

Porfólio IBM com abordagem para backup e recuperação

Por John Webster

15 de agosto de 2012



Evaluator Group

Permitindo que você tome as melhores decisões de tecnologia



O backup e a recuperação de sistemas de TI, aplicativos e dados sempre foram os principais requisitos de TI para corporações de médio a grande porte. Ao mesmo tempo que percebemos que os administradores de TI podem estar descontentes com a sua capacidade de fazer backup e recuperar todos os dados corporativos que necessitam de proteção, os processos funcionais estão instaurados.

Duas forças principais, agora convergentes no datacenter, impõem uma pressão crescente nos administradores de TI para reconstruírem seus processos de proteção de dados: o crescimento acelerado dos dados no escopo de TI da corporação; e a consolidação e virtualização contínuas do ambiente do servidor conduzida pelos projetos de virtualização de servidores com base em VMware.

Aceleração do crescimento de dados

Estamos testemunhando o crescimento acelerado dos dados armazenados em datacenters de médio a grande porte, o que fará com que as relativamente normais taxas de crescimento anual compostas de 50% a 80% pareçam moderadas. Este crescimento acelerado é proveniente de muitas fontes, como:

- Expansão dos negócios e a demanda do gerenciamento de alto nível por respostas rápidas às oportunidades de negócios emergentes
- Necessidade de alavancar tipos adicionais de dados, como dados não estruturados, rich media e dados entre máquinas (RFID, GPS, etc.) que alimentam os novos aplicativos e sistemas de análise.
- Influxo de dados criado fora do datacenter, gerado pela web e por dispositivos de computação móvel
- Necessidade de reter mais informações por mais tempo, conduzida pelas fontes tradicionais de retenção de dados, incluindo auditoria e conformidade, bem como as necessidades de novos usuários de aplicativos para analisar a mudança ao longo do tempo dentro de suas áreas principais (mercados, populações, pesquisa científica, etc.).

Em função da crescente demanda por capacidade de armazenamento que vemos em todos os setores, acreditamos que as práticas atuais de backup e recuperação são insustentáveis. Na verdade, muitas organizações perceberam que os seus métodos de proteção de dados são insustentáveis, principalmente devido ao impacto da crescente demanda por capacidade e de restrições de orçamento.

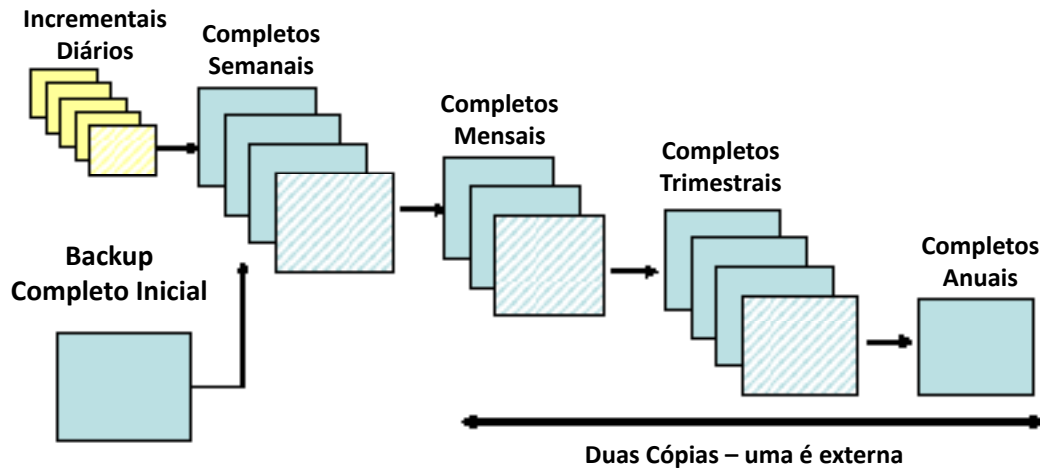


Figura 1: Ciclos de backup de datacenters tradicionais

Os processos de backup não são adequados, gerando maior exposição à perda de dados. A recuperação se torna mais complexa e lenta. Esse aumento causará impactos multifacetados e significativos sobre os processos atuais de proteção de dados (veja a Figura 1 acima) se for permitido que sigam sem resolução:

