

Gerenciamento de desempenho e capacidade: White Paper de práticas recomendadas

Contents

[Introduction](#)

[Visão geral do gerenciamento de capacidade e desempenho](#)

[CPU](#)

[Backplane ou E/S](#)

[Memória](#)

[Tamanhos de interface e tubulação](#)

[Enfileiramento, latência e instabilidade](#)

[Velocidade e distância](#)

[Características do aplicativo](#)

[Melhores práticas de gerenciamento de capacidade e desempenho](#)

[Gerenciamento de nível de serviço](#)

[Análise What-if de rede e aplicativo](#)

[Linha de base e tendências](#)

[Gerenciamento de exceções](#)

[Gerenciamento de QoS](#)

[Coletando e Relatando Informações sobre Capacidade](#)

[Determine suas necessidades](#)

[Definir um processo](#)

[Definir Áreas de Capacidade](#)

[Definir as Variáveis de Capacidade](#)

[Interprete os dados](#)

[Informações Relacionadas](#)

Introduction

A alta disponibilidade da rede é um requisito crítico de missão dentro de uma grande empresa e das redes de provedores de serviço. Os gerentes de rede enfrentam os desafios crescentes para fornecer uma disponibilidade mais alta, incluindo o tempo de inatividade não programado, a falta da experiência, a falta de ferramentas, as tecnologias complexas, a consolidação do negócio e os mercados competitivos. O gerenciamento de capacidade e desempenho ajuda os gerentes de rede a atenderem os objetivos de negócios do novo mundo e obter uma disponibilidade e um desempenho de rede consistentes.

Este documento examina os seguintes tópicos:

- Problemas gerais de capacidade e desempenho, incluindo os riscos e possíveis problemas de capacidade nas redes.
- Melhores práticas de gerenciamento de capacidade e desempenho, incluindo análises hipotéticas, linha de base, tendências, gerenciamento de exceções e gerenciamento de QoS.

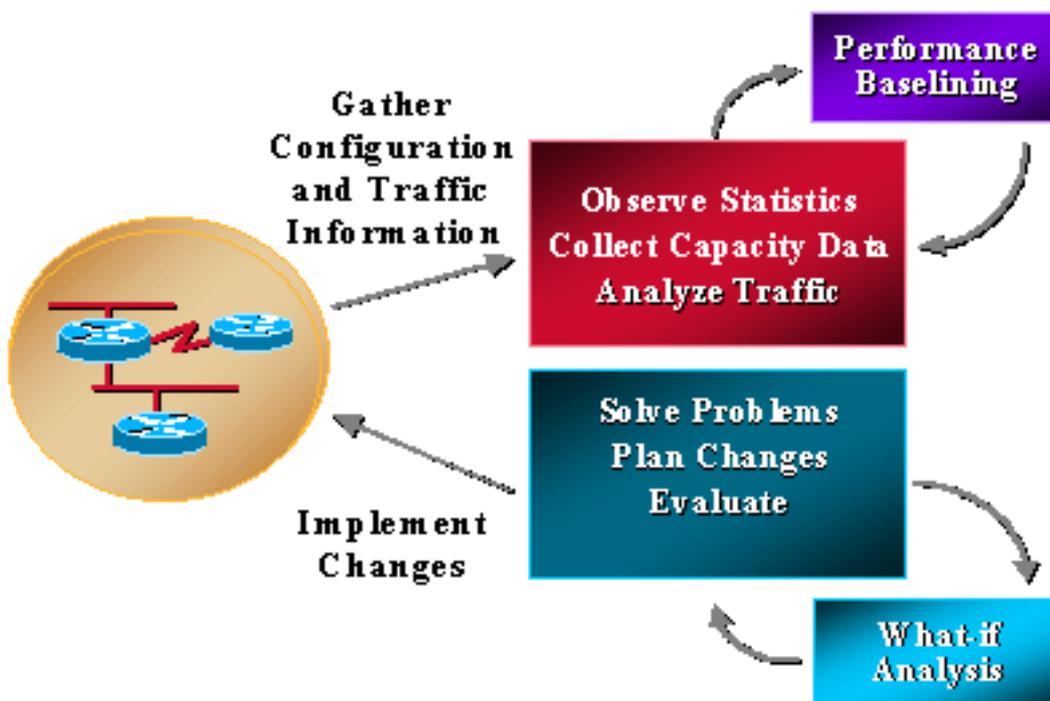
- Como desenvolver uma estratégia de planejamento de capacidade, incluindo técnicas comuns, ferramentas, variáveis MIB e limites usados no planejamento de capacidade.

