

# Gerenciamento de desempenho e capacidade: White Paper de práticas recomendadas

ID do Documento: 20769

Atualizado em: outubro 04, 2005



[Transferência PDF](#)



[Imprimir](#)

[Feedback](#)

## Produtos Relacionados

- [Alta Disponibilidade](#)
- [QoS Congestion Management \(queueing\)](#)

## Índice

[Introdução](#)

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

[CPU](#)

[Backplane ou I/O](#)

[Memória](#)

[Tamanhos da relação e da tubulação](#)

[Enfileiramento, latência, e Jitter](#)

[Velocidade e distância](#)

[Características do aplicativo](#)

[Melhores prática do gerenciamento de desempenho e capacidade](#)

[Service Level Management](#)

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

[Linha de base e tensão](#)

[Gerência da exceção](#)

[Gerenciamento de QoS](#)

[Recolhendo e relatando a informação de capacidade](#)

[Determine suas necessidades](#)

[Defina um processo](#)

[Defina áreas da capacidade](#)

[Defina os variáveis de capacidade](#)

[Interprete os dados](#)

[Informações Relacionadas](#)

[Cisco relacionado apoia discussões da comunidade](#)

## Introdução

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 assuntos:

- Potencialidade geral e problemas de desempenho, incluindo os riscos e as edições da capacidade potencial dentro das redes.
- Melhores prática do gerenciamento de desempenho e capacidade, incluindo análises What-if, linha de base, tensão, gerência da exceção, e gerenciamento de QoS.
- Como desenvolver uma estratégia de planejamento de capacidade, incluindo técnicas comum, ferramentas, variáveis MIB, e pontos iniciais usados no planejamento da capacidade.

## Vista geral do gerenciamento de desempenho e capacidade

O planejamento da capacidade é o processo de determinar os recursos de rede exigidos impedir um desempenho ou um impacto de disponibilidade em aplicativos business-critical. O Gerenciamento de desempenho é a prática controlar do tempo de resposta de serviço de rede, a consistência, e os serviços do qualidade para serviços gerais e individuais e os totais.

**Nota:** Os problemas de desempenho são relacionados geralmente à capacidade. Os aplicativos são mais lentos porque a largura de banda e os dados devem esperar nas filas antes de ser transmitido através da rede. Nas Aplicações de voz, os problemas como o retardo e tremulação afetam diretamente a qualidade da chamada de voz.

A maioria de organizações já recolhem alguns informação e trabalho capacidade-relacionados consistentemente para resolver problemas, planeiam mudanças, e executam a nova função e a funcionalidade de desempenho. Contudo, as organizações não executam rotineiramente a tensão e as análises What-if. A análise What-if é o processo de determinar a influência de uma alteração de rede. Tender é o processo de execução consiste linhas de base da capacidade de rede e os problemas de desempenho e a revisão das linhas de base para a rede tendem para compreender os requisitos de upgrade futuros. O gerenciamento de desempenho e capacidade deve igualmente incluir a gerência da exceção onde os problemas são identificados e resolvido antes dos usuários chame dentro, e o gerenciamento de QoS onde os administradores de rede planeiam, controlam, e identificam problemas de desempenho do serviço individual. O seguinte gráfico ilustra processos do gerenciamento de desempenho e capacidade.

O gerenciamento de desempenho e capacidade igualmente tem suas limitações, relativas tipicamente ao CPU e à memória. Os seguintes são áreas potenciais para o interesse:

- CPU
- Backplane ou I/O
- Memória e buffers
- Tamanhos da relação e da tubulação

- Enfileiramento, latência, e tremor
- Velocidade e distância
- Características do aplicativo

Algumas referências ao planejamento da capacidade e ao Gerenciamento de desempenho igualmente mencionam algo chamado dos “o plano dados” e do “o plano controle”. O plano dos dados é simplesmente capacidade e problemas de desempenho envolvidos com os dados que atravessam a rede quando o plano do controle significar os recursos exigidos manter a funcionalidade apropriada do plano dos dados. A funcionalidade plana do controle inclui despesas gerais do serviço tais como o roteamento, medindo - árvore, manutenções de atividade da relação, e gerenciamento de SNMP do dispositivo. Estes requisitos de placa do controle usam o CPU, a memória, a proteção, o Enfileiramento, e a largura de banda apenas como o tráfego que atravessa a rede. Muitos dos requisitos de placa do controle são igualmente essenciais à funcionalidade total do sistema. Se não têm os recursos que precisam, a rede falha.