- A crescente quantidade de dados incluiu o processo de backup, aumentando a pressão sobre os gerentes da infraestrutura de armazenamento, os quais possivelmente estão se esforçando para garantir que os dados corporativos sejam devidamente protegidos.
- O orçamento destinado à proteção de dados, já responsável por metade dos orçamentos de armazenamento de muitas corporações, continuará a crescer
- O tempo necessário para backup dos dados pode exceder o nível prático do ponto de vista das operações de TI.
- A quantidade de dados a ser protegida pode exigir mais sistemas de backup do que os que podem ser fisicamente acomodados.
- Os novos projetos de TI podem ser afetados à medida que os administradores percebem que estão competindo por recursos de proteção de dados que já estão sobrecarregados.
- Os limites do licenciamento de capacidade podem ser superados – tanto interna quanto externamente – e esgotar o orçamento sem necessidade.

Virtualização do servidor

O backup e a recuperação dos servidores físicos e seus armazenamentos de dados associados podem ser razoavelmente simples quando comparados à entrega do mesmo nível de proteção para servidores virtualizados, bem como as máquinas virtuais e os aplicativos que eles abrangem. O VMware, por exemplo, força os administradores de armazenamento, no mínimo, a modificarem esses processos e a encontrarem novas formas de ampliar o espectro de proteção de dados para servidores virtualizados.

Não há somente um requisito para fazer o backup de um servidor virtualizado inteiro, é preciso também direcionar o backup e restaurar os objetivos – do tempo de recuperação (*recovery time objectives* ou RTO) e do ponto de recuperação (*recovery point objectives* ou RPO) – para cada máquina virtual e/ou aplicativo em execução em cada máquina virtual. A facilidade e a velocidade com que novas máquinas virtuais podem ser implementadas aumentam rapidamente a complexidade e o custo para a proteção de ambientes mistos de servidores físicos e virtuais.

Proteção de dados no portfólio IBM

Levando em consideração a convergência das forças mencionadas acima, é necessária agora uma abordagem holística para implementar a proteção de dados em diversos níveis. A utilização de uma abordagem de nível de sistema reduz o volume de dados de backup por meio do uso sistemático de tecnologias de armazenamento que maximizem sua eficiência e minimizem a complexidade, a saber:

- Estrutura de software que gerencia a proteção de dados
- Compactação e deduplicação
- Processos de captura instantânea e replicação que estão na categoria de Proteção Contínua de Dados (*Continuous Data Protection* ou CDP)
- Controle de cópias via FlashCopy Manager
- Definição automática de camada que inclui a integração com um archive ativo
- Virtualização
- Nuvem via Active Cloud Engine

O objetivo dos administradores de armazenamento que adotam esta abordagem é estabelecer uma estratégia de proteção de dados que simplifique e conseqüentemente reduza a dificuldade de gerenciamento, ou seja, menos processos por unidade de dados primários por meio de automação. Agora, veremos mais detalhadamente os recursos integrados ao portfólio de produtos de proteção de dados da IBM que contemplam a próxima geração de estratégias de proteção de dados necessárias para dar suporte ao crescimento de dados e ao avanço da virtualização dos servidores.

Estrutura do software de proteção de dados - TSM (Tivoli Storage Manager)

É melhor começar esta discussão sobre o portfólio de produtos com a principal estrutura do software de proteção de dados da IBM: Tivoli Storage Manager. Chamamos de TSM uma estrutura de aplicativo de software de gerenciamento porque é na verdade uma coleção de produtos para proteção de dados que abrange muitas áreas com versões específicas para determinados usos. Aqui, destacamos três produtos destinados ao aperfeiçoamento do processo de backup no contexto do rápido crescimento de dados e da capacidade de executar a proteção de dados em escala de petabytes, que envolvem o avanço da virtualização de servidores dentro do círculo de proteção de dados corporativos.

