

## WHITE PAPER

# A Compressão em Tempo Real Promove a Otimização do Armazenamento

Patrocinado pela: IBM

Laura DuBois

Mai de 2012

## SUMÁRIO EXECUTIVO

Não é segredo que orçamentos para armazenamento estão sob pressão constante de forças opostas: por um lado, as forças econômicas estão pressionando os orçamentos para que fiquem achatados ou, em muitos casos, encolham como uma porcentagem da receita da empresa. Por outro lado, a infraestrutura luta para manter o ritmo do crescimento das informações, pressionada por muitas variáveis, tanto sociais quanto econômicas. As empresas não têm outra escolha senão adaptar a sua infraestrutura de armazenamento aos níveis sem precedentes em que as informações estão crescendo.

Na maior parte dos casos, a ineficiência da infraestrutura de computação foi enfrentada através da virtualização do servidor. Vários bastidores de servidores ociosos foram sendo substituídos aos poucos pelos ambientes de computação virtualizados ultradensos que ocupam menos espaço, sobrecarregando os fornecedores de armazenamento ao lidar com o desafio de tornar o armazenamento de dados eficiente e economicamente sustentável.

Os fornecedores de armazenamento reagiram prontamente ao desenvolver um conjunto de soluções direcionadas para diminuir o ritmo do consumo de armazenamento. As tecnologias de otimização de armazenamento, como são chamadas, procuram reduzir a área ocupada pelos equipamentos de armazenamento ao remover redundância e desperdício e ao otimizar a alocação dos dados. Tradicionalmente, os conjuntos de otimização de armazenamento consistiam em thin provisioning, deduplicação e pós-compactação. Thin provisioning procura reduzir o espaço desperdiçado que era alocado, mas não utilizado por causa do excesso de provisionamento. Similarmente, a deduplicação tem como alvo dados redundantes ou duplos, criando instâncias únicas de dados. Para dados que não podem ser “afinados” nem “deduplicados”, a compressão tem como alvo a redução da área de armazenamento geral ao comprimir os dados para diminuí-los. A promessa é que ao implantar uma ou mais — ou todas — essas tecnologias, as empresas podem reduzir seus gastos gerais com armazenamento, retardar compras futuras e obter um retorno melhor para seus investimentos.

Na realidade, no entanto, toda otimização coloca certa pressão sobre o sistema em termos de desempenho ou de eficiência. Historicamente, o que mais prejudica nessa mistura é a compressão, por causa de sua natureza inerente de pós-processamento intensivo de computação. Com frequência isso conduziu as empresas a implantar thin provisioning pequeno e/ou deduplicação, mas deixando a compressão fora da mistura quando se trata de armazenamento primário ativo.

## NESTE INFORME

Neste informe nós examinamos como a Compressão em Tempo Real da IBM está virando o jogo ao oferecer uma solução de otimização robusta, eficiente e econômica. Além disso, a IBM está demonstrando que compressão tem lugar em otimização de armazenamento e complementa de modo desejável outras técnicas de otimização em uso atualmente. A IBM estima que ao implantar as soluções com Compressão em Tempo

Real, como o IBM Storwize V7000, as empresas podem economizar 50% ou mais na quantidade de espaço físico necessário e reduzir o crescimento geral do armazenamento em quase 30%. De acordo com a IBM, isto tipicamente resulta em uma redução de 30 a 40% em custos por gigabyte (GB) para configurações de armazenamento primário, sem comprometer nenhuma capacidade ou desempenho.

O IBM Real-time Compression é um componente essencial do IBM Smarter Storage que, por sua vez, faz parte da estratégia Smarter Computing da IBM, uma abordagem desenvolvida em anos de liderança e inovação no segmento de armazenamento e uma tecnologia avançada que ajuda a direcionar o projeto e a implantação de sistemas de armazenamento. Smarter Storage permite que as organizações assumam o controle de seus armazenamentos, de modo que possam se concentrar em compreender melhor seus dados, agregando mais valor para os negócios.