## CPU

O CPU é usado tipicamente pelo plano do controle e os dados aplanam em todo o dispositivo de rede. No gerenciamento de desempenho e capacidade, você deve assegurar-se de que o dispositivo e a rede tenham o suficiente CPU a funcionar em todas as vezes. CPU insuficiente pode frequentemente desmoronar uma rede porque os recursos inadequados em um dispositivo podem impactar a toda a rede. CPU insuficiente pode igualmente aumentar a latência desde que os dados devem esperar para ser processado quando não há nenhum switching de hardware sem o CPU principal.

## Backplane ou I/O

O backplane ou o I/O referem a quantidade total de tráfego que um dispositivo pode segurar, descrita geralmente em termos do tamanho ou da potencialidade de placa-mãe do BARRAMENTO. A placa-mãe insuficiente conduz normalmente aos pacotes descartado, que podem conduzir às retransmissões e ao tráfego adicional.

## Memória

A memória é um outro recurso que tenha requisitos de placa do plano e do controle dos dados. A memória é exigida para a informação tal como tabelas de roteamento, tabelas ARP, e outras estruturas de dados. Quando os dispositivos executados fora da memória, algumas operações no dispositivo puderem falhar. A operação podia afetar processos planos do controle ou processos planos dos dados, segundo a situação. Se os processos planos do controle falham, a toda a rede pode degradar. Por exemplo, isto pode acontecer quando a memória extra é exigida para a convergência do roteamento.

## Tamanhos da relação e da tubulação

Os tamanhos da relação e da tubulação referem a quantidade de dados que podem ser enviados simultaneamente em toda a uma conexão. Isto frequentemente é referido incorretamente como a velocidade de uma conexão, mas os dados realmente não viajam em velocidades diferentes de um dispositivo a outro. A velocidade do silicone e a ajuda da capacidade do hardware determinam a largura de banda disponível baseada nos media. Além, os mecanismos de software enlatam dados do “regulador de pressão” para conformar-se às alocações de largura de banda específicas para um serviço. Você veem tipicamente este nas redes de provedor de serviços para o Frame

Relay ou o ATM que têm inerentemente capacidades da velocidade de 1.54kpbs a 155mbs e mais altamente. Quando há umas limitações de largura de banda, os dados estão enfileirados em um transmitir fila. Um transmitir fila pode ter os mecanismos de software diferentes para dar a prioridade a dados dentro da fila; contudo, quando há uns dados na fila, deve esperar dados existentes antes que possa enviar aos dados para fora a relação.

## Enfileiramento, latência, e Jitter

O Enfileiramento, latência, e treme igualmente desempenho da influência. Você pode ajustar o transmitir fila para afetar o desempenho em maneiras diferentes. Por exemplo, se a fila é grande, a seguir os dados esperam mais por muito tempo. Quando as filas são pequenas, os dados estão deixados cair. Isto é chamado taildrop e é aceitável para aplicativos de TCP/IP desde que os dados serão retransmitidos. Contudo, a Voz e o vídeo não executam bem com a queda da fila ou mesmo a latência significativa da fila que exigem a atenção especial aos tamanhos da largura de banda ou da tubulação. O atraso da fila pode igualmente ocorrer com filas de entrada se o dispositivo não tem os recursos suficientes para enviar imediatamente o pacote. Isto pode ser devido ao CPU, à memória, ou aos buffers.

A latência descreve o tempo de processamento normal do tempo onde é recebida até que o tempo o pacote é enviado. Os switch de dados e o Roteadores modernos normais têm extremamente - a latência baixa (< 1ms) em condições normais sem confinamentos de recurso. Os dispositivos modernos com os processadores do sinal digital a converter e os pacotes da voz analógica da compressa podem tomar mais por muito tempo, mesmo até 20ms.

