

White paper

# O valor econômico da combinação de Data Domain e IDPA (Integrated Data Protection Appliances)

Validando a eficiência de custo das soluções de appliance para backup Dell EMC

Por Vinny Choinski, analista sênior do ESG Lab;  
e Christophe Bertrand, analista sênior  
Junho de 2018

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

## Índice

Índice.....	2
Introdução.....	3
Benefícios da arquitetura de Data Domain e IDPA .....	4
Benefícios econômicos do Data Domain e IDPA.....	5
A eficiência da desduplicação é importante .....	6
Velocidade e escalabilidade .....	8
Reflexões sobre a economia .....	9
A grande verdade.....	11

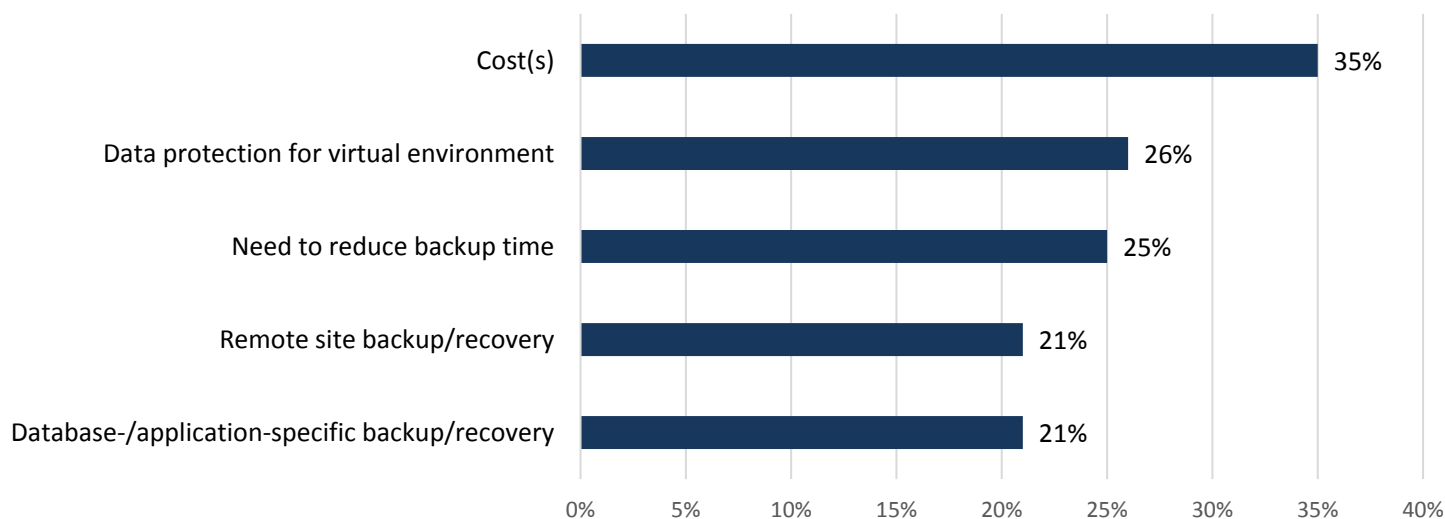
## Introdução

Para superar as inovações e ultrapassar o ritmo da concorrência, as organizações precisam seguir um caminho consistente para manter sua infraestrutura moderna. A TI está sob constante pressão para oferecer uma infraestrutura otimizada para novas iniciativas de negócios e fornecer suporte a aplicativos, tudo enquanto tenta conter ou até mesmo reduzir os custos. Na verdade, os respondentes da pesquisa em andamento do ESG mencionam de maneira consistente a redução de custos como um dos principais geradores de negócios que afetam seus gastos de TI. Em uma pesquisa recente, quando indagados sobre como suas respectivas organizações pretendiam conter os custos em 2017, 27% dos respondentes afirmaram que comprariam novas tecnologias com um melhor retorno sobre o investimento.<sup>1</sup>

Para aprofundamento em um grupo específico, 35% dos gerentes de TI responsáveis pela implementação de tecnologias e processos de proteção de dados mencionam repetidamente o custo como um dos principais desafios (consulte a Figura 1). Esse grupo parece se deparar com problemas relacionados a cargas de trabalho, principalmente desafios ligados à proteção de ambientes virtualizados e escritórios remotos, também especialmente problemáticos. No entanto, uma outra história surge quando observamos as classificações gerais dos desafios da proteção de dados. Custos e virtualização são as principais preocupações, mas os três fatores que vêm em seguida têm natureza *operacional* e *funcional*. São desafios relacionados a desempenho, backup e recuperação de arquitetura distribuída e proteção de aplicativos e banco de dados.<sup>2</sup>