---

### **Situação do Mercado**

Vivemos em um mundo pós-PC. Os dispositivos de computação móvel continuam a proliferar em toda parte, inclusive nas empresas. Até 2015, o mercado de computação móvel consistirá em mais de 243 milhões de tablets e mais de 1,5 bilhões de smartphones. De acordo com o IDC, cerca de 9.57 zettabytes (ZB) de informações são consumidos anualmente, dos quais 1.2 ZB são dados de mídia desestruturada que estão crescendo a um índice anual de crescimento composto de 62% (CAGR - taxa composta de crescimento anual). À medida que a tendência nas empresa, prosumer (produtor + consumidor) e do consumidor mudam para a criação de dados e acesso em trânsito dos mesmos, assim também muda a demanda por mais armazenamento. A pesquisa do IDC mostra que os sistemas mundiais de armazenamento em disco externo apresentaram um crescimento ano a ano de 7,7%, totalizando pouco menos de US\$ 6,6 bilhões no quarto trimestre de 2011. A capacidade total dos sistemas de armazenamento em disco em 2011 alcançou 6.279 petabytes, crescendo 22,4% ano a ano.

Com a demanda das empresas por capacidade de armazenamento no mundo todo projetada para crescer a um CAGR de mais de 43% de 2008 a 2013, a demanda por dados supera amplamente o fornecimento de armazenamento. Além disso, a lentidão persistente na economia significa que as empresas não podem mais garantir a acomodação desse crescimento de dados utilizando as abordagens tradicionais de armazenamento. Já passou o tempo de tratar armazenamento somente com mecanismos adequados de proteção e de desempenho tais como os grupos RAID e gerenciadores de volume. Os custos por GB por disco podem estar caindo, mas os custos de infraestrutura estão subindo, forçando as empresas a descartar as abordagens tradicionais para armazenamento de dados em favor de abordagens mais modernas e melhores.

Os fornecedores de armazenamento seguiram amplamente o exemplo estabelecido pelos fornecedores de virtualização de servidor ao criar um ecossistema de tecnologias que aprimora a utilização da infraestrutura de armazenamento, tornando-a mais eficiente no processo. A prova da adoção crescente pelo mercado de tais tecnologias está no fato de que atualmente quase todas as soluções de armazenamento são concebidas com uma ou mais tecnologias de otimização de armazenamento que podem ser implantadas prontamente. Embora possa ser um desafio medir o tamanho do mercado de otimização de armazenamento como uma fração do mercado geral de armazenamento, a taxa de adoção dessas tecnologias significa que, com o tempo, todos os armazenamentos instalados serão otimizados de um modo ou de outro.

### **Por que a otimização de armazenamento não é mais opcional?**

Enquanto o motivador primário para otimização de armazenamento pode ser um grande crescimento de dados, a otimização de armazenamento também é influenciada em grande parte pelas ineficiências introduzidas por como os dados são criados, acessados e/ou armazenados. Algumas dessas ineficiências possuem um elemento humano, mas outras são subprodutos de sobrecarga de tecnologia.

- ☒ A quantidade de dados estáticos (ou seja, dados que mal são acessados regularmente) continua a crescer a um ritmo muito mais rápido do que a quantidade de dados “quentes” ativos (ou seja, dados que são acessados o tempo todo).
- ☒ Há uma desenfreada duplicação de dados, que são criados graças a tecnologias mais modernas. Por exemplo, a virtualização de servidores pode criar imagens de sistemas operacionais que são quase idênticos uns aos outros. Outro exemplo está em usuários que criam cópias duplicadas de imagens, arquivos e outros dados.
- ☒ Alguns dos formatos utilizados para a criação de conteúdo estático são nativamente ineficientes (isto é, o formato cria uma grande quantidade de espaço vazio). Isto resulta em espaço redundante no disco.
- ☒ O armazenamento de dados estruturados, tais como bancos de dados, pode resultar em ineficiências no que diz respeito a armazenamento por causa dos elementos de metadados que suportam vários atributos para esses dados.
- ☒ As empresas podem ter tipos distintos de armazenamento (e, às vezes, de vários fornecedores) que criaram silos na infraestrutura. Devido a diferentes orientações de interoperabilidade ou considerações de desempenho, as empresas podem achar difícil implantar esses bens de forma homogênea, resultando em que se tornem severamente subutilizados.
- ☒ A classificação por níveis parece uma forma ideal para dimensionar de maneira correta a infraestrutura de armazenamento, mas na prática a falta de mecanismos automáticos de hierarquização apresenta um desafio para os administradores de armazenamento moverem dados de um nível para outro. Isso resulta em uma utilização ineficiente dos níveis de armazenamento e um menor benefício de hierarquização.

