

White paper

O valor econômico do Data Domain

Validando a economia das soluções Data Domain

Por Vinny Choinski, analista sênior do ESG Lab;
e Jason Buffington, analista principal

Maio de 2017

Este white paper do ESG foi autorizado pela
Dell EMC e é distribuído sob licença do ESG.

Índice

Índice.....	2
Introdução.....	3
Benefícios da arquitetura do Data Domain	4
Benefícios econômicos do Data Domain	5
A eficiência da desduplicação é importante	6
Velocidade e escalabilidade	8
Considerações sobre economia	9
A grande verdade.....	10

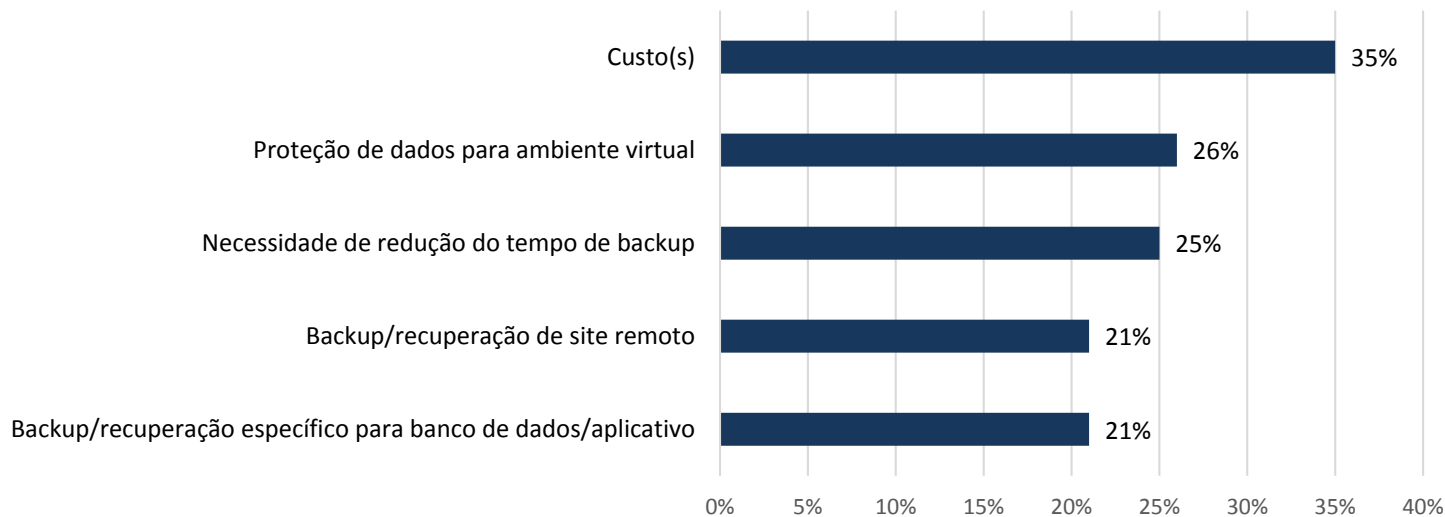
Introdução

Para superar as inovações e ultrapassar o ritmo da concorrência, as organizações precisam seguir por um caminho consistente para manter a infraestrutura moderna. A TI está sob constante pressão para oferecer uma infraestrutura otimizada para novas iniciativas de negócios e para dar suporte a aplicativos, tudo enquanto tenta conter ou até mesmo reduzir os custos. Na verdade, os respondentes da pesquisa em andamento do ESG mencionam constantemente a redução de custos como um dos principais fatores que impactam os gastos da TI. Quando perguntados em uma pesquisa recente sobre como as organizações pretendem conter os custos no próximo ano, 27% dos respondentes disseram que vão comprar novas tecnologias com um melhor retorno sobre o investimento.¹

Para aprofundamento em um grupo específico, 35% dos gerentes de TI com a tarefa de implementação de tecnologias e processos de proteção de dados rotineiramente mencionam o custo como um dos principais desafios (consulte a Figura 1). Este grupo parece encontrar problemas centralizados nas cargas de trabalho, principalmente desafios relacionados à proteção de ambientes virtualizados e escritórios remotos, especialmente problemáticos também. No entanto, ao observar as classificações gerais dos desafios da proteção de dados, surgem questões maiores. Custos e virtualização são as principais preocupações, mas as três considerações reportadas a seguir são de caráter *operacional e funcional*. São desafios relacionados a desempenho, backup e recuperação de arquitetura distribuída e proteção de aplicativos e banco de dados.²

Figura 1. Desafios atuais de tecnologia e processos de proteção de dados

Quais das seguintes opções você caracterizaria como desafios atuais da sua organização em relação às tecnologias e processos de proteção de dados? (Percentual de respondentes, N=387)



Fonte: Enterprise Strategy Group, 2017

Esses desafios também são consistentes com as principais exigências dos líderes de TI relacionadas a uma maior velocidade, agilidade, confiabilidade e controle dos custos, todos reportados consistentemente nos relatórios anteriores do ESG sobre a modernização da proteção de dados. A análise do ESG confirma que as inconsistências entre as dificuldades enfrentadas pelos implementadores e as exigências dos líderes resultaram frequentemente em pouquíssimas

¹Fonte: Relatório de pesquisa do ESG, [2017 IT Spending Intentions Survey](#), março de 2017.

