



IBM Netezza

数据仓库一体机的创始者和领导者



DIREC

IBM中国有限公司

北京市朝阳区北四环中路27号盘古大观写字楼25层

邮编: 100101

电话: 010-63618888

传真: 010-63618555

目录 / contents

STORY

- 4 第一章 关于Netezza
- 5 第二章 Netezza数据设备架构介绍
- 15 第三章 Netezza第四代设备TwinFin系统介绍
- 22 第四章 全球客户及成功案例



第一章 关于Netezza

Netezza是全球领先的数据仓库一体机提供者,致力于提供最优秀的数据仓库。公司成立于2001年,由于表现优异,公司收入连续多年不断上升,截至2011年初,年收入已经高达1.88亿美元。2010年9月20日公司被IBM以17亿美元收购。

公司主要服务于世界500强等特大型企业,在北美、欧洲、中东、非洲、日本及其他亚太地区都设有办事处。服务于零售业、金融服务、电信业、政府、数字媒体、能源、健康和生命科学以及制造业等各种行业。

Netezza的一体机能够提供四种高性能:超级性价比;极强大的可扩展性;使用简捷;提供高性能的高级分析。公司始终坚持以客户为中心的创新,提供业界最好的技术,使复杂的分析计算变得简单,使客户在数据分析领域独具竞争优势,成为您的最佳合作伙伴。



第二章 Netezza数据设备架构介绍

概述

任何企业的成功都取决于是否能够及时获得最有效的信息，并作出正确的决策。机会的流逝，时间和资源消耗都可能使企业处于风险之中。但是，找到关键的信息以便引导企业做出最正确的决策往往意味着需要分析数十亿条的数据纪录和PB级的海量数据，无论是预测结果还是确定某种趋势，或者选择最佳方案，都需要通过对海量模糊的数据进行分析。那些能够及时获得此种智能信息的公司通常能够比竞争对手做出更快的反应，制定更好的决策。

分析方法的不断创新能为企业提供新的情报财富，企业各个方面的业务也将因此而受益。但是当人们迫切需要重要信息时，提供这些重要信息的平台往往是用户最后才想到的事情。在用户眼中它应该象电灯开关一样简单、可靠和直接，能够简单快速的处理人们几乎无法想像的工作量。同时，该平台应该能够长时间可靠地为企业提供服务，在提供最低总体拥有成本的同时，能够承受随更多用户数据量无尽增长所带来的负荷量，并保持其相应的性能。

简单的设备 绝佳的性能

Netezza平台的出现彻底改变了业界数据仓库及分析设备的标准，其设备能够在现在及未来为企业提供行业领先的高性能价格比、易于使用的解决方案。该平台创立了高级分析方法中的新领域，能极其快速高效的运行庞大的数据处理任务。对于需要商业智能的用户及其企业来说，即使面对不断增长的业务需求，该系统都能始终为其提供最好的智能服务。


Netezza以提供最市场上最优性能价格比的产品为设计原则，因此其数据仓库和分析设备产品采用了革命性的设计理念。该设备是为高速分析而建立，其强大的功能并非来自最强劲最昂贵的IT组件，而是如何将合适的组件组装起来并将其性能发挥到极致。大规模并行处理(MPP)流将多核CPU与Netezza独特的FPGA加速流技术引擎(FASTTM)相结合，从而能提供连那些非常昂贵的系统都无法匹配甚至接近的性能。而且作为一个使用非常简单的设备，用户只需要输入指令，系统便能以令人惊讶的速度将结果直接反馈给用户，而不需要进行索引或对系统做出任何调整和优化。设备的简单性也使应用的开发简单化，使快速创新和高性能分析能力为广泛的用户和处理过程提供服务。

本文介绍了Netezza系统的非对称性大规模并行处理(AMPPTM)架构，并概述了该系统如何组织查询和分析工作使系统以前所未有的速度运转。我们将了解Netezza系统软件和硬件是如何协同配合以将其每个关键组件的性能发挥到极致，以及这个为处理数万个用户海量数据查询而设计的系统是如何工作的。总之，Netezza系统是一个独特的数据仓库和分析平台，具有无可比拟的性价比，并且为应对来自当前和未来的挑战做好了准备。