As tecnologias de otimização de armazenamento são destinadas diretamente ao aprimoramento da utilização dos ativos ao alterar a colocação e a organização dos dados. Como uma camada adicional da tecnologia, a otimização de armazenamento é, em última análise, projetada para superar as ineficiências introduzidas por seres humanos ou pela própria tecnologia:

- ☒ **Hierarquização automatizada** é a capacidade de mover apenas porções ativas de dados para níveis mais elevados (e mais caros) enquanto que a maioria menos ativa dos dados permanece em níveis mais baixos de armazenamento. O princípio básico é que enquanto os dados podem crescer, uma porcentagem cada vez maior de dados é de dados estáticos ou fixos que não precisam residir em camadas de alto desempenho em todos os momentos.
- ☒ **A virtualização de armazenamento** é a capacidade de reunir recursos de armazenamento diferentes e muitas vezes de vários fornecedores sob uma gestão e uma forma de visualização única.
- ☒ **Thin provisioning** é a capacidade de definir uma unidade de armazenamento (sistema completo, conjunto de armazenamento ou volume) com um tamanho de capacidade lógico que é maior do que a capacidade física atribuída à unidade de armazenamento. O host ou dispositivo que acessam a unidade de armazenamento veem a capacidade lógica e não a capacidade física.

- ☒ Deduplicação é a capacidade de analisar um bloco de dados único ou arquivo ou uma série de blocos de dados ou arquivos procurando por padrões comuns e substituí-los ou direcioná-los para uma única instância daquele padrão, reduzindo assim a duplicação de tais padrões na base de armazenamento. Devido à natureza de bloquear o acesso, a deduplicação é oferecida principalmente em armazenamento baseado em arquivo.
- ☒ A compressão é a capacidade de comprimir os dados para que os blocos se tornem menores. A utilização de compressão permite que esses dados consumam muito menos armazenamento em comparação com outras tecnologias de otimização, como a deduplicação ou o thin provisioning.

Como os motivadores e técnicas de otimização de armazenamento demonstram, a otimização de armazenamento não é uma abordagem que serve para tudo, ao contrário de algumas outras tecnologias de armazenamento, tais como replicação, clones de matriz e assim por diante. A natureza diversa dos dados em si torna algumas técnicas de otimização de armazenamento mais adequadas para determinados tipos de dados do que outras. No entanto, ao utilizar tecnologias de otimização de armazenamento como um pacote em uma infraestrutura de armazenamento compartilhado, as empresas podem obter importantes benefícios tangíveis, sem comprometer a qualidade do serviço.

---

## **Compressão em Tempo Real**

Antes de destacar os benefícios da Compressão em Tempo Real, vamos primeiro examinar porque a compressão é muitas vezes excluída da lista quando se trata de otimização de armazenamento. Muitos fornecedores são rápidos em apontar que, ao contrário do thin provisioning ou da deduplicação, a compressão pode ter resultados mistos e, por vezes, adversos. Isso ocorre em parte porque a abordagem tradicional força os fornecedores para conseguir um equilíbrio delicado entre a compressibilidade e a penalidade de desempenho, resultando em desempenho abaixo do esperado. É por isso que geralmente é implementada somente para dados que são utilizados com pouca frequência.

### ***Desafios da Utilização da Compressão***

A abordagem tradicional e comumente adotada é a de comprimir os dados fora da produção. Este tipo de compressão, que também é conhecido como compressão pós-processo, entra em ação depois que os dados já foram gravados no disco. Isto é muito semelhante a vários utilitários com base em servidores que compactam arquivos e pastas, assim que são criados. Este método é inerentemente ineficiente, pois seus algoritmos consomem ciclos relevantes de computação, são intensivos para disco e requerem um "delta" adicional de espaço para armazenar os dados não comprimidos. Em um servidor, na maioria das vezes, pode-se ir mais longe com esse impacto em função do poder do processamento adicional, mas em uma série de armazenamento dedicado, uma sobrecarga adicional desse tipo pode facilmente produzir uma degradação perceptível na qualidade do serviço, o que é automaticamente observado no servidor e nos aplicativos. Como resultado, os fornecedores são rápidos em não valorizar a compressão na maioria das situações, a menos que o conjunto de dados seja insignificante ou o sistema de armazenamento como um todo esteja subutilizado.