²Fonte: Relatório de pesquisa do ESG, [Data Protection Modernization](#), a ser publicado

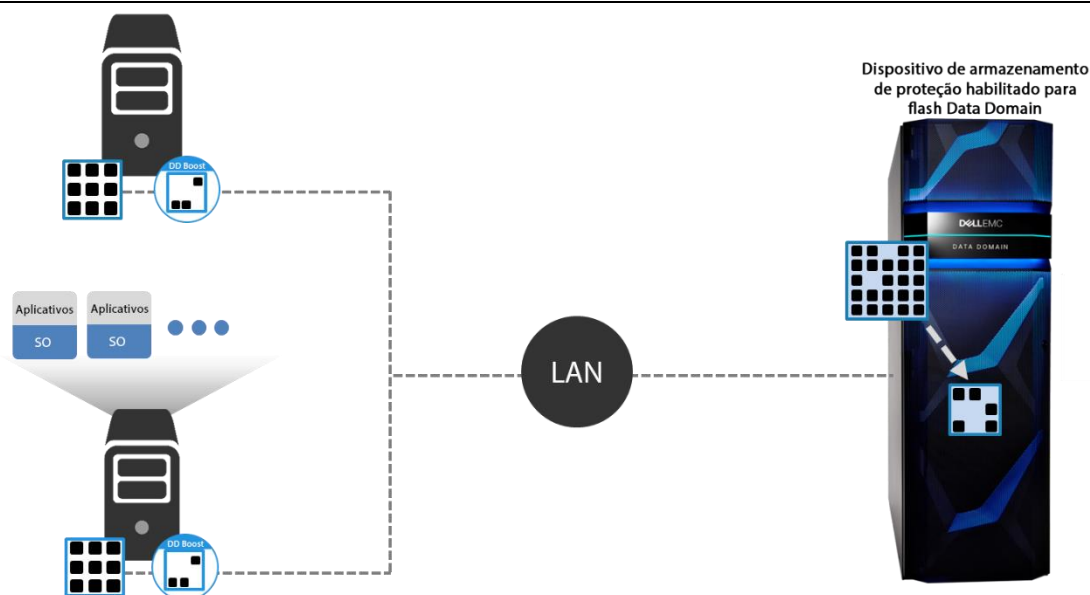
transformações e modernizações atuais. Os implementadores não conseguiram agir de acordo com as exigências dos líderes devido a impedimentos técnicos em seus ambientes de proteção de dados.³

Este artigo discute como os sistemas Data Domain e as soluções integradas na arquitetura do Data Domain, como o Dell EMC Integrated Data Protection Appliance, ajudam no fornecimento da agilidade de que os implementadores de TI precisam para transformar suas infraestruturas para atender as exigências dos líderes seniores de TI, solucionar os desafios atuais de proteção de dados e, mais importante, reduzir o custo do armazenamento dos dados de proteção. As descobertas do ESG são baseadas em uma auditoria e análise dos principais indicadores de desempenho (KPIs): dados do mundo real de sistemas implementados, incluindo o preço original de compra, custos ambientais, capacidade/utilização e desempenho.

Benefícios da arquitetura do Data Domain

Em 2006, a EMC tomou a decisão de evitar a agregação de um código de deduplicação de dados em suas soluções existentes de proteção de dados, marcando uma mudança estratégica na estratégia de proteção de dados. Essa mudança começou com o aproveitamento da tecnologia a partir de duas aquisições-chave. A primeira compra ofereceu uma tecnologia de deduplicação baseada em origem, e a segunda foi o Data Domain, com tecnologia de deduplicação de destino. As tecnologias dessas duas aquisições tornaram-se componentes fundamentais para as soluções de proteção de dados da EMC. Agora sob a marca Dell EMC, os sistemas Data Domain, inclusive o Integrated Data Protection Appliance, são dispositivos de proteção de dados habilitados para flash, totalmente integrados e de uso específico desenvolvidos para reduzir a capacidade necessária de armazenamento em disco para retenção e proteção de dados. Com deduplicação de dados baseada em origem e destino nativamente integrada na arquitetura, esses sistemas possibilitam a conclusão de mais backups em menos tempo, fornecem restaurações com mais rapidez e de forma mais confiável e reduzem a quantidade da capacidade de armazenamento necessária para proteção de dados. A geração mais recente dos sistemas Data Domain introduziu a inclusão de flash para metadados, permitindo o desempenho em escala e um acesso e restauração de máquinas virtuais vinte vezes mais rápidos diretamente do Data Domain, em comparação com as gerações anteriores.

