



***Menos de um ano desde o anúncio do release do IBM Netezza Data Warehouse Appliance, mais de 100 clientes já adotaram o appliance, incluindo:***

- **Blue Cross Blue Shield de Massachusetts**
- **BlueKai**
- **Catalina Marketing**
- **Con-way Freight**
- **DataLogix**
- **Epsilon, uma Empresa Alliance Data**
- **interCLICK**
- **Intercontinental Exchange**
- **Japan Medical Data Center**
- **Kelley Blue Book**
- **Marshfield Clinic and Marshfield Clinic Research Foundation**
- **MediaMath**
- **MetroPCS**
- **Micro Ad**
- **MyLife.com**
- **NYSE Euronext**
- **Pacific Northwest National Laboratory**
- **Premier, Inc.**
- **The Nielsen Company**
- **Wind Telecom**

 Índice	PÁGINA
<b>1</b> Introdução	<u>4</u>
<b>2</b> OLTP (Processamento de Transações On-line) e Data Warehousing	<u>6</u>
<b>3</b> Desempenho de Consulta	<u>8</u>
<b>4</b> Simplicidade de Operação	<u>17</u>
<b>5</b> Valor	<u>23</u>
<b>6</b> Comparando o Exadata e o IBM Netezza Data Warehouse Appliance	<u>28</u>
<b>7</b> Conclusão	<u>29</u>

A Netezza se concentra na tecnologia projetada para consultar e analisar grandes dados. Os inovadores appliances de data warehouse da empresa estão revolucionando o mercado. Com o desejo de explorar dados a custos reduzidos de Operação e de propriedade, muitos de nossos clientes trocaram seu data warehouse da Oracle. A Oracle trouxe atualmente ao mercado o Exadata. Uma máquina que aparentemente faz tudo o que o IBM Netezza Data Warehouse Appliance faz, e também processa Transações on-line. A força do Exadata está simplificando e eliminando o custo dos sistemas OLTP (Processamento de Transações On-line) dos clientes Oracle. Ele consolida diversos sistemas, remove a duplicação e substitui a complexidade por um único ambiente escalável OLTP de alto desempenho. No entanto, esses benefícios são comprometidos quando as organizações tentam consolidar seus bancos de dados de analítica no Exadata. A Netezza possui um histórico de entrega de valor aos clientes Oracle. Devido à eliminação da carga de trabalho de análise dos bancos de dados OLTP da Oracle, atualmente, mais de 200 organizações conseguem obter o melhor dos dois mundos.

Essa avaliação do Exadata e do IBM Netezza Data Warehouse Appliance como plataformas de data warehouse foi criada a partir de um ponto de vista ousado da Netezza, no entanto, para garantir a credibilidade, decidimos seguir o conselho de Philip Howard, Diretor de Pesquisa da Bloor Research e Curt Monash, Presidente da Monash Research.

A inovação exige que pensemos e façamos as coisas de maneira diferente, resolvendo problemas através de novas abordagens. A Netezza se concentra exclusivamente nas necessidades e desejos dos clientes relacionados ao data warehousing. O IBM Netezza Data Warehouse Appliance fornece um desempenho excelente para as consultas no data warehouse de nossos clientes. O IBM Netezza Data Warehouse Appliance oferece simplicidade aos clientes. Qualquer pessoa com um conhecimento básico de SQL e Linux possui as qualificações necessárias para realizar as poucas tarefas administrativas que são exigidas para manter níveis de serviço consistentes em cargas de trabalho que mudam dinamicamente. O desempenho com simplicidade do IBM Netezza Data Warehouse Appliance reduz seus custos de propriedade e execução de um data warehouse. E, mais importante, nossos clientes criam um novo valor para o negócio com a implementação de aplicativos analíticos que antes consideravam fora de alcance.

***“A Netezza foi parte da inspiração para o Exadata. O Teradata foi parte da inspiração para o Exadata”, reconheceu Larry Ellison em 27 de janeiro de 2010.  
“Gostaríamos de agradecê-los pelo apoio e pelo incentivo para que entrássemos nos negócios de hardware”.***

“A Netezza foi parte da inspiração para o Exadata. O Teradata foi parte da inspiração para o Exadata”, reconheceu Larry Ellison em 27 de janeiro de 2010. “Gostaríamos de agradecê-los pelo apoio e pelo incentivo para que entrássemos nos negócios de hardware”.<sup>1</sup> Esse comentário, feito com o habitual entusiasmo de Larry Ellison, envolve uma séria questão: apenas a melhor tática para chamar a atenção da Oracle. O Exadata representa uma direção estratégica para a Oracle, adaptando seu sistema de gerenciamento de banco de dados de OLTP e associando-o ao sistema de armazenamento paralelo massivo da Sun. A Oracle lançou o Exadata V2 com a promessa de desempenho máximo para o processamento de consultas analíticas e Transações on-line. É evidente que a Oracle supera o OLTP. Mas o data warehousing e analytics possuem demandas bastante diferentes de software e hardware em comparação com o OLTP. As credenciais de data warehousing do Exadata exigem um exame minucioso, especialmente em relação à simplicidade e valor.

Este White Paper começa com uma revisão das diferenças entre o processamento de Transações on-line e o processamento de consultas e análises em um data warehouse. Em seguida, ele aborda o Exadata e o IBM Netezza Data Warehouse Appliance a partir das perspectivas de desempenho de consulta, simplicidade de operação e valor.

Apenas pedimos aos leitores que façam como nossos clientes e parceiros: deixem de lado as noções de como um sistema de gerenciamento de banco de dados deveria funcionar, estejam abertos a novos modos de pensar e preparem-se para fazer menos, e não mais, para obter um resultado melhor.

## ...preparem-se para fazer menos, e não mais, para obter um resultado melhor.

Um aviso: A Netezza não possui acesso direto a uma máquina do Exadata. Fomos bem-sucedidos no feedback detalhado que recebemos de muitas organizações que avaliaram ambas as tecnologias e escolheram o IBM Netezza Data Warehouse Appliance. Devido ao tamanho do Oracle e sua ênfase no Exadata, as informações disponíveis publicamente sobre o Exadata são incrivelmente escassas. Os casos de uso mencionados pela Oracle fornecem poucos dados à discussão, que por si só diz respeito a diversos seguidores do segmento de mercado, por exemplo, Information Week. As informações compartilhadas neste documento foram disponibilizadas de maneira franca. Quaisquer imprecisões resultantes de nossos erros não ocorreram com a finalidade de disponibilizar informações erradas.

<sup>1</sup> Consulte [http://oracle.com.edgesuite.net/ivt/4000/8104/9238/12652/lobby\\_external\\_flash\\_clean\\_480x360/default.htm](http://oracle.com.edgesuite.net/ivt/4000/8104/9238/12652/lobby_external_flash_clean_480x360/default.htm)

# OLTP (Processamento de Transações On-line) e Data Warehousing

Os sistemas OLTP executam diversas Transações curtas. O escopo de cada transação é pequeno, limitado a um ou a apenas alguns registros e, portanto, é esperado que os dados sejam frequentemente armazenados em cache. Embora os sistemas de OLTP processem grandes volumes de consultas de banco de dados, eles se concentram na gravação (ATUALIZAÇÃO, INSERÇÃO e EXCLUSÃO) em um conjunto de dados atual. Esses sistemas normalmente são específicos para um processo ou função de negócios, por exemplo, o gerenciamento do saldo atual de uma conta-corrente. Seus dados normalmente são estruturados em 3NF (third normal form). Os tipos de transação de sistemas OLTP são estáveis e seus requisitos de dados são bem compreendidos, portanto, as estruturas de dados secundárias, como índices, podem localizar registros em disco de maneira útil, antes de transferi-los da memória para o processamento.

Em comparação, os sistemas de data warehouse são caracterizados por Operações de leitura (SELEÇÃO) de banco de dados predominantemente intensas em um conjunto de dados atual e histórico. Enquanto uma operação OLTP acessa uma pequena quantidade de registros, uma consulta de data warehouse pode verificar uma tabela de um bilhão de linhas e unir seus registros com os registros de diversas outras tabelas. Além disso, é impossível prever quais usuários de negócios de consultas solicitarão dados em seus armazéns, reduzindo o valor do armazenamento em cache e indexando estratégias. As opções para o dimensionamento do data warehouse variam de 3NF a modelos dimensionais, como esquemas tipo “flocos de neve” e “estrela”.

Os dados em cada sistema que suprem um data warehouse típico são estruturados de modo a refletir as necessidades de um processo de negócios específico. Antes que os dados sejam carregados no data warehouse, eles são limpos, deduplicados e integrados.

Este documento divide o data warehouse em primeira ou segunda geração. Embora essa classificação possa não suportar um exame minucioso mais profundo, ela reflete como muitos de nossos clientes falam de seus caminhos evolutivos para gerar um valor cada vez maior de seus dados.

Os armazéns de dados de primeira geração normalmente são carregados durante a noite. Eles fornecem informações aos seus negócios através de um corpo estável de painéis e relatórios com base em SQL desenvolvidos lentamente. Como esses simples armazéns são, de alguma maneira, semelhantes aos sistemas OLTP (seus requisitos de dados e carga de trabalho são compreendidos e estáveis), as organizações frequentemente adotam os mesmos produtos de gerenciamento de banco de dados que usam para o OLTP.

Com o produto vem a prática: os administradores de banco de dados analisam os requisitos de dados e índices de desenvolvimento de cada relatório para acelerar a recuperação de dados. A chegada das técnicas e da tecnologia OLTP parece ser um sucesso, até que os volumes de dados no data warehouse ultrapassam os volumes normalmente gerenciados nos sistemas Transacionais.