**Figura 1. Os atuais desafios de tecnologia e processos de proteção de dados**

**Which of the following would you characterize as challenges with your organization's current data protection processes and technologies? (Percent of respondents, N=387)**



Fonte: Enterprise Strategy Group

Esses desafios também são consistentes com as principais determinações dos líderes de TI, que estão relacionadas a maiores níveis de velocidade, agilidade, confiabilidade e controle dos custos, todos consistentemente mencionados nos relatórios anteriores do ESG sobre a modernização da proteção de dados. A análise do ESG confirma que muitas vezes as inconsistências entre as dificuldades enfrentadas pelos implementadores e as ordens dos líderes resultaram na concretização de pouquíssimas transformações ou modernizações. Os implementadores não conseguiram agir de acordo com as ordens dos líderes devido a barreiras técnicas em seus ambientes de proteção de dados.<sup>3</sup>

<sup>1</sup> Fonte: relatório de pesquisa do ESG, [2017 IT Spending Intentions Survey](#), março de 2017.

<sup>2</sup> Fonte: pesquisa de levantamento do ESG, [Data Protection Modernization Trends](#), dezembro de 2016.

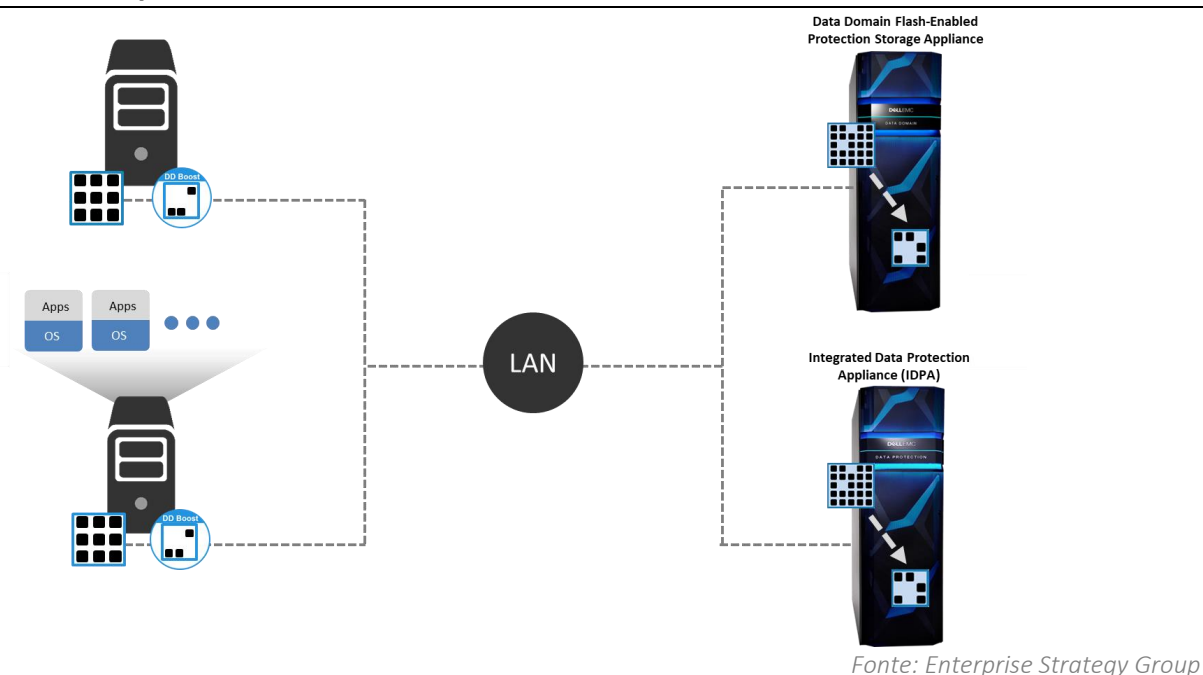
<sup>3</sup> Fonte: relatório de pesquisa do ESG, [2015 Trends in Data Protection Modernization](#), setembro de 2015.