TSM FastBack

O FastBack fornece proteção contínua de dados (CDP) para armazenamento com base em bloco. (A proteção contínua de dados em nível de arquivo é oferecida via FastBack for Files). O FastBack trabalha em nível de bloco e implementa o backup incremental a partir do disco principal para o disco secundário. Como os blocos são atualizados, somente as alterações para os blocos de dados primários são refletidas no secundário, cópia backup para disco, reduzindo significativamente a quantidade de dados transferidos no processo de backup e disponibilizando, de imediato, uma cópia atual dos dados para quando precisarem ser recuperados. O FastBack inclui a deduplicação de destino (pós-processamento), controles de política para automatizar os processos CDP e suporte para backup de escritórios remotos.

TSM FlashCopy Manager

O TSM FlashCopy Manager gerencia as capturas instantâneas do FlashCopy com base em sistemas de armazenamento IBM e em processos de recuperação que utilizam o FlashCopy, assim como as capturas instantâneas XIV. Automatizado, com base em políticas e integrado com o VMware VADP para as capturas instantâneas das máquinas virtuais (veja abaixo). Ele também suporta ambientes de servidor agrupado como o IBM HACMP e o Microsoft Cluster Server.

Com o FlashCopy Manager, as cópias point-in-time (captura instantânea) dos dados são feitas sem interrupções nas operações de produção. Ele também pode ser utilizado para criar cópias de clones de dados, tanto para backup e teste quanto para restaurar cópias, conforme necessário. Os grupos de consistência também recebem suporte para manter uma relação entre os diversos FlashCopy. Esta opção também garantirá a consistência da ordem de gravação entre os VDisks dentro de um grupo de consistência FlashCopy.

TSM for Virtual Environments

O backup e a recuperação dos servidores físicos e seus armazenamentos de dados associados podem ser razoavelmente simples quando comparados à entrega do mesmo nível de proteção para servidores virtualizados. O VMware força os administradores de TI, no mínimo, a modificarem esses processos e a encontrarem novas formas de ampliar o espectro de proteção de dados para servidores virtualizados. Não há somente um requisito para fazer o backup de um servidor virtualizado inteiro, é preciso também direcionar o backup e restaurar os objetivos –do tempo de recuperação (recovery time objectives ou RTO) e do ponto de recuperação (recovery point objectives ou RPO) – para cada máquina virtual e/ou aplicativo em execução em cada máquina virtual. A facilidade e a velocidade com que novas máquinas virtuais podem ser implementadas aumentam rapidamente a complexidade e o custo para a proteção de ambientes mistos de servidores físicos e virtuais.

Começando com a Versão 4, o VMware vSphere inclui o vStorage API for Data Protection (VADP) que permite que aplicativos de backup e recuperação como o TSM realizem backup centralizado das máquinas virtuais vSphere sem LAN, e substitui a estrutura do VCB (VMware Consolidated Backup) das versões anteriores do VMware que exigem a implementação de agentes para cada máquina virtual (marca "A" na Figura 2 abaixo). Ele também oferece o recurso para que o TSM seja mais profundamente integrado ao VMware.