Neste século, as corporações e órgãos do setor público consideram normais as taxas de crescimento de dados de 30% a 50% ao ano. As tecnologias e práticas bem-sucedidas no mundo do OLTP comprovam ser cada vez menos aplicáveis ao data warehousing. O índice como auxílio à recuperação de dados é um exemplo disso. Como o sistema de banco de dados processa tarefas para carregar os dados, ele também permanece ocupado atualizando seus diversos índices. Com grandes volumes de dados, isso se torna um processo bastante lento, fazendo com que as tarefas de carregamento invadam suas janelas de processamento distribuídas. Apesar de trabalhar por várias horas, a equipe técnica perde níveis de serviço estabelecidos com os negócios. A produtividade é afetada enquanto as unidades de negócios aguardam que os relatórios e dados sejam disponibilizados.

## **As tecnologias e práticas bem-sucedidas no mundo do OLTP comprovam ser menos aplicáveis ao data warehousing...**

As organizações estão redefinindo o modo como necessitam e desejam explorar seus dados. Este documento refere-se a esse desenvolvimento com data warehouse da segunda geração. Esses novos armazéns, que gerenciam grandes conjuntos de dados com facilidade, servem como a memória corporativa. Quando interrogados, eles evocam novamente os eventos registrados nos anos anteriores. Essas memórias distantes aumentam a exatidão de aplicativos de análise preditiva. Feeds de fluxo constante estão substituindo cargas de lotes durante a noite, reduzindo a latência entre o registro e a análise de um evento dentro do contexto de milhões de outros eventos. Além do SQL simples usado para preencher relatórios e painéis, o data warehouse processa regressões lineares, Naïve Bayes e outros algoritmos matemáticos de analítica avançada. Observar um aumento repentino nas vendas de um produto com alta margem de lucro em apenas cinco lojas fará com que um fornecedor compreenda o que aconteceu e por quê. Esse conhecimento informa estratégias para promover uma atividade de vendas semelhante em todas as 150 localizações da loja. O sistema de computação que suporta o data warehouse deve ser capaz de gerenciar essas variações repentinas na demanda sem interromper os relatórios e painéis regulares. Os usuários de negócios estão exigindo liberdade para explorar seus dados no momento e da maneira que quiserem. Seu desejo por rapidez não dá lugar às tecnologias das quais o desempenho depende do trabalho coordenado dos administradores.

Para analisar esses volumes de terabyte para petabyte, é necessário um computador capaz de distribuir dados por diversos nós de cálculo, cada um analisando um conjunto de dados menor em paralelo, com todos os outros nós de estações na grade. Sistemas de Data Warehouse capazes de atingir um desempenho de consulta rápido e consistente com cargas de trabalho em constante mudança possibilitam uma nova classe de aplicativos, criando insights valiosos e incentivando a adoção de tomadas de decisão orientadas a dados.

## **Desempenho de Consulta com o Oracle Exadata**

Ao adquirir a Sun, a Oracle chegou à conclusão que a Netezza alcançou uma década antes: os sistemas de data warehouse obtiveram a mais alta eficiência quando todas as partes, software e hardware, foram otimizadas de acordo com as metas. O Exadata foi criado a partir de dois subsistemas conectados por uma rede rápida: um sistema de armazenamento inteligente comunicando-se através de InfiniBand com um Oracle Database 11g V2 com RAC (Clusters de Aplicativos Reais). Um único sistema de rack inclui uma camada de armazenamento de 14 servidores de armazenamento, chamados de células do Exadata, em uma grade de MPP (processamento paralelo massivo), correlacionada com o banco de dados Oracle RAC executando como um cluster de disco compartilhado de oito nós simétricos de multiprocessamento.

**Ao adquirir a Sun, a Oracle chegou à conclusão que a Netezza alcançou uma década antes: os sistemas de data warehouse obtiveram a mais alta eficiência quando todas as partes, software e hardware, foram otimizadas de acordo com as metas.**

A Oracle considera a camada de data warehouse do Exadata inteligente porque ela processa projeções de SQL, filtragem de junção simples e de restrição<sup>2</sup>, antes de colocar o conjunto de dados resultante na rede para processar o recebimento de dados pelo Oracle RAC. Essa técnica é chamada de varredura inteligente. No entanto, o processamento dos dados na camada MPP do Exadata depende de diversos fatores, incluindo o tipo de tabela usada para armazenar dados, como os dados recentes chegaram ao data warehouse e outros fatores discutidos abaixo.

<sup>2</sup> A Technical Overview of the Sun Oracle Exadata Storage Server and Database Machine – white paper da Oracle, outubro de 2009.

Dados gerenciados em estruturas Oracle, chamadas de IOTs (tabelas organizadas por índice), não são processados pela varredura inteligente. Nas páginas de 3 a 20 do Conceitos do Oracle® Database 11g Versão 2 (11.2), de fevereiro de 2010, a documentação recomenda “Tabelas organizadas por índice são úteis quando partes de dados relacionadas precisam ser armazenadas em conjunto ou quando dados devem ser armazenados fisicamente em uma ordem específica. Esse tipo de tabela é frequentemente usada para aplicativos OLAP, espaciais ou de recuperação de informações. As páginas de 19 a 53 do Guia do Administrador Oracle® Database 11g Release 2 (11.2), de fevereiro de 2010, declaram “Tabelas organizadas por índice são adequadas para estruturas de índice específicas de aplicativos de modelagem. Por exemplo, aplicativos de recuperação de informações baseados em conteúdo que contêm dados de texto, imagem e áudio exigem índices invertidos que podem ser modelados com eficácia usando-se tabelas organizadas por índice”. Os clientes da Netezza executam esses aplicativos de data warehouse da segunda geração no IBM Netezza Data Warehouse Appliance. Por exemplo, o algoritmo de edição de distância Levenshtein não tem valor em aplicativos que variam desde a limpeza dos dados de nome e endereço para campanhas de marketing até análises complexas de texto realizadas por órgãos de segurança nacional, conforme examinam detalhadamente os nomes de pessoas de interesse listadas na “porta dos dados de entrada”. As organizações que seguem o conselho da Oracle e gerenciam dados espaciais, de texto, imagem e áudio em IOTs, perceberão que seus aplicativos têm um desempenho abaixo do ideal, considerando que o Exadata não é capaz de trazer a eficiência do processamento paralelo massivo para seus dados.

Os sistemas de data warehouse de primeira geração, que frequentemente executam como bancos de dados OLTP em servidores SMP, simplesmente carecem de um processamento eficiente para executar todas as consultas que os negócios precisam. O ODS (data warehouse operacionais) representa uma estrutura adotada por arquitetos para solucionar esse problema. O ODS é regularmente atualizado por diversos sistemas operacionais, integrando dados e fornecendo serviços transacionais e analíticos aos aplicativos. Os sistemas de data warehouse de segunda geração que operam na arquitetura MPP oferecem às organizações a oportunidade de consolidar os ODSs no data warehouse. O Exadata não consegue trazer a eficiência da varredura inteligente aos blocos que contêm uma única transação ativa (INSERÇÃO, ATUALIZAÇÃO, EXCLUSÃO), e um bloco do Exadata pode conter diversos registros. Essa peculiaridade de arquitetura da Oracle gera complexidade para os clientes que consideram o Exadata um ODS.

Embora a Oracle forneça 511 funções na última versão de seus bancos de dados, apenas 319 dessas funções podem ser processadas na camada de armazenamento MPP do Exadata. A condicionalidade da varredura inteligente abrange os atributos de uma função. Por exemplo, o Exadata não irá explorar as funções de data/hora usando o atributo `months_between(d,sysdate) = 0`, `months_between(d,current_date) = 0` ou `months_between(d,to_date('01-01-2010','DD-MM-YYYY')) = 0`<sup>3</sup>.

<sup>3</sup> <http://antognini.ch/2010/08/exadata-storage-server-and-the-query-optimizer-%E2%80%93-part-4/>

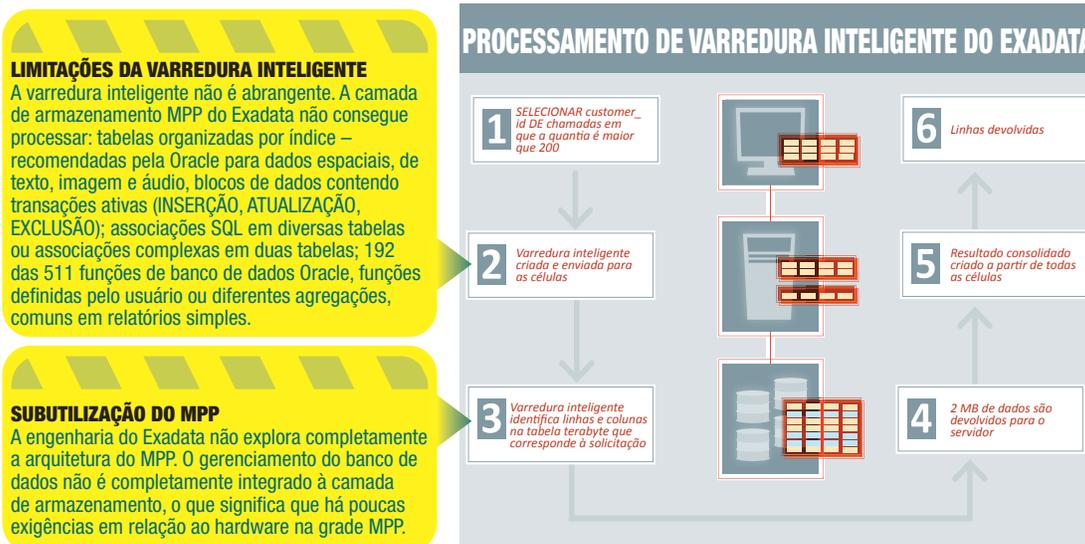
A camada de data warehouse do Exadata fornece filtros Bloom para implementar simples junções entre uma tabela maior e uma menor. Qualquer coisa mais complexa que isso não poderá ser processada no MPP. Consultas analíticas frequentemente exigem associações mais complexas que aquelas suportadas pelo Exadata. Considere o caso real de um fornecedor internacional que precisa de um insight sobre o valor do dólar nas vendas de lojas localizadas no Reino Unido. Uma simples consulta SQL exige uma associação entre três tabelas: vendas, moeda e lojas.

```
seleciona sum(sales_value * exchange_rate) us_dollar_sales
em vendas, moeda, lojas
em que sales.day = currency.day
e stores.country = 'UK'
e currency.country = 'USA'.
```

O Exadata não consegue processar essa associação de três tabelas em sua camada MPP e, em vez disso, precisará mover de maneira ineficiente todos os dados necessários para os cálculos localizados na rede para o Oracle RAC.