Netezza产品深入分析

架构设计原则

Netezza数据仓库设备将数据库，数据处理和数据存储集成在一个紧凑、优化的系统中，并随时满足数据量增长的需要。其系统架构是基于以下设计原则而建立的，而这些核心方法是Netezza公司在数据仓库和分析



设备领域中提供最佳性价比系统的重要标志。

数据处理尽可能接近数据源

Netezza系统架构的设计基于计算机科学的一个基本原则：当对一个大型数据集进行操作时，除非绝对需要，否则不要轻易移动数据。Netezza公司将这一原则发挥至极致，利用现场可编程逻辑门阵列(FPGA)组件尽早的将多余的数据从数据流中过滤掉，以减少其对硬盘空间的占用。该数据过滤过程在接近数据源的地方进行，从而消除了I/O瓶颈，并将下游组件如CPU、内存和网络从多余的数据处理中解放出来，从而使系统性能有了根本性的显著提高。

平衡的大规模并行架构

Netezza的架构结合了SMP(对称多处理)和MPP(大规模并行处理)的优点，建立了一个能以极快的速度分析PB量级数据的设备。该架构中的每一个组件，包括处理器、FPGA、内存和网络连接都经过了精心的挑选和优化，在硬盘物理条件许可的情况下，以最低的成本和能耗快速地处理数据。Netezza通过软件协调这些组件以流水线的方式对数据进行并发操作，从而充分获取和利用每个MPP的节点上最大吞吐量。除了性能以外，这个平衡的架构还能够为超过1000的并发处理流提供线性可扩展性，以及非常低的总体拥有成本。

用于高级分析的平台

MPP和在接近数据源的地方进行数据处理的方法也同样适用于针对于大型数据集所进行的高级分析。Netezza系统能够轻松地将复杂的非SQL算法嵌入到MPP流的处理组件中，并且没有并行程式或网格程式所具有的典型复杂性。这种针对庞大的数据量能够以“流水线”方式对复杂数据进行分析处理的能力，能够消除将数据转移到单独硬件的延迟和开销，为数据仓库和高级分析方法的融合提供了一个理想的平台，同时其性能也提高了几个数量级。

简单的设备

Netezza使日常的操作变得自动化和简化，使得用户无需对复杂的底层平台进行了解。无论何时，当需要与其他设计需求进行取舍时，简单性都是我们所采取的首要原则。与其他解决方案不同的是，Netezza平台不需要用户对其进行任何调整和优化，只要安装完成便能立即高速处理用户的查询指令以及各种级别的工作负荷。即使是一些通常需要消耗时间较长的任务，比如系统的安装升级，以及确保高可用性和业务连续性等任务都进行了很大程度的简化，以节省宝贵的时间和资源。

加快创新和性能提升

从长远来看，Netezza设计架构的主要目标就是要比竞争对手更快地实现系统性能价格比的提升并增加创新功能。由于Netezza使用开放的刀片式组件，从而能够迅速采用更先进的技术。通过使用FPGA的涡轮效果——一种平衡的硬件配置加上与其紧耦合的智能软件，使系统的性能得到了全面提升，远远高于其单个组件所能达到的系统性能提升程度。事实证明，Netezza公司自推出其系统以来，其系统性能每两年就能提高4倍(比摩尔定律翻一番)，远远超过其他竞争对手的水平。