Este artigo discute como os sistemas Dell EMC Data Domain e as soluções convergentes baseadas na arquitetura Data Domain, como o Dell EMC IDPA (Integrated Data Protection Appliance), ajudam a entregar a agilidade de que os implementadores de TI precisam para transformar suas infraestruturas a fim de atender as determinações dos líderes sênior de TI, solucionar os desafios atuais de proteção de dados e, principalmente, reduzir o custo de armazenamento de dados de proteção. As descobertas do ESG são baseadas em uma auditoria e análise dos KPIs (Key Performance Indicator, principal indicador de desempenho): dados reais de sistemas implementados, inclusive o preço original de compra, custos ambientais, capacidade/utilização e desempenho.

## Benefícios da arquitetura de Data Domain e IDPA

Em 2006, a EMC tomou a decisão de evitar a vinculação do código de deduplicação de dados em suas soluções existentes de proteção de dados, o que marcou uma mudança fundamental em sua estratégia de proteção de dados. Essa mudança começou com o aproveitamento da tecnologia a partir de duas aquisições cruciais. A primeira compra garantiu a tecnologia de deduplicação baseada em origem, e a segunda foi o Data Domain, com tecnologia de deduplicação de destino. As tecnologias dessas duas aquisições se transformaram em componentes fundamentais nas soluções de proteção de dados da EMC. Agora, sob a marca Dell EMC, os sistemas Data Domain, inclusive o IDPA, são appliances de proteção de dados habilitados para flash, totalmente integrados e de uso específico projetados para reduzir a quantidade necessária de armazenamento em disco para reter e proteger dados. Com a deduplicação de dados baseada em origem e destino integrada nativamente na arquitetura, esses sistemas viabilizam a realização de mais backups em menos tempo, fornecem restaurações mais rápidas e mais confiáveis, e reduzem a quantidade da capacidade de armazenamento necessária para a proteção de dados. A geração recente de sistemas Data Domain e IDPA lançou a inclusão de flash para metadados, viabilizando o desempenho em escala e operações de acesso e restauração instantâneos de máquinas virtuais diretamente a partir do Data Domain e IDPA que são vinte vezes mais rápidas do que as gerações anteriores.

**Figura 2. Visão geral da arquitetura do Data Domain e IDPA**



Os principais recursos de arquitetura do Data Domain e IDPA incluem:

- **SISL (Stream-informed Segment Layout):** o SISL permite que os sistemas Data Domain e IDPA 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 e IDPA 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.
- **Desduplicação em linha:** os sistemas Data Domain e IDPA executam a deduplicação na CPU e na memória conforme 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 clients 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 clients e o sistema Data Domain e IDPA. Isso também ajuda a liberar recursos na rede do cliente e no sistema Data Domain e IDPA para um melhor desempenho de deduplicação no destino.
- **Data Invulnerability Architecture:** a combinação entre Data Domain e IDPA foi desenvolvida de modo a garantir que os dados possam ser recuperados de maneira confiável. A Data Invulnerability Architecture oferece verificação em linha de leitura e de gravação, 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 durante todo o ciclo de vida nos sistemas Data Domain e IDPA.

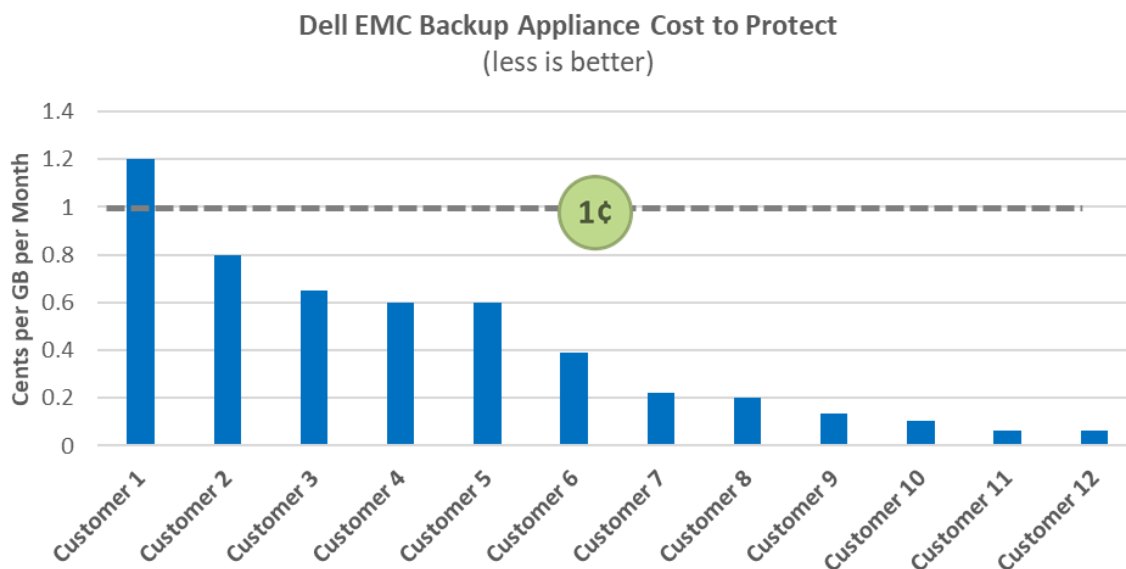
## Benefícios econômicos do Data Domain e IDPA