Visão geral do gerenciamento de capacidade e desempenho

O planejamento de capacidade é o processo de determinar os recursos de rede necessários para evitar um impacto no desempenho ou na disponibilidade de aplicativos críticos para os negócios. O gerenciamento de desempenho é a prática de gerenciar o tempo de resposta, a consistência e a qualidade do serviço de rede para serviços individuais e gerais.

Observação: os problemas de desempenho geralmente estão relacionados à capacidade. Os aplicativos são mais lentos porque a largura de banda e os dados devem esperar em filas antes de serem transmitidos pela rede. Em aplicações de voz, problemas como atraso e jitter afetam diretamente a qualidade da chamada de voz.

A maioria das organizações já coleta algumas informações relacionadas à capacidade e trabalha de forma consistente para resolver problemas, planejar alterações e implementar novas funcionalidades de capacidade e desempenho. No entanto, as empresas não executam rotineiramente análises de tendências e hipóteses. A análise de hipóteses é o processo de determinar o efeito de uma alteração de rede. Tendências é o processo de execução de linhas de base de capacidade de rede e problemas de desempenho e revisão das linhas de base de tendências de rede para entender requisitos de atualização futuros. O gerenciamento de capacidade e desempenho também deve incluir o gerenciamento de exceções, onde os problemas são identificados e resolvidos antes da chamada dos usuários, e o gerenciamento de QoS, onde os administradores de rede planejam, gerenciam e identificam problemas individuais de desempenho de serviço. O gráfico a seguir ilustra os processos de gerenciamento de capacidade e desempenho.



O gerenciamento de capacidade e desempenho também tem suas limitações, geralmente relacionadas à CPU e à memória. A seguir estão possíveis áreas de preocupação:

- CPU

- Backplane ou E/S
- Memória e buffers
- Tamanhos de interface e tubulação
- Enfileiramento, latência e instabilidade
- Velocidade e distância
- Características do aplicativo

Algumas referências ao planejamento de capacidade e ao gerenciamento de desempenho também mencionam algo chamado "plano de dados" e "plano de controle". O plano de dados é simplesmente um problema de capacidade e desempenho envolvido com os dados que passam pela rede, enquanto o plano de controle significa recursos necessários para manter a funcionalidade adequada do plano de dados. A funcionalidade do plano de controle inclui sobrecarga de serviço, como roteamento, spanning tree, manutenção de atividade da interface e gerenciamento SNMP do dispositivo. Esses requisitos de plano de controle usam CPU, memória, buffer, enfileiramento e largura de banda exatamente como o tráfego que atravessa a rede. Muitos dos requisitos do plano de controle também são essenciais para a funcionalidade geral do sistema. Se eles não tiverem os recursos necessários, a rede falhará.

CPU

A CPU é normalmente usada pelo plano de controle e pelo plano de dados em qualquer dispositivo de rede. No gerenciamento de capacidade e desempenho, você deve garantir que o dispositivo e a rede tenham CPU suficiente para funcionar o tempo todo. A CPU insuficiente pode frequentemente recolher uma rede porque os recursos inadequados em um dispositivo podem afetar toda a rede. A CPU insuficiente também pode aumentar a latência, pois os dados devem esperar para serem processados quando não houver switching de hardware sem a CPU principal.

Backplane ou E/S

Backplane ou E/S se refere à quantidade total de tráfego que um dispositivo pode suportar, normalmente descrita em termos de tamanho de BARRAMENTO ou capacidade de backplane. Um painel traseiro insuficiente normalmente resulta em pacotes descartados, o que pode levar a retransmissões e tráfego adicional.

Memória

A memória é outro recurso que tem requisitos de plano de dados e plano de controle. A memória é necessária para informações como tabelas de roteamento, tabelas ARP e outras estruturas de dados. Quando os dispositivos ficam sem memória, algumas operações no dispositivo podem falhar. A operação pode afetar os processos do plano de controle ou do plano de dados, dependendo da situação. Se os processos do plano de controle falharem, a rede inteira poderá ser degradada. Por exemplo, isso pode acontecer quando é necessária memória extra para a convergência de roteamento.

Tamanhos de interface e tubulação

Os tamanhos de interface e tubulação se referem à quantidade de dados que pode ser enviada simultaneamente em qualquer conexão. Geralmente, essa velocidade é chamada incorretamente de velocidade de uma conexão, mas os dados realmente não viajam em velocidades diferentes de um dispositivo para outro. A velocidade do silício e a capacidade do hardware ajudam a determinar a largura de banda disponível com base na mídia. Além disso, os mecanismos de

software podem "acelerar" os dados para se adequarem a alocações de largura de banda específicas para um serviço. Geralmente, isso é visto em redes de provedores de serviços para frame-relay ou ATM que têm inerentemente capacidades de velocidade de 1,54 kbps a 155 mbs e superiores. Quando há limitações de largura de banda, os dados são enfileirados em uma fila de transmissão. Uma fila de transmissão pode ter diferentes mecanismos de software para priorizar dados dentro da fila; no entanto, quando houver dados na fila, ele deverá aguardar os dados existentes antes de encaminhar os dados para fora da interface.