灵活的配置和极强的可扩展性

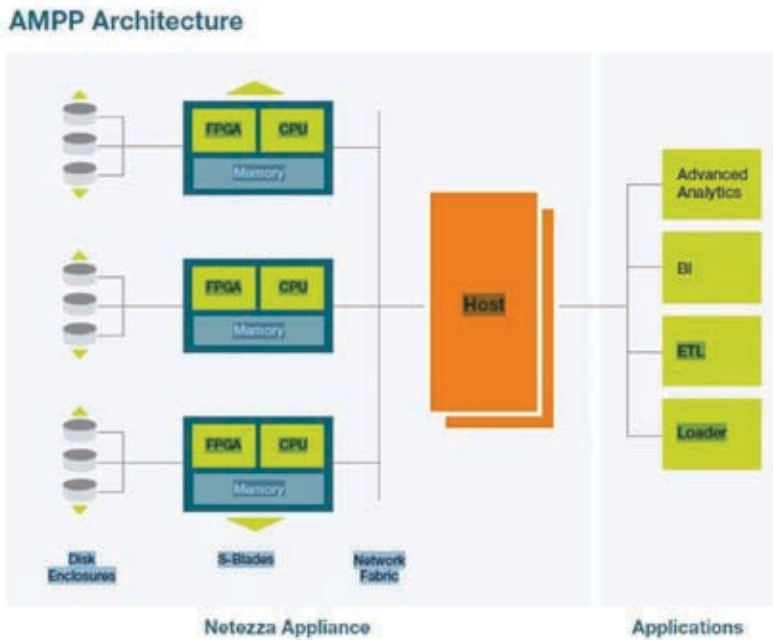
Netezza的模块化方式能够将用于查询的用户数据容量从几百个GB扩展到数十个PB。系统架构的高适应性能够满足不同的数据仓库和分析需求，开放式刀片组件的使用能够方便的调整磁盘-处理器-内存的配置比率，以满足性能或存储优先的应用需求。同样的架构还支持基于内存的系统，为关键的应用任务提供极为快速、实时的分析能力。

以下的章节将介绍Netezza如何将这些方法应用到实际工作中。

系统构建模块

Netezza系统性能的主要优势来自其独特的AMPP处理架构，该架构将SMP前端与一个无共享的MPP后端相

结合完成查询处理。该架构将经过精心挑选的各个组件集成在一起组成了平衡的整体系统。通过每个处理组件对多个数据流进行操作，并尽早过滤掉多余的数据。最多可有多达一千多个MPP处理组件共同工作，有效分解和处理工作负荷。



现在让我们来深入Netezza设备的内部，了解其主要模块和工作原理：

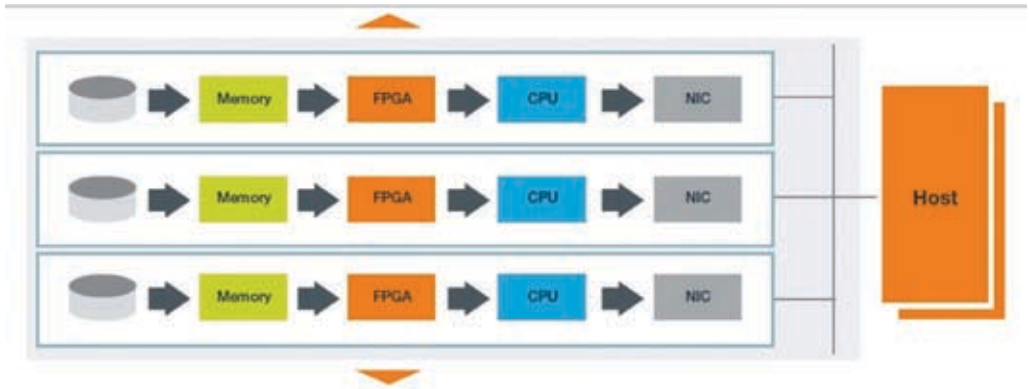
- **Netezza主机：** SMP主机是由多个主备配置的高性能Linux服务器组成，满足系统高可用性需要。其中激活的主机为外部工具和应用程序提供标准化的接口；将SQL查询语句汇编成可执行的代码段—称为snippets；建立优化的查询计划并将snippets分配给MPP处理节点来进一步执行。

- **Snippet Blades (S-Blades):** S-Blades是智能处理节点，构成了设备的涡轮MPP引擎。每个S-Blade都作为一个独立的服务器，其中包含功能强大的多核CPU、多引擎FPGA和GB级别的RAM。所有节点都实现负载均衡配置，并同时工作以提供最佳性能。CPU内核拥有充足的缓存来运行复杂的算法，以满足对大量数据和高级分析应用的需要。
- **磁盘封装:** 磁盘柜中堆积着RAID高密度，高性能磁盘。每个磁盘包含一个数据库表中的一部分数据。磁盘柜通过高速互联连接到S-Blades，使所有磁盘同时工作，以最快的速度，将数据流传送到S-Blades上。
- **网络结构:** 所有系统组件通过高速网络结构而进行互联。Netezza设备运行定制是基于IP的协议，该协议可以充分利用网络带宽并消除拥塞，即使在持续突发的网络流量下也能很好地工作。经过优化的网络可以扩展到一个节点以上，并允许各个节点同时传输大量数据到其他所有节点。