O Jitter descreve a lacuna inter-pacote para aplicativos fluentes, incluindo a Voz e o vídeo. Se os pacotes chegam em horas diferentes com sincronismo diferente da lacuna inter-pacote, a seguir o tremor é alto e a Qualidade de voz degrada. O Jitter é principalmente um fator do retardo de enfileiramento.

## Velocidade e distância

A velocidade e a distância são igualmente um fator no desempenho da rede. As redes de dados têm uma velocidade consistente do encaminhamento de dados baseada na velocidade da luz. Esta é aproximadamente 100 milhas por milissegundos. Se uma organização está executando um aplicativo de cliente/servidor internacionalmente, a seguir podem esperar um atraso correspondente do encaminhamento de pacote. A velocidade e a distância podem ser um fator tremendo no desempenho do aplicativo quando os pedidos não são aperfeiçoados para o desempenho da rede.

## Características do aplicativo

As características do aplicativo são a última área que afeta a capacidade e o desempenho. As edições tais como tamanhos de janela pequenos, Keepalives do aplicativo, e a quantidade de dados enviados sobre a rede contra o que é exigida podem afetar o desempenho de um aplicativo em muitos ambientes, especialmente WAN.

## Melhores prática do gerenciamento de desempenho e capacidade

Esta seção discute os cinco melhores prática principais do gerenciamento de desempenho e

capacidade em detalhe:

- [Service Level Management](#)
- [Análise What-if da rede e do aplicativo](#)
- [Linha de base e tensão](#)
- [Gerência da exceção](#)
- [Gerenciamento de QoS](#)

## [Service Level Management](#)

O Service Level Management define e regula outros processos exigidos do gerenciamento de desempenho e capacidade. As gerentes de rede compreendem que precisam o planejamento da capacidade, mas enfrentam a realização do orçamento e prover de pessoal das limitações que impedem uma solução completa. O Service Level Management é uma metodologia provada que as ajudas com edições do recurso definindo uma responsabilidade em dois sentidos que pode entregar-se e criadora para um serviço amarrem a esse que pode entregar-se. Você pode realizar este em duas maneiras:

- Crie um Service Level Agreement entre usuários e a organização de rede para um serviço que inclua o gerenciamento de desempenho e capacidade. O serviço incluiria relatórios e recomendações manter a qualidade do serviço. Contudo, os usuários devem ser preparados para financiar o serviço e todas as elevações exigidas.
- A organização de rede define seu serviço do gerenciamento de desempenho e capacidade e então tenta o financiamento para esse serviço e promove-o numa base casuística.

Em todos os casos, a organização de rede deve começar definindo um planejamento da capacidade e um serviço de gerenciamento do desempenho que inclua que aspectos do serviço podem atualmente fornecer e o que é planejado no futuro. Um serviço completo incluiria uma análise What-if para alterações de rede e mudanças do aplicativo, linha de base e tensão para variáveis de desempenho definidos, gerência da exceção para a capacidade e variáveis de desempenho definidos, e gerenciamento de QoS.

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

Execute uma análise What-if da rede e do aplicativo para determinar o resultado de uma alteração planejada. Sem uma análise What-if, as organizações tomam riscos significativos mudar o sucesso e a disponibilidade da rede total. Em muitos casos, as alterações de rede conduziram ao colapso congestivo que causa a muitos o tempo ocioso da máquina das horas de produção. Além, uma quantidade repentina de introduções de aplicativo falha e causa o impacto a outros usuários e aplicativos. Estas falhas continuam em muitas organizações de rede, contudo são completamente evitáveis com algumas ferramentas e algumas etapas de planejamento adicionais.

Você precisa normalmente alguns processos novos de executar uma análise What-if da qualidade. A primeira etapa é identificar níveis de risco para todas as mudanças e exigir uma análise What-if mais detalhada para umas alterações de risco mais altas. O nível de risco pode ser um campo requerido para todas as submissões da mudança. Umas mudanças mais altas do nível de risco exigiriam então uma análise What-if definida da mudança. Uma análise What-if da rede determina a influência das alterações de rede em edições dos recursos de controle plano da utilização de rede e da rede. Uma análise What-if do aplicativo determinaria o sucesso de aplicativo do projeto, os requisitos de largura de banda, e as todas as edições dos recursos de rede. As tabelas a seguir são exemplos de requisitos de teste da atribuição e da correspondência

do nível de risco:

Nível de Risco	Definição	Mude recomendações do planeamento
1	<ul style="list-style-type: none"> <li>• Impacto potencial elevado ao número grande dos usuários (500+) ou de serviço crítico do negócio devido aos novos produtos, ao software, à topologia, ou à introdução da característica.</li> <li>• A mudança envolve o tempo ocioso da máquina previsto da rede.</li> </ul>	<ul style="list-style-type: none"> <li>• Valide o laboratório da solução nova. A validação de laboratório inclui o teste de solução documentado e validação e análise What-if que mostram o impacto a infraestrutura existente. Nós recomendamos pilotos da solução. As soluções novas exigem a conclusão de um documento de suporte das operações.</li> <li>• Execute a revisão de projeto de Cisco NSA.</li> <li>• Crie o plano de recuo.</li> <li>• Crie o plano de implementação</li> <li>• Crie o processo de alteração.</li> </ul>
2	<ul style="list-style-type: none"> <li>• Impacto potencial elevado ao número grande dos usuários (500+) ou serviço</li> </ul>	<ul style="list-style-type: none"> <li>• Execute a análise What-if para</li> </ul>

	<p>crítico do negócio devido a um grande aumento do tráfego ou os usuários, as mudanças da espinha dorsal, ou as mudanças de roteamento.</p> <ul style="list-style-type: none"> <li>• A mudança pode exigir algum tempo ocioso da máquina.</li> </ul>	<p>determinar o impacto ao ambiente existente (deve ser feito no ambiente de laboratório).</p> <ul style="list-style-type: none"> <li>• Mudanças de roteamento do teste e da revisão para a funcionalidade.</li> <li>• Crie o plano de recuo.</li> <li>• Execute a revisão de projeto para o roteamento principal ou as mudanças da espinha dorsal.</li> <li>• Crie o plano de implementação</li> <li>.</li> <li>• Crie o processo de alteração.</li> </ul>
3	<ul style="list-style-type: none"> <li>• Impacto potencial médio ao número menor de usuários ou de serviço de negócio devido a alguma alteração sem padronização.</li> <li>• Inclui novos produtos, software, topologia, adição de características ou de novos usuários, tráfego aumentado, ou a topologia não padronizada.</li> <li>• A mudança pode exigir algum tempo ocioso da máquina.</li> </ul>	<ul style="list-style-type: none"> <li>• Execute a análise de engenharia da solução nova (pode exigir a validação de laboratório).</li> <li>• Crie o plano de implementação</li> <li>.</li> <li>• Crie o processo de alteração.</li> </ul>
4	<ul style="list-style-type: none"> <li>• Abaixar o serviço potencial ou o impacto do usuário.</li> <li>• Inclui adicionar os módulos de rede de molde padrão novos, tais como a construção ou os switch de</li> </ul>	<ul style="list-style-type: none"> <li>• Crie o plano de implementação</li> <li>.</li> <li>• Crie o processo de alteração.</li> </ul>

	<p>servidor/Hubs no Roteadores.</p> <ul style="list-style-type: none"> <li>• Inclúes que traz acima locais MACILENTOS novos ou serviços de acesso testados adicionais.</li> <li>• Todas as mudanças do nível de risco 3 foram provadas tecnicamente no ambiente de produção.</li> <li>• A mudança pode exigir algum tempo ocioso da máquina.</li> </ul>	
5	<ul style="list-style-type: none"> <li>• Nenhum impacto do usuário ou do serviço.</li> <li>• Inclui adicionar usuários individuais às mudanças da rede e de configuração padrão tais como a senha, a bandeira, o SNMP, ou os outros parâmetros de configuração padrão.</li> <li>• Nenhum tempo ocioso da máquina.</li> </ul>	<ul style="list-style-type: none"> <li>• Processo de alteração opcional.</li> </ul>

Uma vez que você define onde você precisa a análise What-if, você pode definir o serviço.

