



BACKUP – Análise, Variáveis e Complexidade (1 / 5)

A criação de uma arquitetura de backup é uma tarefa complexa devido a heterogeneidade dos elementos que o compõe. O objetivo principal é realizar backup dos sistemas em produção mantendo-os disponíveis para a necessidade da recuperação dos dados armazenados.

Para este objetivo fundamental ser atingido diversos vetores técnicos e organizacionais devem ser considerados e, como não poderia deixar de ser, temos uma questão de concorrência por disponibilidade de recursos.

Para que decisões possam ser tomadas com acurácia e coerência é necessário que todos os fatores envolvidos no backup corporativo sejam entendidos e levados em consideração. Vamos a seguir abordar estes vetores.

Recursos de Infra-estrutura

A infra-estrutura de backup é composta por diversos componentes como os drives, mídias, a library, o servidor de backup, as interfaces de rede, a rede de dados, o storage de Backup e inclusive o hardware dos clientes que efetuam backup. Podemos analisar estes recursos sob 3 pontos de vista: arquitetura, dimensionamento e utilização.

Sob o ponto de vista de arquitetura estes recursos devem ser combinados ou integrados de maneira que possibilitem a utilização de sua máxima capacidade bem como ofereçam flexibilidade de utilização.

É importante ressaltar que em um backup corporativo, diferentemente do backup de um determinado servidor ou conjunto discreto de servidores, as condições de trabalho não podem ser precisadas ou totalmente pré-estabelecidas.

A infra-estrutura do backup corporativo deve ser criada para ser totalmente adaptativa possibilitando se conformar as mais diferentes realidades bem como se reconfigurar continuamente às necessidades dos servidores, volumes e tecnologias. Por isto é importante a elaboração, de fato, de uma arquitetura de backup e não somente o dimensionamento e aquisição de equipamentos.

A diferença básica entre as duas abordagens (arquitetura e dimensionamento) é que quando olhamos o problema sob o ponto de vista de arquitetura, nós focamos em como vamos estruturar a solução para atender a demanda atual e como que os recursos que dimensionamos poderão se adaptar a condições não planejadas.

Sob o ponto de vista de dimensionamento, também existe previamente alguma avaliação de arquitetura, mas obrigatoriamente quem tem foco em dimensionamento, não se compromete que, além destes recursos serem configurados de modo a oferecer a capacidade de vazão necessária para que os volumes de backups sejam backupeados dentro das janelas definidas, estes mesmos recursos sejam re-utilizados e continuamente adaptados e integrados com novos recursos que se fizerem necessários.

De qualquer forma, quando estamos lidando com backup corporativo, além de ser necessária a elaboração de um projeto de arquitetura, também é imprescindível o correto dimensionamento dos recursos.

Uma variável complexa e importante é a vazão de dados que será necessária, porque ela resume, através de uma proporção, duas outras variáveis que são, o volume a ser backupeado e a janela de tempo para isto ser feito.

O último fator é a utilização e nesse caso estamos lidando com a vazão de dados que resume na forma de resultado o que foi planejado, mas, não se pode ter uma infra-estrutura de backup sem contínua medição da sua utilização.



BACKUP – Análise, Variáveis e Complexidade (2 / 5)

Para determinar a vazão real é muito importante estar atento que este atributo não é determinado pela capacidade instalada no servidor e na Library de backup, ou seja, não é determinada pela quantidade de drives, interfaces e processadores. **A vazão é um número mínimo obtido percorrendo-se uma cadeia integrada que parte do cliente indo até a library de fitas.**

Estamos lidando com uma cadeia de componentes que se inicia no disco do cliente e passa pelo barramento, interface de rede, rede de dados, interface do servidor de backup, Cpu, memória, barramento, interface de storage, library, drive e mídia.

É uma cadeia longa, com muitos componentes e a melhor performance vai ser determinada pelo pior componente (Bottleneck / Gargalo). Isto significa também que em uma mesma instalação poderemos ter números diferentes em função do cliente.

Este fator é o que determina as vazões reais serem menores do que as especificadas pelos fabricantes dos equipamentos.