Na compressão tradicional, quando os aplicativos fazem várias atualizações de dados, tais mudanças são gravadas no disco de forma descompactada. Posteriormente, outra operação tem de ser agendada e comprimir os dados com base em sua localização física em um volume, seja qual for a sua relação com outros blocos de dados que o aplicativo possa estar acessando.

Os resultados produzidos por mecanismos de compressão tradicionais também variam muito. Tais mecanismos de compressão leem um pedaço fixo de dados pré-determinados e produzem uma saída variável, dependendo da compressibilidade desses dados. Além disso, as taxas de compressão dependem dos tamanhos dos blocos: quanto maior o bloco, maior a taxa de compressão e maior a sobrecarga do desempenho. Blocos de tamanhos menores resultam em taxas menores de compressão. O meio-termo entre os tamanhos do bloco e a sobrecarga de desempenho não é tão bom no final.

Um efeito colateral da compressão é que resulta em fragmentação. Devido ao pós-processo e à natureza variável dos mecanismos de compressão tradicionais, a compressão “degenera” a continuidade dos dados ao longo do tempo. Os dados comprimidos são espalhados em pedaços em todo o volume e necessitam de coleta de lixo frequente. O impacto deste efeito é o impacto no desempenho ao longo do tempo.

### ***Compressão em tempo real revigora o papel da compressão na otimização do armazenamento***

Atualmente, a maioria dos fornecedores oferece compressão como um recurso opcional licenciável, mas recomenda-se firmemente que os administradores habilitem apenas em conjuntos de dados selecionados. Na maioria das situações, a adoção de compressão para dados primários ativos na infraestrutura de armazenamento em geral continua a ser limitada. Uma vez que cada penalidade no desempenho do sistema de armazenamento pode ter um efeito cascata sobre o resto dos aplicativos, a maioria das empresas opta por deixar a compressão desativada.

Uma nova abordagem, conhecida como Compressão em Tempo Real, pode estar colocando em evidência novamente o papel da compressão na otimização do armazenamento. A Compressão em Tempo Real torna viável a utilização da compressão sem penalidades operacionais. Ela promete taxas de compressão de até cinco vezes para dados em produção. Além disso, pela redução de dados primários, ela também comprime todas as cópias derivadas desses dados, como backups, arquivos, instantâneos e réplicas. A IBM implementou a tecnologia no sistema de armazenamento Storwize V7000, eliminando a necessidade de aplicativos externos.

O IDC espera que outros fornecedores sigam a liderança da IBM e se afastem da compressão tradicional em repouso e desenvolvam tecnologias semelhantes mas competitivas de compressão. Isso irá validar e reforçar o papel de compressão como um componente essencial da eficiência. De fato, à medida que os processadores de armazenamento se tornam mais rápidos e mais poderosos, os fornecedores de armazenamento podem direcionar ciclos de compressão para núcleos ou processadores dedicados no controlador de armazenamento, além de ciclos de deduplicação.

### ***O que é a Compressão em Tempo Real e quais são os seus Benefícios?***

Ao contrário das tecnologias tradicionais de compressão que comprimem os dados como uma operação pós-processo, a Compressão em Tempo Real opera em dados primários ativos à medida que estão sendo acessados. Isso expande a esfera de compressão a um conjunto muito maior de cargas de trabalho com resultados previsíveis e mensuráveis. Além disso, essa compressão está “sempre ligada”, o que significa que pode ser habilitada em cargas de trabalho ativas e não necessita de períodos previstos para pós-processamento, ao contrário de seus antecessores.