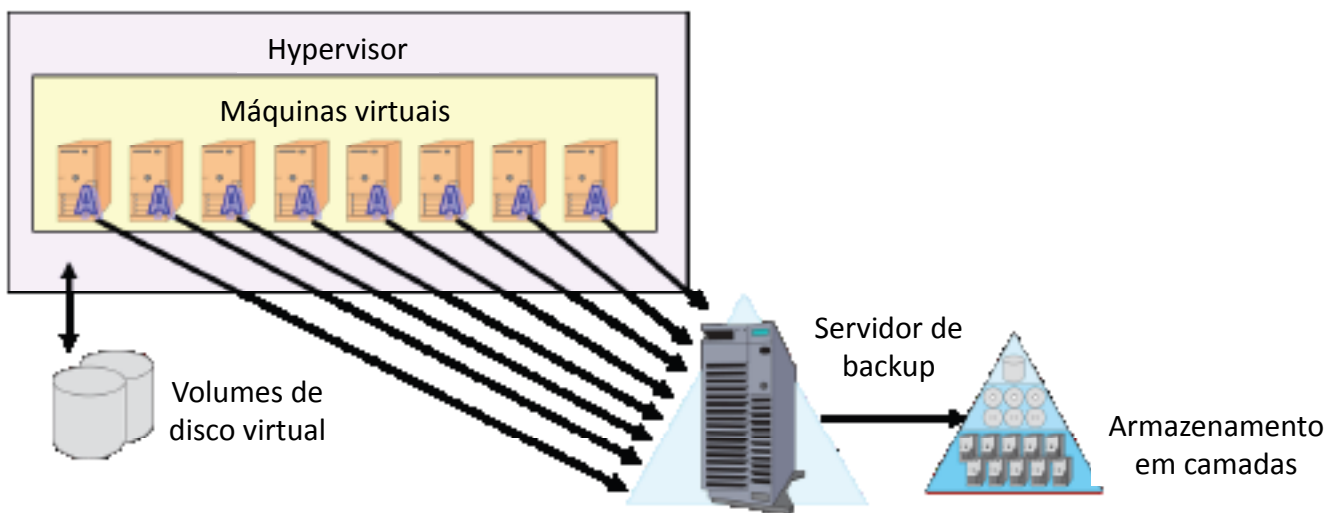


Figura 2. Implementação do VADP pelo vSphere do VMware

Com o TSM VE (TSM for Virtual Environments), o VADP é suportado enquanto nenhum processamento de backup é necessário dentro das máquinas virtuais convidadas, tampouco há a necessidade por parte do TSM de implementar agentes ou de fazer o download e instalar o software adicional do VMware. O TSM VE suporta um processo de cópia em uma única etapa, da origem para o destino, que inclui backups incrementais da máquina virtual utilizando o rastreamento de mudança de blocos (veja a Figura 3 abaixo). Também há suporte para a restauração incremental e completa de arquivos, volumes e todas as máquinas virtuais. A recuperação de dados é disponibilizada imediatamente ao mesmo tempo que é copiada em segundo plano. As políticas de backup automatizadas são automaticamente aplicadas às novas máquinas virtuais à medida que forem incluídas em um servidor ESX.

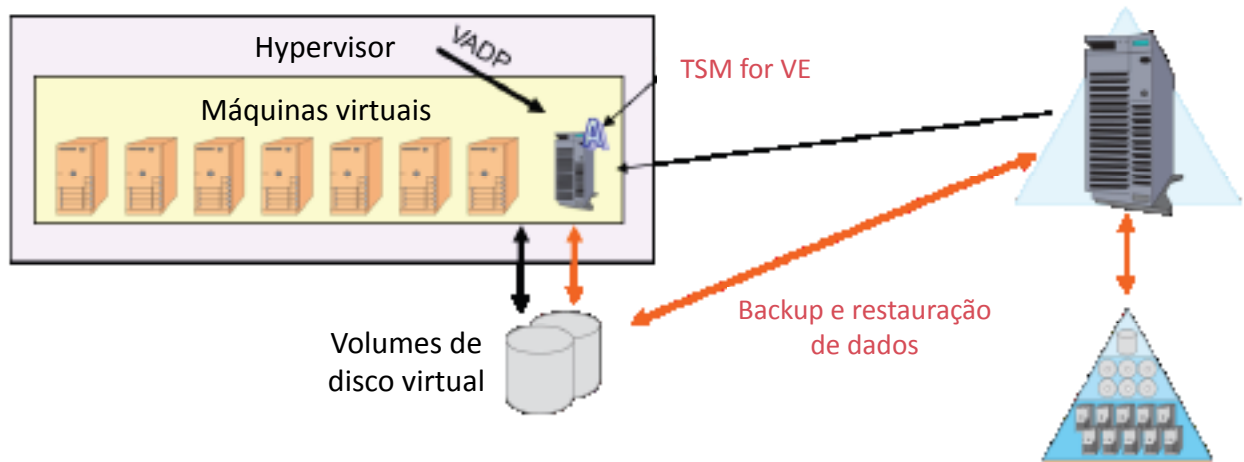


Figura 3. Backup de máquina virtual com o TSM VE

Capturas instantâneas e replicação no IBM XIV e Storwize V7000

Tanto as matrizes do IBM XIV quanto do Storwize V7000 suportam as funções de proteção de dados internos e automatizados, incluindo as capturas instantâneas e os processos de replicação completa de dados que criam cópias internas e externas do sistema.