Enfileiramento, latência e instabilidade

O enfileiramento, a latência e o jitter também afetam o desempenho. Você pode ajustar a fila de transmissão para afetar o desempenho de diferentes maneiras. Por exemplo, se a fila for grande, os dados aguardarão mais. Quando as filas são pequenas, os dados são descartados. Isso é chamado de taildrop e é aceitável para aplicações TCP, já que os dados serão retransmitidos. No entanto, voz e vídeo não funcionam bem com queda de fila ou mesmo com latência de fila significativa que exige atenção especial à largura de banda ou tamanhos de tubulação. O atraso da fila também pode ocorrer com filas de entrada se o dispositivo não tiver recursos suficientes para encaminhar imediatamente o pacote. Isso pode ser devido à CPU, memória ou buffers.

A latência descreve o tempo de processamento normal desde o momento em que é recebida até o momento em que o pacote é encaminhado. Os roteadores e switches de dados modernos normais têm latência extremamente baixa (< 1ms) sob condições normais sem restrições de recursos. Dispositivos modernos com processadores de sinal digital para converter e compactar pacotes de voz analógicos podem demorar mais, até até 20 ms.

O jitter descreve a lacuna entre pacotes para aplicações de streaming, incluindo voz e vídeo. Se os pacotes chegarem em momentos diferentes com temporização de intervalo entre pacotes diferente, o jitter será alto e a qualidade de voz será reduzida. O jitter é principalmente um fator de atraso na fila.

Velocidade e distância

Velocidade e distância também são um fator no desempenho da rede. As redes de dados têm uma velocidade consistente de encaminhamento de dados com base na velocidade da luz. Isso é aproximadamente 160 quilômetros por milissegundo. Se uma organização estiver executando uma aplicação cliente-servidor internacionalmente, ela pode esperar um atraso de encaminhamento de pacotes correspondente. A velocidade e a distância podem ser um fator importante no desempenho do aplicativo quando os aplicativos não estão otimizados para o desempenho da rede.

Características do aplicativo

As características do aplicativo são a última área que afeta a capacidade e o desempenho. Questões como tamanhos pequenos de janelas, manutenções de atividades de aplicativos e a quantidade de dados enviados pela rede versus o necessário podem afetar o desempenho de um aplicativo em muitos ambientes, especialmente WANs.

Melhores práticas de gerenciamento de capacidade e desempenho

Esta seção discute as cinco melhores práticas principais de gerenciamento de capacidade e desempenho em detalhes:

- [Gerenciamento de nível de serviço](#)
- [Análise de hipóteses de rede e aplicativo](#)
- [Linha de base e tendências](#)
- [Gerenciamento de exceções](#)
- [Gerenciamento de QoS](#)

Gerenciamento de nível de serviço

O gerenciamento de nível de serviço define e controla outros processos de gerenciamento de capacidade e desempenho necessários. Os gerentes de rede entendem que precisam de planejamento de capacidade, mas enfrentam restrições de orçamento e de pessoal que impedem uma solução completa. O gerenciamento de nível de serviço é uma metodologia comprovada que ajuda com problemas de recursos, definindo um produto final e criando responsabilidade bidirecional para um serviço vinculado a esse produto final. Isso pode ser feito de duas maneiras:

- Crie um contrato de nível de serviço entre os usuários e a organização de rede para um serviço que inclua gerenciamento de capacidade e desempenho. O serviço incluiria relatórios e recomendações para manter a qualidade do serviço. No entanto, os usuários devem estar preparados para financiar o serviço e todas as atualizações necessárias.
- A organização de rede define seu serviço de gerenciamento de capacidade e desempenho e depois tenta financiar esse serviço e atualizações caso a caso.

Em qualquer caso, a organização da rede deve começar definindo um serviço de planejamento de capacidade e gerenciamento de desempenho que inclua quais aspectos do serviço eles podem fornecer atualmente e o que está planejado no futuro. Um serviço completo incluiria uma análise de hipóteses para alterações de rede e alterações de aplicativos, linha de base e tendências para variáveis de desempenho definidas, gerenciamento de exceções para variáveis de capacidade e desempenho definidas e gerenciamento de QoS.

