数据仓库设备软件相结合，以从每个 MPP 节点提取最大吞吐量。来自存储阵列的专用高速互连使数据能够在流出磁盘时立即交付至内存。使用智慧的算法将压缩数据高速缓存至内存中，确保直接从内存中提供最常访问的数据，而无需访问磁盘。在 FPGA 内并行运行的 FAST 引擎会以物理速度解压并过滤出 95% - 98% 的表数据，仅保留相关数据以应答查询。流中剩余的数据会由同样并行运行的 CPU 内核同时处理。在 IBM Netezza 数据仓库设备中运行的千余个此类并行片段处理器上会重复该过程。由此使性能比价格昂贵许多的系统超出多个数量级。

IBM 的大规模并行处理架构



为 S-Blade 加速：IBM FAST 引擎的力量

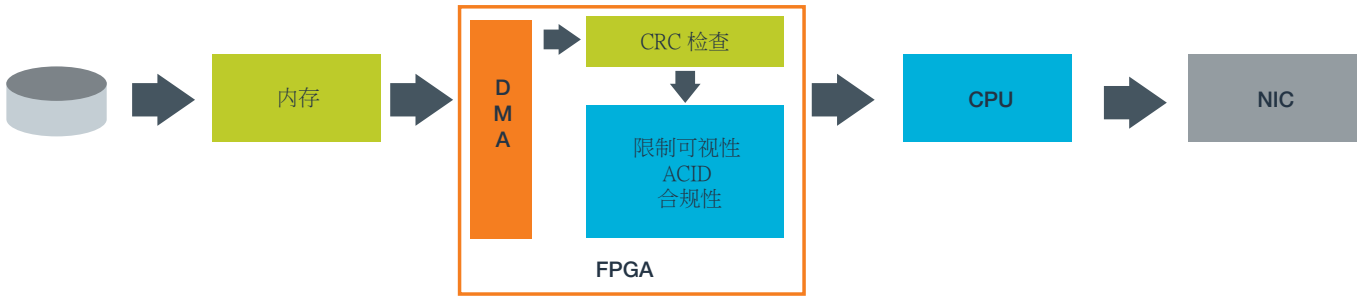
FPGA 是 IBM Netezza 数据仓库设备的性价比优势的关键推动因素。每个 FPGA 都包含嵌入式引擎，以对数据流执行过滤和变换功能。这些 FAST 引擎可动态重新配置，使其能够通过软件进行修改或扩展。这些引擎通过在查询执行期间提供的参数针对每个片段进行定制，并以极高的速度对由直接存储器存取 (DMA) 模块交付的数据流进行操作。

FAST 引擎包括：

- “压缩”引擎： 这是 IBM Netezza 数据仓库设备的创新，将系统性能提升 4 到 8 倍。[3] 该引擎能够以网速解压数据，即时将磁盘上的每个块变换为内存中的 4 - 8 个块。由此显著加速任何数据仓库中最缓慢的组件 - 磁盘。
- “投影”和“限制”引擎： 这些引擎通过基于 SQL 查询中的 SELECT 和 WHERE 子句中的参数分别过滤出列和行，进一步提升性能。
- “可视性”引擎： 该引擎在以流式方法速度保持 ACID（原子性、一致性、隔离性和耐久性）合规性中扮演关键的角色。它可过滤出查询不应“看到”的行，例如，属于尚未成交的交易的行。

IBM FAST 引擎还可为将来通过对 IBM Netezza 数据仓库设备软件的增强所添加的创新功能提供可扩展的框架。这些新功能有望进一步提升系统性能、安全性和可靠性。

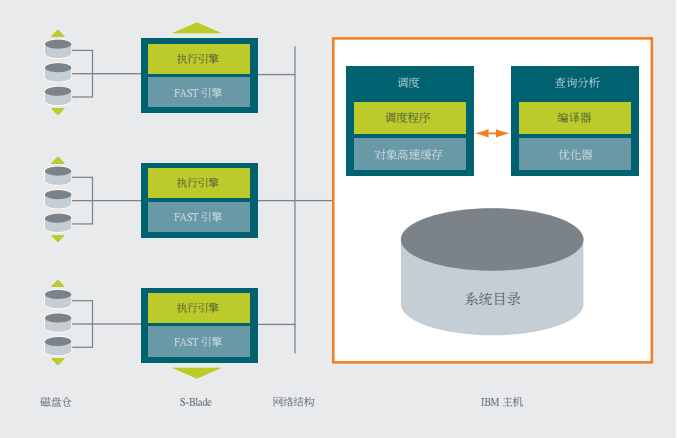
IBM FAST 引擎



在 IBM 上编排查询

IBM Netezza 数据仓库设备硬件组件和智能系统软件紧密结合。该软件旨在充分利用设备的硬件功能并整合多种创新以提供指数级的性能收益，无论是针对简单的查询、复杂的特殊查询还是深入的分析。在本节中，我们将仔细查看构建到系统中的智能。