XIV

Captura instantânea e cópia de volume

Snapshot copy – cria uma cópia instantânea de um volume. As capturas instantâneas do XIV são implementadas como um “redirect on write” (ROW) no qual as novas cópias de dados são criadas por meio de atualizações de metadados residentes na memória. As cópias de capturas instantâneas do XIV permitem milhares de capturas (excepcionais) com espaço eficiente e desempenho praticamente idêntico aos volumes padrão. Este recurso permite que os administradores de TI criem sofisticados fluxos de trabalho de dados procedentes da produção com uma complexidade operacional mínima, tais como a autoatualização de desenvolvimento e de conjuntos de dados de QA.

Cópia de volume – cria uma cópia independente de volume completo de outro volume. As cópias de volume do XIV são disponibilizadas imediatamente após sua criação ao mesmo tempo que os dados são copiados por meio de uma tarefa em segundo plano.

Remote Mirroring

O XIV Remote Mirroring cria cópias de dados point-in-time geograficamente dispersos utilizando espelhos de captura instantânea. Ele também suporta, ao mesmo tempo, a captura instantânea sincronizada de um grupo de volume ou de consistência, em ambos os sistemas de origem e de destino. Os espelhos de captura instantânea do XIV permitem a implementação de processos de continuidade de negócios que podem proteger, simultaneamente, os aplicativos contra dano lógico e falha no datacenter.

Espelho síncrono – mantém uma cópia completa de um grupo de volume ou de consistência em um sistema remoto do XIV, em que as gravações estão associadas a ambos os sistemas antes do reconhecimento de E/S do host (objetivo do ponto de recuperação ou RPO zero). Uma vez que o desempenho do WAN geralmente afeta a latência de E/S do host, a distância recomendada entre os locais do IBM XIV sincronicamente espelhados é de 100 km.

Espelho assíncrono – mantém uma cópia completa de um grupo de volume ou de consistência em um sistema remoto do XIV, mas permite um objetivo do ponto de recuperação (RPO) entre 30 segundos e 24 horas pelo reconhecimento de E/S do host antes que a gravação seja transmitida para o local remoto. O espelhamento assíncrono do XIV é alimentado por capturas instantâneas do mesmo. Uma captura instantânea do volume de origem é periodicamente criada e comparada com a captura mais recente no local remoto do XIV. A variação entre a captura instantânea atual e a anterior é então transmitida para o local remoto e aplicada à cópia espelhada. Pode ser definido um RPO de 30 segundos a 24 horas para cada grupo de consistência ou de volume espelhado. Além disso, as variações de captura instantânea podem ser transmitidas do cache do XIV ao invés do disco. A implementação das cópias assíncronas reduz a sobrecarga do sistema e pode diminuir significativamente a quantidade de dados transferidos entre os locais.

Grupos de consistência

Os volumes do XIV podem ser agrupados logicamente para operações de proteção e recuperação de desastres de dados que exijam consistência e fidelidade na ordem da gravação entre os volumes de origem e réplica. Os agrupamentos de consistência podem ser aplicados às capturas instantâneas point-in-time e a volumes replicados de forma síncrona / assíncrona.

Storwize V7000

FlashCopy

O FlashCopy faz uma cópia point-in-time de um volume de origem para um volume de destino. O FlashCopy é independente do host e suporta a criação da cópia no armazenamento interno do Storwize V7000 e entre todos os diferentes sistemas de armazenamento externos conectados ao Storwize V7000.

O FlashCopy, como funcionalidade de cópia instantânea (snapshot), é implementado como uma tecnologia com eficiência no consumo de espaço usando copy-on-write. Ele cria dois conjuntos de ponteiros para todos os dados dentro de um VDisk – um na origem, um no destino.