注意: 所有系统组件都配有冗余。无论主机处于主/备的状态，设备中的所有其他组件均可实现热插拔。用户数据采用完全镜像保护，确保了99.99%以上的可用性。

优越性能的来源——深入了解S-Blade

现在让我们将目光投向众多Snippet处理器中的一个：采用商用组件和Netezza公司软件相结合的方式从每个MPP的节点获取最大吞吐量。存储阵列中专门的高速互连模块确保数据以最快的速度从硬盘传送到内存。另外通过使用智能算法，被压缩数据尽可能存放于缓存中，以确保最常用的数据能马上从内存中获取而不需要访问磁盘。在FPGA中的FAST引擎并行工作，以物理速度解压并过滤出95%-98%的表数据，而保留只与当次查询结果有关的数据，并由多个并行的CPU核对其进行同时处理。所有这些措施使得Netezza设备的性能甚至要比那些昂贵的同类系统还高出几个数量级。



S-Blades的涡轮技术: 强大的Netezza FAST引擎

Netezza实现低价格、高性能设备的关键技术之一是采用了FPGA。每个FPGA包含多个嵌入式引擎，对数据流进行过滤和转换功能。该快速引擎可实现动态配置，并能通过软件进行修改或扩展。软件可根据从查询执行过程中获得的参数，通过内存访问(DMA)模块对该引擎进行快速定制修改。



FAST引擎包含

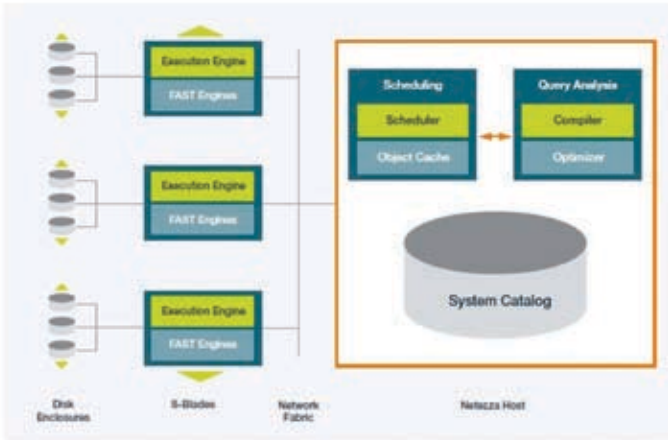
- **压缩引擎。**作为Netezza公司的一项创新技术，能将系统性能提高4-8倍。该引擎以线速解压数据，并立即将磁盘上的数据块转化为4-8块，然后存储在内存中，其结果是极大的提高了任何数据仓库中速度最慢的组件——磁盘的速度。
- **项目和限制引擎。**该引擎根据SQL查询语句中的SELECT和WHERE子句中的参数分别筛选出相关的列和行，从而进一步提高了系统性能。
- **可见度引擎。**该引擎在Netezza系统中以流速度维护数据的ACID特性(原子性，一致性，隔离性和持久性)方面发挥着关键的作用。它过滤掉不需要显示出来的行，例如：那些还没有最终提交的数据行。

Netezza FAST引擎还提供了一个可扩展的框架，以确保该引擎能够在将来通过软件的优化而增加新的创新功能。这些新的功能确保Netezza能进一步提升系统性能，并增加系统安全性和可靠性。

Netezza系统中的查询调度

Netezza公司的硬件系统和智能系统软件是紧密交织在一起的。该软件的设计充分利用硬件设备的能力，并包含了很多技术创新以实现指数化的性能增益。无论是面对简单查询、复杂的随机查询或者深度分析计算，系统均能很好地完成工作。在本章节中，我们将了解在该系统中智能功能是如何一步步建立起来的。

Software Architecture



Netezza软件组件包括

- 先进的并行优化器: 该优化器对查询逻辑进行转译以提高运行效率，并确保充分利用每个处理结点上的部件
- 智能调度器: 确保无论有多少工作负荷，系统都能在最高吞吐量的状态下运行
- 增强的Snippet处理器: 同时高效执行多个查询和复杂的分析功能
- 智能网络: 能快速轻松的将大量数据在Netezza系统中传输