Vamos ver em mais detalhes como as arquiteturas do Data Domain e IDPA resultam em 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 de custo. A deduplicação de dados implementada na origem ou no lado do client 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 eficiência da rede aprimorada permite um aumento dos dados de backup com a utilização da infraestrutura de rede existente e possivelmente elimina ou posterga a necessidade de upgrades caros da rede. Obviamente, quanto menor for o volume de dados a ser transferido, melhor 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 e IDPA são compatíveis com deduplicação tanto no lado da origem quanto no lado do destino, proporcionando aos clientes a flexibilidade para implementar a deduplicação onde fizer mais sentido nos ambientes deles.

Conforme apresentado na Figura 3, a análise de dados reais feita pelo ESG, incluindo hardware, software, energia, refrigeração e deduplicação, demonstra que os sistemas Data Domain e IDPA 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 12 clientes analisados pelo ESG. Essa variação relativamente ampla na 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 e IDPA.

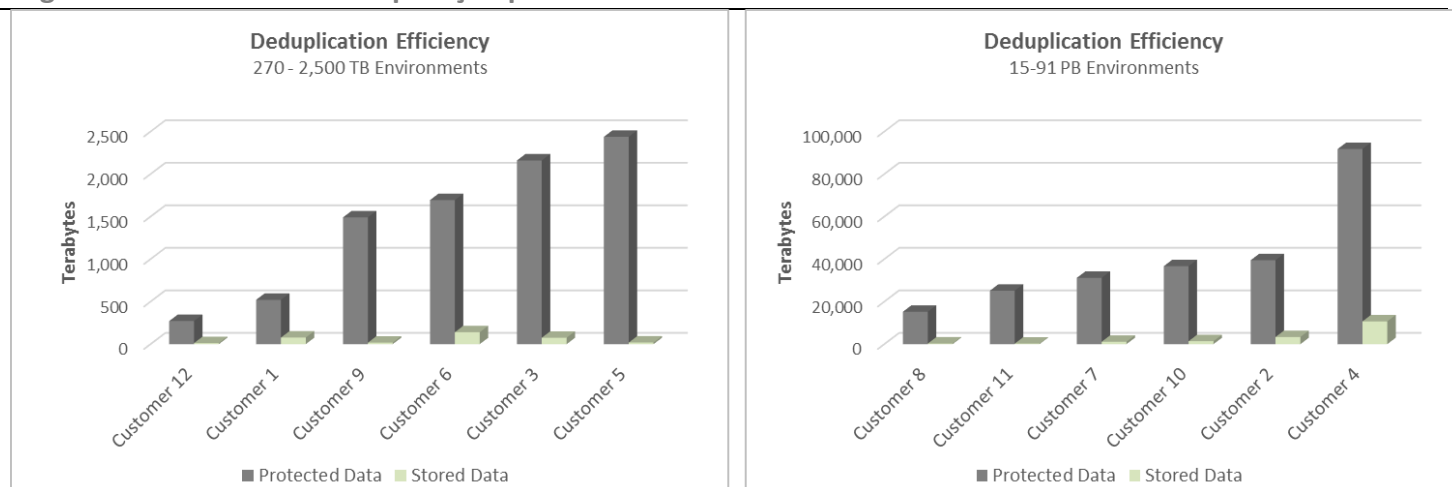
**Figura 3. O retorno econômico do Data Domain e IDPA**