Figura 2. Visão geral da arquitetura do Data Domain



Fonte: Enterprise Strategy Group, 2017

Os principais recursos de arquitetura do Data Domain incluem:

³Fonte: Relatório de pesquisa do ESG, [2015 Trends in Data Protection Modernization](#), setembro de 2015.

- **SISL (Stream-Informed Segment Layout)** O SISL permite que os sistemas Data Domain realizem 99% do processamento de deduplicação na CPU e na RAM. Isso significa que os sistemas não dependem do número de discos para melhorar o desempenho.
- **Segmentação de tamanho variável:** Para taxas de deduplicação ideais, os sistemas Data Domain aproveitam a segmentação de tamanho variável para fragmentar fluxos baseados na estrutura natural dos dados. Isso permite que o sistema determine se o segmento é exclusivo antes de compactá-lo ou armazená-lo.
- **Deduplicação em linha:** Os sistemas Data Domain efetuam deduplicação na CPU e na memória à medida em que o fluxo de backup é recebido pelo sistema. Isso significa que apenas dados exclusivos são enviados e armazenados em disco. Isso elimina a necessidade de uma área de transferência em disco e de recursos de computação para deduplicação pós-processo.
- **Software Data Domain Boost:** O software Data Domain Boost distribui partes do processo de deduplicação para os clientes de aplicativo ou para o servidor de backup. Com o DD Boost, apenas dados exclusivos atravessam a conexão entre o servidor de backup ou os clientes e o sistema Data Domain. Isso também ajuda a liberar recursos na rede do cliente e no sistema Data Domain para um melhor desempenho de deduplicação no destino.
- **Data Domain Data Invulnerability Architecture:** O Data Domain é desenvolvido para garantir a recuperabilidade dos dados de forma confiável. A Data Invulnerability Architecture oferece verificação de leitura e de gravação em linha, fornecendo proteção e recuperação automática para problemas de integridade de dados durante a inclusão e a recuperação dos dados. A detecção contínua de falhas e a autocorreção garantem a continuidade da recuperabilidade dos dados por todo o ciclo de vida no sistema Data Domain.