让我们从用户输入查询指令开始来了解这些组件是如何协同工作的。了解技术的读者会看到，Netezza公司在处理查询时将完全不同于其他数据仓库系统。

制定优化的查询计划

主机首先编译查询指令，并针对Netezza的AMPP架构创建一个优化的查询执行计划。Netezza公司智能优化器功能是该系统最大的优势之一。该优化器利用系统中所有的MPP节点，根据查询指令中涉及的所有数据表格

收集详细的最新信息，产生针对该次查询的实时统计数据。由于大部分信息都是在查询执行期间捕获的，因此实际开销非常低。由于Netezza设备组件之间能够进行相互通讯，使得优化器能够针对每一次操作更加准确的评估磁盘、处理器和网络开销。通过获取准确的数据，优化器得以制定查询计划并充分发挥各个组件的能力。

智能优化器: 优化表连接次序体现智能优化器功能的一个例子是优化复杂的连接操作次序。例如，如果将多个小数据表格合并成一个大的表格，优化器将保持大表格处于分布存储状态，同时将所有小表格数据通过广播发给每个单独的SPU。该方法最大程度地减少了数据移动，同时利用MPP系统架构并行处理连接操作。

在系统运行之前，优化器利用统计数据转换查询指令，以尽量减少磁盘I/O和数据移动 -- I/O和数据移动均为降低数据仓库系统性能的关键因素。优化器执行的转换操作包括:

- 确定正确的连接次序
- 重写表达式
- 消除冗余的SQL操作

转换成snippets

编译器将查询计划转换成可执行代码段，即snippets。该查询代码段由设备中的Snippet处理器在所有的数据流上并发执行。

每个snippet都包含两个部分: 一个是编译后的代码，由单个CPU核执行；以及FPGA参数的集合，用于定制FAST引擎从而对该snippet有关的数据进行过滤。实际上，针对每个snippet的定制工作使得Netezza能够实时地为每一个独立的查询进行硬件优化配置。

编译器智能: 对象缓存

主机采用了对象缓存功能以进一步提高查询性能。这是一个面向已经汇编过的Snippet代码的巨大的缓存存储器，并支持不同的参数。例如，一个子句为"where name='bob'"的snippet可能与另一个子句为"where name='Jim'"的Snippet使用相同的代码，只是查询条件中的名字不同。这种方法可消除99%以上的snippets汇编步骤。



第三章 Netezza第四代设备TwinFin系统介绍

Netezza公司的TwinFin. 数据仓库和分析设备:

我们针对数据分析所采用的方法已获得专利，并经过验证，内嵌在易于使用的设备中。该方法在以物理速度处理大量数据的同时，最大限度地减少了数据的移动。整个设备运行速度极快且价格低廉，让我们的客户得以进行那些在过去不可能实现的数据分析工作。通过使用Netezza数据分析设备，我们的客户不仅可以克服其现有系统的局限性，并且简化了其处理数据的方法。

TwinFin系统是Netezza公司推出的第四代设备，以更优的性价比重新设定了数据仓库和BI分析系统的标准。TwinFin是目的性很强且基于标准而建立的数据设备，在架构上将数据库、服务器和存储功能进行了集成，是一个独立并易于管理的系统。TwinFin设备针对快速分析PB量级的数据而设计，其性能比其他传统数据仓库供应商的设备性能高出10-100倍，而其成本却比同类产品低的多。

Netezza公司TwinFin的特点:

- 领先的价格和性能优势—极具竞争力的价格，同时系统性能超出同类竞争对手的10-100倍。
- 可扩展性——数据量可从1TB以下扩展到PB量级。
- 支持数千计用户的使用，并支持高度复杂、混合的工作负荷。
- 刀片式架构——行业领先的多核Intel刀片服务器，在实施过程中结合了磁盘存储和Netezza FPGAs专利数据过滤技术。
- 行业标准接口(SQL, ODBC, JDBC, OLE DB)



- 完全兼容市场领先的BI工具，应用软件以及基础设施。
- 设备易用性——无需索引或调整；少量的运行管理工作。
- 大型企业级的可靠性和可用性——超过99.99%的正常运行时间
- 绿色——低功耗和散热要求且占地面积紧凑
- 针对高级分析的平台——好几个数量级的性能优势
- 极快的加载速度——每小时超过2TB
- 极快的备份速度——数据速率超过4TB每小时的高速备份以及存储性能

