**Backup**

***TIPOS DE BACKUP (Estratégia Concursos)***

* ***Backup Completo -*** *Também chamado de Total, Normal ou Full, trata-se do backup que faz uma cópia de todos os dados de uma unidade. Se uma organização possui a política de realizar backup completo diariamente, todos os dados serão copiados todos os dias, mesmo que não tenham sido alterados. A principal vantagem é que aumenta a chance de recuperação de dados íntegros, além de realizar uma operação menos sofisticada/complexa e de exigir um menor tempo para recuperar dados.*
* ***Backup Incremental -*** *Trata-se de uma cópia de todos os dados que foram criados ou modificados desde o último backup normal ou incremental anterior. Ele surgiu para sanar algumas deficiências do Backup Completo, como sempre copiar todos os dados a cada operação mesmo que nenhuma alteração tenha sido realizada. Após o backup, ele desmarca o atributo de arquivamento (Bit Archive = 0), informando que aquele arquivo já sofreu backup.*
* ***Backup Diferencial -*** *Também conhecido como Backup Incremental Cumulativo, trata-se de uma cópia de todos os dados que foram criados ou modificados desde o último backup completo ou incremental. Percebam que o Backup Diferencial copiará todos os dados que foram modificados desde o último backup completo ou incremental anterior e não removerá o atributo de arquivamento (Bit Archive = 1).*
* ***Backup de Cópia -*** *Trata-se de uma cópia de todos os arquivos selecionados, mas não os marca como arquivos que passaram por backup, isto é, não alteram a Flag/Bit Archive. Em geral, é utilizado de forma emergencial e normalmente é usada caso se queira fazer o backup de arquivos entre um backup completo e um backup incremental. Galera, esse é o tipo de backup mais raro em prova até porque ele é extremamente simples.*
* ***Backup Diário -*** *Trata-se de uma cópia de todos os arquivos selecionados que foram modificados no dia da execução do backup diário. Os arquivos não são marcados como arquivos que passaram por backup, isto é, não alteram a Flag/Bit Archive. Ele utiliza como critério apenas a data do backup. Quanto a esse backup, assim como o anterior, não há muito o que detalhar. Ele é tão simples de entender que – junto com o anterior – são os mais raros de cair em prova.*

**Outras definições**

* **Diferencial: (Acumulativo)** Um backup diferencial copia arquivos criados ou alterados desde o último backup normal ou incremental. **Não**marca os arquivos como arquivos que passaram por backup. Portanto, o Backup diferencial também pode ser usado do segundo dia em diante, mas com a diferença que, ao**não marcar o arquivo**, o sistema entende que tem que copiar de novo, por isso ocorre a **acumulação**de arquivos copiados. A operação de **backup diferencial**é **semelhante a um incremental na primeira vez em que é realizada**, na medida em que irá copiar todos os dados alterados desde o backup anterior. No entanto, cada vez que é executado após o primeiro backup, serão copiados todos os dados alterados **desde o backup completo** anterior e não com relação ao último backup.
* **Incremental**: Um backup incremental copia somente os arquivos criados ou alterados desde o último backup normal ou incremental e os**marca** como arquivos que já passaram pelo backup. O Backup **Incremental**deve ser feito do segundo dia em diante, para copiar apenas os novos arquivos criados.
* **Backup incremental contínuo (ou progressivo)** - "Funciona como o backup incremental. O que faz um backup incremental para sempre diferente de um backup incremental é a disponibilidade dos dados. Como você deve se lembrar, a recuperação de um backup incremental requer o backup completo, e cada backup subsequente até o backup que você precisa recuperar. **A diferença do backup incremental para sempre é que ele automatiza o processo de recuperação, de modo que você não tem que descobrir quais conjuntos de backups precisam ser recuperados.** Em essência, o processo de recuperação de dados do incremental para sempre torna-se transparente e imita o processo de recuperação de um backup completo."
* **Hot Backup**: cópias de segurança realizadas com o sistema em execução.
* **Cold Backup**: cópias de segurança realizadas com o sistema fora de execução.
* **Snapshot**significa “foto instantânea”, na tecnologia é uma cópia de dados que possui c**omportamento semelhante a um backup.** As funções Snapshot fornecem cópias instantâneas de dados que se **parecem e se comportam como backups completos sem consumir o espaço em disco equivalente**. As imagens instantâneas (snapshots) são ideais para várias aplicações, incluindo testes de backup, ou de desenvolvimento, análise da informação e mineração de dados. No instante da criação de um snapshot, uma imagem de dados é tirada sem afetar os aplicativos que estão rodando. Com base na imagem, as alterações dos dados serão copiadas para o volume instantâneo quando novas gravações ocorrerem. Com design cópia-na-gravação, o snapshot Infortrend protege dados de produção de alterações acidentais, exclusões e corrupções com os requisitos de capacidade mínima e sobrecarga de desempenho. Ao acessar uma cópia instantânea no ponto de recuperação desejado, os usuários podem restaurar imediatamente a disponibilidade do sistema pela interrupção de dados.