Fonte: Enterprise Strategy Group

### A eficiência da deduplicação é importante

O ESG começou a explorar o valor econômico gerado pela economia da capacidade de deduplicação do Data Domain e IDPA por meio de auditoria e análise dos dados de suporte CallHome de 12 clientes ativos da Dell EMC. Conforme exibido na Figura 4, os volumes de dados que precisavam de proteção nos ambientes dos clientes variaram de aproximadamente 270 TB a mais de 90 PB. Os clientes abrangeram diversos setores, inclusive tecnologia, indústria, seguros e saúde. A seleção de clientes em diferentes setores foi traçada para capturar os resultados de deduplicação em diferentes tipos de conjuntos de dados. As barras cinza na Figura 4 mostram o volume de dados protegidos para cada ambiente. As barras verdes mostram o volume de dados armazenados após a deduplicação.

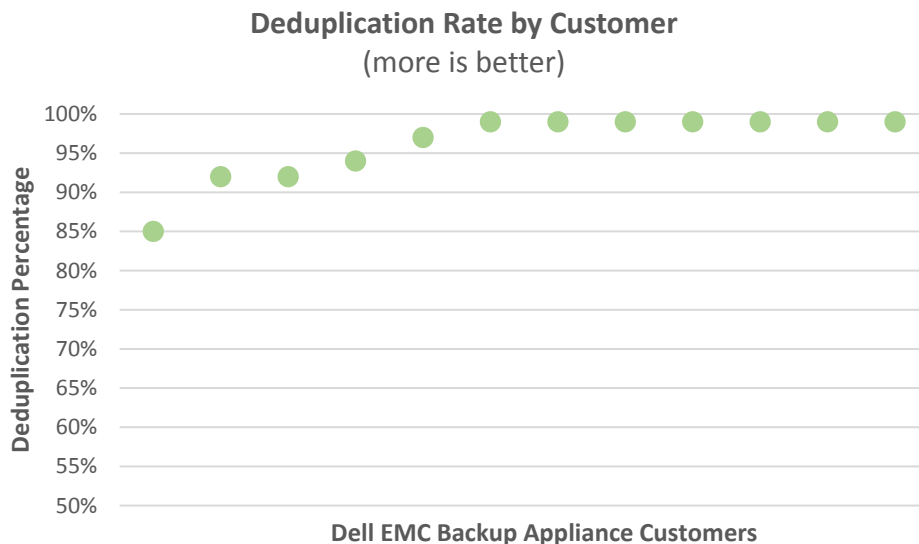
**Figura 4. Eficiência de deduplicação por cliente**

Fonte: Enterprise Strategy Group

- **Desduplicação:** a maior taxa de desduplicação observada nos dados analisados foi de 126:1. Esse resultado veio do ambiente do cliente identificado pelo primeiro ponto de dados no lado esquerdo da Figura 4. A taxa média de desduplicação para todos os clientes analisados foi de 41:1.
- **Proteção:** Até mesmo o cliente com a menor taxa de desduplicação observada, 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 outra visão sobre o efeito de desduplicação do Data Domain e IDPA. Essa figura mostra o percentual de desduplicação alcançado para cada um dos mesmos 12 ambientes dos clientes. As taxas de desduplicação variam de 85% a 99%. Normalmente, quanto maior for o tempo do recebimento de dados no ambiente pelo Data Domain e IDPA, maior será a taxa de desduplicação. Isso porque a probabilidade de encontrar os mesmos padrões de dados ao longo do tempo é maior, sendo necessário apenas armazenar dados exclusivos.

O recurso de segmentação de tamanho variável da arquitetura do Data Domain e IDPA é um componente-chave para atingir esse nível de desduplicação. Com a segmentação de tamanho variável, as soluções Data Domain e IDPA podem ser mais facilmente alinhadas aos padrões naturais, inclusive com marcadores de registro de data e hora de bancos de dados, nas estruturas de dados enviadas ao dispositivo para proteção. A segmentação de tamanho variável produz níveis significativamente mais altos de desduplicação do que as arquiteturas de tamanho fixo. Ela também viabiliza uma escalabilidade mais eficaz em um único pool de armazenamento. Isso também resulta em menos dispositivos para gerenciar, economizando nos custos operacionais.