Todas as atualizações para a fonte VDisk são refletidas na cópia VDisk por um processo de copy-on-write automatizado que ocorre como uma operação em segundo plano. Pode ser estabelecida uma prioridade pelo administrador de TI que controla a largura da banda interna do Storwize V7000 (até 64 MB/s) utilizada pelas operações do FlashCopy. Estas incluem:

FlashCopy em cascata – o destino de um FlashCopy é utilizado como origem para réplicas adicionais do FlashCopy

Múltiplas reversões de destino do FlashCopy – para operações de restauração, os destinos do FlashCopy podem se tornar pontos de restauração sem desfazer a relação de origem e destino e sem esperar pela conclusão da cópia original. Há também suporte para diversos destinos e diversos pontos de reversão. Podem ser criados até 256 VDisks de destino a partir de uma única origem VDisk.

FlashCopy incremental – utiliza o rastreamento de mudança de blocos para atualizar um disco virtual que foi atualizado desde que se iniciou um FlashCopy

FlashCopy de espaço eficiente – permite que um VDisk totalmente alocado seja mapeado para um VDisk de espaço eficiente, criando efetivamente uma cópia provisionada do VDisk original

Replicação remota utilizando o Metro Mirror e o Global Mirror

O Storwize V7000 oferece dois produtos para replicação remota:

Metro Mirror – fornece uma cópia remota via replicação síncrona.

Global Mirror – fornece uma cópia remota via replicação assíncrona.

Além disso, o Storwize V7000 oferece suporte ao Multiple Cluster Mirroring. As configurações suportadas incluem cascadeamento (bunker) e replicação em arquitetura estrela (star), completamente conectados entre nós. Este nível de espelhamento em cluster corresponde àqueles oferecidos pelos sistemas avançados de armazenamento, incluindo os sistemas IBM DS8000, EMC VMAX e Hitachi USP/VSP.

Capturas Instantâneas do Storwize V7000 Unified e Replicação pelo Lado do Arquivo

O armazenamento com base em arquivo do Storwize V7000 também suporta capturas instantâneas de arquivos de espaço eficiente. Podem ser feitos até 256 capturas instantâneas por sistema de arquivos. Também pode ser utilizada a replicação assíncrona pela Ethernet por proteção. A replicação entre dois locais é criptografada para proteção.

Compactação e deduplicação

IBM ProtecTier

O sistema IBM ProtecTIER é uma solução escalável de backup em disco que realiza a deduplicação e compactação em linha para melhorar a eficiência do armazenamento dedicado à proteção de dados e processos de recuperação de desastres. Como um destino de backup para o software de backup de sistemas abertos, o sistema ProtecTIER pode ser configurado como uma VTL (Virtual Tape Library) emulando as principais bibliotecas e unidades de fita, como a comunicação do destino NAS sobre o protocolo CIFS ou como um dispositivo controlado OpenStorage Technology (OST) da Symantec NetBackup.

O ProtecTIER oferece uma maior disponibilidade por meio de seu cluster de nó duplo, bem como a replicação remota entre os sistemas ProtecTier geograficamente separados, a qual é suportada ao utilizar comunicações assíncronas integradas sobre o IP. As configurações de replicação remota incluem “muitos para muitos” para até quatro replicações simultâneas entre os sistemas, e “muitos para um”, em que até 12 origens podem replicar para um único destino ProtecTIER.

Três modelos, incluindo as configurações Appliance e Gateway, atendem diferentes requisitos de clientes em relação à proteção de dados de sistemas abertos em ambientes corporativos pequenos, médios e grandes. Há também um modelo (System z) de mainframe que utiliza um controlador FICON frontal a um servidor ProtecTIER.