O principal diferenciador é que na Compressão em Tempo Real o mecanismo de compressão processa um fluxo de dados variáveis com base nos padrões de dados que são realmente gravados. A Compressão em Tempo Real aproveita a “localidade temporal” e não a localização física: Os dados que são acessados em conjunto são comprimidos juntos, seja qual for a sua localização física. Portanto, quando as aplicações fizerem atualizações relacionadas a diferentes partes do volume, essas atualizações serão tratadas em conjunto e de forma contígua. Isto é semelhante à operação do sistema real porque tira proveito da estrutura dos dados, do tamanho dos dados e da relação dos dados com outros dados. Essa consciência da carga de trabalho minimiza a quantidade de operações de compressão e/ou descompressão, resultando em menos IOPS (operações de entrada e saída por segundo) de disco e menos sobrecarga no controlador de armazenamento. Isso aumenta os índices de compressão e a eficiência, sem comprometer o desempenho.

---

### **Implementação do IBM Storwize V7000**

A IBM adquiriu a tecnologia de Compressão em Tempo Real através da aquisição da empresa Storwize em 2010. Inicialmente, a IBM ofereceu essa tecnologia através de dispositivos desenvolvidos especificamente para comprimir de modo transparente em ambientes NAS. A IBM ainda vende esses dispositivos atualmente, mas também deu um passo à frente ao implementar a tecnologia dentro da sua plataforma Storwize V7000. A plataforma Storwize V7000 utilizará a mesma tecnologia do Mecanismo de Compressão de Acesso Aleatório (RACE) que os dispositivos dedicados utilizam.

A IBM planeja introduzir a compressão no Storwize V7000 como um componente licenciável opcional através da atualização de um firmware. Assim, os usuários existentes de Storwize V7000 conseguirão selecionar compressão como um atributo ao criar novos volumes. Os administradores também têm a opção de converter volumes existentes utilizando espelhamento de volume para comprimir volumes mal provisionados e eliminar o espaço não utilizado no processo. A IBM inicialmente planeja aceitar até 200 volumes comprimidos, mas pode aumentar esse limite com os lançamentos posteriores. A interface gráfica com o usuário do Storwize V7000 apresentará informações sobre desempenho relacionado com compressão e permitirá que os administradores gerenciem e monitorem compressão a partir de um único console. A Compressão em Tempo Real aprimora a funcionalidade do Storwize V7000 sem criar sobrecarga administrativa adicional.

### ***Como a Compressão no Storwize V7000 se compara com a Abordagem Tradicional?***

Conforme observado anteriormente, as abordagens tradicionais para compressão vêm tendo uma má reputação por seu desempenho imprevisível, sobrecarga e a natureza pesada de pós-processamento. A Compressão em Tempo Real, por outro lado, altera a equação ao criar uma camada de otimização em linha e eficiente. O impacto imediato da implementação da Compressão em Tempo Real é que o sistema de armazenamento produz resultados com menos IOPS durante a rotina de compressão. Cargas de trabalho ativas, tais como bancos de dados e sistemas de e-mail muitas vezes executam pequenas atualizações nos dados existentes. A análise da IBM sugere a melhoria relevante quando comparada com as abordagens tradicionais para essas cargas de trabalho (consulte a Tabela 1).

## TABELA 1

Comparação da Compressão Tradicional com a Compressão em Tempo Real da IBM

Compressão Tradicional	Compressão em Tempo Real da IBM
“pedaço” de 1 MB	Atualização de 100 Bytes
Leitura de 1 MB	Leitura de 0 MB
Descompressão de 1 MB	Descompressão de 0 MB
Atualização de 100 Bytes	Atualização de 0 Byte
Compressão de 1 MB	Compressão de 100 Bytes
Gravação de 1 MB	Gravação de < 100 Bytes
E/S total por operação de compressão: 2 MB	E/S total por operação de compressão: < 100 Bytes

Fonte: IBM, 2012

A Compressão em Tempo Real Storwize V7000 alcança índices de compressão semelhantes aos dos dispositivos de Compressão em Tempo Real da IBM (consulte a Tabela 2). Uma vez que a Compressão em Tempo Real pode ser implementada com uma variedade de tipos de dados maior do que a compressão tradicional, os benefícios potenciais podem ser maiores. Taxas de compressão previsíveis tornam mais fácil para as empresas planejar os orçamentos dedicados ao sistema de armazenamento.