Você pode executar uma análise What-if da rede com as ferramentas de modelo ou com um laboratório que imite o ambiente de produção. As ferramentas de modelo são limitadas por como jorra o aplicativo compreende as edições dos recursos de dispositivo e por desde que a maioria de alterações de rede são dispositivos novos, o aplicativo não pode compreender o efeito da mudança. O melhor método é construir alguma representação da rede de produção em um laboratório e testar o software desejado, a característica, o hardware, ou a configuração sob a carga usando geradores de tráfego. Escapar rotas (ou a outra informação de controle) da rede de produção no laboratório igualmente aumenta o ambiente de laboratório. Teste exigências de recursos adicionais com os tipos de tráfego diferentes, incluindo o SNMP, a transmissão, o Multicast, cifrado, ou o tráfego compactado. Com as todas estas metodologias diferentes, analise as exigências de recursos de dispositivo durante situações de estresse potencial tais como a convergência de rota, o não sincronismo de link, e os reinícios do dispositivo. As questões de utilização de recurso incluem áreas normais dos recursos de potencialidade tais como o CPU, a memória, a utilização de backplane, os bufferes, e o Enfileiramento.

Os aplicativos novos devem igualmente executar uma análise What-if para determinar o sucesso de aplicativo e os requisitos de largura de banda. Você executa normalmente esta análise em um ambiente de laboratório usando um analisador de protocolo e um simulador de retardo WAN para compreender o efeito da distância. Você precisa somente um PC, um hub, um dispositivo MACILENTO do atraso, e um roteador de laboratório conectado na rede de produção. Você pode simular a largura de banda no laboratório estrangulando o tráfego usando o Formatação de tráfego genérico ou a taxa limite no roteador de teste. O administrador de rede pode trabalhar conjuntamente com o grupo de aplicativo para compreender requisitos de largura de banda,



edições do windowing, e edições de desempenho potencial para o aplicativo no LAN e nos ambientes de WAN.

Execute uma análise What-if do aplicativo antes de distribuir qualquer aplicativo de negócio. Se você não faz este, o grupo de aplicativo responsabiliza a rede pelo desempenho ruim. Se você pode de algum modo exigir uma análise What-if do aplicativo para disposições novas através do processo de gerenciamento de alteração, você pode ajudar a impedir distribuições sem sucesso e a compreender melhor aumentos repentinos no consumo de largura de banda para o servidor cliente e as requisições de arquivos de lote.

## Linha de base e tensão

A linha de base e a tensão permitem que os administradores de rede planeiem e as upgrades de rede completas antes que um problema de potencialidade cause o tempo ocioso da máquina ou os problemas de desempenho da rede. Compare a utilização de recurso durante períodos de tempo sucessivos ou destile a informação para baixo ao longo do tempo em um base de dados e permita que os planejadores ver parâmetros da utilização de recurso pela últimos hora, dia, semana, mês, e ano. Em qualquer dos casos, alguém deve rever a informação um semanal, quinzenal, ou no mensal. O problema com linha de base e tensão é que exige uma quantidade de informação opressivamente rever nas redes grandes.

Você pode resolver este problema é diversas maneiras:

- A abundância da construção da capacidade e do interruptor no ambiente de LAN assim que na capacidade é não uma edição.
- Divida a informação da tendência em grupos e concentre-a no requisito de alta disponibilidade ou nas áreas crítica da rede, tais como as sites WAN crítico ou o centro de dados LAN.
- Os mecanismos de relatório podem destacar as áreas que caem acima de um determinado limiar de atenção especial. Se você executa áreas de disponibilidade crítica primeiramente, você pode significativamente reduzir a quantidade de informação exigida para a revisão.

Com todos os métodos anteriores, você ainda precisa de rever a informação em uma base periódica. A linha de base e a tensão são um esforço pró-ativo e se a organização tem somente recursos para o suporte reagente, os indivíduos não lerão os relatórios.