Storwize V7000 Random Access Compression Engine

O RACE (Random Access Compression Engine) da IBM é fornecido como um recurso para o Storwize V7000. O RACE compactará dados de bloco na forma como são recebidos pelo Storwize V7000 e, em seguida, os armazenará em forma compactada. O RACE atua no Storwize V7000 à medida que os dados são gravados no disco e, dessa forma, evita a necessidade de primeiro armazenar dados descompactados, enquanto aguarda a compactação como uma tarefa em segundo plano. O RACE foi concebido para ser utilizado com dados primários ativos, tais como bancos de dados de produção e sistemas de email, ao contrário de processos de deduplicação comuns aos dispositivos de backup.

Proteção de Dados Assistido por Nuvem com Active Cloud Engine

O Active Cloud Engine da IBM é um recurso integrado do sistema de arquivos do GPFS da IBM (versão 3.2 e este acima) e fornece dois recursos avançados que aumentam a eficiência ambiental de proteção de dados e reduzem a complexidade acerca de arquivos do sistema de armazenamento IBM Storwize V7000 Unified e IBM SONAS. A primeira é uma função de gerenciamento de arquivos com base em política que controla o posicionamento inicial e a migração do arquivo entre todas as camadas de armazenamento internas pré-definidas do Storwize V7000 e do SONAS, assim como a migração de arquivo externo para um servidor IBM TSM a fim de proteger dados e arquivamento (veja a Figura 4 abaixo). Ela também pode ser utilizada para mover arquivos para outros sistemas de armazenamento com base em arquivos ou para fita se utilizado com um cliente Tivoli Storage Manager instalado. O Active Cloud Engine automatizou a movimentação de dados e a exclusão é baseada nas políticas definidas pelo administrador

de TI e em um amplo conjunto de metadados pré-definidos. As políticas são definidas na interface gráfica com o usuário do Active Cloud Engine. Um dos resultados da utilização deste recurso de migração automatizada de arquivo é a menor inclusão de dados no backup de aplicativo de produção, reduzindo o volume dos dados de backup com o passar do tempo.

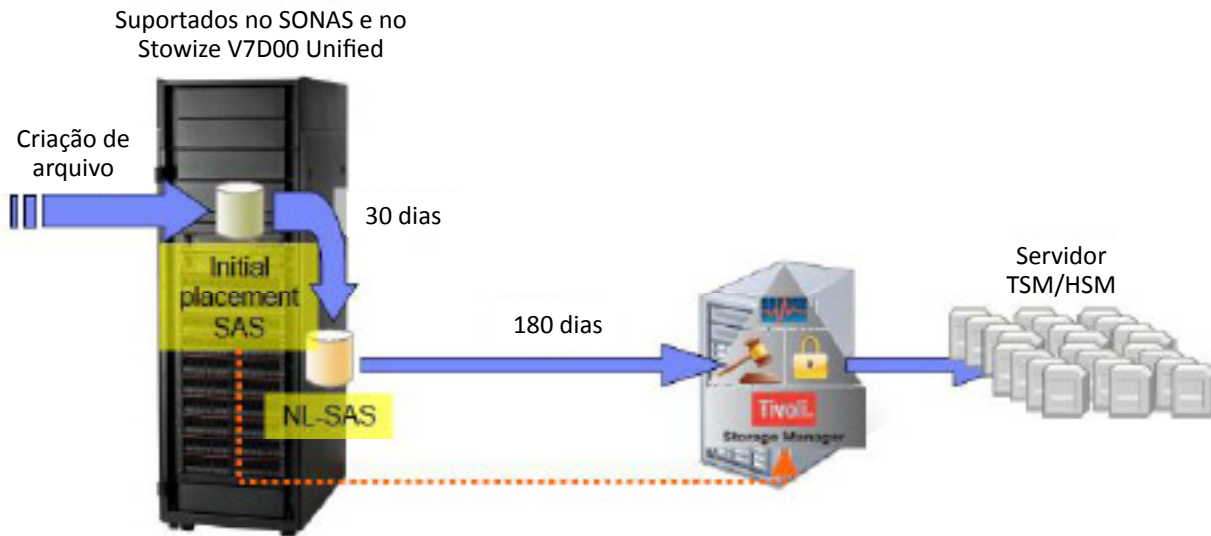


Figura 4. Active Cloud Engine

O segundo recurso do *Cache Cluster* fornece a capacidade de armazenar os arquivos em locais remotos e gerenciar a serialização das atualizações entre os sites locais e remotos. Uma modificação em um arquivo em um site remoto pode ser propagada de volta para o site local com base nas políticas definidas pelo administrador de TI. Esse recurso pode reduzir a necessidade de compartilhamentos de backup em escritórios remotos, onde o NAS é utilizado.