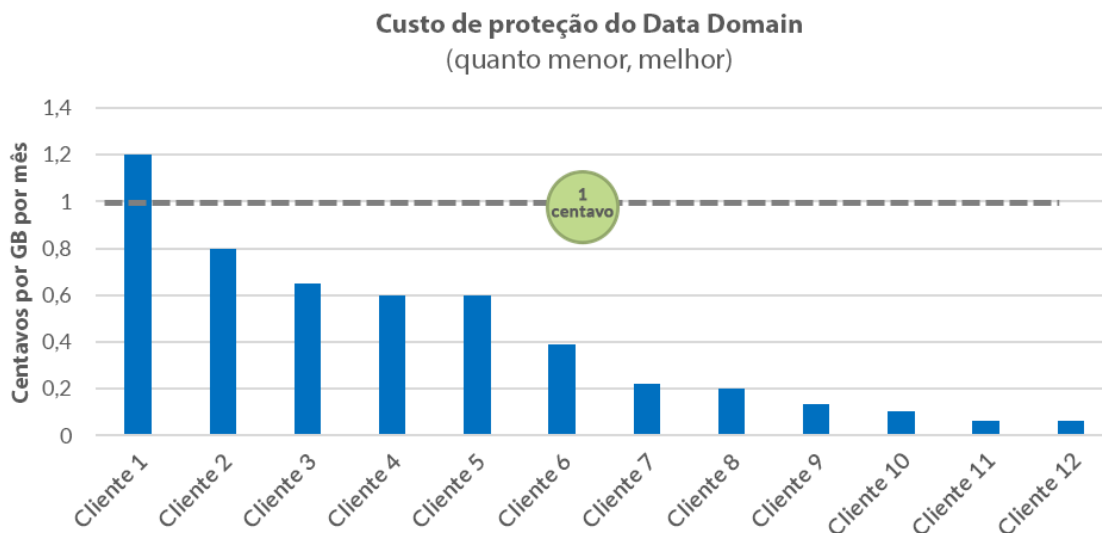
Benefícios econômicos do Data Domain

Vamos ver em mais detalhes como a arquitetura do Data Domain pode representar benefícios econômicos para os participantes nos negócios. A deduplicação de dados permite aos clientes armazenar mais dados com a mesma quantidade de espaço de disco físico. Isso reduz os requisitos de capacidade de armazenamento e o custo. A deduplicação de dados implementada na origem ou no lado do cliente também ajuda com a economia de capacidade, mas com um benefício adicional de melhoria do desempenho de backup. Com a deduplicação no lado da origem, apenas blocos de dados exclusivos são enviados da origem para o destino durante a operação de backup, reduzindo significativamente o tráfego de rede. Essa melhoria na eficiência da rede permite um aumento dos dados de backup com a utilização da infraestrutura de rede existente e elimina ou posterga, possivelmente, a necessidade de upgrades caros da rede. Obviamente, quanto menor for o volume de dados a ser transferido, mais rápido será o desempenho do backup. Durações mais curtas de backup também permitem aos clientes aumentar a frequência dos backups, reduzindo o risco de perda de dados, o que pode ser extremamente dispendioso para uma organização. Com o software DD Boost, os sistemas Data Domain dão suporte tanto para deduplicação do lado da origem quanto do lado do destino, proporcionando aos clientes uma flexibilidade para implementação da deduplicação onde quer que faça mais sentido para os ambientes desses clientes.

Conforme mostrado na Figura 3, a análise do ESG dos dados no mundo real, incluindo hardware, software, energia, refrigeração e deduplicação, demonstra que os sistemas Data Domain são facilmente capazes de fornecer armazenamento para ambientes de proteção de dados por menos de um centavo por GB por mês.

Observe como o custo da proteção varia entre 0,062 e 1,2 centavo por gigabyte para os doze clientes analisados pelo ESG. Essa relativamente ampla variedade em economia ocorre principalmente devido ao volume da economia da capacidade de deduplicação que está sendo obtida pelos clientes da Dell EMC nos ambientes de produção. O restante deste relatório analisa mais detalhadamente o custo da proteção com foco em como a economia da deduplicação pode ser aumentada com uma combinação completa do software Dell EMC Data Protection Suite com o hardware Data Domain.

Figura 3. O valor econômico do Data Domain

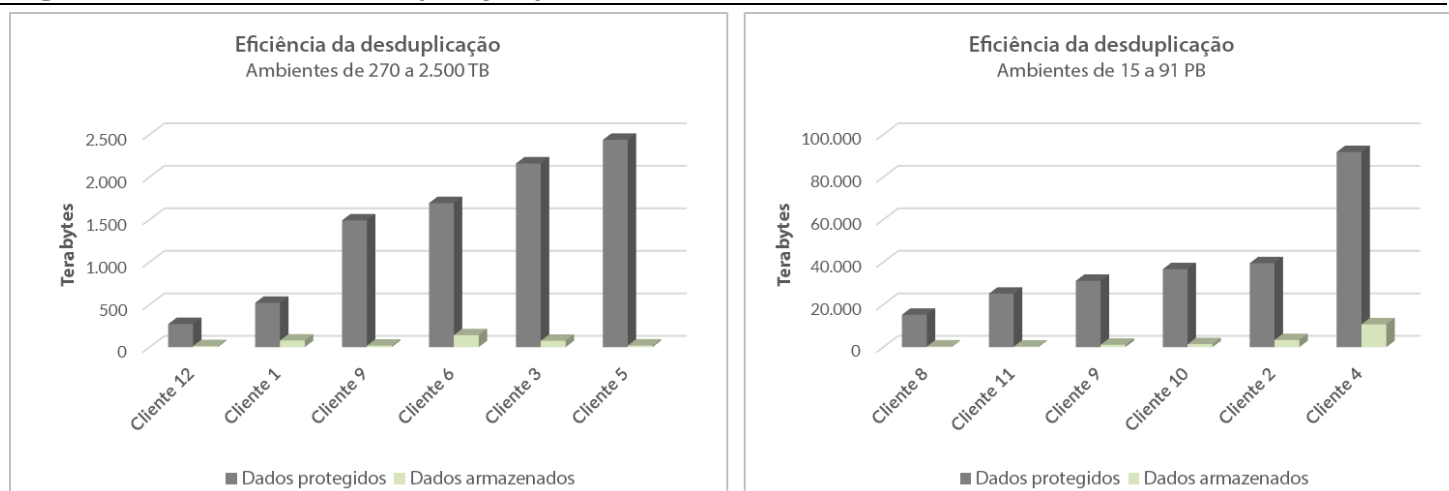


Fonte: Enterprise Strategy Group, 2017

A eficiência da deduplicação é importante

O ESG começou a sua exploração sobre o valor econômico da economia da capacidade da deduplicação do Data Domain por meio de auditoria e análise dos dados de suporte call home de 12 clientes ativos da Dell EMC. Conforme mostrado na Figura 4, os ambientes dos clientes variam de aproximadamente 270 TB a mais de 90 PB de dados a serem protegidos. Os clientes abrangeram diversos setores, inclusive o da tecnologia, indústria, seguros e saúde. A seleção de clientes em diferentes setores foi designada para capturar os resultados da deduplicação em diferentes tipos de conjuntos de dados. As barras cinza da Figura 4 mostram o volume de dados sendo protegido para cada ambiente. As barras verdes mostram o volume de dados armazenado após a deduplicação.