Além disso, devemos estar atentos, porque estamos lidando com um sistema de equilíbrio hiperestático, onde uma variável influencia as outras. Sob o ponto de vista de utilização, nossa obrigação é configurarmos a programação do backup para atingirmos a capacidade máxima de cada componente. Para isto temos que garantir que o fluxo de dados vá do cliente à mídia sem interrupção e sem gargalo e em sua máxima capacidade.

Para obter isso, temos que saturar os diversos componentes da cadeia e sempre ter certeza que esta saturação é sustentável. Capacidade ociosa significa sub-utilização de recursos, perda de tempo valioso na janela de backup e principalmente, no caso de backup corporativo, sub-utilização de recursos compartilhados.

Este último elemento que adicionamos, a gestão do compartilhamento de recursos é que torna o problema do backup corporativo mais complexo e diferente dos backups discretos.

No backup corporativo, além da questão da performance, temos que administrar com extrema atenção a competição por recursos.

Na equação do backup corporativo, o excesso de capacidade de alguns recursos em relação a outros ou um desbalanceamento na programação dos backups, pode fazer com que a vazão de todos os backups caiam, prejudicando a eficiência do sistema.

Posteriormente descosturaremos cada um dos tópicos que formam os recursos de infra-estrutura necessários para implementação de uma arquitetura de Backup.

Janelas de Backup

A janela, ou seja, o tempo disponível para “backupear” um determinado volume de dados parece sempre ser o foco da programação e do dimensionamento do backup. Isto não deixa de ser uma verdade, entretanto, quando lidamos com backup corporativo, não pode existir somente uma janela de backup, que normalmente é relacionada a um evento de calendário (dia, semana, mês, etc.).

A determinação de uma janela esta ligada ao ciclo de atualização de dados de uma aplicação – Volatilidade da Base de Dados. É claro que o problema é simplificado pelo fato de termos servidores funcionalmente equivalentes bem como aplicações que por acompanharem o mesmo processo de negócios da empresa vão ter ciclos muito similares de atualização.



BACKUP – Análise, Variáveis e Complexidade (3 / 5)

Todo o backup é uma cópia de segurança de dados que será utilizada quando uma perda ocorrer. Assim, literalmente, estaremos usando um backup para re-estabelecer a condição operacional de um sistema a partir de uma imagem temporal anterior. Para muitas aplicações significa que poderá haver a perda de dados entre o momento da imagem armazenada no backup e o momento corrente da falha, paralisação e/ou perda de dados.

Cada aplicação tem um determinado ciclo de operação e funcionamento que determina o período onde os seus dados são consultados e atualizados. Algumas aplicações têm um ciclo contínuo de operação, ou seja, funcionam 24 x 7, outras tem um ciclo discreto onde existe horário de total ou quase total inatividade. O backup, bem programado, deve ser posicionado para capturar e preservar um volume de atualizações que represente a menor perda aceitável. Assim, não existe um conceito absoluto de quando um backup deve ser realizado.

De fato esta também é uma equação com diversas variáveis, mas que o resultado, sempre, é o que se admite perder. Um backup será utilizado para recuperar uma situação de perda de dados. A perda de dados diz respeito com a atualização e não com os dados estáticos ou não modificados.

Para solucionar este problema existem diversas estratégias, técnicas e recursos e o conjunto desta análise chamamos muito “simplicadamente” de Política de Backup, já que se trata de um termo “pequeno” para um problema complexo. Normalmente, se erra ao se fazer uma política de backup, porque o foco da política não é apenas determinar como será o backup diário, semanal e mensal (completo, incremental ou diferencial) e qual a política de retenção dos dados. Quando estabelecemos uma Política de Backup não podemos nos deter simplesmente na tarefa de salvar dados em fitas digitais (DAT). Esta tarefa é uma das estratégias que usamos de preservação de dados mas não é a única. Devemos, sobretudo, analisar e especificar como poderemos, através dos Backups, preservar a disponibilidade das aplicações através da segurança dos dados armazenados.