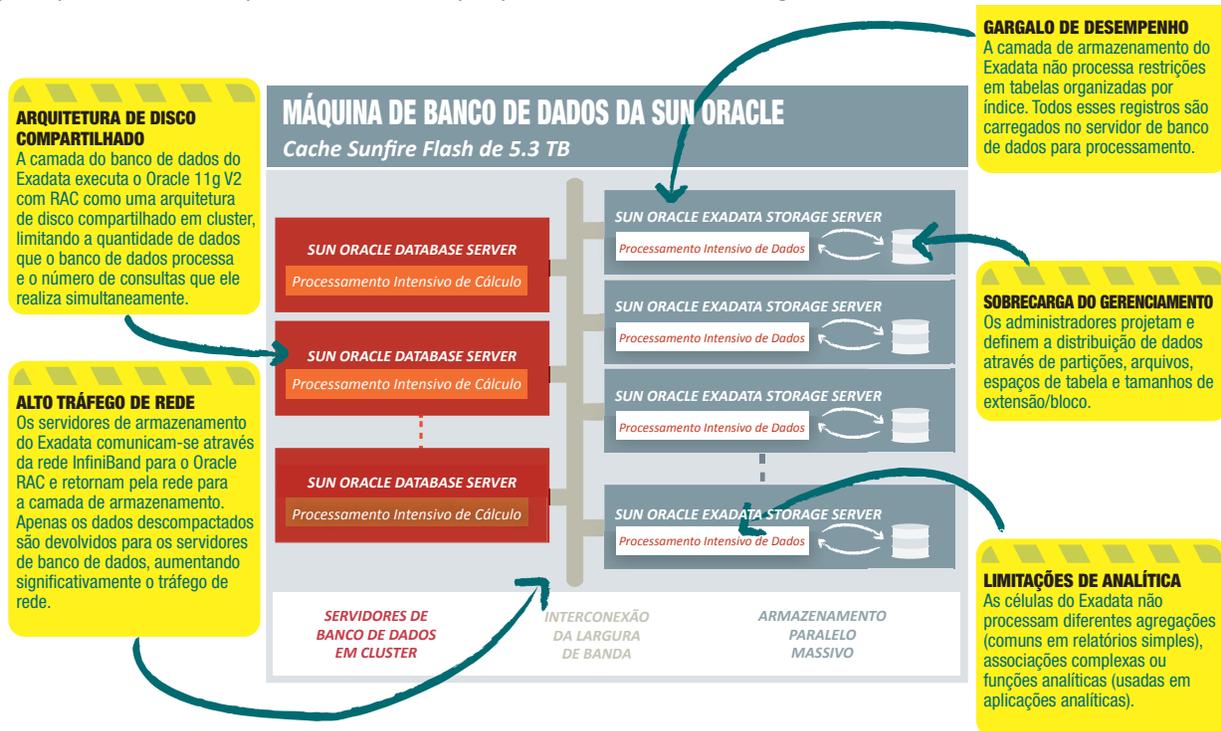
As UDFs (funções definidas pelo usuário) desempenham um papel central no data warehouse de segunda geração. As UDFs ampliam a funcionalidade de consulta para além do SQL e possibilitam o processamento analítico integrado. O Exadata não consegue processar as UDFs no MPP, mas deve transmitir a tarefa de processamento para o Oracle RAC, um banco de dados OLTP que não foi projetado para e não é capaz de analisar grandes conjuntos de dados.

A camada de armazenamento do Exadata não consegue processar diferentes agregações, que são comuns até mesmo em relatórios simples, e é amplamente usada nas consultas analíticas.



Os servidores de armazenamento do Exadata não podem se comunicar entre si. Em vez disso, toda a comunicação é realizada através da rede InfiniBand para o Oracle RAC e, em seguida, retornada pela rede para a camada de data warehouse. A arquitetura é benéfica para o processamento de Transação on-line, em que cada Transação, com um escopo de um ou alguns registros, pode ser atendida quando movida de um pequeno conjunto de dados do data warehouse para o banco de dados. A análise de conjuntos de dados muito grandes obriga o Exadata a mover esses grandes volumes entre suas camadas de data warehouse e banco de dados, afetando assim o desempenho da consulta de maneira negativa.

A Oracle posiciona sua utilização da InfiniBand de comutação de 40 Gb/s como uma vantagem em relação ao IBM Netezza Data Warehouse Appliance. Na realidade, o Exadata precisa dessa rede dispendiosa porque o sistema é ineficiente e desequilibrado. Os servidores de armazenamento do Exadata trabalham muito pouco. Desse modo, são colocados mais dados na rede do que seria necessário para que sejam recebidos para processamento pelo Oracle RAC, que por sua vez, realiza um grande trabalho.



Cada disco na camada de data warehouse do Exadata é compartilhado por todos os nós na grade que executa o Oracle RAC. Esse armazenamento comum gera um risco de que uma página seja lida por um nó enquanto está sendo atualizada por outro. Para lidar com isso, a Oracle exige a coordenação entre os nós. Cada nó verifica a atividade do disco de seus peers para evitar conflitos. Os técnicos da Oracle chamam essa atividade de ping de bloco. Há desperdício de ciclos de cálculo consumidos enquanto cada nó verifica a atividade de disco de seus peers ou perdas, enquanto um nó espera ocioso para que outro nó conclua uma Operação. Por serem úteis para o OLTP, em um banco de dados especialmente projetado para data warehousing, esses ciclos não seriam desperdiçados, mas sim utilizados em consultas de processamento, explorações de dados e execuções de análises.

Considere o caso de uma organização interessada na análise de dados de comércio de ações para identificar correlações em um setor do mercado de ações. Seu algoritmo calcula o coeficiente de correlação de Spearman ( $\rho$  de Spearman), medindo a dependência estatística entre duas variáveis pela avaliação de como o relacionamento entre elas pode ser descrito. Essa análise gera um insight valioso relacionado ao fato de ações específicas influenciarem ou não o comportamento de outras ações no mesmo setor do mercado dentro de uma janela de um a dez minutos. O conjunto de dados é amplo: uma longa história sobre dados de mercado e comércio diário, a ser analisada em relação a seu registro de histórico enquanto o mercado se movimenta. A análise exige uma associação Cartesiana entre todas as ações no setor enquanto calcula um Preço Médio Ponderado por Volume e um valor de Retorno de Fechamento Anterior para a ação que está sendo investigada. Os resultados são transmitidos para a função do coeficiente de correlação de Spearman para calcular a Covariância Populacional e o desvio padrão de todas as combinações de ações para o período. Depois de várias horas realizando testes, o cliente ainda estava esperando por um resultado do Oracle 10g, cujo desempenho precário o excluiu como tecnologia em um caso de negócios com uma janela máxima de 10 minutos.

No Exadata, o Oracle 11g R2 é correlacionado com um sistema de armazenamento MPP e uma rede rápida. No entanto, a varredura inteligente não consegue processar associações Cartesianas, a primeira etapa desse processo analítico, e não agrega nenhum valor a esse cálculo. O Exadata colocará todos os registros na rede e os enviará pelo Oracle RAC. A pergunta é: qual o nível de eficiência do Oracle RAC em comparação ao Oracle 10g? No processamento de algoritmos, a Oracle deve criar e gerenciar conjuntos de dados temporários e gravá-los fora da memória para armazenamento. O cache em flash do Exadata pode desempenhar alguma função aqui, mas o tamanho dos conjuntos de dados e a complexidade dos algoritmos farão com que os processos de banco de dados sejam gravados no disco. Esse fluxo do Oracle RAC é retornado pela rede ainda impedida com os dados enviados pela camada de armazenamento MPP, que estão em fila e não processados, esperando pela atenção do Oracle RAC totalmente ocupado. A conexão de rede do Exadata entre os servidores e o disco compartilhado é um gargalo. O Oracle RAC, projetado como um banco de dados OLTP, é um gargalo.