Muitas soluções de gerenciamento de rede fornecem a informação e os gráficos em variáveis dos recursos de potencialidade. Infelizmente, a maioria de povos usam somente estas ferramentas para o suporte reagente a um problema existente; isto derrota a finalidade da linha de base e da tensão. Duas ferramentas que são eficazes em fornecer a informação da tendência da capacidade para redes Cisco são o produto dos acordos para sanidade da rede e o Produtos INS EnterprisePRO. Em muitos casos, as organizações de rede executam linguagens de script simples para recolher a informação de capacidade. Estão abaixo alguns exemplos de relatório que foram recolhidos através do script para a utilização do enlace, a utilização CPU, e o desempenho do sibilo. Outros variáveis de recurso que podem ser importantes de tender incluem a memória, a profundidade de fila, o volume de transmissão, o buffer, a notificação de congestionamento do Frame Relay, e a utilização de backplane. Refira estes à tabela para obter informações sobre da utilização do enlace e da utilização CPU:

### Utilização do enlace

Recurso	Endereç	Segment	Utilizaçã	Pico de
---------	---------	---------	-----------	---------

	o	o	o média (%)	utilizaçã o (%)
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 polling	Utilização média (%)	Pico de utilização (%)
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 do enlace

Recurso	Endereço	AvResT (Senhora) 09-09-98	AvResT (Senhora) 09-09-98	AvResT (Senhora) 09-09-98	AvResT (Senhora) 10-01-98
AADR01	10.190.56.1	469.1	852.4	461.1	873.2
ABNR01	10.190.52.1	486.1	869.2	489.5	880.2
APRR01	10.190.54.1	490.7	883.4	485.2	892.5
ASAR01	10.196.170.1	619.6	912.3	613.5	902.2
ASRR01	10.196.178.1	667.7	976.4	655.5	948.6
ASYR01S					503.4
AZWR01	10.177.32.1	460.1		444.7	
BEJR01	10.195.18.1	1023.7	1064.6	1184	1021.9

### Gerência da exceção

A gerência da exceção é uma metodologia valiosa para a capacidade e problemas de desempenho de identificação e de resolução. A ideia é receber violações da notificação de capacidade e do limiar de desempenho a fim investigar e fixar imediatamente o problema. Por

exemplo, um administrador de rede pôde receber um alarme para a alta utilização da CPU em um roteador. O administrador de rede pode registrar no roteador para determinar porque o CPU é tão alto. Pode então executar alguma configuração de solução que reduz o CPU ou cria uma lista de acesso que impede o tráfego que causa o problema, especialmente se o tráfego não parece ser crítico para negócio.

Você pode configurar a gerência da exceção para mais problemas críticos que usam razoavelmente simplesmente comandos de configuração de rmon em um roteador ou que usam umas ferramentas mais avançadas tais como o gerente do nível de serviço netsys conjuntamente com o SNMP, o RMON ou os dados de Netflow. A maioria de ferramentas de gerenciamento da rede têm a capacidade de ajustar pontos iniciais e alarmes em violações. O aspecto importante do processo da gerência da exceção é fornecer perto da notificação do tempo real da edição. Se não, o problema pode desaparecer antes de qualquer um observou que a notificação esteve recebida. Isto pode ser feito dentro de um NOC se a organização tem a monitoração consistente. Se não, nós recomendamos a notificação de pager.

O seguinte exemplo de configuração fornece a notificação da aumentação e do limiar de queda para o CPU de roteador a um arquivo de registro que possa ser revisto de forma consistente. Você pode estabelecer comandos rmon similares para violações de limiar críticas da utilização do enlace ou outros limiares 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 de Qualidade de Serviço envolve criar e monitorar classes de tráfego específicas dentro da rede. Um tráfego fornece mais desempenho consistente para os grupos de aplicativo específicos (definidos dentro das classes de tráfego). Os parâmetros de modelagem de tráfego fornecem a flexibilidade significativa na prioridade e no modelagem de tráfego para classes de tráfego específicas. Estas características incluem capacidades tais como o Committed Access Rate (CAR), o Weighted Random Early Detection (WRED), e o Enfileiramento tornado mais pesado feira baseado classe. As classes de tráfego são criadas normalmente com base em slas de desempenho para mais aplicativos críticos do negócio e em requisitos do aplicativo específicos tais como a Voz. o tráfego NON-crítico ou do NON-negócio seria controlado igualmente em uma maneira que não pode afetar aplicativos de prioridade mais elevada e serviços.

Criar classes de tráfego exige uma compreensão da linha de base da utilização de rede, de requisitos do aplicativo específicos, e de prioridades do aplicativo de negócio. Os requisitos do aplicativo incluem o conhecimento dos tamanhos do pacote, das edições do intervalo, das exigências do tremor, das exigências da explosão, das requisições de arquivos de lote, e das edições de desempenho geral. Com este conhecimento, os administradores de rede podem criar os planos e as configurações do modelagem de tráfego que fornecem um desempenho do aplicativo mais consistente através de uma variedade de topologias LAN/WAN.