**Desduplicação:** processo usado para diminuir a quantidade de dados armazenados.

A ideia é fazer com que, no sistema de arquivos, nunca exista dois ou mais computadores armazenando os mesmos dados. Com isso, **ganha-se espaço em disco e diminuição no tempo de backup**.

**Dados redundantes são evitados**, ou seja, temos uma redução do espaço em disco.

Também gera uma economia para as empresas de TI, que não precisam pagar por mais espaço no armazenamento em nuvem, por exemplo.

A **Deduplicação** é usada no backup do tipo **incremental**. A Deduplicação é um processo no qual elimina a duplicação de dados, logo os arquivos do usuário não serão recopiados novamente em outras rotinas de backup. A Deduplicação na origem remove os dados redundantes em ambiente de produção, antes mesmo de serem enviados para o servidor de backup. Por consectário lógico, eles não serão enviados pela rede, diminuindo, assim, o tráfego da rede e melhorando o desempenho no armazenamento. A Deduplicação não só permite economizar em custos de armazenamento, mas **também acelera as comunicações entre sites**(LOCAL-NUVEM ou LOCAL-NUVEM-LOCAL), não trafegando por exemplo várias cópias dos mesmos dados através do link de internet.

* **Deduplicação in-line -** A Deduplicação in-line elimina os dados redundantes durante o processo de backup, antes mesmo de serem gravados. O processo in-line geralmente é mais demorado visto que os dados serão armazenados somente após a Deduplicação dos mesmos.
* **Deduplicação pós-processamento -** Na deduplicação pós-processamento o backup completo é realizado normalmente e a analise e remoção das redundâncias nos dados são efetuadas em um processo posterior separadamente do backup. Isso garante um backup mais rápido, porém necessita de mais espaço de armazenamento.
* **Deduplicação de Origem -** A deduplicação na origem remove os dados redundantes em ambiente de produção, antes mesmo de serem enviados para o servidor de backup. A Deduplicação na origem usa softwares de que comunicam com o servidor de backup comparando os novos dados com os dados já armazenados e caso o servidor ou appliance já tenha uma cópia previamente armazenada desses dados, os mesmos não serão enviados pela rede, como consequência diminuindo o trafego da rede e melhorando o desempenho no armazenamento. A desvantagem de utilizar a deduplicação na origem é que ela utiliza o poder de processamento do próprio servidor de produção. A deduplicação na origem é altamente recomendada em ambientes virtualizados devido a grande quantidade de dados redundantes geradas nos arquivos de disco .vmdk da maquina virtual. Outro cenário recomendado é para backup de sites menores e remotos onde o aumento do uso da CPU no processo de backup não causa grande impacto nos servidores.
* **Deduplicação de Destino -**  Na deduplicação de destino o processamento e remoção dos dados redundantes são feitos após o seu envio pela rede, ou seja, direto no appliance ou servidor de backup. Este tipo de deduplicação causa muito menos impacto no processamento do ambiente de produção, porém utiliza muito a largura de banda, visto que todos os dados são enviados para backup sem nenhuma compressão ou deduplicação.  A Deduplicação de destino é ideal para ambientes de produção com altas carcas de trabalho ou que possuem uma janela de backup limitada como por exemplo servidores de banco de dados.

= **IN-LINE**: remove as redundâncias **DURANTE O PROCESSO** de backup. “antes de salvar”

= PÓS-PROCESSAMENTO: faz backup de tudo, **DEPOIS REMOVE**;

= **DE ORIGEM**: remove duplicatas **ANTES** de fazer o backup;

= **DE DESITINO:** remove dados duplicados **APÓS SALVAR,** separado da produção;

* RPO (*Recovery Point Objective*): A quantidade máxima de dados que pode ser perdida quando um serviço é restaurado após uma interrupção. O objetivo do ponto de recuperação é expresso como duração de tempo antes da falha. Por exemplo, o objetivo do ponto de recuperação de um dia pode ser suportado por cópias de segurança diárias e então até 24 horas de dados podem ser perdidos. O objetivo do ponto de recuperação para cada serviço de TI deve ser negociado, acordado e documentado, sendo usado como requisito para o desenho do serviço e planos da continuidade do serviço de TI.
* RTO (*Recovery Time Objective*): O tempo máximo permitido para a recuperação de um serviço de TI após uma interrupção. O nível de serviço a ser fornecido pode ser menor que as metas de nível de serviço normais. O objetivo do tempo de recuperação para cada serviço de TI deve ser negociado, acordado e documentado. Veja também análise de impacto no negócio.