**Figura 5. Percentual de desduplicação por cliente**

Fonte: Enterprise Strategy Group

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

- **Taxas de desduplicação:** as taxas de desduplicação auditadas e analisadas variaram de 85% a 99%, com uma média de 96%. Isso significa que apenas uma fração dos dados de produção (1% ou menos nas melhores situações) precisa ser armazenada nos sistemas Data Domain e IDPA para proteção.
- **Menos de 90%:** apenas um cliente entre os que tiveram os dados analisados teve uma taxa de desduplicação inferior a 90%. Novamente, esse cliente representa o ambiente altamente distribuído com pools de proteção distribuídos.
- **90% e acima:** os demais clientes analisados atingiram taxas de desduplicação superiores a 90%, e a maioria dos ambientes dos clientes estava funcionando a uma taxa de desduplicação de 99%.

## Velocidade e escalabilidade

O aumento da capacidade de armazenamento por meio de desduplicação nem sempre vem acompanhado de um alto desempenho. A combinação de Data Domain e IDPA resolve esse desafio ao aproveitar dois elementos-chave de desempenho: o SISL (Stream-Informed Segment Layout) e o software DD Boost.

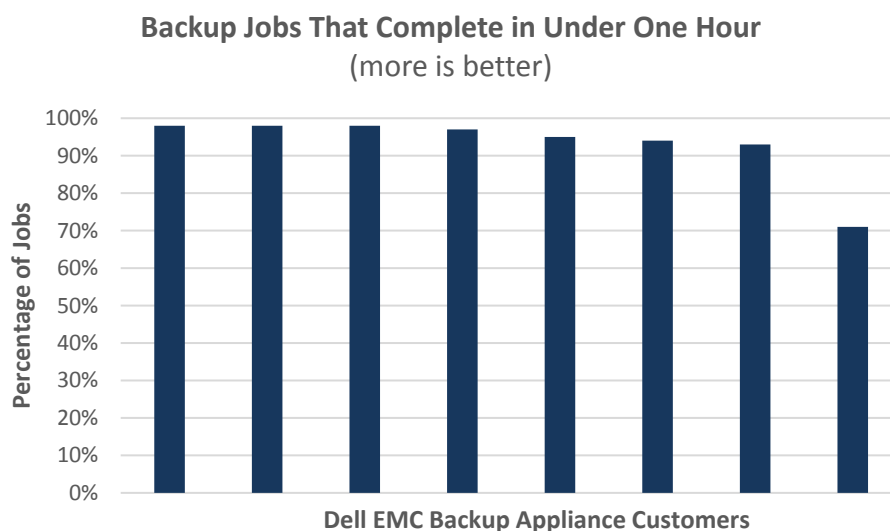
O SISL é um elemento arquitetural das plataformas Data Domain e IDPA que permite que 99% do processo de desduplicação ocorra na CPU e na RAM. Isso significa que a solução não depende de um alto número de eixos de disco para o desempenho. Como resultado, soluções com menos espaço ocupado podem atingir o mesmo desempenho de configurações com maior número de eixos. Além disso, como os sistemas Data Domain e IDPA usam CPUs Intel, eles se beneficiam dos aprimoramentos de desempenho de cada nova versão dos processadores Intel.

O software DD Boost também ajuda a concretizar a entrega de desempenho por meio da distribuição de partes do processo de desduplicação para os clients de aplicativos ou para o servidor de backup. Com o DD Boost, apenas dados exclusivos precisam ser movidos do servidor de backup ou dos clients para o sistema Data Domain e IDPA. O DD Boost



reduz em até 99% o volume de dados movido. O benefício líquido das conclusões mais rápidas dos trabalhos de backup foi auditado pelo ESG. Conforme apresentado na Figura 6, normalmente a maioria dos clientes conclui os trabalhos de backup em até 1 hora.