Definir uma *Janela de Backup* consiste em conciliar o ciclo de atividade da aplicação, com a necessidade de salvamento de dados versus o impacto de realizar a atividade de backup. Não podemos ignorar que realizar um backup sempre vai ter um impacto sobre o funcionamento da aplicação e vice e versa, isto é, a performance do backup será impactado pela concorrência da aplicação.

Quando tratamos de backup corporativo este é um fator importante porque estamos compartilhando recursos entre diversas aplicações e um backup que seja mais demorado do que deveria vai reter recursos que podem fazer falta para outra tarefa. No backup corporativo temos sempre que olhar o conjunto e buscar maximizar a utilização dos recursos.

Cada aplicação vai ter o seu próprio equilíbrio entre segurança de dados e impacto sobre performance, além do que também será beneficiada pelas Tecnologias de Armazenamento das quais se utilizam. No entanto é preciso entendermos que, mesmo para aplicações baseadas em *Storage Redundante* (SFT), ex.: Raid-1 ou Raid-5, independentemente do maior nível de segurança, tais sistemas não estão imunes a incidentes de origem lógica (e não física).

Ao definirmos uma janela de backup, temos que considerar o Ciclo de Operação da Aplicação para obter o entendimento de qual será a estratégia e janela ideal, além disso iremos determinar qual o Limite de Tolerância da Aplicação, visto em caso de problemas na execução da programação diária, pode ser necessário fazer ou refazer backups fora da janela ideal.

Sempre existirá a tendência de alinhar a política de backup com eventos de calendário, como dias e semanas, isso é natural, mas não é correto. Em sistemas que funcionam 24 x 7 não existe o conceito de dia. Em sistema que funcionam mundialmente na Internet idem. Sistemas comerciais que funcionam 7 x 5 podem ter a necessidade de ter pelos menos 2 backups neste período uma vez que ninguém gostaria de perder todo um dia de vendas e faturamento, por exemplo.

Após analisarmos todas as aplicações, suas necessidades e definirmos os volumes de dados e as janelas, obteremos uma espécie de coincidência de tarefas para um determinado horário. Chamaremos este horário, onde existe uma intensa concorrência de recursos (de tal forma que o atraso em uma tarefa impacta diretamente as demais) de Janela Crítica de Produção, ou Janela de Produção.



BACKUP – Análise, Variáveis e Complexidade (4 / 5)

Devemos lembrar que esta é apenas uma convenção e ressaltar que somente backups muito importantes e totalmente estáveis, seguros e otimizados (performance) podem ser executados nesta faixa de horário. Em backup corporativo todo o horário do dia é uma janela de produção de alguma aplicação.

Ao levantarmos as necessidades de cada aplicação, devemos ser muito rigorosos para apurar não somente a vontade do desenvolvedor ou usuário da aplicação, mas, realmente entender a aplicação para podermos ter flexibilidade de programação do backup.

Se dúvidas, **quem deve definir como será o backup é a área de produção e não a de aplicação**. É a área de produção que define os recursos de armazenamento para cada aplicação e é ela que conhece os riscos de perda de dados. A área de aplicação e usuária são fundamentais para definir o ciclo de operação e atualização de dados que, em função de quão é crítico cada processo e volume de dados, vão determinar a técnica de salvamento e o tipo de mídia mais adequado.

Política de Backup

A determinação da Política de Backup passa pelo entendimento da aplicação, do seu ciclo de operação e dos seus volumes. Em função disso, devemos definir a estratégia e a periodicidade de backup de dados.

Em relação a estratégia temos : o Backup pode ser **Completo** (Total / Full) quando todos os dados são gravados / salvos, o **Diferencial** quando somente os dados que mudaram depois do último backup são salvos ou **Incremental** quando os dados modificados entre o último backup completo e cada incremental são salvos, pode se entender como Backup Cumulativo.

Estas 3 estratégias determinam **tempos muito diferentes para realizar o backup**, mas, principalmente **afetam o tempo de recuperação de dados**. Por exemplo, o backup mais rápido sempre será o **Diferencial**, mas a operação de recuperação, dependendo do que se perde envolverá muito mais volumes de armazenamento. Entretanto, pode ser uma alternativa muito atraente **quando se tem um grande volume de atualização de dados**.