# **Para tudo, exceto consultas simples o Exadata deve mover grandes conjuntos de dados de sua camada de armazenamento para sua camada de banco de dados, aumentando as dúvidas em relação a sua adequação como uma plataforma para um data warehouse moderno.**

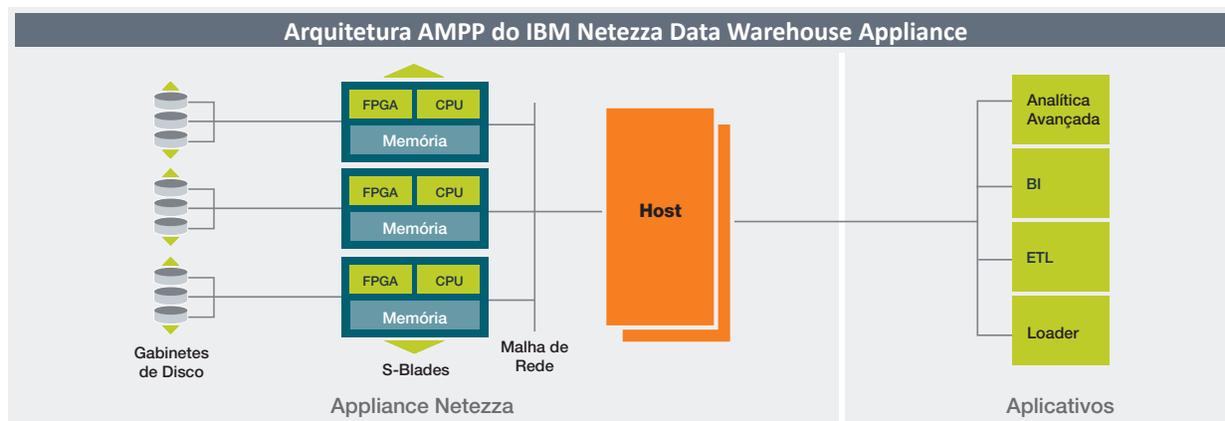
Para o data warehousing, a camada de armazenamento MPP do Exadata é limitada principalmente a execução de filtros de nível determinado (restrições de nível de linha com base na cláusula “onde” da instrução SQL e projeções de nível de coluna com base na cláusula “selecionar”) e fornecimento de energia de processamento de CPU para descompactação de dados armazenados no disco conforme é realizada a varredura. Além disso, o processamento na camada MPP do Exadata é condicional, obrigando os administradores a explorarem mais detalhadamente a complexidade da máquina. Todas as associações complexas, agregações e lógicas de filtragem restantes (o trabalho “pesado” da análise de dados e warehousing) executam então os resultados de todos os nós do Exadata nos nós principais do SMP em cluster que estão sendo executados no Oracle 11g RAC. O Exadata apresenta oportunidades interessantes para CIOs que buscam consolidar diversos sistemas OLTP em uma única plataforma. Para tudo, exceto consultas simples, o Exadata deve mover grandes conjuntos de dados de sua camada de armazenamento para sua camada de banco de dados, aumentando as dúvidas em relação a sua adequação como uma plataforma para um data warehouse moderno.

## **Desempenho de Consulta com o IBM Netezza Data Warehouse Appliance**

O IBM Netezza Data Warehouse Appliance foi projetado do começo ao fim como uma plataforma de data warehousing. A Netezza emprega uma arquitetura AMPP (Processamento Paralelo Assimétrico Massivo). Um host<sup>4</sup> de Multiprocessamento Simétrico enfrenta uma grade de nós de Processamento Paralelo Massivo. O IBM Netezza Data Warehouse Appliance explora essa grade MPP para processar o trabalho “pesado” dos dados de análise e warehousing.

Um nó na grade do IBM Netezza Data Warehouse Appliance é chamado de S-Blade (Snippet-Blade), um servidor independente que contém CPUs (unidades de processamento central) de vários núcleos. Cada CPU é equipada com um FPGA (Field Programmable Gate Array) de vários mecanismos e gigabytes de memória de acesso aleatório. Como as CPUs possuem suas próprias memórias, elas permanecem concentradas exclusivamente na análise dos dados, e nunca são desviadas para rastrear atividades de outros nós, como acontece com o ping de bloco no Oracle RAC.

<sup>4</sup> O IBM Netezza Data Warehouse Appliance possui dois hosts SMP para redundância, mas apenas um é ativado de cada vez.



Um FPGA é um chip semiconductor equipado com uma grande quantidade de portas internas programáveis para implementar praticamente qualquer função lógica, e particularmente eficiente no gerenciamento de tarefas de processamento de fluxo. Fora da Netezza, os FPGAs são usados em aplicativos como processamento de sinal digital, imagens médicas e reconhecimento de fala. Os engenheiros da Netezza desenvolveram máquinas de software dentro dos FPGAs de nossos appliances para acelerar o processamento de dados antes que ele atinja a CPU. Dentro de cada rack do Exadata, a Oracle dedica 14 servidores de armazenamento de oito passagens para realizar menos do que a Netezza com 48 FPGAs integrados aos nossos servidores blade. Cada FPGA (apenas um quadrado de 1x1 de silício) realiza seu trabalho com uma eficiência enorme, consumindo pouca energia e gerando pouco calor.

**A Netezza não se baseou em um sistema antigo com falhas conhecidas e o reformulou-o com uma nova camada de data warehouse inteligente.**

**O IBM Netezza Data Warehouse Appliance foi projetado como uma plataforma otimizada para data warehousing.**

A comunicação entre nós na grade MPP da Netezza ocorre em uma malha de rede que opera em um protocolo customizado com base em IP, utilizando completamente a largura de banda entre seções e eliminando o congestionamento, mesmo em contínuos tráfegos de rede intermitente. A rede é otimizada para escalar mais de mil nós, enquanto permite que cada nó copie grandes transferências de dados para cada nó simultaneamente. Essas transferências possibilitam uma grande eficiência às tarefas de processamento típicas de data warehousing e analítica avançada. Assim como as instruções SQL beneficiam-se do processamento na arquitetura MPP do IBM Netezza Data Warehouse Appliance, os algoritmos complexos de cálculo também são beneficiados no núcleo da analítica avançada. Gerações anteriores de tecnologias separam fisicamente o processamento de aplicativos do processamento de banco de dados, apresentando ineficiências e restrições enquanto grandes conjuntos de dados são retirados e devolvidos do data warehouse para as plataformas de processamento analítico. A Netezza traz os cálculos complexos da analítica avançada para sua grade MPP, executando os algoritmos em cada CPU de maneira física próxima aos dados, fazendo com que a movimentação de dados seja redundante e elevando o desempenho. Os algoritmos beneficiam-se da execução em diversos nós da grade MPP da Netezza, livres das restrições impostas nos sistemas em cluster menos escaláveis.

O IBM Netezza Data Warehouse Appliance demonstra como uma verdadeira arquitetura MPP se sobressai nos cálculos do coeficiente de correlação de Spearman. Uma verdadeira carga de trabalho e um verdadeiro conjunto de dados. O cliente carrega um grande volume de dados de troca no IBM Netezza Data Warehouse Appliance e transmite constantemente os dados de fluxo dos mercados ativos para o data warehouse. O IBM Netezza Data Warehouse Appliance executa cada etapa da consulta em paralelo utilizando todos os seus recursos de hardware e software. O armazenamento inteligente da Netezza seleciona apenas as linhas necessárias para aquele setor de mercado e projeta apenas as colunas necessárias para a avaliação. O resultado da associação é diretamente enviado à implementação do código da análise estatística, que o IBM Netezza Data Warehouse Appliance transfere por download para cada processador em sua grade MPP, executando em paralelo os cálculos complexos. Os resultados de cada nó na grade MPP são retornados pela rede ao host para a coleta final, e devolvidos ao aplicativo que os solicitou. O IBM Netezza Data Warehouse Appliance conclui a análise em alguns minutos, isto é, dentro da janela de 10 minutos da organização, e depois executa regularmente pelo tempo que o mercado estiver aberto e enquanto a organização quiser avaliar o comportamento de diferentes ações em mercados dinâmicos. Além do setor de serviços financeiros, o coeficiente de correlação de Spearman é útil para estatísticos que trabalham em econometria, controle de qualidade, bioinformática e marketing.

A Netezza não se baseou em um sistema antigo com falhas conhecidas e o reformulou-o com uma nova camada de armazenamento inteligente. O IBM Netezza Data Warehouse Appliance foi projetado como uma plataforma otimizada para data warehousing. O IBM Netezza Data Warehouse Appliance executa incondicionalmente todas as consultas em sua grade paralela massiva, proporciona um desempenho de maneira generosa e facilita a vida dos programadores, administradores e usuários.

Um cliente da Netezza, do setor de serviços financeiros, usou a abordagem Lean para analisar a quantidade de recursos necessária para gerenciar seu data warehouse Oracle. Ele aprendeu ao criar e manter índices, agregações, visualizações materializadas e data marts **que mais de 90% do trabalho de sua equipe de TI era considerado desperdício ou processamento sem valor agregado.**

## Simplicidade de Operação com o Oracle Exadata

A Oracle confia em administradores de banco de dados qualificados, com experiência e treinados para projetar bancos de dados para fornecer desempenho às consultas. Uma técnica de projeto é o particionamento físico de dados, o que significa que as consultas podem ser restritas às partições conhecidas para conter dados relevantes. As partições são uma herança das raízes do OLTP da Oracle. No data warehouse de segunda geração, os administradores não têm como saber quais consultas e análises serão executadas no dia seguinte ou na próxima semana. O Exadata inclui o Gerenciamento de Armazenamento Automático da Oracle, que automatiza a divisão das partições em todos os discos disponíveis. No entanto, uma equipe de administração ainda precisa criar partições, configurar e gerenciar grupos de discos para data warehouse compartilhado em instâncias, selecionar e implementar espelhamento de duas ou três passagens e configurar os tamanhos das Unidades de Alocação. Além disso, a configuração do Exadata exige que os administradores criem e gerenciem espaços de tabela, espaços de índice, espaços temporários e extensões. Conforme observado pelo analista do setor, Curt Monash, “Melhor ainda seria um sistema que não dependesse em grande parte do particionamento complexo para obter um bom desempenho”.<sup>5</sup>

<sup>5</sup> Curt Monash em <http://www.dbms2.com/2009/09/21/notes-on-the-oracle-database-11g-release-2-white-paper/>

Conforme descrito na seção anterior, independentemente de o Exadata processar dados condicionalmente em sua camada de data warehouse MPP, essa complicação exige que os administradores, mesmo aqueles com experiência em versões anteriores do Oracle, participem de uma série de aulas de orientação. A Oracle recomenda: “Certifique-se de que seus DBAs tenham um conhecimento atualizado (11g R2). Além do conhecimento em Exadata”. A obtenção de conhecimento atualizado exige que cada administrador participe de pelo menos de quatro cursos, totalizando 18 dias. Os administradores novos em Oracle RAC precisam estudar por mais cinco dias para aprender as complexidades de sua administração.