Figura 4. A eficiência da deduplicação por cliente



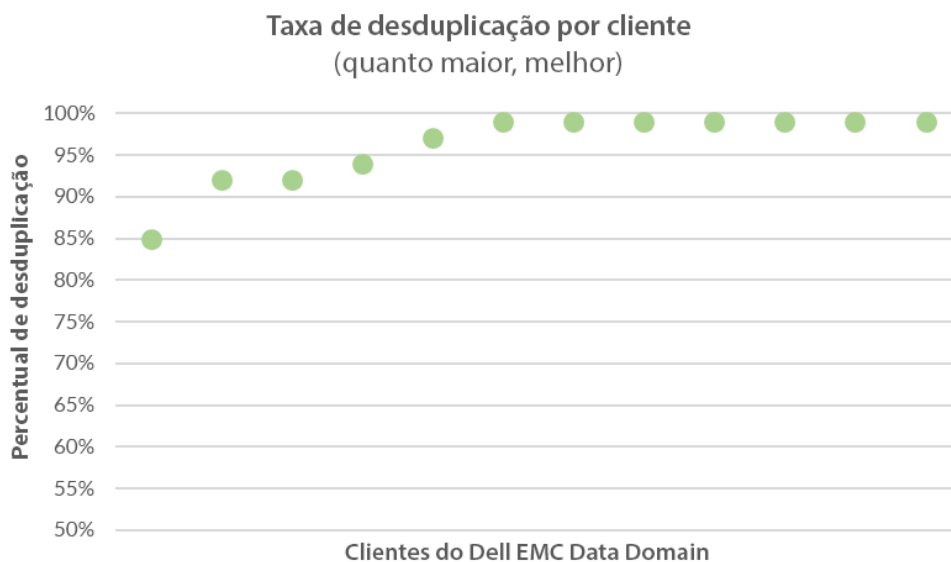
Fonte: Enterprise Strategy Group, 2017

- **Desduplicação:** A maior razão de desduplicação observada nos dados analisados foi de 126:1. Esse resultado foi decorrente do aproveitamento que um cliente teve com o software de backup e recuperação Data Protection Suite com um dispositivo Data Domain, sendo identificado pelo primeiro ponto de dados no lado esquerdo da Figura 4. A razão média de desduplicação para todos os clientes analisados foi de 41:1.
- **Proteção:** Até mesmo o cliente com a menor razão observada de desduplicação, de aproximadamente, 9:1, poderia proteger quase 92 PB de dados usando apenas 10 PB de capacidade. Esse cliente é identificado pelo último ponto de dados no lado direito da Figura 4. O cliente faz parte do setor industrial e tem um ambiente altamente distribuído.

A Figura 5 fornece uma outra visão sobre o efeito da desduplicação do Data Domain. Essa figura mostra o percentual de desduplicação obtido para cada um dos mesmos 12 ambientes dos clientes. As taxas de desduplicação variam de 85% a 99%. Tipicamente, quanto maior for o tempo do recebimento de dados no ambiente pelo Data Domain, maior será a taxa de desduplicação. Isso porque a probabilidade de encontrar os mesmos padrões de dados ao longo do tempo é maior e precisa apenas de armazenamento de dados exclusivos.

O recurso de segmentação de tamanho variável da arquitetura do Data Domain é um componente-chave para a obtenção desse nível de desduplicação. Com a segmentação de tamanho variável, a solução Data Domain pode ser alinhada mais facilmente com os padrões naturais, inclusive com marcadores de registro de data e hora de bancos de dados, nas estruturas de dados enviadas para o dispositivo para proteção. A segmentação de tamanho variável produz níveis significativamente mais altos de desduplicação do que as arquiteturas de tamanho fixo. Ela também permite uma escalabilidade mais eficiente em um único pool de armazenamento. Isso também resulta em menos dispositivos para gerenciamento, economizando nos custos operacionais.

Figura 5. Porcentagem de desduplicação por cliente



Fonte: Enterprise Strategy Group, 2017

Outro benefício da alta desduplicação de segmentação de tamanho variável surge quando os dados são replicados entre os dispositivos do Data Domain. Menos dados significa menos dados para serem replicados. Isso não significa apenas um melhor desempenho da replicação, mas também menores requisitos de largura de banda de WAN para o processo de replicação.