Análise What-if de rede e aplicativo

Executar uma análise de hipóteses de rede e aplicativo para determinar o resultado de uma alteração planejada. Sem uma análise de hipóteses, as organizações assumem riscos significativos para mudar o sucesso e a disponibilidade geral da rede. Em muitos casos, as alterações na rede resultaram em colapso congestivo, causando muitas horas de inatividade da produção. Além disso, uma quantidade surpreendente de introduções de aplicativos falha e causa impacto a outros usuários e aplicativos. Essas falhas continuam em muitas organizações de rede, mas podem ser totalmente evitadas com algumas ferramentas e algumas etapas adicionais de planejamento.

Normalmente, você precisa de alguns processos novos para executar uma análise de hipóteses de qualidade. A primeira etapa é identificar níveis de risco para todas as alterações e exigir uma análise mais detalhada do tipo e se para alterações de risco mais alto. O nível de risco pode ser um campo obrigatório para todos os envios de alterações. Alterações de nível de risco mais alto exigiriam uma análise de hipóteses definida da alteração. Uma análise hipotética da rede determina o efeito das alterações na utilização da rede e problemas de recursos do plano de controle da rede. Uma análise hipotética do aplicativo determinaria o sucesso do aplicativo do projeto, os requisitos de largura de banda e quaisquer problemas de recursos de rede. As tabelas a seguir são exemplos de atribuição de nível de risco e requisitos de teste correspondentes:

Nível de Risco	Definição	Alterar Recomendações de Planejamento
1	<ul style="list-style-type: none"> • Alto impacto potencial para um grande número de usuários (mais de 500) ou serviços essenciais aos negócios devido à introdução de novos produtos, softwares, topologias ou recursos. • A alteração envolve o tempo de inatividade esperado da rede. 	<ul style="list-style-type: none"> • Validar o laboratório da nova solução. A validação do laboratório inclui testes e validação documentados da solução e análise de hipóteses mostrando o impacto na infraestrutura existente. Recomendamos pilotos de solução. Novas soluções exigem a conclusão de um documento de suporte a operações. • Realizar a revisão do projeto do Cisco NSA. • Crie um plano de recuo. • Criar plano de implementação • Criar processo de alteração.
2	<ul style="list-style-type: none"> • Alto impacto potencial para um grande número de usuários (mais de 500) ou serviço essencial para os 	<ul style="list-style-type: none"> • Executar análise de hipóteses para determinar o

	<p>negócios devido a um grande aumento de tráfego ou usuários, alterações de backbone ou alterações de roteamento.</p> <ul style="list-style-type: none"> • A alteração pode exigir algum tempo de inatividade. 	<p>impacto no ambiente existente (deve ser feito em um ambiente de laboratório).</p> <ul style="list-style-type: none"> • Testar e revisar as alterações de roteamento quanto à funcionalidade. • Crie um plano de recuo. • Realizar a revisão do projeto para alterações importantes no roteamento ou no backbone. • Criar plano de implementação . • Criar processo de alteração.
3	<ul style="list-style-type: none"> • Impacto potencial médio para um número menor de usuários ou serviços comerciais devido a qualquer alteração fora do padrão. • Inclui novo produto, software, topologia, adição de recursos ou novos usuários, aumento do tráfego ou topologia fora do padrão. • A alteração pode exigir algum tempo de inatividade. 	<ul style="list-style-type: none"> • Executar análise de engenharia da nova solução (pode exigir validação de laboratório). • Criar plano de implementação . • Criar processo de alteração.
4	<ul style="list-style-type: none"> • Menor impacto potencial sobre o serviço ou o usuário. • Inclui a adição de novos módulos de rede de modelo padrão, como switches/hubs de servidor ou de 	<ul style="list-style-type: none"> • Criar plano de implementação . • Criar processo de alteração.

	<p>construção em roteadores.</p> <ul style="list-style-type: none"> • Inclui a criação de novos locais de WAN ou serviços adicionais de acesso comprovados. • Todas as alterações de nível de risco 3 foram tecnicamente comprovadas no ambiente de produção. • A alteração pode exigir algum tempo de inatividade. 	
5	<ul style="list-style-type: none"> • Nenhum impacto no usuário ou no serviço. • Inclui a adição de usuários individuais à rede e alterações de configuração padrão, como senha, banner, SNMP ou outros parâmetros de configuração padrão. • Sem tempo de inatividade. 	<ul style="list-style-type: none"> • Processo de alteração opcional.

Depois de definir onde você precisa da análise de hipóteses, você pode definir o serviço.