软件架构



IBM Netezza 数据仓库设备软件组件包括：

- 复杂的并行优化器，用于变换查询以提升运行效率，并确保每个处理节点中的每个组件都得到充分利用
- 智能的调度程序，用于保持系统以峰值吞吐量运行，而不论工作负载如何
- 经过特殊加速的片段处理器，用于高效地同时执行多个查询和复杂的分析功能
- 智慧的网络，用于通过 IBM Netezza 数据仓库设备轻松移动大量数据

制定经过优化的查询计划...

当用户提交查询时，主机会对其进行编译，并创建专为 IBM 的 AMPP 架构优化的查询执行计划。智能的 IBM 优化是该系统最强大的优势之一。该优化操作利用系统中的所有 MPP 节点，收集关于查询中所引用的每个数据库表的详细的最新统计信息。其中大部分度量值是在查询执行期间捕获的，具有极低的开销，生成按查询个性化的及时统计信息。就本质而言，具有能够彼此通信的集成组件的 IBM Netezza 数据仓库设备支持基于成本的优化，以更准确地度量与操作相关联的磁盘、处理和网络成本。通过依赖于准确的数据而不是仅仅依赖于各种试探方法，优化器可以生成查询计划，非常高效地利用所有组件。

优化智能：计算连接顺序

优化智能的示例之一是确定复杂连接中的最佳连接顺序的能力。例如，将多个小型表连接至大型事实表时，优化器可以选择将小型表完整地广播到每个 S-Blade，同时使大型表广泛分布于所有片段处理器之间。该方法可最大限度减少数据移动，同时利用 AMPP 架构来实现并行连接。

IBM 优化利用这些统计信息在开始处理查询前先对其进行变换，从而最大限度减少磁盘 I/O 和数据移动，以去除可能降低数据仓库系统性能的因素。变换操作包括：

- 确定正确的连接顺序
- 重写表达式
- 从 SQL 操作中除去冗余

将其转换为片段...

编译器可将查询计划转化为可执行的代码段（称为片段），这是可由分段处理器跨设备中的所有数据流并行执行的查询段。每个片段都具有两个元素：由单个 CPU 内核执行的编译代码和为该特定片段定制 FAST 引擎过滤的一组 FPGA 参数。这种逐个片段的定制方法使 IBM Netezza 数据仓库设备实际上能够提供针对个别查询实时优化的硬件配置。