- **Taxas de deduplicação:** As taxas de deduplicação auditadas e analisadas variaram de 85% a 99%, com uma média de 96%. Isso significa que apenas uma fração dos dados de produção (nas melhores situações 1% ou menos) precisa ser armazenada no sistema Data Domain para proteção.
- **Abaixo de 90%:** Apenas um cliente dos dados analisados teve uma taxa de deduplicação inferior a 90%. Novamente, esse cliente representa o ambiente altamente distribuído com pools distribuídos de proteção.
- **90% e acima:** Os demais clientes analisados obtiveram taxas de deduplicação superiores a 90%, e a maioria dos ambientes dos clientes estava funcionando a uma taxa de deduplicação de 99%.

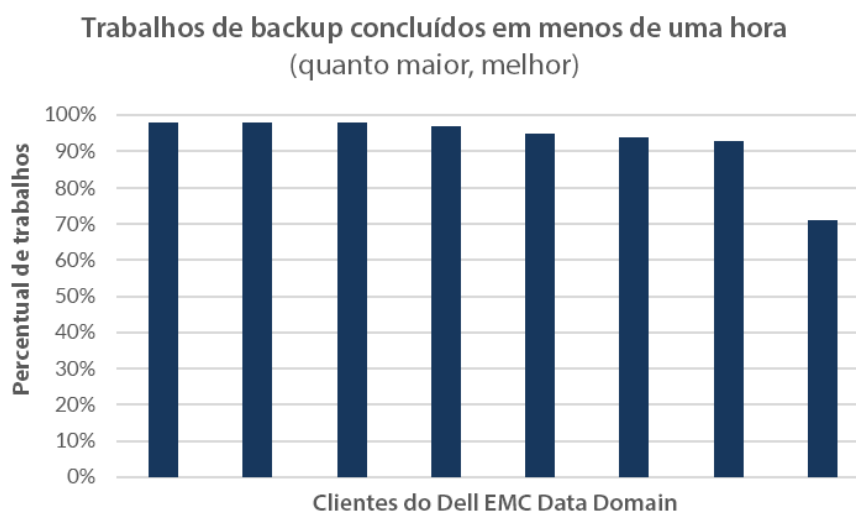
Velocidade e escalabilidade

O aumento da capacidade de armazenamento por meio da deduplicação nem sempre anda de mãos dadas com o alto desempenho. O Data Domain resolve esse desafio com o aproveitamento de dois elementos-chave de desempenho: O SISL (Stream-Informed Segment Layout) e o software DD Boost.

O SISL é um elemento de arquitetura da plataforma do Data Domain que permite que 99% do processo de deduplicação ocorra na CPU e na RAM. Isso significa que a solução não depende de uma contagem alta de rotação de disco para a obtenção de desempenho. Como resultado, soluções de menor espaço ocupado baseadas na mesma controladora do Data Domain podem alcançar o mesmo desempenho que as configurações de maior capacidade. Além disso, como as controladoras do Data Domain utilizam CPUs Intel, elas se beneficiam das melhorias de desempenho de cada nova versão dos processadores Intel.

O software DD Boost também contribui para o desempenho por meio da distribuição de partes do processo de deduplicação para os clientes de aplicativos ou para o servidor de backup. Com o DD Boost, apenas dados exclusivos precisam ser movidos do servidor de backup ou dos clientes para o sistema Data Domain. O DD Boost reduz o volume de dados movido em até 99%. O benefício líquido das conclusões mais rápidas dos trabalhos de backup foi auditado pelo ESG. Conforme mostrado na Figura 6, a maioria dos clientes rotineiramente conclui os trabalhos de backup em uma hora ou menos.