Você pode executar uma análise de hipóteses de rede com ferramentas de modelagem ou com um laboratório que imite o ambiente de produção. As ferramentas de modelagem são limitadas pela maneira como o aplicativo compreende os problemas de recursos do dispositivo e como a maioria das alterações de rede são dispositivos novos, o aplicativo pode não entender o efeito da alteração. O melhor método é construir alguma representação da rede de produção em um laboratório e testar o software, o recurso, o hardware ou a configuração desejada sob carga usando geradores de tráfego. As rotas de vazamento (ou outras informações de controle) da rede de produção para o laboratório também melhoram o ambiente do laboratório. Testar requisitos de recursos adicionais com diferentes tipos de tráfego, incluindo tráfego SNMP, broadcast, multicast, criptografado ou compactado. Com todas essas metodologias diferentes, analise os requisitos de recursos do dispositivo durante possíveis situações de estresse, como convergência de rota, oscilação de link e reinicializações de dispositivo. Os problemas de utilização de recursos incluem áreas de recursos de capacidade normal, como CPU, memória, utilização de painel traseiro, buffers e enfileiramento.

Novos aplicativos também devem executar uma análise de hipóteses para determinar o sucesso do aplicativo e os requisitos de largura de banda. Você normalmente executa essa análise em um ambiente de laboratório usando um analisador de protocolo e um simulador de atraso de WAN para entender o efeito da distância. Você só precisa de um PC, hub, dispositivo de atraso de WAN e roteador de laboratório conectado à rede de produção. Você pode simular a largura de banda no laboratório, acelerando o tráfego usando modelagem de tráfego genérica ou limitação de taxa no roteador de teste. O administrador de rede pode trabalhar em conjunto com o grupo de aplicativos para entender os requisitos de largura de banda, problemas de janelamento e possíveis problemas de desempenho para o aplicativo em ambientes de LAN e WAN.

Execute uma análise de hipóteses de um aplicativo antes de implantar qualquer aplicativo de

negócios. Se você não fizer isso, o grupo de aplicativos culpará a rede pelo mau desempenho. Se, de alguma forma, for possível exigir uma análise de hipóteses de um aplicativo para novas implantações por meio do processo de gerenciamento de alterações, você poderá ajudar a evitar implantações malsucedidas e entender melhor os aumentos repentinos no consumo de largura de banda para os requisitos de lote e cliente-servidor.

Linha de base e tendências

As linhas de base e tendências permitem que os administradores de rede planejem e concluem atualizações de rede antes que um problema de capacidade cause tempo de inatividade da rede ou problemas de desempenho. Comparar a utilização de recursos durante períodos de tempo sucessivos ou reduzir o volume de informações ao longo do tempo em um banco de dados e permitir que os planejadores exibam os parâmetros de utilização de recursos para a última hora, dia, semana, mês e ano. Em ambos os casos, alguém deve revisar as informações semanalmente, quinzenalmente ou mensalmente. O problema com a criação de linhas de base e tendências é que ela requer uma quantidade enorme de informações para serem analisadas em redes grandes.

Você pode resolver esse problema de várias maneiras:

- Crie muita capacidade e faça switching no ambiente de LAN para que a capacidade não seja um problema.
- Divida as informações de tendência em grupos e concentre-se em áreas críticas ou de alta disponibilidade da rede, como locais de WAN críticos ou LANs de data center.
- Os mecanismos de relatório podem destacar áreas que ficam acima de um determinado limite para atenção especial. Se você implementar primeiro as áreas críticas de disponibilidade, poderá reduzir significativamente a quantidade de informações necessárias para revisão.

Com todos os métodos anteriores, você ainda precisa revisar as informações periodicamente. As linhas de base e tendências são um esforço pró-ativo e, se a organização tiver apenas recursos para suporte reativo, os indivíduos não lerão os relatórios.

Muitas soluções de gerenciamento de rede fornecem informações e gráficos sobre variáveis de recursos de capacidade. Infelizmente, a maioria das pessoas só usa essas ferramentas para suporte reativo a um problema existente; isso anula o objetivo da criação de linha de base e tendências. Duas ferramentas eficientes no fornecimento de informações sobre tendências de capacidade para redes Cisco são o produto Concord Network Health e os produtos INS EnterprisePRO. Em muitos casos, as organizações de rede executam linguagens de script simples para coletar informações de capacidade. Abaixo estão alguns relatórios de exemplo que foram coletados via Script para utilização de link, utilização de CPU e desempenho de ping. Outras variáveis de recurso que podem ser importantes para a tendência incluem memória, profundidade da fila, volume de broadcast, buffer, notificação de congestionamento do Frame Relay e utilização do painel traseiro. Consulte estas tabelas para obter informações sobre a utilização do link e da CPU:

Utilização de link

Recurso	Endereço	Segmento	Utilização média (%)	Utilização de Pico (%)
JTKR01S2	10.2.6.1	128 Kbps	66.3	97.6