## TABELA 2

Índices de Compressão Observados pela IBM em Ambientes dos Clientes

Aplicativos		Compressão Observada
Bancos de dados		Até 80%
Servidores Virtuais (VMware)	Sistema Operacional Linux virtual	Até 70%
	Sistema Operacional Windows virtual	Até 50%
Escritório	2003	Até 75%
	2007 ou posteriores	Até 20%
CAD/CAM		Até 70%

Fonte: IBM, 2012

Ao possibilitar a compressão em volumes novos ou existentes no Storwize V7000, as empresas podem extrair maior capacidade utilizável de seus investimentos existentes. Isso permite que muitas empresas achatem ou reduzam seus gastos em armazenamento para a maioria das configurações comuns.

Um benefício adicional da utilização da Compressão em Tempo Real com o Storwize V7000 é que mesmo volumes extraídos de armazenamento virtualizado externo podem ser comprimidos. Em muitos casos, isso representa quase o dobro da quantidade de capacidade utilizável para investimentos modestos em CAPEX e menores custos de OPEX.

---

## **Desafios e Oportunidades para o Storwize V7000**

Nos tempos atuais, quando os orçamentos de armazenamento são limitados ou reduzidos, as organizações de TI continuam a procurar maneiras de armazenar de modo eficiente e rápido o crescimento de dados a custos mais baixos. A Compressão em Tempo Real, embora atraente, ainda está na fase inicial de adoção. Muitas organizações continuam a considerar a compressão como uma tecnologia intensiva de pós-processamento de computação.

Um dos principais desafios que as organizações enfrentam atualmente ao utilizar a Compressão em Tempo Real é o do risco versus a recompensa. As organizações certamente exigem respostas para perguntas como a seguinte: Quanto devo pagar para comprimir os meus dados e que economias vou ver como resultado? Quis são os meus riscos ao comprimir os meus dados? Quais são as penalidades do desempenho? Conseguirei manter os índices de compressão ou eles se degeneram ao longo do tempo? Ser capaz de quantificar a economia será um grande fator para decidir se vale a pena adquirir ou não o software. As organizações podem recorrer à ferramenta Compressimator da IBM para obter uma estimativa dos benefícios esperados da compressão para ambientes específicos. Utilizando os resultados e as taxas de crescimento esperados, as organizações podem obter economias potenciais através da implantação do Storwize V7000 com a compressão ou a habilitação da compressão em volumes existentes. A IBM pode proativamente utilizar esta ferramenta para fornecer aos clientes existentes e futuros clientes uma visualização do potencial de economia.

A tarefa que se apresenta para a IBM é ajudar as organizações a remover a sua visão de compressão como uma tarefa inerente de pós-processo. A IBM pode continuar a obter reconhecimento entre seus clientes ao orientá-los sobre as práticas recomendadas e utilizar casos para a implantação da Compressão em Tempo Real, criando documentos que descrevam a utilização da compressão juntamente com outras tecnologias de otimização de armazenamento, tais como thin provisioning, hierarquização de armazenamento e deduplicação.

## **CONCLUSÃO E ORIENTAÇÃO ESSENCIAL**

As tecnologias de otimização de armazenamento chegaram para ficar. Graças à Compressão em Tempo Real, a compressão tem o seu lugar no portfólio de otimização de armazenamento. A IBM deu um passo na direção certa ao oferecer essa tecnologia em um de seus principais produtos de armazenamento.

A natureza previsível, confiável e linearmente escalável da Compressão em Tempo Real irá impulsionar a sua adoção em ambientes com cargas de trabalho diversas. Isto, por sua vez, conduzirá as empresas a saírem de suas zonas de conforto, tornando a Compressão em Tempo Real em uma das tecnologias obrigatórias em seus ambientes. Ser capaz de quantificar, prever e medir as economias apresentadas pelas tecnologias de otimização de armazenamento como a Compressão em Tempo Real irá equipar melhor as empresas para enfrentar o crescimento explosivo dos dados.

---

## **Direito de Copyright**

Publicação Externa de Informação e Dados da IDC — Qualquer informação da IDC a ser utilizado em propaganda, comunicados da imprensa ou materiais promocionais requer aprovação prévia por escrito do Vice-Presidente ou Diretor Regional adequado da IDC. Um rascunho do documento proposto deve acompanhar qualquer pedido relacionado. A IDC se reserva ao direito de negar a aprovação do uso externo não importando o motivo.

Copyright 2012 IDC. A reprodução sem a permissão por escrito é completamente proibida.