Figura 6. Desempenho do trabalho de backup



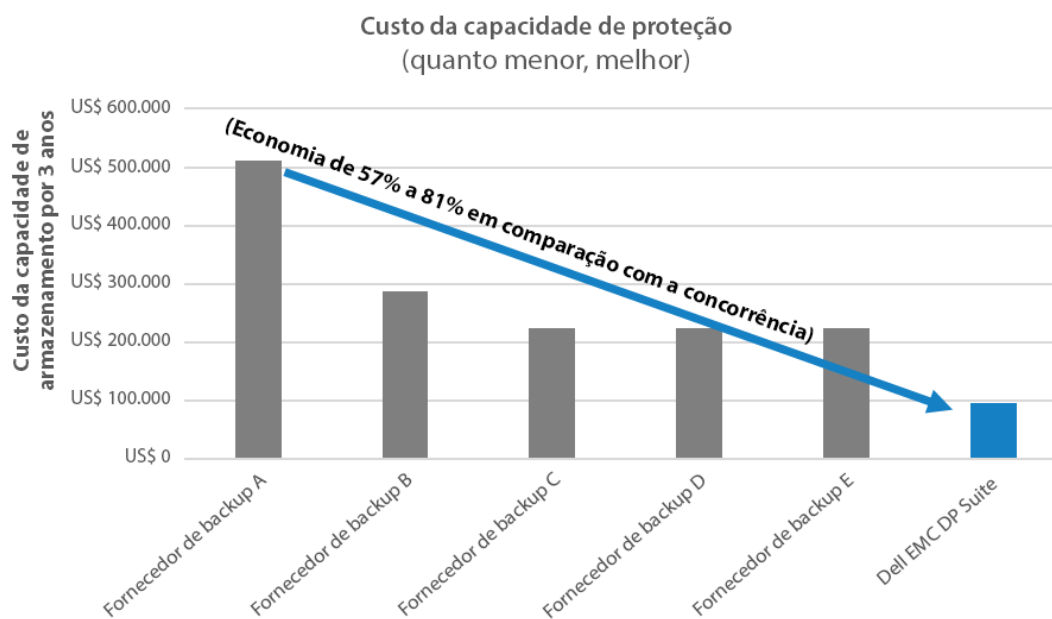
Fonte: Enterprise Strategy Group, 2017

- **Em menos de 15 minutos:** Para um cliente do setor de tecnologia, 76% de todos os trabalhos de backup foram concluídos em menos de 15 minutos.
- **Em menos de 1 hora:** Para oito dos doze clientes analisados, de 70% a 98% de todos os trabalhos de backup foram concluídos em menos de uma hora (conforme mostrado na Figura 6).
- **Em menos de 4 horas:** Para um cliente do setor de seguros, todos os trabalhos de backup foram concluídos em menos de quatro horas.
- **Mais de 4 horas:** Para dois dos maiores ambientes (em termos de número e tamanho de aplicativos), a conclusão de todos os trabalhos de backup demorou mais de quatro horas.

Considerações sobre economia

O ESG concluiu a exploração do valor econômico da economia da capacidade de desduplicação do Data Domain por meio de auditoria e análise dos dados de suporte call home de mais de 15.000 sistemas Data Domain implementados em todo o mundo. Essa amostra foi coletada nos sistemas Data Domain que gravam o software para backup que está sendo utilizado e as taxas de desduplicação que estão sendo obtidas. Os dados mostram que a eficiência da desduplicação do Data Domain pode ser levada ainda mais adiante quando uma solução completa de proteção de dados da Dell EMC é implementada. Conforme mostrado na Figura 7, a combinação do Data Domain com o Data Protection Suite aumentou a economia da desduplicação e reduziu o custo da capacidade de proteção de 57% a 81%.

Figura 7. Economizando mais com o Dell EMC Data Domain e o Data Protection Suite



Fonte: Enterprise Strategy Group, 2017

A solução Data Domain oferece outros benefícios econômicos além do custo da economia de capacidade mostrado na Figura 7. A análise do ESG dos dados dos clientes do Data Domain do mundo real também revelou os seguintes benefícios:

- **Redução de largura de banda da rede:** Como a maioria dos aplicativos corporativos, as soluções de backup e recuperação dependem dos recursos da rede para funcionar. Isso significa tanto os recursos da LAN para backup e restauração local quanto os recursos da WAN para recuperação de desastres e continuidade dos negócios.

Soluções ineficientes de proteção de dados podem consumir rapidamente recursos valiosos da rede e até impactarem na conectividade e na produtividade dos usuários. Com o aproveitamento da deduplicação do Data Domain, um fabricante multinacional conseguiu reduzir sua utilização de largura de banda de proteção de dados local e remota em 98%.

- **Desempenho:** A deduplicação eficiente dos dados pode ter um grande impacto na melhoria do desempenho geral da proteção de dados. Com a implementação do DD Boost, que reduz o volume de dados necessário a ser transferido entre o cliente e o sistema Data Domain, um fabricante de equipamentos pesados conseguiu melhorar sua velocidade de backup e restauração em 50%. O mesmo cliente melhorou seu preparo para recuperação de desastres em 90%.
- **Redução do espaço ocupado pelo datacenter:** Os requisitos de espaço físico, especialmente a quantidade necessária para dar suporte ao espaço, a energia e a refrigeração de um datacenter modernos, não são baratos. Pode ser difícil, e algumas vezes impossível, expandir fisicamente um datacenter sem ter de mudar de local. Com o Data Domain, uma cadeia de lojas de departamento eliminou a fita física em seu ambiente de proteção de dados e liberou três fileiras inteiras de espaço de rack altamente valiosas do datacenter.

A grande verdade

As principais exigências de proteção de dados dos líderes de TI estão focadas na melhoria da confiabilidade e da agilidade essenciais da(s) solução(ões) em uso. A exigência que segue logo atrás é a redução de custo, também vista como uma principal prioridade entre os implementadores de proteção de dados. Esses desafios não devem ser vistos como contraditórios ou mutuamente exclusivos. Na verdade, eles podem ser todos resolvidos com a melhoria das soluções de proteção de dados desenvolvidas tanto para eficiência quanto para confiabilidade e capacidade.