JYKR01S0	10.2.6.2	128 Kbps	66.3	97.8
FMCR18S4/4	10.2.5.1	384 Kbps	51.3	109.7
PACR01S3/1	10.2.5.2	384 Kbps	51.1	98.4

Utilização da CPU

Recurso	Endereço de pesquisa	Utilização média (%)	Utilização de Pico (%)
FSTR01	10.28.142.1	60.4	80
NERT06	10.170.2.1	47	86
NORR01	10.73.200.1	47	99
RTCR01	10.49.136.1	42	98

Utilização de link

Recurs o	Endereço	AvRes T (mS) 09-09- 98	AvRes T (mS) 09-09- 98	AvResT (mS) 09- 09-98	AvResT (mS) 10- 01-98
AADR. 01	10.190.56 .1	469.1	852.4	461.1	873.2
ABNR0 1	10.190.52 .1	486.1	869.2	489.5	880.2
ABR01	10.190.54 .1	490.7	883.4	485.2	892.5
ASAR0 1	10.196.17 0.1	619.6	912.3	613.5	902.2
ASR01	10.196.17 8.1	667.7	976.4	655.5	948.6
ASYR0 1S					503.4
AZWRT 01	10.177.32 .1	460.1		444.7	
BEJR0 1	10.195.18 .1	1023. 7	1064. 6	1184	1021.9

Gerenciamento de exceções

O gerenciamento de exceções é uma metodologia valiosa para identificar e resolver problemas de capacidade e desempenho. A ideia é receber notificações de violações de limite de capacidade e desempenho para investigar e corrigir imediatamente o problema. Por exemplo, um administrador de rede pode receber um alarme de alta utilização da CPU em um roteador. O administrador de rede pode fazer login no roteador para determinar por que a CPU está tão alta. Ela pode executar alguma configuração corretiva que reduza a CPU ou criar uma lista de acesso que impeça o tráfego que causa o problema, especialmente se o tráfego não parecer ser crítico para os negócios.

Você pode configurar o gerenciamento de exceções para problemas mais críticos de forma justa, usando os comandos de configuração RMON em um roteador ou usando ferramentas mais avançadas como o gerenciador de nível de serviço Netsys em conjunto com dados SNMP, RMON ou Netflow. A maioria das ferramentas de gerenciamento de rede tem a capacidade de definir limites e alarmes em violações. O aspecto importante do processo de gerenciamento de exceções é fornecer notificação quase em tempo real do problema. Caso contrário, o problema pode desaparecer antes que alguém note que a notificação foi recebida. Isso pode ser feito em um NOC se a organização tiver monitoramento consistente. Caso contrário, recomendamos a notificação por pager.

O exemplo de configuração a seguir fornece notificação de elevação e queda de limiar para a CPU do roteador para um arquivo de registro que pode ser revisto de forma consistente. Você pode configurar comandos RMON semelhantes para violações de limite de utilização de link crítico ou outros limites SNMP.

```
rmon event 1 trap CPUtrap description
"CPU Util >75%"rmon event 2 trap CPUtrap description
"CPU Util <75%"rmon event 3 trap CPUtrap description
"CPU Util >90%"rmon event 4 trap CPUtrap description
"CPU Util <90%"rmon alarm 75 lsystem.56.0 10 absolute rising-threshold
75 1 falling-threshold 75 2rmon alarm 90 lsystem.56.0 10 absolute rising-threshold
90 3 falling-threshold 90 4
```

Gerenciamento de QoS

O gerenciamento da qualidade de serviço envolve a criação e o monitoramento de classes de tráfego específicas dentro da rede. Um tráfego fornece um desempenho mais consistente para grupos de aplicativos específicos (definidos dentro de classes de tráfego). Os parâmetros de modelagem de tráfego fornecem flexibilidade significativa na priorização e na modelagem de tráfego para classes específicas de tráfego. Esses recursos incluem recursos como Taxa de Acesso Comprometida (CAR - Committed Access Rate), Detecção Antecipada Aleatória Ponderada (WRED - Weighted Random Early Detection) e Enfileiramento Ponderado Baseado em Classe. As classes de tráfego são normalmente criadas com base em SLAs de desempenho para aplicativos mais críticos de negócios e requisitos de aplicativos específicos, como voz. O tráfego não crítico ou não comercial também seria controlado de forma que não pudesse afetar aplicativos e serviços de prioridade mais alta.