Um cliente da Netezza, do setor de serviços financeiros, usou a abordagem Lean<sup>6</sup> para analisar a quantidade de recursos necessária para gerenciar seu data warehouse Oracle. Ele aprendeu ao criar e manter índices, agregações, visualizações materializadas e data marts que mais de 90% do trabalho de sua equipe de TI era considerado desperdício ou processamento sem valor agregado. O custo desse desperdício se converte em custos de licença de hardware e software desnecessários, terabytes de data warehouse desperdiçados, ciclos de carregamento de dados e desenvolvimentos prolongados, longos períodos de indisponibilidade de dados, dados antiquados, carregamentos e consultas com desempenho precário e custos administrativos excessivos.

O Exadata faz muito pouco para simplificar o gerenciamento de um data warehouse Oracle. Os administradores devem gerenciar diversas camadas de servidores, cada uma com imagens de sistema operacional, firmware, sistemas de arquivo e software a serem mantidos. A Oracle sugere que os DBAs esperam gastar 26% menos tempo com o gerenciamento do 11g<sup>7</sup> (a versão do banco de dados no Exadata) do que gastam com implementações mais antigas do 10g. Se isso for confirmado na prática e o Exadata reduzir em 1/4 o tempo que os clientes desperdiçam com administração inútil, a Oracle terá dado um passo na direção certa. Os appliances da Netezza foram projetados para não desperdiçar o tempo dos clientes. “A equipe de DBA apenas auxilia o ambiente e gerencia o modelo de segurança de alto nível para o appliance. E é isso. Eles não precisam fazer mais nada (por exemplo, o conceito de indexar é estranho para eles quando se trata da Netezza).”<sup>8</sup>

<sup>6</sup> Com raízes na fabricação, “Lean” é uma prática que usa ferramentas e técnicas do Six Sigma para analisar os gastos desnecessários de recursos e identificar atividades que não agregam qualquer valor ao produto ou ao serviço para serem eliminadas.

<sup>7</sup> <http://www.dbms2.com/2009/09/21/notes-on-the-oracle-database-11g-release-2-white-paper/>

<sup>8</sup> Cliente usando Oracle para OLTP e Netezza para data warehousing mencionado no fórum do Linked-In Exadata Vs Netezza em [http://www.linkedin.com/groupAnswers?viewQuestionAndAnswers=&gid=2602952&discussionID=11385070&sid=1275353329699&t\\_rk=ug\\_qa\\_q&goback=.ana\\_2602952\\_1275353329699\\_3\\_](http://www.linkedin.com/groupAnswers?viewQuestionAndAnswers=&gid=2602952&discussionID=11385070&sid=1275353329699&t_rk=ug_qa_q&goback=.ana_2602952_1275353329699_3_)

Os usuários de negócios não exigem apenas que suas consultas sejam concluídas rapidamente, mas também esperam um desempenho consistente. Um relatório que foi concluído em cinco segundos ontem e em três minutos hoje provavelmente gerará um chamado solicitando resposta da equipe de helpdesk de TI. O data warehouse está inevitavelmente sujeitos às demandas de cargas de trabalho dinâmicas e variadas. Os dados que chegam de sistemas OLTP através de tarefas em lote ou feeds de fluxo são carregados, as tarefas administrativas como backup, restauração e limpeza que são executadas no plano de fundo, fora da visão dos negócios e painéis, estão em constante atualização. Ao mesmo tempo, os aplicativos com uso intensivo de cálculos, como aqueles que fazem a previsão de quais solicitações ou trocas podem ser fraudulentas ou irregulares, criam uma carga “pesada” e repentina na infraestrutura do data warehouse. O fornecimento de desempenho consistente aos negócios faz com que sejam necessários dois requisitos do data warehouse: desempenho de consulta consistente e gerenciamento de carga de trabalho eficaz, o que simplifica a alocação da energia de cálculo disponível para todas as tarefas que precisam desse serviço, geralmente com base nas prioridades acordada com os negócios.

Para gerenciar cargas de trabalho no Oracle é necessário que os administradores aprendam diversos parâmetros de ajuste, muitos deles com um alto grau de dependência dos outros. A situação é mais complicada no Exadata, em que determinados parâmetros devem ser definidos com o mesmo valor para todos os processadores na grade. Mais do que um sistema de controle descendente, os administradores devem trabalhar com configurações de parâmetro individual em alterações experimentais de maneira ascendente e observar o efeito em um ambiente de produção. A filosofia da Oracle em relação ao gerenciamento de carga de trabalho tem êxito para o processamento de transações on-line, em que a carga que cada aplicativo coloca nos bancos de dados e servidores, pode ser descrita em um ambiente de desenvolvimento. Os efeitos das configurações de parâmetros são anotados e compreendidos e os ajustes incrementais são realizados. Depois que um sistema estável for obtido, as configurações de parâmetro poderão ser colocadas em prática.

Não existe “carga normal” nos armazéns de segunda geração. Os usuários de negócios irão liberar análises que exigem cálculos em conjuntos de dados massivos conforme os eventos ocorrem no mundo exterior. Os analistas não irão esperar por um momento calmo enquanto outras cargas diminuem. A natureza dinâmica do data warehouse expõe a fragilidade da abordagem de ajuste da Oracle para o gerenciamento de cargas de trabalho. A obtenção e manutenção de um desempenho consistente para grandes comunidades de usuários, com diferentes aplicativos e requisitos de dados, por meio de cargas aumentadas e reduzidas, é uma tarefa complexa. Mesmo uma equipe de administradores da Oracle com experiência e altamente treinada precisará se esforçar para manter um data warehouse estável capaz de fornecer desempenho consistente a tarefas de carga e relatórios regulares enquanto deixa espaço para variações inesperadas, mas importante, nas atividades do analista.

**“O modo como fizemos uma prova de conceito com eles [Netezza] foi da seguinte maneira: eles nos enviaram uma caixa, nós a colocamos em nosso datacenter e a conectamos à nossa rede. Dentro de 24 horas, estávamos em funcionamento. Eu não estou exagerando. Foi fácil assim”.**

O Oracle RAC é uma tecnologia complexa e seus parâmetros de ajuste são misteriosos. A complexidade do Exadata é confirmada pela quantidade de dias de treinamento que a Oracle recomenda mesmo para os administradores experientes em versões anteriores de seus bancos de dados.

#### **Simplicidade de Operação com o IBM Netezza Data Warehouse Appliance**

Os clientes da Netezza confirmam prontamente em registro público que nossos appliances são simples de instalar e usar. “O modo como fizemos uma prova de conceito com eles [Netezza] foi da seguinte maneira: eles nos enviaram uma caixa, nós a colocamos em nosso datacenter e a conectamos à nossa rede”, disse ele. “Dentro de 24 horas, estávamos em funcionamento. Eu não estou exagerando. Foi fácil assim”<sup>9</sup>. Esse comentário é de Joseph Essas, Vice-Presidente de Tecnologia na eHarmony, Inc., uma empresa que já está usando o banco de dados da Oracle e o software RAC.

Reduzir o tempo para obter produtividade é um bom começo. A filosofia da Netezza é trazer simplicidade a todas as fases do data warehousing. A primeira tarefa que um cliente enfrenta é carregar seus dados. O IBM Netezza Data Warehouse Appliance automatiza a distribuição de dados. O conhecimento de projetos de prova de conceito é que os clientes carregam seus dados para a Netezza usando distribuição automática, executam suas consultas e comparam os resultados com seus ambientes da Oracle altamente receptivos. Para tudo, exceto as consultas mais simples, a distribuição automática é boa o suficiente para que o IBM Netezza Data Warehouse Appliance supere a Oracle. Posteriormente, os clientes poderão analisar todas as suas consultas para identificar aquelas que podem ser aceleradas pela redistribuição de dados em diferentes chaves. O IBM Netezza Data Warehouse Appliance facilita essa tarefa.

<sup>9</sup> <http://www.itworld.com/software/61575/eharmony-finds-data-warehouse-match-netezza>

## Há algo a ser dito para uma abordagem

- SEM ajuste/monitoramento de interconexão de cluster (GES & GCS)
- SEM ajuste/conhecimento específico em RAC (DBAs com experiência em RAC são menos do que uma conveniência)
- SEM dimensionamento e configuração de dbspace/tablespace
- SEM refazer/configurar e definir o tamanho de log físico
- SEM registros/configuração e dimensionamento de log lógico
- SEM configuração e dimensionamento de página/blocos para tabelas
- SEM configuração e dimensionamento de extensão para tabelas
- SEM alocação e monitoramento de espaço temporário
- SEM integração das recomendações do kernel de OS
- SEM manutenção dos níveis de correção recomendados do OS
- SEM sessões JAD para configurar host/rede/armazenamento
- SEM dicas de consulta (por exemplo, first\_rows) e otimizadores (por exemplo, optimizer\_index\_cost\_adj)
- SEM pacote de estatísticas (estatísticas, acertos do cache, monitoramento de evento de espera)
- SEM ajustes de memória (SGA, buffers de bloco, etc.)
- SEM planejamento/criação/manutenção de índices
- Estratégias de particionamento simples: HASH ou ROUND ROBIN

Todas as consultas enviadas ao IBM Netezza Data Warehouse Appliance são automaticamente processadas em sua grade paralela massiva sem qualquer envolvimento dos administradores do banco de dados. As consultas e análises entram no IBM Netezza Data Warehouse Appliance por meio da máquina host, na qual o otimizador, o compilador e o planejador as decompõem em diversas partes ou fragmentos diferentes e distribui essas instruções à grade MPP de nós de processamento, ou S-Blades. Todos esses, em seguida, processam suas cargas de trabalho simultaneamente em relação à suas fatias de dados gerenciadas localmente.