Dependendo do volume de dados que são modificados o Backup **Incremental** pode ser uma alternativa muito atraente, conciliando velocidade e simplicidade.

Apesar de parecer o pior caso, o Backup Completo ou Total, na maioria das vezes é a única alternativa, uma vez que nem todas as aplicações foram feitas ou preparadas para os backups parciais (incremental e diferencial). Quando somente o backup completo é a alternativa torna-se necessário usar uma infraestrutura de backup de maior desempenho ou ainda, empregar tecnologias mais sofisticadas de armazenamento, como a Duplexação de Disco / Servidores, além de conexões de alta velocidade.

Além disso, com a gradativa queda de preços das tecnologias para armazenamento de dados, existem diversas alternativas de mídias on-line para geração de cópias de segurança, em detrimento dos custos tentadores das mídias off-line.

Todos estes recursos permitem implementarmos a Política de Backup baseada em estratégias eficazes capazes de salvaguardar a segurança e disponibilidade das aplicações.

O ponto de partida para a definição da política de backup é a necessidade de minimizar a perda de dados e otimizar a sua recuperação. A estratégia do backup vai depender da disponibilidade de infra-estrutura de *Storage*, dos volumes de atualização e dos processos críticos do sistema.

Por exemplo, em sistemas de missão crítica que usam banco de dados relacionais, podemos usar mídia on-line para salvar os "archive logs" possibilitando assim rapidez na recuperação e minimizando as perdas. Para servidores de arquivo ou sistema de correio eletrônico que armazenam arquivos planos, a estratégia pode ser o uso dos archive logs de file system, transaction logs dos aplicativos, HSM ou backup diferenciais.



BACKUP – Análise, Variáveis e Complexidade (5 / 5)

Sempre haverá a possibilidade de termos recursos na aplicação para geração de logs de atualização ou eventualmente ser necessário que a própria aplicação tenha uma funcionalidade específica para a preservação dos dados.

Assim, fechando este ponto, não é possível definir e implementar um backup sem analisar e conhecer cada aplicação alvo do Backup.

Para finalizar este assunto é importante voltar ao princípio. Definir uma Política de backup é uma atividade principal da área de produção. Não é uma atividade de suporte e não é uma atividade da área de aplicação. A política tem ligação com a infra-estrutura de armazenamento existente nos *DataCenters*.

Nem toda a aplicação merece e deve ter o mesmo recurso de armazenamento, isto é um desperdício e uma volta a um conceito que levou ao ocaso do Mainframe. No Mainframe independente do tipo e importância do processo todas as aplicações recebiam e competiam por um pedaço do mesmo equipamento.

O ambiente distribuído nos permite utilizar e integrar recursos de diferentes tecnologias e adequando desta forma a equação de custo x benefício. O mesmo vale para a tecnologia de armazenamento. A grande habilidade do profissional de storage é utilizar o recurso (estratégia e tecnologia) mais adequado para cada aplicação e o grande benefício da arquitetura é suportar e integrar isso.

Administração de Recursos

Como foi dito no princípio, os recursos do backup administrativo iniciam no próprio servidor hospedeiro da aplicação cliente. Este servidor é parte da infra-estrutura de backup porque ele irá executar a principal atividade no backup que é extrair os dados a serem salvaguardados.

Por esta razão, administrar o backup corporativo é lidar um ambiente dinâmico onde as condições podem mudar dia a dia. Não temos condições de garantir que as condições nas quais foi feito o backup de hoje serão as mesmas do backup de amanhã. Tudo pode mudar, visto que trata-se de variáveis dinâmicas.

Administrar um backup corporativo significa tomar diariamente novas decisões em função da realidade corrente. Não existe programação estática no backup corporativo, seja pela inclusão de novos clientes que podem requerer a reprogramação de vários backups como também a alteração do volume de dados e/ou performance dos cliente pode alterar significativamente os resultados.

É necessário então, uma administração contínua dos recursos e uma tomada constante de decisões visando readequar os recursos, estratégias e / reprogramação da própria realidade do *dia-a-dia* das empresas, tornando imprescindível para tal uma equipe de gestão para a atividade em questão.