A criação de classes de tráfego exige uma compreensão básica da utilização da rede, dos requisitos específicos de aplicativos e das prioridades dos aplicativos de negócios. Os requisitos de aplicativos incluem conhecimento do tamanho dos pacotes, problemas de tempo limite, requisitos de instabilidade, requisitos de intermitência, requisitos de lote e problemas gerais de desempenho. Com esse conhecimento, os administradores de rede podem criar planos e configurações de modelagem de tráfego que fornecem um desempenho de aplicativo mais consistente em uma variedade de topologias LAN/WAN.

Por exemplo, uma organização tem uma conexão ATM de 10 megabits entre dois locais principais. Às vezes, o link fica congestionado devido a transferências de arquivos grandes, o que causa degradação no desempenho do processamento de transações on-line e qualidade de voz ruim ou inutilizável.

A organização configurou quatro classes de tráfego diferentes. A voz recebeu a prioridade mais alta e pôde manter essa prioridade, mesmo se estourasse sobre a taxa de volume de tráfego estimada. A classe de aplicativos críticos recebeu a próxima prioridade mais alta, mas não foi permitido estourar o tamanho total do link menos os requisitos de largura de banda de voz

estimados. Quando estourar, será descartado. O tráfego de transferência de arquivos recebeu simplesmente uma prioridade mais baixa e todo o tráfego restante cabia em algum lugar no meio.

A organização agora precisa executar o gerenciamento de QoS nesse link para determinar a quantidade de tráfego que cada classe está recebendo e medir o desempenho dentro de cada classe. Se a organização não fizer isso, pode haver escassez de recursos para algumas classes ou os SLAs de desempenho podem não ser atendidos dentro de uma classe específica.

Gerenciar configurações de QoS ainda é uma tarefa difícil devido à falta de ferramentas. Um método é usar o gerenciador de desempenho de Internet (IPM) da Cisco para enviar tráfego diferente pelo link que se enquadra em cada uma das classes de tráfego. Você pode monitorar o desempenho de cada classe e o IPM oferece análise de tendências, em tempo real e salto por salto para identificar áreas com problemas. Outros ainda podem confiar em um método mais manual, como investigar o enfileiramento e os pacotes descartados em cada classe de tráfego com base nas estatísticas da interface. Em algumas organizações, esses dados podem ser coletados via SNMP ou analisados em um banco de dados para linhas de base e tendências. Algumas ferramentas também existem no mercado que enviam tipos de tráfego específicos através da rede para determinar o desempenho de um serviço ou aplicativo específico.

Coletando e Relatando Informações sobre Capacidade

A coleta e a comunicação de informações sobre a capacidade devem estar ligadas às três áreas recomendadas de gestão da capacidade:

- Análise de hipóteses, que se concentra na alteração da rede e como essa alteração afeta o ambiente
- Linha de base e tendências
- Gerenciamento de exceções

Em cada uma dessas áreas, desenvolva um plano de coleta de informações. No caso de análises de hipóteses de rede ou aplicativo, você precisa de ferramentas para imitar o ambiente de rede e entender o efeito da alteração em relação a possíveis problemas de recursos no plano de controle do dispositivo ou no plano de dados. No caso de linha de base e tendências, você precisa de instantâneos para dispositivos e links que mostram a utilização atual de recursos. Em seguida, você analisa os dados ao longo do tempo para entender os possíveis requisitos de atualização. Isso permite que os administradores de rede planejem corretamente as atualizações antes que surjam problemas de capacidade ou desempenho. Quando surgem problemas, você precisa do gerenciamento de exceções para alertar os administradores de rede para que eles possam ajustar a rede ou corrigir o problema.

Esse processo pode ser dividido nas seguintes etapas:

1. Determine suas necessidades.
2. Defina um processo.
3. Definir áreas de capacidade.
4. Defina as variáveis de capacidade.
5. Interprete os dados.

Determine suas necessidades

O desenvolvimento de um plano de gerenciamento de capacidade e desempenho requer a

compreensão das informações necessárias e da finalidade dessas informações. Divida o plano em três áreas obrigatórias: uma para análise de hipóteses, linha de base/tendências e gerenciamento de exceções. Em cada uma dessas áreas, descubra quais recursos e ferramentas estão disponíveis e o que é necessário. Muitas organizações não conseguem implantar ferramentas porque consideram a tecnologia e os recursos das ferramentas, mas não consideram as pessoas e a experiência necessárias para gerenciar as ferramentas. Inclua as pessoas e o conhecimento necessários em seu plano, bem como melhorias no processo. Essas pessoas podem incluir administradores de sistema para gerenciar as estações de gerenciamento de rede, administradores de banco de dados para ajudar na administração do banco de dados, administradores treinados para usar e monitorar as ferramentas e administradores de rede de nível superior para determinar políticas, limites e requisitos de coleta de informações.