Conclusão

À medida que se acelera o crescimento de dados armazenados, é exercida uma pressão crescente sobre os administradores de TI para que façam um uso muito mais eficiente da infraestrutura e dos recursos de armazenamento do que já fizeram até então. A IBM aplica os princípios de eficiência de armazenamento para o ambiente de proteção de dados corporativos das seguintes maneiras:

Técnicas de backup “sempre incrementais” (TSM FastBack) removem grandes quantidades de dados redundantes dos processos de backup. Técnicas de compactação e deduplicação podem reduzir drasticamente os volumes de dados para o armazenamento em disco primário e secundário. Ambos eliminam um impedimento importante para proteger os dados em escala de petabytes.

A migração de volumes de dados raramente acessados e inativos de dispositivos de armazenamento primário a secundário (Active Cloud Engine) também reduz o volume de backups de dados de produção. Além disso, permite aos administradores de TI o planejamento de backups de dados esporádicos raramente acessados, reduzindo o tempo total e simplificando o suporte administrativo dedicado à proteção de dados corporativos.

O uso de capturas instantâneas de espaço eficiente (FlashCopy) pode reduzir a exigência de capacidade dedicada à proteção de dados em até 75%, de acordo com avaliações da IBM. Isso também remove um impedimento importante para assegurar que os dados estão protegidos em escala de petabytes.

Daqui para frente, acreditamos que a IBM continuará a produzir mais integração com plataformas de virtualização de servidor em seu portfólio de produtos de proteção de dados, dos quais o Tivoli TSM for Virtual Environments é um exemplo. Acreditamos também que o armazenamento da IBM fará uso crescente do software residente de armazenamento para aprimorar a eficiência e a funcionalidade do ambiente de proteção de dados corporativos.

Sobre a Evaluator Group

A Evaluator Group Inc. dedica-se a ajudar **profissionais de TI** e fornecedores a criarem e implementarem estratégias que aproveitem ao máximo o valor de suas informações digitais e de armazenamento. Os serviços da Evaluator Group entregam **análises detalhadas e imparciais** sobre arquiteturas, infraestruturas e gerenciamento de armazenamento para profissionais de TI. Desde 1997, a Evaluator Group presta serviços para milhares de usuários finais e fornecedores por meio de avaliações de produto e de mercado, análise competitiva e treinamento. www.evaluatorgroup.com. Siga-nos no Twitter: @evaluator_group

Copyright 2012 Evaluator Group, Inc. Todos os direitos reservados.

Nenhuma parte desta publicação pode ser reproduzida ou transmitida de qualquer forma ou por qualquer meio, eletrônico ou mecânico, incluindo fotocópia e gravação, ou armazenado em um banco de dados ou sistema de recuperação para qualquer finalidade sem o consentimento expresso da Evaluator Group Inc. As informações contidas neste documento estão sujeitas a alterações sem aviso prévio. A Evaluator Group não assume nenhuma responsabilidade por erros ou omissões. A Evaluator Group não faz neste documento garantias expressas ou implícitas em relação ao uso ou operação dos produtos aqui descritos. Em nenhuma hipótese, a Evaluator Group será responsável por quaisquer danos indiretos, específicos, insequentes ou incidentais decorrentes ou associados a qualquer aspecto desta publicação, mesmo que avisados da possibilidade de tais danos. A Evaluator Series é uma marca comercial da Evaluator Group, Inc. Todas as outras marcas comerciais são de propriedade de suas respectivas empresas.

Este documento foi desenvolvido com recursos da IBM. Embora o documento utilize material publicamente disponível de vários fornecedores, incluindo a IBM, ele não reflete necessariamente as opiniões de tais fornecedores sobre as questões abordadas neste documento.