Um fragmento que chega a cada S-Blade do IBM Netezza Data Warehouse Appliance inicia a leitura dos dados compactados do disco para a memória. O FPGA então lê os dados dos buffers de memória e, utilizando seu Mecanismo de Compactação, os descompacta, transformando instantaneamente cada bloco do disco no equivalente a 4-8 blocos de dados dentro do FPGA. A engenharia da Netezza acelera o componente mais lento em qualquer data warehouse: o disco. Em seguida, dentro do FPGA os dados passam para o Mecanismo de Projeto, que filtra as colunas com base nos parâmetros especificados na cláusula SELEÇÃO da consulta SQL que está sendo processada. Apenas os registros que cumprem a cláusula SELEÇÃO são aprovados para serem enviados ao Mecanismo de Restrição, em que as linhas que não são necessárias para processar a consulta são bloqueadas para passar pelas portas, com base nas restrições especificadas na cláusula ONDE. O Mecanismo de Visibilidade mantém a conformidade ACID (Atomicidade, Consistência, Isolamento e Durabilidade) nas velocidades de fluxo. Todo esse trabalho, a limpeza constante de colunas e linhas desnecessárias, é obtida em um FPGA eficiente em termos energéticos, medindo somente 6,5 cm<sup>2</sup>, aproximadamente. Se o IBM Netezza Data Warehouse Appliance não precisar mover dados, eles não serão movidos.

Estando o pré-processamento do FPGA concluído, ele envia apenas o conjunto de registros resultante recortado de volta para a memória S-Blade, onde a CPU executará operações de banco de dados de nível mais alto, como classificações, junções e agregações, fazendo isso em paralelo com todas as outras CPUs na grade MPP. A CPU também pode aplicar algoritmos complexos integrados ao código do fragmento para processamento de analítica avançada. A CPU finalmente monta todos os resultados intermediários a partir da entrada do fluxo de dados e produz um resultado para o fragmento, enviado pela malha da rede para outros S-Blade ou para o host, conforme indicado pelo código do fragmento. Quando os dados solicitados por uma JUNÇÃO não são colocados em um nó, a malha de rede entre nós do IBM Netezza Data Warehouse Appliance redistribui de maneira eficiente e simples, posteriormente no ciclo de processamento, depois que o banco de dados concluiu as restrições e projeções. Alguns algoritmos altamente complexos exigem comunicação entre nós para calcular suas respostas. O IBM Netezza Data Warehouse Appliance explora a mensagem fazendo com que a interface comunique resultados temporários e produza o resultado final.

E, como os blocos de dados originais compactados ainda estão na memória, eles podem ser automaticamente reutilizados em consultas posteriores que exigem dados semelhantes por meio do cache de tabela do IBM Netezza Data Warehouse Appliance. Um mecanismo automatizado que não exige qualquer treinamento ou envolvimento do DBA.

# **Apenas três meses depois de mudar para a Netezza, Apenas três meses depois de mudar para a Netezza, um cliente relatou que sua equipe forneceu mais aplicativos analíticos do que puderam nos três anos anteriores com a Oracle.**

Como o IBM Netezza Data Warehouse Appliance aplica o paralelismo completo a todas as tarefas, seu sistema de gerenciamento de carga de trabalho desempenha uma função crítica no controle de quanto dos recursos de computação do aplicativo é disponibilizado para cada tarefa. Na arquitetura de appliance da Netezza, um componente de software controla todos os recursos do sistema: processadores, discos, memória, rede. Essa elegância é a base do Sistema de Gerenciamento de Carga de Trabalho do IBM Netezza Data Warehouse Appliance. O Sistema de Gerenciamento de Carga de Trabalho do IBM Netezza Data Warehouse Appliance simplifica para que os administradores aloquem recursos de cálculo para os usuários e grupos com base nas prioridades acordadas com a empresa e mantenham tempos de resposta consistentes para diversas comunidades.

O IBM Netezza Data Warehouse Appliance elimina o trabalho inútil de ajustar o banco de dados. Equipado para tomar suas próprias decisões inteligentes, os appliances da Netezza não exigem qualquer ajuste e pouca administração de sistema. As poucas tarefas administrativas necessárias para manter desempenho consistente em cargas de trabalho dinâmicas e em alteração estão ao alcance de qualquer pessoa com experiência em Linux e SQL. A única coisa necessária é que o administrador aloque os recursos do IBM Netezza Data Warehouse Appliance para os grupos dentro da comunidade do usuário e controle manualmente o Sistema de Gerenciamento de Carga de Trabalho. Livre de ciclos constantes de administração de banco de dados, a equipe técnica pode se envolver com os negócios para investigar novas maneiras de explorar dados que gerem valor. Apenas três meses depois de mudar para a Netezza, um cliente relatou que sua equipe forneceu mais aplicativos analíticos do que puderam nos três anos anteriores com a Oracle. Processar os aplicativos analíticos próximo ao local onde os dados são gerenciados, explorando a mesma plataforma MPP que é usada para processar SQL, representa uma oportunidade real para que as organizações aumentem drasticamente o valor que obtêm com os dados.

Com enormes volumes de dados e cargas de trabalho que frequentemente fazem uso intenso de cálculos e são complexas, os hardwares e softwares necessários para executar um data warehouse representam uma despesa significativa. A experiência de diversas organizações usando bancos de dados OLTP para criarem seus armazéns de primeira geração é que os custos de treinamento e contratação de equipe técnica especializada não apenas ultrapassa os custos de ativos, mas ocorre ano após ano. Esses custos sobrecarregam os projetos de data warehouse e prejudicam enormemente o calor que eles criam para a organização. Como a análise de desperdício conduzida pelo cliente de serviços financeiros tanto da Netezza quanto da Oracle é destacada, usar o Oracle para data warehousing exige muito trabalho. Os clientes da Netezza declaram que as tarefas de administração que exigem técnica são simplesmente desnecessárias. Sabendo-se disso, é incontestável que operar um banco de dados Oracle exige que os administradores passem a maior parte de seus tempos cuidando e alimentando a tecnologia subjacente, enquanto os clientes da Netezza passam a maior parte de seu tempo criando valor para explorar seus dados.

A nova camada de armazenamento do Exadata inclui outra camada de complexidade para que os administradores ajustem e gerenciem. Como o Exadata é muito novo e, por isso, apenas algumas soluções de data warehouse que usam a tecnologia estão em produção, as projeções relacionadas ao custo de propriedade são prematuras. No entanto, os clientes devem esperar que a obtenção de um alto e consistente desempenho do Exadata trará custos substanciais de treinamento, projeto de banco de dados e administração.

Enquanto a inclusão de uma nova camada de data warehouse reduz o gargalo de rendimento do disco para o banco de dados da Oracle, a engenharia do Exadata é mais adaptada ao processamento paralelo massivo do que à exploração completa da arquitetura. A falha da Oracle em integrar gerenciamento de dados completamente à camada de data warehouse do Exadata, significa que é exigido muito pouco do hardware em sua grade MPP. Isso faz com que o custo de aquisição do Exadata seja mais alto. Os clientes pagam pelo hardware que nunca será completamente explorado pelo software. Esses custos vão se acumulando durante o tempo de vida do data warehouse. Os clientes pagam pela utilização insuficiente do espaço em seus datacenters, o que poderia retornar um grande valor se fosse usado para armazenar um sistema de computação mais eficiente.

Avaliando os Sistemas		Netezza	Oracle
Item		IBM Netezza Data Warehouse Appliance 12	Exadata v2 (SAS)
Desempenho e Arquitetura	MPP	<ul style="list-style-type: none"> <li>• MPP Verdadeiro</li> <li>• Otimizado para Data Warehousing e Analítica</li> </ul>	<ul style="list-style-type: none"> <li>• Nós de armazenamento paralelos e híbridos e nós principais em cluster SMP</li> <li>• Otimizado para OLTP (processamento de transações)</li> </ul>
	Arquitetura de Hardware	<ul style="list-style-type: none"> <li>• S-Blades de processamento completo</li> <li>• (1 CPU principal + 1 FPGA principal / 1 unidade de disco)</li> <li>• Nó de host SMP usado principalmente para interface de aplicativos/usuários</li> <li>• Redistribuição independente entre blades</li> </ul>	<ul style="list-style-type: none"> <li>• Armazenamento inteligente</li> <li>• (1 CPU principal / 1,5 unidade de disco)</li> <li>• Nós de cluster SMP operando em Oracle 11g RAC</li> <li>• InfiniBand</li> <li>• (Nós do Exadata para cluster SMP)</li> <li>• Utilização do nó principal em todas as redistribuições de dados</li> </ul>
	Fluxo de Dados	<ul style="list-style-type: none"> <li>• Desempenho do FPGA auxilia na descompactação do S-Blade, filtragem de predicado, reforço da segurança de nível de linha</li> <li>• Mais de 95% do trabalho feito em S-Blades</li> </ul>	<ul style="list-style-type: none"> <li>• Nós do Exadata principalmente usados para descompactação e filtragem de predicado</li> <li>• A maioria do trabalho de Data Warehousing e Analítica realizado no nó principal do SMP</li> </ul>
	Analítica do Banco de Dados	<ul style="list-style-type: none"> <li>• Plataforma MPP completamente utilizada para analítica</li> <li>• Funções definidas pelo usuário, agregações e tabelas</li> <li>• Suporte às linguagens: C/C++, Java, Python, R, Fortran</li> <li>• Suporte ao paradigma SQL, Matriz, Grade, Hadoop</li> <li>• Conjunto integrado de mais de 50 analíticas principais (completamente paralelo)</li> <li>• Software livre: Suporte para bibliotecas GNU Scientific &amp; CRAN</li> <li>• Desenvolvimento Integrado Env.: Eclipse &amp; R GUI c/ assistentes</li> </ul>	<ul style="list-style-type: none"> <li>• Processamento de analítica limitado ao cluster de nó principal apenas</li> <li>• Função definida pelo usuário e agregados</li> <li>• Suporte às linguagens: C/C++, Java</li> <li>• Suporte ao paradigma SQL, Matriz (secundária)</li> <li>• Funções básicas de analítica</li> </ul>
	Escala	<ul style="list-style-type: none"> <li>• Desempenho linear e escalabilidade de dimensão de dados</li> <li>• Gerenciamento de carga de trabalho de classe corporativa, com recursos completos e outros recursos</li> </ul>	<ul style="list-style-type: none"> <li>• Desempenho não linear e escalamento de dimensão de dados, desempenho e gargalo de e/s no cluster de nó principal</li> </ul>
Simplicidade	Integração e Gerenciamento de Appliances	<ul style="list-style-type: none"> <li>• Sem ajustes, sem indexação, sem partições</li> <li>• Sistema balanceado desenvolvido para fornecer um desempenho com o melhor preço</li> </ul>	<ul style="list-style-type: none"> <li>• Alta dependência de desempenho de ajuste</li> <li>• O desempenho depende de habilidades de design de banco de dados físico, incluindo índices e partições</li> </ul>