Definir um processo

Você também precisa de um processo para garantir que a ferramenta seja usada com sucesso e de forma consistente. Você pode exigir melhorias no processo para definir o que os administradores de rede devem fazer quando ocorrem violações de limite ou que processo seguir para criar linha de base, criar tendências e atualizar a rede. Depois de determinar os requisitos e recursos para o planejamento de capacidade bem-sucedido, você pode considerar a metodologia. Muitas organizações optam por terceirizar esse tipo de funcionalidade para uma organização de serviços de rede, como a INS, ou criar a especialização internamente, pois consideram o serviço uma competência central.

Definir Áreas de Capacidade

O plano de planejamento da capacidade deve também incluir uma definição das zonas de capacidade. Estas são áreas da rede que podem compartilhar uma estratégia comum de planejamento de capacidade: por exemplo, a LAN corporativa, os escritórios de campo da WAN, os locais críticos da WAN e o acesso discado. Definir diferentes áreas é útil por vários motivos:

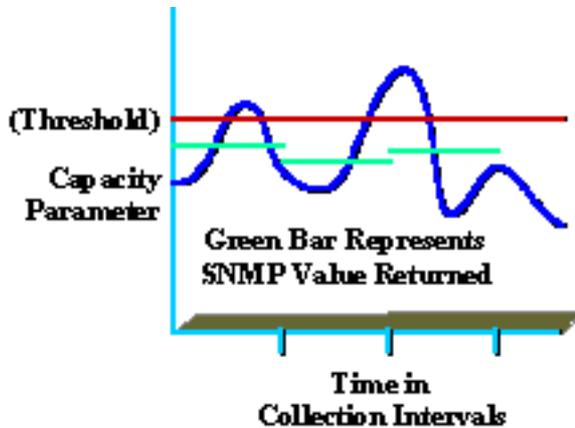
- Áreas diferentes podem ter limites diferentes. Por exemplo, a largura de banda da LAN é muito mais barata do que a largura de banda da WAN, portanto os limites de utilização devem ser menores.
- Áreas diferentes podem exigir o monitoramento de variáveis MIB diferentes. Por exemplo, os contadores FECN e BECN no Frame Relay são críticos para entender os problemas de capacidade do Frame Relay.
- Pode ser mais difícil ou demorado atualizar algumas áreas da rede. Por exemplo, os circuitos internacionais podem ter prazos de entrega muito mais longos e precisam de um nível correspondente de planejamento mais alto.

Definir as Variáveis de Capacidade

A próxima área importante é definir as variáveis a serem monitoradas e os valores de limite que exigem ação. A definição das variáveis de capacidade depende significativamente dos dispositivos e meios físicos usados na rede. Em geral, parâmetros como CPU, memória e utilização de link são valiosos. No entanto, outras áreas podem ser importantes para tecnologias ou requisitos específicos. Isso pode incluir profundidades de fila, desempenho, notificação de congestionamento de frame-relay, utilização de painel traseiro, utilização de buffer, estatísticas de netflow, volume de broadcast e dados RMON. Tenha em mente seus planos de longo prazo, mas comece com apenas algumas áreas importantes para ajudar a garantir o sucesso.

Interprete os dados

Entender os dados coletados também é fundamental para fornecer um serviço de alta qualidade. Por exemplo, muitas empresas não entendem totalmente os níveis de utilização média e de pico. O diagrama a seguir mostra um pico de parâmetro de capacidade com base em um intervalo de coleta SNMP de 5 minutos (mostrado em verde).



Mesmo que o valor relatado seja menor que o limite (mostrado em vermelho), os picos ainda podem ocorrer dentro do intervalo de coleta que estejam acima do valor limite (mostrado em azul). Isso é significativo porque durante o intervalo de coleta, a organização pode estar experimentando valores de pico que afetam o desempenho ou a capacidade da rede. Tome cuidado para selecionar um intervalo de coleta significativo que seja útil e que não cause sobrecarga excessiva.

Outro exemplo é a utilização média. Se os funcionários estiverem apenas no escritório de oito a cinco, mas a utilização média for 7X24, as informações poderão induzir em erro.

Informações Relacionadas

- [Suporte Técnico - Cisco Systems](#)

Sobre esta tradução

A Cisco traduziu este documento com a ajuda de tecnologias de tradução automática e humana para oferecer conteúdo de suporte aos seus usuários no seu próprio idioma, independentemente da localização.

Observe que mesmo a melhor tradução automática não será tão precisa quanto as realizadas por um tradutor profissional.

A Cisco Systems, Inc. não se responsabiliza pela precisão destas traduções e recomenda que o documento original em inglês ([link fornecido](#)) seja sempre consultado.