A eficiência surge de várias formas, frequentemente baseada no custo da realização dos negócios e na inteoperabilidade.

- A capacidade de fornecimento de armazenamento de proteção de maneira econômica por meio da redução do custo para atender o aplicativo de proteção de dados.
- Inteoperabilidade nas soluções de software de backup, já que apenas uma das quatro organizações tem um único aplicativo de backup.⁴
- A eficiência da escalabilidade e da deduplicação do sistema de armazenamento de proteção, que pode ser usada para reduzir o número, e o custo, dos sistemas que precisam ser implementados e gerenciados.
- A capacidade de aproveitamento do mesmo pool de armazenamento de proteção com capacidade otimizada para backup e arquivamento de dados.
- Inteoperabilidade (e integração) entre o armazenamento de proteção e a variedade de plataformas, para que os proprietários de aplicativos (por exemplo, administradores de banco de dados e VAdmins) possam utilizar as próprias ferramentas para proteção/recuperação complementar, enquanto continuam sendo cidadãos responsáveis da TI, aproveitando o armazenamento eficiente da proteção centralizada.

Ao fornecer o que muitos do setor consideram sinônimo de “armazenamento de proteção” e/ou “deduplicação”, a Dell EMC expandiu a plataforma do Data Domain, de simplesmente um componente de uma infraestrutura de proteção de dados para a base de um ecossistema que inclui armazenamento da produção, fornecedores de backup, fornecedores de

⁴Fonte: Relatório de pesquisa do ESG, *Data Protection Modernization*, a ser publicado.

arquivamento e aceleradores de plug-in para aplicativos (ou seja, DD Boost), entre outros elementos. O Dell EMC Integrated Data Protection Appliance é uma extensão desse ecossistema com software de proteção e lógica analítica pré-configurados. Desse modo, há um ímpeto significativo para continuar a inovação onde a demanda do mercado atual se encontra, resultando em maiores ganhos de eficiência nos mecanismos de desduplicação do Data Domain, melhorias com suas integrações e expansões das capacidades para além da desduplicação no local, como classificação em nuvem, replicação de/para aplicativos hospedados na nuvem, appliances virtuais, recuperação de desastres para a nuvem, etc.

Um dos mais importantes elementos diferenciados de qualquer plataforma de desduplicação é o rigor da garantia da integridade dos dados dentro do repositório, já que há muita dependência da integridade de cada block exclusivo dentro do sistema. Desse modo, é essencial que as organizações investiguem com atenção o nível de zelo com as verificações do sistema e a frequência dessas verificações em relação ao armazenamento subjacente para garantia de integridade. A abordagem da Dell EMC em relação a isso tem sido uma característica da plataforma do Data Domain desde o início, referida como Data Invulnerability Architecture (DIA).

Com base em uma auditoria da coleta de dados em campo de mais de 15.000 ambientes de produção, o ESG confirmou que a combinação do hardware Dell EMC Data Domain com o software Data Protection Suite reduz o custo da capacidade de proteção de 57% a 81%, em comparação com os ambientes do Data Domain com software de backup da concorrência. Organizações que estão hesitando em fazer um investimento no Dell EMC Data Domain com base no “preço” se beneficiariam se reconsiderassem os benefícios econômicos ao longo do tempo, inclusive o custo do tempo reduzido de inatividade e da perda de dados, a melhoria do desempenho e, mais importante, o custo da proteção a um valor inferior a um centavo por GB por mês.

Todas as marcas comerciais são propriedade de suas respectivas empresas. As informações contidas nesta publicação foram obtidas de fontes que o ESG (The Enterprise Strategy Group) considera confiáveis, mas não são garantidas pelo ESG. Esta publicação pode conter opiniões do ESG, que estão sujeitas a mudanças de tempos em tempos. Esta publicação é protegida por direitos autorais do The Enterprise Strategy Group, Inc. Qualquer reprodução ou redistribuição integral ou parcial desta publicação, seja em formato impresso, eletrônico ou em qualquer outro formato, para pessoas não autorizadas a recebê-la e sem o consentimento expresso do The Enterprise Strategy Group, Inc. representa uma violação da lei de direitos autorais dos Estados Unidos e estará sujeita a uma ação por danos civis e, se aplicável, a processo criminal. Em caso de dúvida, entre em contato com o Departamento de atendimento ao cliente do ESG nos Estados Unidos pelo telefone +1 (508) 482-0188.



O Enterprise Strategy Group é uma empresa de pesquisa, análise e estratégia de TI conhecida mundialmente pelo oferecimento de percepções e inteligência práticas para a comunidade global de TI.

© 2017 by The Enterprise Strategy Group, Inc. Todos os direitos reservados.