**Figura 6. Desempenho do trabalho de backup**

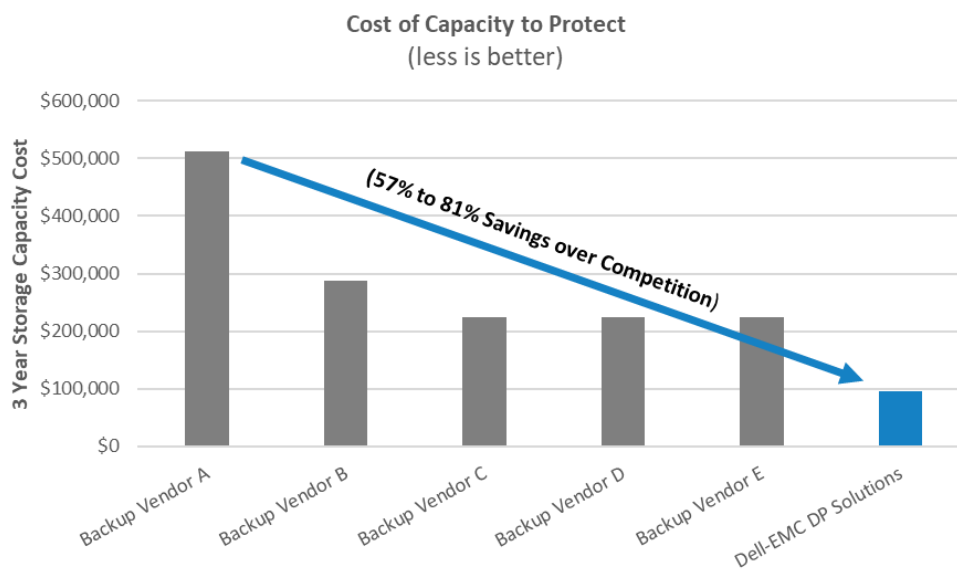


Fonte: Enterprise Strategy Group

- **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 8 dos 12 clientes analisados, de 70% a 98% de todos os trabalhos de backup foram concluídos em menos de 1 hora (conforme exibido 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 4 horas.
- **Mais de 4 horas:** para 2 dos maiores ambientes (em termos de número e tamanho de aplicativos), a conclusão de todos os trabalhos de backup demorou mais de 4 horas.

## Reflexões sobre a economia

O ESG realizou a exploração do valor econômico da economia da capacidade de desduplicação do Data Domain e IDPA por meio de auditoria e análise dos dados de suporte CallHome de mais de 15.000 appliances de backup Dell EMC Data Domain implementados em todo o mundo. Essa amostra foi coletada a partir de sistemas que também gravaram o software para backup que está sendo usado e as taxas de desduplicação que estão sendo alcançadas. Os dados mostram que a eficiência de desduplicação dos appliances de backup Dell EMC pode ser levada ainda mais adiante com uma solução completa de proteção de dados Dell EMC. Conforme apresentado na Figura 7, a combinação do hardware de appliance Dell EMC com o software de proteção Data Protection Suite ampliou a economia da duplicação e reduziu de 57% a 81% os custos de capacidade de armazenamento em 3 anos comparado a soluções de outros fornecedores.

**Figura 7. Ampliando a economia do Data Domain e IDPA com o Data Protection Suite**

Fonte: Enterprise Strategy Group

As soluções Data Domain e IDPA oferecem outros benefícios econômicos além da economia de capacidade de armazenamento apresentada na Figura 7. A análise de dados reais dos clientes feita pelo ESG também revelou os seguintes benefícios:

- **Redução de largura de banda da rede:** assim como a maioria dos aplicativos corporativos, as soluções de backup e recuperação dependem de recursos de rede para funcionar. Isso significa tanto recursos de LAN para backup e restauração locais quanto recursos de WAN para recuperação de desastres e continuidade de negócios. Soluções ineficientes de proteção de dados podem consumir rapidamente recursos valiosos da rede e até mesmo afetar negativamente a conectividade e a produtividade de usuários. Ao aproveitar a deduplicação do Data Domain e IDPA, uma fabricante multinacional conseguiu reduzir em 98% sua utilização de largura de banda local e remota para proteção de dados.
- **Desempenho:** a deduplicação eficiente de dados pode ter um grande impacto sobre a melhoria do desempenho geral da proteção de dados. Com a implementação do DD Boost, que reduz o volume de dados que precisa ser transferido entre o cliente e o sistema Data Domain, um fabricante de equipamentos pesados conseguiu melhorar em 50% suas velocidades de backup e restauração. O mesmo cliente melhorou seu preparo para recuperação de desastres em 90%.
- **Redução do espaço ocupado pelo datacenter:** o espaço físico, especialmente a quantidade necessária para viabilizar os requisitos de espaço, energia e refrigeração de um datacenter moderno, não é barato. Pode ser difícil, e algumas vezes impossível, expandir um datacenter fisicamente sem precisar mudar de local. Por exemplo, com o Data Domain uma cadeia nacional de lojas de departamento eliminou a fita física em seu ambiente de proteção de dados e liberou três fileiras inteiras de espaço altamente valioso de rack no datacenter.