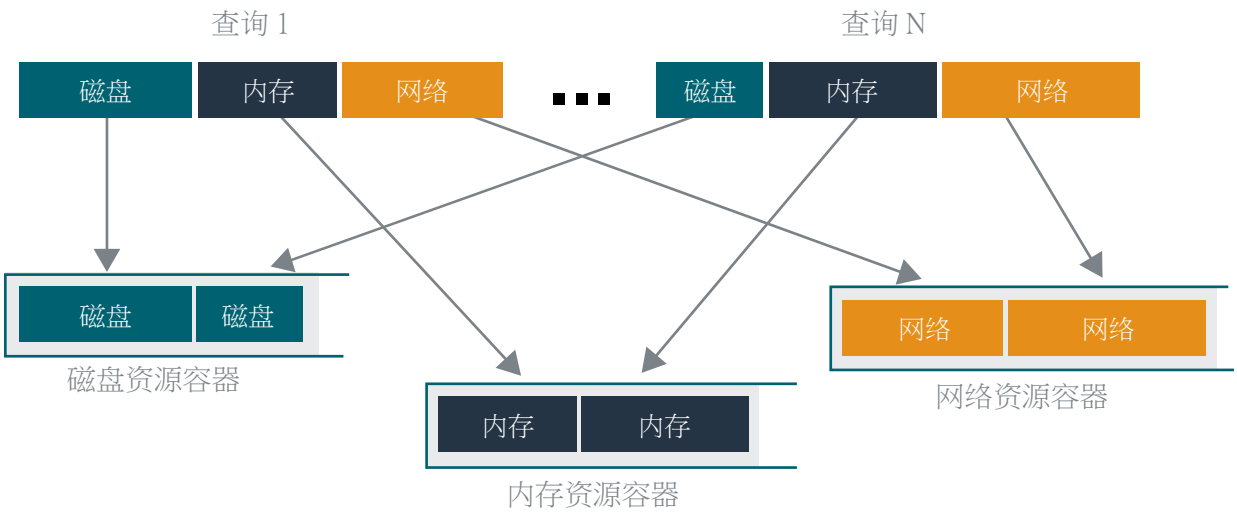
编译器中的智能：对象高速缓存

主机使用对象高速缓存功能来进一步加速查询性能。这是先前编译的片段代码的大型高速缓存，该代码支持参数变体。例如，带有子句“where name='bob'”的片段使用的编译代码可能与带有子句“where name=' jim'”的片段所使用的编译代码相同，但是其设置反映的名称不同。该方法可消除超过 99% 的片段的编译步骤

对其进行调度以适时运行...

IBM 调度程序可平衡复杂工作负载间的执行情况，以满足不同用户的目标，同时维持最高的利用率和吞吐量。它会考虑多个因素，包括查询优先级、大小和资源可用性，以确定何时在 S-Blade 上执行片段。该设备架构使调度程序能够从系统的每个组件收集更多有关资源可用性的最新的准确度量值。该调度程序使用复杂的算法，这些算法通过利用近 100% 的磁盘带宽并确保内存和网络资源不会因过载而导致系统崩溃并损失效率，从而最大化系统吞吐量。这是 IBM 架构的重要特性，可确保系统即使在极高的负载下也能保持以峰值吞吐量运行。

调度程序中的智能：无资源过载



当调度程序亮绿灯时，片段会通过智能网络结构广播至所有片段处理器。

并行执行...

每个 S-Blade 上的所有片段处理器现在都具有执行自己的部分片段所需的指令。除主机调度除程序外，片段处理器具有自己的智能抢先调度程序，允许同时执行来自多个查询的片段。调度程序会考虑查询的优先顺序以及为发出查询的用户或组预留的资源，以确定何时调度执行特定片段及执行的持续时间。当这一时刻到来时，好戏就上演了：

- 每个片段处理器上的处理器内核都会使用查询片段中包含的参数来配置 FAST 引擎并设置数据流。
- 片段处理器会将表数据从磁盘阵列读入内存，利用称为 ZoneMap™ 加速的 IBM Netezza 数据仓库设备创新来减少磁盘扫描。片段处理器还会在访问磁盘上的数据块之前先查询高速缓存，避免在内存中已存在数据时进行扫描。
- FPGA 随后会对数据流进行操作。它首先通过以网速解压数据流，将数据流加速 4 – 8 倍。[4]
- 随后，FAST 引擎会过滤出与查询不相关的任何数据。剩余数据会流回内存以供 CPU 内核进行并行处理。该数据通常只占原始流的极小一部分 (2% – 5%)，从而显著缩短处理器内核所需的执行时间。
- 处理器内核会拾取数据流，并对其执行核心的数据库操作，例如排序、连接和聚集。它还会应用片段处理器中嵌入的复杂算法，以进行高级分析处理。
- 来自每个片段处理器的结果会在内存中进行组合，以生成整个片段的子结果。该过程会在千余个片段处理器间同时重复，并行执行成百上千或成千上万个查询片段。

ZoneMap 加速：IBM 反索引

ZoneMap 加速利用数据仓库中行的自然顺序将性能提升多个数量级。该技术避免扫描列值在查询开始和结束范围之外的行。例如，如果某个表包含两年的每周记录（近 100 周），而查询正在寻找仅其中一周的数据，那么 ZoneMap 加速可以将性能提高达 100 倍。不同于索引，ZoneMaps 可针对每个数据库表自动创建并更新，不会发生任何管理开销。

接着返回结果！

现在，所有片段处理器都具有需要组合的结果。片段处理器使用智能网络结构与主机以及在彼此间进行灵活的通信，执行中间计算和聚集。

主机会对从片段处理器接收到的中间结果进行组合、编译最终结果集并将其返回至用户应用程序。与此同时，其他查询正在流经系统，处于完成的各个阶段。

网络中的智能：可预测的性能和可扩展性

IBM 的定制网络协议是专为与大容量数据仓储相关联的数据量和流量模式而设计的。IBM 协议可确保最大限度地利用网络带宽而不会使其过载，从而实现接近线路速率的可预测的性能。