Por exemplo, uma organização tem uma conexão ATM 10-megabit entre duas sites principal. O link torna-se às vezes congestionado de grandes transferências de arquivo, que causa a degradação do desempenho para o Transaction Processing em linha e a Qualidade de voz deficiente ou inusável.

A organização estabelece quatro classes de tráfego diferentes. A Voz foi dada a prioridade mais alta e reservada manter essa prioridade mesmo se estourou sobre a taxa calculada do volume de tráfego. A classe do aplicativo crítico foi dada a prioridade mais alta seguinte mas não foi permitida estourar sobre o tamanho do enlace total menos as exigências calculadas da largura de banda de voz. Quando estoura, estará deixada cair. O tráfego de transferência de arquivo foi dado simplesmente uma baixa prioridade e todo tráfego restante coube em algum lugar no meio.

A organização agora tem que executar o gerenciamento de QoS neste link para determinar quanto tráfego cada classe está tomando e para medir o desempenho dentro de cada classe. Se a organização não faz esta, a inanição pode ocorrer para algumas classes ou as slas de desempenho não podem ser encontradas dentro de uma classe particular.

Controlar configurações de QoS é ainda umas tarefas difíceis devido à falta das ferramentas. Um método é usar o gerente do desempenho de Internet de Cisco (IPM) para enviar o tráfego diferente através do link que cai em cada um das classes de tráfego. Você poderia monitora então o desempenho para cada classe e o IPM fornece a tensão, a análise em tempo real, e a análise do salto a salto para localizar áreas do problema. Outro podem ainda confiar em mais método manual como a investigação do Enfileiramento e dos pacotes descartado dentro de cada classe de tráfego baseada em estatísticas da relação. Em algumas organizações, estes dados podem ser recolhidos através do SNMP ou ser analisados gramaticalmente em um base de dados de linhas de base e em uma tensão. Algumas ferramentas igualmente existem no mercado que enviam tipos de tráfego específicos através da rede para determinar o desempenho para um serviço particular ou um aplicativo.

## Recolhendo e relatando a informação de capacidade

Recolher e relatar a informação de capacidade devem ser ligados às três áreas recomendadas do gerenciamento de capacidade:

- Análise What-if, que se centra em torno da alteração de rede e como a mudança afeta o ambiente
- Linha de base e tensão
- Gerência da exceção

Dentro de cada um destas áreas, desenvolva um plano de coleção de informação. No caso das análises What-if da rede ou do aplicativo, você precisa ferramentas de imitar o ambiente de rede e de compreender a influência das edições potenciais relativas a do recurso da mudança dentro do plano do controle de dispositivo ou do plano dos dados. No caso da linha de base e da tensão, você precisa os snapshots de dispositivos e os links que mostram a utilização dos recursos atual. Você revê então os dados ao longo do tempo para compreender requisitos de upgrade potenciais. Isto permite que os administradores de rede planeiem corretamente elevações antes que a capacidade ou os problemas de desempenho elevarem. Quando os problemas elevaram, você precisa a gerência da exceção de alertar os administradores de rede assim que podem ajustar a rede ou fixar o problema.

Este processo pode ser dividido nas seguintes etapas:

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

## Determine suas necessidades

Desenvolver um plano do gerenciamento de desempenho e capacidade exige a compreensão da informação que você precisa e a finalidade dessa informação. Rache o plano em três áreas exigidas: um cada para a análise What-if, a linha de base/a tensão, e a gerência da exceção. Dentro de cada um destas áreas, descubra que recursos e ferramentas estão disponíveis e o que é precisado. Muito a falha da organização com disposições da ferramenta porque consideram a tecnologia e as características das ferramentas mas não consideram os povos e a experiência necessários controlar as ferramentas. Inclua os povos e a experiência em seu plano, assim como as melhorias de processo exigidos. Estes povos podem incluir administradores de sistema para controlar as estações de gerenciamento de rede, os administradores da base de dados ajudar com a administração da base de dados, os administradores treinados para usar e monitorar as ferramentas, e os administradores de rede de mais alto nível para determinar políticas, pontos iniciais, e exigências da coleção de informação.