## A grande verdade

As principais orientações de proteção de dados dos líderes de TI estão focadas no aprimoramento dos níveis básicos de confiabilidade e agilidade das soluções em uso. A orientação que vem logo atrás delas é a redução de custo, também vista como uma das principais prioridades entre os implementadores de proteção de dados. Esses desafios não devem ser vistos como contraditórios ou mutuamente exclusivos. Na verdade, todos eles podem ser resolvidos com o aprimoramento das soluções de proteção de dados desenvolvidas com foco tanto na eficiência quanto na confiabilidade e capacidade.

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

- A capacidade de fornecer armazenamento de proteção de maneira econômica por meio da redução do custo para atender o aplicativo de proteção de dados.
- Interoperabilidade entre as soluções de software de backup, já que apenas 1 em 4 organizações tem um único aplicativo de backup.<sup>4</sup>
- A escalabilidade e a eficiência de desduplicação do sistema de armazenamento de proteção, que podem ser usadas para reduzir o número (e o custo) dos sistemas que precisam ser implementados e gerenciados.
- A capacidade de aproveitar o mesmo pool de armazenamento de proteção com capacidade otimizada para backup e arquivamento de dados.
- Interoperabilidade (e integração) entre o armazenamento de proteção e a variedade de plataformas, para que os proprietários de aplicativos (p. ex., administradores de banco de dados e vAdmins) possam utilizar as próprias ferramentas para proteção/recuperação complementar, enquanto continuam sendo cidadãos responsáveis da TI e aproveitando o armazenamento eficiente da proteção centralizada.
- A capacidade de acelerar a implementação com soluções integradas de proteção de dados, reduzindo o número de sistemas que precisa ser implantado, simplificando o licenciamento com uma abordagem abrangente e reduzindo os problemas de interoperabilidade.
- A capacidade de simplificar a administração geral de ambientes de proteção de dados com gerenciamento e relatórios centralizados — além de atualização/patches de etapa única para todos os componentes de proteção de dados no appliance — o que reduz o tempo necessário para manter o ambiente.
- A capacidade de integrar recursos mais amplos, como nuvem (retenção a longo prazo, recuperação de desastres em nuvem), pesquisa e lógica analítica, que reduz mais uma vez os tempos de implementação e simplifica o gerenciamento.

Ao oferecer o que muitos do setor consideram sinônimo de “armazenamento de proteção” e/ou “desduplicação”, a Dell EMC deixou de simplesmente fornecer componentes de proteção de dados e passou a fornecer um ecossistema completo que inclui armazenamento da produção, fornecedores de backup, fornecedores de arquivamento e plug-ins aceleradores para aplicativos (ou seja, DD Boost), entre outros elementos. Agora o Dell EMC IDPA faz parte desse ecossistema com software de proteção, servidores e armazenamento, gerenciamento a partir de uma interface de usuário, recursos em nuvem e pesquisa e lógica analítica, tudo incluído e pré-configurado. Desse modo, há um ímpeto significativo para manter a inovação à frente da demanda atual do mercado, resultando em maiores ganhos de eficiência nos mecanismos de desduplicação, aprimoramentos em suas integrações e expansões de suas capacidades além da simples desduplicação no

<sup>4</sup> Fonte: pesquisa de levantamento do ESG, *Data Protection Modernization Trends*, dezembro de 2016.

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 no repositório, já que há muita dependência da integridade de cada block exclusivo no sistema. Sendo assim, é essencial que as organizações investiguem com atenção o nível de zelo e a frequência das verificações do sistema em relação ao armazenamento subjacente para a garantia de integridade. A abordagem da Dell EMC para isso tem sido uma característica marcante de sua plataforma de appliance de backup desde sua criação até o que hoje em dia é conhecido por DIA (Data Invulnerability Architecture).

Com base em uma auditoria de dados coletados em campo a partir de mais de 15.000 ambientes de produção, o ESG confirmou que a combinação do hardware Dell EMC Data Domain e IDPA com o software Data Protection Suite proporcionar uma redução de 57% a 81% no custo da capacidade de proteção comparado aos ambientes de appliance de backup Dell EMC que usam software de backup da concorrência. Organizações que estão indecisas quanto a fazer um investimento no Dell EMC ou IDPA com base no “preço” se beneficiariam se reconsiderassem os benefícios econômicos ao longo do tempo, inclusive o custo das reduções de tempo de inatividade e perda de dados, o aprimoramento 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 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.

© 2018 por The Enterprise Strategy Group, Inc. Todos os direitos reservados.