Logo, o RPO deve ser maior que o RTO, senão ocorrerá perdas de dados a mais do que é suportado pela empresa.

* **Hot site** é um ambiente de processamento de dados secundário, distanciado do Datacenter primário, que espelha todos os equipamentos de TI, que interligados de forma redundante e contingêncial com o site primário, opera de forma a manter todos os dados originais da instituição replicados no site oposto, passíveis de serem prontamente utilizados em caso de desastre. Ora, para serem prontamente utilizados eles necessitam ficar disponíveis e em operação o tempo todo.
* **Cluster** - é uma tecnologia capaz de fazer computadores mais simples trabalharem em conjunto, como se formassem uma máquina só, no intuito de processar uma tarefa. Cada computador que faz parte do cluster recebe o nome de *nó* (ou *node*). Esses "nós" podem até mesmo ser compostos por computadores simples, como PCs de desempenho mediano. Por exemplo, uma loja na internet pode utilizar um cluster de alta disponibilidade para garantir que suas vendas possam ser realizadas 24 horas por dia e, ao mesmo tempo, aplicar balanceamento de carga para suportar um expressivo aumento eventual no número de pedidos causados por uma promoção.

**Replicação**

* **Replicação Remota Síncrona -** A Replicação síncrona grava dados para Storages primários e secundados ao mesmo tempo. Portanto, os dados permanecem idênticos e atuais em ambas as fontes. O processo funciona rapidamente com pequena margem de erro. A vantagem da replicação síncrona é que elimina o risco de perda acidental de dados. A desvantagem é que requer comunicação de baixa latência porque o site secundário deve confirmar que cada pacote foi recebido sem erro. Quanto mais longe o local secundário é do primário, mais difícil é conseguir. O objetivo do ponto de recuperação da replicação síncrona é quase “zero”.

* **Replicação Remota Assíncrona -** Na replicação remota assíncrona, existe um delay antes dos dados serem gravados no Storage secundário. Isso por que a replicação assíncrona é desenhada para trabalhar em distancias mais longas e requer menos alocação de banda, o que a torna uma solução melhor em casos de recuperação de desastres. Porém na replicação assíncrona temos riscos de perder dados durante uma interrupção do sistema devido ao fato dos dados no dispositivo de destino não estarem sincronizados em tempo real com o dispositivo de origem. A replicação assíncrona não é confiável e não garante que deu certo uma vez que não aguarda a resposta para continuar executando. Os dados chegam ao destino de replicação do host de replicação com um atraso, variando de quase instantâneo a minutos ou mesmo horas. Se a replicação estiver sendo feita para um data center geograficamente separado, a replicação assíncrona é considerada eficaz, pois trabalha de forma coesa com latência da rede e é tolerante à largura de banda. Por esse motivo, alguns especialistas em armazenamento chamam esse método de “Armazenar e encaminhar”. O objetivo do ponto de recuperação da replicação assíncrona o RPO (Recovery point objective) é de 15 minutos a algumas horas.

**Plano de serviço**

Você precisa levar em consideração muitos fatores ao criar seu serviço de backup e recuperação. Entre os fatores a serem considerados, incluem-se:

* Prioridades de backup e recuperação rápidos - Recovery Time Objective (RTO).
* A frequência com que os dados mudam.
* Restrições de tempo na operação de backup.
* Mídia de armazenamento.
* Requisitos de retenção de dados.
* Prevalência dos dados recuperados – Recovery Point Objective (RPO).

**Qual é o melhor?**  
  
Depende realmente do projeto.  
Para replicação de longa distância, a replicação assíncrona pode ser a única opção.  
Para a replicação de dados sem erros, síncrono é a melhor resposta.  
Assíncrono é mais econômica, quando comparado à replicação síncrona.

**Macete Fitas LTO** - Não precisa decorar a capacidade de todas as fitas **LTO**, basta entender que a fita LTO1 armazena 100GB e que as próximas versões simplesmente dobram o tamanho de armazenamento da anterior.

* LTO1 - 100GB
* LTO2 - 200GB
* LTO3 - 400GB
* LTO4 - 800GB

E assim sucessivamente.