Para compensar o desempenho precário do Oracle RAC como um banco de dados de analítica, o Exadata contém uma grande parte de hardware. A aquisição e a operação da máquina no datacenter são caras. A energia e o resfriamento de um único rack do Exadata consomem quase 60% a mais da energia do que um único rack do IBM Netezza Data Warehouse Appliance.

Como os custos destroem o valor, é uma questão fundamental saber se o Exadata ajuda os clientes a criarem valor. Os armazéns de dados de primeira geração desempenham uma função importante ao manter a organização informada do passado recente, e ainda assim os dados liberam um grande potencial por meio de analítica avançada e outros recursos de armazéns de segunda geração, discutidos anteriormente neste documento. O Oracle RAC equipado com o data warehouse tradicional não comprovou ser um sucesso nessa função de dados. A camada de armazenamento do Exadata não consegue processar junções complexas, diferentes agregações e funções analíticas. É difícil visualizar como duas tecnologias, individualmente incapazes para analisar detalhadamente conjuntos de dados muito grandes com um alto desempenho, conseguirão fazê-lo ao se conectarem por meio de uma rede rápida e integrada no mesmo rack.

### **Valor com o IBM Netezza Data Warehouse Appliance**

Os engenheiros da Netezza integram o gerenciamento de dados e análise profunda com grades não compartilhadas paralelas massivas. Um resultado dessa inovação é a simplicidade para nossos clientes, o que se traduz diretamente em custos incrivelmente reduzidos de propriedade e operação dos armazéns de dados, como não seria possível com produtos de banco de dados orientados a OLTP, como os da Oracle.

As demandas de armazéns de dados mudaram para além do simples processamento SQL para uma exploração completa dos dados, o que exige que o data warehouse seja capaz de executar análises espaciais, modelos de previsão, gráficos investigativos e outros aplicativos analíticos. Para ilustrar, uma organização que deseja avaliar o risco criado por catástrofes, como um incêndio e uma inundação e um grande portfólio de propriedades com seguro desenvolve um aplicativo usando o Oracle Spatial e Oracle 10g para identificar ativos de seguro em áreas propensas a riscos. Analisar mais de 4 milhões de locais exige junções espaciais e o processamento de algoritmos espaciais que fazem uso intensivo de cálculos. Avaliar o risco de incêndio exige 43 bilhões de cálculos em 10 mil polígonos, enquanto a avaliação do risco de inundações exige 30 bilhões de cálculos em 7 mil polígonos. A Oracle levaria mais de uma hora para concluir cada avaliação. Os ventos mudam de direção de maneira inesperada e intensa. E as cheias surgem. O comportamento dinâmico de riscos naturais indica que perfis de riscos podem mudar incrivelmente dentro de uma hora. A empresa precisa ter maior agilidade para que possa analisar a situação sempre que achar necessário para seus negócios. De maneira diferente do Oracle 10g ou do Exadata, a Netezza executa junções espaciais e algoritmos que fazem uso intensivo de cálculos de análise espacial em sua grade de processamento paralelo massivo. A Netezza conclui a avaliação de risco de incêndio em 1,6 segundo e a avaliação de risco de enchente em 4,4 segundos, dando à empresa a agilidade para obter valor total de seus dados.

O IBM Netezza Data Warehouse Appliance foi criado a partir do zero para o processamento paralelo massivo de SQL e aplicativos de análise avançada. A Netezza deixa as organizações livres de linguagens proprietárias. Os parceiros e clientes podem possuir aplicativos existentes para o IBM Netezza Data Warehouse Appliance ou optar por desenvolver novos aplicativos analíticos na linguagem de sua escolha, incluindo C++, Java, Python, R e Fortran. As organizações que usam C podem aproveitar mais de mil funções analíticas disponíveis como software livre na biblioteca GNU Scientific.<sup>10</sup> As organizações que usam R também podem usar mais de 2 mil pacotes publicamente disponíveis no CRAN (Rede de Archive R Abrangente). Além disso, os parceiros e clientes da Netezza podem optar por trabalhar com MapReduce / Hadoop como um mecanismo de ingestão altamente escalável para pré-processar grandes conjuntos de dados gerados por aplicativos da Web públicos e logs da Web antes que sejam carregados no IBM Netezza Data Warehouse Appliance para análise.

<sup>10</sup> [www.gnu.org/software/gsl/](http://www.gnu.org/software/gsl/)

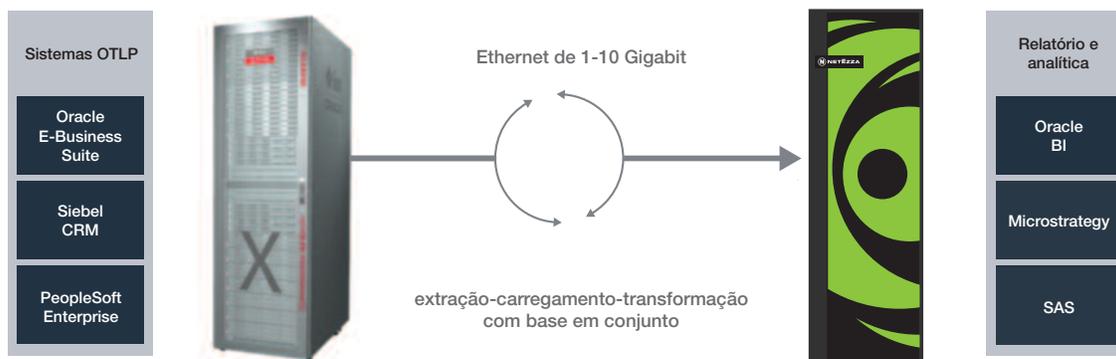
# 6

## Combinando o Exadata com o IBM Netezza Data Warehouse Appliance

A seção 2 discute as diferenças entre o processamento de transação on-line e o data warehousing. As otimizações de OLTP, estratégias de cache de dados e ajustes de desempenho têm pouca relevância para o data warehousing. Com seu banco de dados projetado para OLTP, o Exadata oferece aos CIOs a oportunidade de consolidar diversos sistemas transacionais separados, cada um com seu próprio hardware e banco de dados Oracle, em um único ambiente gerenciável.

Essa consolidação provavelmente trará oportunidades para reduzir os custos e melhorar o desempenho dos estágios ETL (extração-transformação-carregamento), usados para alimentar o data warehouse de sistemas de origem. Conforme incluem mais fontes aos seus armazéns de primeira geração, algumas organizações criam inadvertidamente uma rede complexa de ETL. Muitos dos clientes da Oracle criam seus próprios aplicativos ETL usando o banco de dados Oracle com PL/SQL e outras sintaxes e interfaces proprietárias a esse sistema de banco de dados. Enquanto o processamento com base em linha de PL/SQL pode ter sido rápido o suficiente para transformar volumes de dados relativamente baixos, típicos de armazéns de primeira geração, os grandes conjuntos de dados comuns em armazéns de segunda geração exigem um desempenho maior disponível no processamento com base em conjunto. O software GoldenGate da Oracle e outros produtos de fornecedores especialistas de ETL oferecem esse recurso. Embora estejam além do escopo deste documento, o Netezza Migrator e o Netezza Data Virtualizer também podem contribuir para o sucesso dos projetos de reengenharia de ETL e para acelerar o desempenho da consulta.

O Exadata, executando em diversos aplicativos Oracle, transmite os dados do IBM Netezza Data Warehouse Appliance para Análise



A Netezza surgiu como a principal alternativa ao Oracle para data warehousing. Mover data warehouse e data marts do Oracle para a Netezza cria novas oportunidades, e não riscos. A maioria dos clientes da Netezza já trilhou esse caminho, e muitos deles fazendo parcerias com as empresas de integração de sistemas com grande histórico de migrações bem-sucedidas.

O Exadata é uma evolução da plataforma OLTP da Oracle. O sistema de gerenciamento de banco de dados da Oracle foi criado para OLTP, em que volumes de dados são relativamente modestos, se comparados aos armazéns de dados. A atividade do banco de dados para um sistema OLTP pode ser avaliada antes que seja colocado em produção. Os administradores têm tempo para projetar, testar e otimizar cada recuperação de dados da transação. As soluções de data warehouse devem processar imediatamente qualquer consulta que os negócios precisarem para solicitar seus dados. As tecnologias que exigem uma mediação do administrador não são adequadas à tarefa. Solicitar essa tecnologia em uma função que não seja o processamento de transações causa uma enorme tensão nas pessoas e processos equipados para gerenciar e operar um data warehouse.