## Defina um processo

Você igualmente precisa um processo de assegurar-se de que a ferramenta esteja usada com sucesso e consistentemente. Você pode exigir melhorias de processo para definir que administradores de rede devem fazer quando as violações de limiar ocorrem ou que processo a seguir para a linha de base, tendendo, e promovendo a rede. Uma vez que você determina as exigências e os recursos para o planejamento da capacidade bem sucedido, você pode considerar a metodologia. Muitas organizações escolhem externalizar este tipo de funcionalidade a uma organização de serviços de rede tal como o INS ou construir a em-casa da experiência porque consideram o serviço uma competência de núcleo.

## Defina áreas da capacidade

O plano para o planejamento da capacidade deve igualmente incluir uma definição de áreas da capacidade. Estas são áreas da rede que pode compartilhar de uma estratégia de planejamento de capacidade comum: por exemplo, a LAN corporativa, os escritórios de campo MACILENTOS, as sites WAN crítico, e o acesso de discagem de entrada. Definir áreas diferentes é útil por vários motivos:

- As áreas diferentes podem ter pontos iniciais diferentes. Por exemplo, a largura de banda de LAN é muito mais barata do que a largura de banda de WAN assim que os limiares de utilização devem ser mais baixos.
- As áreas diferentes podem exigir a monitoração de variáveis MIB diferentes. Por exemplo, os contadores FECN e BECN no Frame Relay são críticos em problemas de potencialidade compreensivos do Frame Relay.
- Pode ser mais difícil ou demorado para promover algumas áreas da rede. Por exemplo, os circuitos internacionais podem ter um tempo de ligação muito mais longo e precisar uma correspondência de mais alto nível do planejamento.

## Defina os variáveis de capacidade

A área importante seguinte está definindo as variáveis para monitorar e os valores de limiar que exigem a ação. Definir os variáveis de capacidade depende significativamente dos dispositivos e dos media usados dentro da rede. Em parâmetros gerais tais como o CPU, a memória e a utilização do enlace são valiosas. Contudo, outras áreas podem ser importantes para Tecnologias

ou exigências específicas. Estes podem incluir profundidades de fila, desempenho, notificação de congestionamento do Frame Relay, utilização de backplane, utilização do buffer, estatísticas de Netflow, volume de transmissão, e dados RMON. Mantenha na mente seus planos de longo prazo, mas no começo com somente algumas áreas principal ajudar a assegurar o sucesso.

## [Interprete os dados](#)

Compreender os dados recolhidos é igualmente chave a proporcionar um serviço de alta qualidade. Por exemplo, muitas organizações não compreendem inteiramente níveis do pico e da utilização média. O seguinte diagrama mostra um pico do parâmetro da capacidade baseado 5 em um intervalo de coleção do minuto SNMP, (mostrado no verde).

Mesmo que o valor relatado seja menos do que os picos do ponto inicial (mostrado no vermelho) ainda podem ocorrer dentro do intervalo de coleção que estão acima do valor de limiar (mostrado no azul). Isto é significativo porque durante o intervalo de coleção, a organização pode experimentar os valores de pico que afetam o desempenho ou a capacidade da rede. Seja cuidadoso selecionar um intervalo de coleção significativo que sejam útil e que não cause despesas gerais excessivas.

Um outro exemplo é utilização média. Se os empregados estão somente no escritório de oito a cinco, mas a utilização média é 7X24, a informação pode ser enganadora.

## [Informações Relacionadas](#)

- [Suporte Técnico - Cisco Systems](#)

Era este documento útil? [Sim nenhum](#)

Obrigado para seu feedback.

[Abra um caso de suporte](#) (exige um [contrato de serviço Cisco](#).)

## **Cisco relacionado apoia discussões da comunidade**

[Cisco apoia a comunidade](#) é um fórum para que você faça e responda a perguntas, sugestões da parte, e colabora com seus pares.

Refira [convenções dos dicas técnicas da Cisco](#) para obter informações sobre das convenções usadas neste documento.

Atualizado em: outubro 04, 2005

ID do Documento: 20769