流量在三个不同区域顺畅流动：

- 以广播方式从主机至片段处理器（1 到 1000 以上）
- 从片段处理器至主机（1000 以上到 1），在 S-Blade 中聚集，处于系统机架级别
- 在片段处理器之间（1000 以上到 1000 以上），大规模数据自由流动，以进行中间处理

为有需要的所有人提供随需应变的信息

最佳解决方案有时并不是最大或最昂贵的解决方案，而是具有最智慧的设计。IBM 早已认识到流式方法处理相比于其他分析和数据仓储系统使用的传统计算架构，具有与生俱来的优势。由此产生的紧凑型设备具有的性能令其他大得多的系统相形见绌，可以针对庞大数据量以及成千上万名并行用户的混合工作负载迅速运行复杂的算法。其他功能可以对处理性能进行补充，从而使 IBM 成为帮助企业成功的独一无二的平台：

- 使用简易性： 正如一体化设备所应该做到的那样，IBM Netezza 数据仓库设备可自我管理，并始终以峰值吞吐量运行。系统软件可确保在无人干预的情况下实现这一点。
- 在整个企业内制定更有效的决策： 嵌入式功能以最少量的开发工作将新一代分析方法带入数据库中。无需单独的服务器硬件或者在大规模的数据传输上浪费时间 – 只会以闪电般的速度产生结果，并能够将关键商业智能带给组织中所有部门内可从中获益的每个人。
- 满足未来需求的敏捷性： 该系统不仅是为了面对当前的挑战而构建，更是为了未来而构建，它可线性扩展至数十 PB 的用户数据，性能加速远超过摩尔定律所描述的传统加速。

IBM Netezza 数据仓库设备使用户及其公司能够制定出最明确的决策，同时确保性能。耳听为虚。要充分认识 IBM Netezza 数据仓库设备，最好的方法就是亲眼见证其运作。我们相信，您将会认为其最大限度利用数据的能力是无与伦比的。

关于 IBM Netezza 数据仓库设备

IBM Netezza 数据仓库设备通过将数据库、服务器和存储器集成到单一、便于管理且只需最低限度的安装和持续管理的设备，彻底改变了数据仓储和高级分析方法，同时产生更快速持久的分析性能。IBM Netezza 系列数据仓库设备通过在设备中数据驻留位置直接整合所有分析活动显著简化了业务分析，从而实现业内领先的性能。请访问 **ibm.com**/software/data/netezza 以了解我们的数据仓库设备系列如何在每一个步骤中消除复杂性，并帮助您为组织实现切实的业务价值。要查看最新的数据仓库和高级分析博客及视频等，请访问：thinking.netezza.com

关于 IBM 数据仓储和分析解决方案

IBM 可提供最广泛且最全面的数据仓储、信息管理和业务分析软件、硬件和解决方案产品服务组合，以帮助将信息资产的价值最大化，并发现新的洞察，以更快地制定明智决策并优化业务成果。

了解更多信息

要了解有关 IBM 数据仓储和分析解决方案的更多信息，请与您的 IBM 销售代表或 IBM 业务合作伙伴联系，或者访问：**ibm.com**/software/data/netezza

国际商业机器中国有限公司
北京市朝阳区工体北路甲 2 号
盈科中心 IBM 大厦 25 层
邮编:100027

IBM 主页位于:
ibm.com

IBM, the IBM logo, ibm.com and Netezza are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at

ibm.com/legal/copytrade.shtml

Microsoft and SQL Server are trademarks of Microsoft Corporation in the United States, other countries or both.

Other company, product and service names may be trademarks or service marks of others.

[1] Gordon Moore 作为 Intel 的创始人之一,于 1965 年曾预测芯片上的晶体管数量每隔两年将翻一番。软件应用程序通常依赖于这些处理器改良来长期提升性能。

[2] 基于内部 IBM Netezza 性能基准测试以及从 Netezza 客户处获取的结果。

[3] 基于对设备中的 FPGA 进行的内部 IBM Netezza 性能基准测试。

[4] 如上。

© Copyright IBM Corporation 2012

请回收再利用

IMW14584-CNZH-01