性能、简单、价值

性能

Netezza公司的TwinFin系统以好几个数量级的性能优势超越了其他同类分析系统，而该优势来自于其独特的大规模并行处理架构(MPP)。该架构融合了开放式刀片服务器、磁盘存储器以及Netezza公司的FPGA专利数据过滤技术。这种技术融合为高度复杂的混合工作负荷提供了超级快速的查询性能，支持上万用户的快速查询，并提供了高速的紧密分析能力和能够扩展到PB量级数据的模块扩展能力。

简单

TwinFin系统作为一款简单易用，高性能的设备，其不需要索引、不需要对系统进行调整或其他操作。该系统可随时进行数据加载及执行查询任务，并通过ODBC、JDBC和OLE DB等标准接口集成业界所有领先的ETL、BI和分析应用。

TwinFin是一款集成了所有硬件、软件和存储介质的设备，因此大大缩短了部署周期。同时TwinFin还免除了其他传统数据库所必须的管理开销，用户可以将本来用于该分析设备的资源分配给其他计划或策略，从而对公司的盈利产生积极的影响。



价值

作为一款商业型设备，Netezza公司的TwinFin堪称低成本数据分析设备的首选，但其价值远远超过了您最初花费的购买价格。Netezza的设备不论是在内部资源还是在执行成本方面都只需要最低的日常支持，从而其总拥有成本很低。对于用户来说，没有任何其他不为用户所知的额外费用。

Netezza能够让用户快速评估重要的对企业利润有积极影响的BI策略。使用Netezza设备，意味着在面对任何可能的市场机遇或威胁时，您的公司能够根据最精准的情报做出更快更准确的反应。

在公司需要尽可能地对不断变化的市场条件以及日益增加的分析需求进行快速反应的时候，一款简单、易于安装、能够对千兆级数据进行快速运行分析的系统就更具意义。

这是如何实现的呢？

已取得专利的流体系结构

Netezza公司的TwinFin设备虽然采用刀片服务器和存储技术，但仍保持了Netezza以贴近存储设备处理数据，从而最大程度上减少数据移动的基本原则。TwinFin使用的FPGAs就是Netezza为了达到这一目的所特别开发的专项技术。FPGA当数据流经磁盘的时候，就可以将无关数据快速过滤掉。这不仅解决了I/O瓶颈问题，并且将CPU、内存和网络从无关数据处理的工作中解放出来，实现更加卓越的系统性能。

一个标准的单机架承载32TB非压缩数据的Netezza TwinFin设备有12台Snippet刀片或者SBlades。所有的snippets(或称为SQL查询处理可执行代码段)都在这些SBlades中执行。这些SBlades都是智能处理节点，组成了TwinFin设备的MPP引擎。每个Sblade都是一个独立的服务器，包含有强大的多核Intel处理器、Netezza独有的多引擎FPGAs和千兆字节的内存。这些组件都配置均衡，同时工作以提供最佳性能。这也是Netezza公司专利和快速优良系统性能的核心。

架构中的所有组件，包括处理器、FPGA、内存以及网络都经过精心挑选和优化，在磁盘物理条件允许的情



况下，用最低的成本和能耗以最快的速度过滤数据。除了系统拥有均衡的硬件配置外，Netezza软件组织这些组件以流水线的方式对数据进行并发操作，从而充分利用和获取每个SBlade上的最大吞吐量。Netezza的数据库软件包括用于查询指令高效运行转化的先进并发优化器，以及用于减少磁盘数据获取时间的算法。除了最基本的性能以外，这个平衡的架构，不仅能够为一千多并发处理流提供线性的可扩展性，同时其总拥有成本也非常低。



每一个Snippet刀片或Sblade包括:

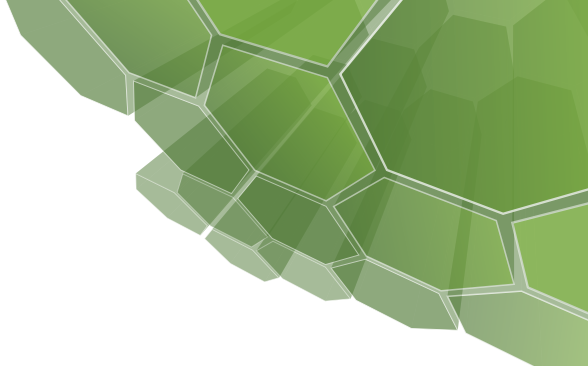
- 一个标准刀片服务器
- 一个特别的卡—Netezza数据库加速器，其附着在刀片服务器上。正是因为这张卡，TwinFin才能够实现无与伦比的杰出性能。



Netezza TwinFin* 描述	单机架系统			单机架系统	
Netezza TwinFin	TwinFin3	Twinfin6	Twinfin12	机架2	机架3+
机架	1	1	1	2	3-10
S-Blades	3	6	12	24	#Racks × 12
用户数据(TB) (未压缩)	8	16	32	64	#Racks × 32
耗电/机架 (最大瓦特/机架)	2,819.4	3,960	7,635	7,400	7,000
散热/机架 (BTU/小时)	9,600	13,500	26,100	25,500	24,000
重量/机架(千克)	453.6	589.7	907.2	907.2	907.2
高度/机架(厘米)	202	202	202	202	202
深度/机架(厘米)	101.6	101.6	101.6	101.6	101.6
宽度/机架(厘米)	64.8	64.8	64.8	64.8	64.8
功 率	200-240VAC 50/60Hz, Single Phase 16A	200-240VAC 50/60Hz, Single Phase 24A	200-240VAC 50/60Hz, SinglePhase 48A 200-208VAC 60Hz, 3- Phase Delta 32A 200-240VAC 50Hz, 3- Phase WYE 24A	200-240VAC 50/60Hz, SinglePhase 48A 200-208VAC 60Hz, 3- Phase Delta 32A 200-240VAC 50Hz, 3- Phase WYE 24A	200-240VAC 50/60Hz, SinglePhase 48A 200-208VAC 60Hz, 3- Phase Delta 32A 200-240VAC 50Hz, 3- Phase WYE 24A
Drops/机架	2	2	2	2	2
安全性	UL/CSA/EN60950				
排放量	FCC Part15,ICES-003,AUS/NZ C-Tick,VCCI andEN55022 Class A;European Immurity:EN55024				

*pending Compliance Certmcaton

其它的Netezza Twinfin产品型号: Twinfin 18, Twinfin24, Twinfin30, Twinfin36, Twinfin48, Twinfin60, Twinfin72, Twinfin84, Twinfin98, Twinfin108, Twinfin120. 型号中的数字表示在该设备中SBlades的数量。



软件	第三方应用
<p>操作系统: Red Hat Linux Advanced Server5.3</p> <p>支持APIs:SQL,OLE DB,OOBC3.5,JDBC V3.0 Ttpe4</p> <p>SQL标准: SQL-92compliant,with SQL-99 extensions</p> <p>高速数据加载/卸载: 与ETL和EAI工具的互操作性达到超过2TB/小时的传输速率</p> <p>备份/恢复: 具备与EMC Legato、IBMTivoli和Veritas的互操作性。传输速率超过4TB/小时</p> <p>数据库可移植性: 可从IBM DB2,Informix,Microsoft SQLServer,MySQL,Oracle,Red Brick,Sybase IQ,Teradata进行数据库移植</p> <p>其他工具: Windows and web-based DB Admin GUI;CLI and high-speed loading/unloading for AIX,HP-UX,Linux,Solaris and Windows</p> <p>所属硬件: 支持以太网和SAN链接</p>	<p>数据整合: Ab initio,Business Objects/SAP, Composite Software, Expressor Software,Golden/Gate Software,Information,IBM Information Server,Orade Sunopsis,WisdomForce</p> <p>数据分析: Bussiness Objects/SAP, Kaldo,KCEN, Quest Software,SAS,SPSS</p> <p>BV报告: Actuate,Business,Objects/SAP,Congnos,an IBM Company,Information Bulders,MicroStrategy,Oracle,QIKTech,Unica</p> <p>业务连续性/兼容性: ENC,IBM DataMirror,IBM Tivoli Storage Manager,Symartec Veritas</p>

第四章 全球客户及成功案例

数字媒体



金融服务



政府部门



卫生和生命科学



零售/消费产品



电信



其它