***“Esse [Netezza] é o primeiro produto de data warehouse com um roadmap de produto de longo prazo que se alinha perfeitamente ao seu próprio roadmap. Nós chamamos isso de banco de dados sob demanda”.***

***– Steve Hirsch, Chief Data Officer na NYSE Euronext***

A Oracle orienta os clientes de que o Exadata possui uma arquitetura semelhante à Netezza, mas melhor, porque o IBM Netezza Data Warehouse Appliance não suporta todos os tipos de dados ou padrões SQL e também não suporta a mineração de dados ou uma alta simultaneidade. Os clientes da Netezza discordam: “Esse [Netezza] é o primeiro produto de data warehouse com um roadmap de produto de longo prazo que se alinha perfeitamente ao seu próprio roadmap. Chamamos isso de banco de dados sob demanda”<sup>11</sup>, disse Steve Hirsch, Chief Data Officer na NYSE Euronext.

<sup>11</sup> [www.netezza.com/customers/nyse-euronext-video.aspx](http://www.netezza.com/customers/nyse-euronext-video.aspx)

Devido às diferentes características de carga de trabalho, alguns clientes tentam executar OLTP e sistemas de data warehouse na mesma infraestrutura. Fazer isso exige ajustes e otimização constantes. Os técnicos são colocados em uma situação difícil: ou eles aceitam o desempenho comprometido para ambos o OLTP e o data warehousing, ou reconfiguram constantemente o banco de dados em uma tentativa inútil de satisfazer demandas conflitantes das diferentes cargas de trabalho. As organizações continuarão a executar o OLTP e sistemas de data warehouse em diferentes plataformas, cada uma especificamente configurada de acordo com as necessidades de suas cargas de trabalho. As organizações que planejam usar o Oracle Exadata for OLTP<sup>12</sup> podem obter o melhor de ambos os mundos, combinando-o com o IBM Netezza Data Warehouse Appliance para data warehousing.

O único data warehouse que realmente importa é o “seu” data warehouse: seus aplicativos operando em seus dados em seu datacenter. Uma PoC (prova de conceito) interna cria a oportunidade para que um departamento de TI investigue detalhadamente uma tecnologia, aprendendo como é e podem usar o IBM Netezza Data Warehouse Appliance para ajudar seus peers de negócios a obterem um maior valor dos dados. Aproveitar ao máximo essa oportunidade exige que a PoC seja gerenciada com a mesma disciplina dedicada a outros projetos. Curt Monash oferece um sábio conselho em seu blog: “Melhores práticas para DBMS POCs de analítica”, incluindo uma consultoria independente e envolvente para orientar o projeto para um resultado bem-sucedido. Para organizações que desejam entender como seus armazéns executam no IBM Netezza Data Warehouse Appliance, sem qualquer custo ou risco, a Netezza oferece um “test drive”: Para agendar um, acesse [www.netezza.com/testdrive](http://www.netezza.com/testdrive).

## ***IBM Netezza Data Warehouse Appliance: para usar e aproveitar.***

### **Saber Mais**

A Netezza tem tanta certeza de que você gostará da solução IBM Netezza Data Warehouse Appliance, que o convidamos a testá-la em nosso site com seus dados – sem qualquer custo. Fazer o “test drive” é fácil.

Saiba mais aqui: [www.netezza.com/testdrive](http://www.netezza.com/testdrive)

<sup>12</sup> Acesse o blog de DBMS2 de Curt Monash <http://www.dbms2.com/2010/01/22/oracle-database-hardware-strategy/> para ver sua discussão sobre a função da tecnologia como o Exadata como uma plataforma para consolidar diversos bancos de dados Oracle de uma corporação, em vez de executar algumas tarefas difíceis de gerenciamento de banco de dados.

## Forneça seu Feedback

O que você acha sobre as ideias e argumentos deste e-book? Conte-nos sobre o que você gostou, não gostou ou quer discutir mais detalhadamente.

- Bate-papo com a Comunidade Netezza: [www.enzeecommunity.com/groups/twinfin-talk](http://www.enzeecommunity.com/groups/twinfin-talk)
- Tweet sobre o documento: [#twinfin](https://twitter.com/twinfin)

## Entre em Contato Conosco

Visite nosso blog: [www.enzeecommunity.com/blogs/nzblog](http://www.enzeecommunity.com/blogs/nzblog)

Visite o Web site da Netezza: [www.ibm.com/software/br/data/netezza](http://www.ibm.com/software/br/data/netezza)

Visite o Web site da Comunidade Netezza: [www.enzeecommunity.com](http://www.enzeecommunity.com)

Envie-nos seus comentários: [www.netezza.com/company/contact\\_form.aspx](http://www.netezza.com/company/contact_form.aspx)

## Sobre o Autor



Phil Francisco, Vice-Presidente, Product Management & Product Marketing, Netezza, possui mais de 20 anos de experiência em desenvolvimento de tecnologia e marketing de tecnologia global. Como Vice-Presidente de Product Management & Product Marketing na Netezza, ele promove novos negócios e estratégias de produtos, administra o portfólio de produtos e orienta os programas de marketing de produtos. Antes da Netezza, Francisco era o Vice-Presidente de Marketing na PhotonEx, uma desenvolvedora líder de sistemas de transporte óptico de 40 Gb/s para os principais fornecedores de rede de telecomunicações. Antes da PhotonEx, Francisco atuou como Vice-Presidente de Marketing do Produto para a Lucent Technologies' Optical Networking Group, onde trabalhou com alguns dos maiores portadores de telecomunicações no planejamento e implementação de soluções de rede óptica. Francisco possui uma patente em arquiteturas de rede óptica avançadas. Ele possui bacharelado em Engenharia Elétrica e em Ciência da Computação, formando-se na Moor School of Electrical Engineering na Universidade da Pensilvânia. Ele obteve seu diploma de mestrado em Engenharia Elétrica na Stanford University e concluiu o Programa de Gerenciamento Avançado na Fuqua School of Business na Duke University. Leia o blog de Phil em: [www.enzeecommunity.com/blogs/nzblog](http://www.enzeecommunity.com/blogs/nzblog).

## Sobre a Netezza

### ***Netezza – Pioneira, Líder, Comprovada***

Com a invenção dos appliances de data warehouse, a Netezza revolucionou e simplificou a análise de negócios para empresas que possuem dados em excesso e se esforçam muito para encontrar a velocidade e eficiência de processamento para analisar e compreender o que todos eles significam.

Atualmente, a Netezza é líder em appliances de data warehouse, combinando armazenamento, processamento, banco de dados e analítica em um único sistema que fornece de 10 a 100 vezes melhor desempenho, com um terço do custo de outros sistemas. Mais uma vez estamos redefinindo os parâmetros de preço/desempenho. Pense na Netezza como uma Ferrari, com o preço e a eficiência de um carro econômico.

Com centenas de clientes, incluindo Natonwide, Neiman Marcus, Orange UK, The Sherwin-Williams Company, Virgin Media e muito mais, além de escritórios no mundo todo, a Netezza (NYSE: NZ) é uma solução comprovada para a elevação de custos e complexidade de data warehousing e analítica. Nosso impressionante ecossistema de parceiros globais, incluindo Ab Inito, Business Objects, Cognos, EMC, IBM, Informatica, Microsoft, MicroStrategy, SAS e muitos outros, e uma lista abrangente de integradores de sistema, revendedores e parceiros desenvolvedores no mundo todo indica que nossos clientes podem ficar tranquilos sabendo que a Netezza irá deixá-los confortáveis com qualquer infraestrutura que eles possam ter.

Apresentamos o primeiro appliance de data warehouse do mundo e desafiamos o status quo. Criamos um novo segmento de mercado e definimos a agenda para um setor inteiro. E agora, impulsionados por nossos clientes, estamos desenvolvendo soluções para resolver desafios empresariais maiores e mais complexos, incluindo federações globais, integração de dados, integração de sistema legado continuidade de negócios, analítica avançada e conformidade. Com nossa forte base financeira, modelo de negócios rentável, foco no crescimento e comunidade extremamente entusiasmada “enzee”, podemos garantir que a Netezza sempre será um grande parceiro para a analítica, crescendo com você conforme seus dados e demandas crescem.



Netezza Corporation  
26 Forest Street  
Marlborough, MA 01752

+1 508 382 8200 TEL  
+1 508 382 8300 FAX

[www.ibm.com/software/br/data/netezza/](http://www.ibm.com/software/br/data/netezza/)

#### Sobre a Netezza Corporation:

A Netezza, uma empresa IBM, é líder global em appliances de data warehouse e análises de negócio que simplificam drasticamente as análises de alto desempenho em toda a corporação. A tecnologia da Netezza permite que organizações processem enormes quantidades de dados capturados em uma velocidade excepcional, fornecendo uma vantagem operacional e competitiva significativa no atual segmento de mercado que usa dados de maneira intensiva, incluindo mídias digitais, energia, serviços financeiros, governos, ciências biológicas e médicas, varejo e telecomunicações. A Netezza tem sua sede em Marlborough, Massachusetts, e possui escritórios na América do Norte, Europa e região da Ásia-Pacífico. Para mais informações sobre a Netezza, visite [www.ibm.com/software/br/data/netezza/](http://www.ibm.com/software/br/data/netezza/)

© Netezza, uma Empresa IBM, 2011. Todos os direitos reservados. Todos os outros nomes de empresas, marcas e produtos contidos aqui podem ser marcas registradas ou marcas comerciais registradas de seus respectivos proprietários.