

Administración de capacidad y rendimiento: Informe oficial de Mejores Prácticas

Contenido

[Introducción](#)

[Descripción general de administración de capacidad y rendimiento](#)

[CPU](#)

[Placa base o E/S](#)

[Memoria](#)

[Tamaños de interfaz y de tubería](#)

[Colas, latencia y fluctuación](#)

[Velocidad y distancia](#)

[Características de aplicación](#)

[Prácticas recomendadas de administración de capacidad y rendimiento](#)

[Administración de nivel de servicio](#)

[Análisis de hipótesis de aplicaciones y redes](#)

[Línea de base y tendencias](#)

[Gestión de excepciones](#)

[Gestión de QoS](#)

[Recopilación e informes de información de capacidad](#)

[Determine sus necesidades](#)

[Definir un proceso](#)

[Definir áreas de capacidad](#)

[Definir las variables de capacidad](#)

[Interpretar los datos](#)

[Información Relacionada](#)

Introducción

Una alta disponibilidad de red es un requisito de misión crítica dentro de las redes del proveedor de servicios y las grandes empresas. Los administradores de red se enfrentan a cada vez mayores desafíos para proporcionar la más alta disponibilidad, incluido el tiempo de inactividad no programado, la falta de experiencia, herramientas insuficientes, tecnologías complejas, la consolidación comercial y los mercados de la competencia. La administración de capacidad y rendimiento ayudan a los administradores de red a alcanzar nuevos objetivos comerciales mundiales y una disponibilidad y funcionamiento de la red constantes.

Este documento examina los siguientes temas:

- Problemas generales de capacidad y rendimiento, incluidos los riesgos y los posibles problemas de capacidad en las redes.
- Mejores prácticas de gestión de capacidad y rendimiento, incluidos análisis condicionales, líneas de base, tendencias, gestión de excepciones y gestión de QoS.

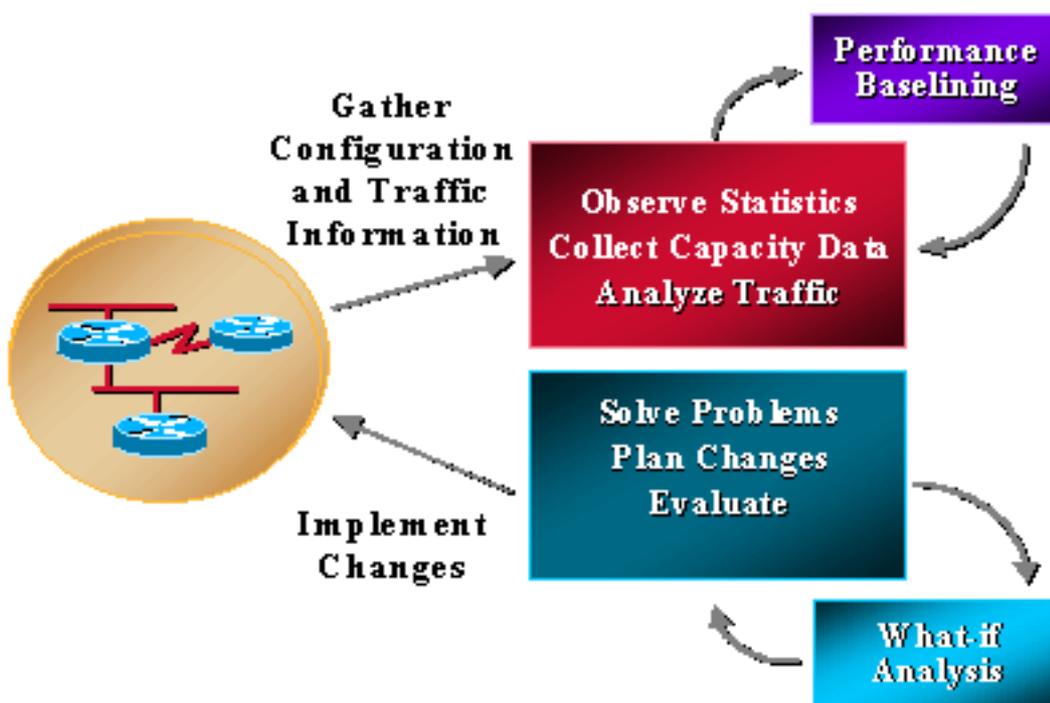
- Cómo desarrollar una estrategia de planificación de capacidad, incluidas las técnicas, herramientas, variables MIB y umbrales comunes utilizados en la planificación de capacidad.

Descripción general de administración de capacidad y rendimiento

La planificación de la capacidad es el proceso de determinación de los recursos de red necesarios para evitar un impacto en el rendimiento o la disponibilidad de las aplicaciones vitales para la empresa. La gestión del rendimiento es la práctica que consiste en gestionar el tiempo de respuesta, la coherencia y la calidad de los servicios individuales y generales.

Nota: Los problemas de rendimiento suelen estar relacionados con la capacidad. Las aplicaciones son más lentas porque el ancho de banda y los datos deben esperar en colas antes de transmitirse a través de la red. En las aplicaciones de voz, problemas como el retraso y la fluctuación afectan directamente a la calidad de la llamada de voz.

La mayoría de las organizaciones ya recopilan información relacionada con la capacidad y trabajan de forma coherente para solucionar problemas, planificar cambios e implementar nuevas funciones de capacidad y rendimiento. Sin embargo, las organizaciones no realizan de forma rutinaria análisis de tendencias y hipótesis. El análisis de hipótesis es el proceso de determinar el efecto de un cambio en la red. La tendencia es el proceso de realizar una recopilación básica de la capacidad de la red y los problemas de rendimiento, así como de revisar las referencias de las tendencias de la red para comprender los requisitos de actualización futuros. La gestión de capacidad y rendimiento también debe incluir la gestión de excepciones, en la que los problemas se identifican y resuelven antes de que los usuarios llamen, y la gestión de QoS, en la que los administradores de red planifican, gestionan e identifican problemas de rendimiento de servicios individuales. En el gráfico siguiente se ilustran los procesos de administración de capacidad y rendimiento.



La gestión de la capacidad y el rendimiento también tiene sus limitaciones, relacionadas normalmente con la CPU y la memoria. Las siguientes son posibles esferas de preocupación:

- CPU
- Placa base o E/S
- Memoria y búferes
- Tamaños de interfaces y tuberías
- Colas, latencia y fluctuación
- Velocidad y distancia
- Características de aplicación

Algunas referencias a la planificación de la capacidad y la gestión del rendimiento también mencionan algo llamado "plano de datos" y "plano de control". El plano de datos consiste simplemente en problemas de capacidad y rendimiento relacionados con los datos que atraviesan la red, mientras que el plano de control implica recursos necesarios para mantener la funcionalidad adecuada del plano de datos. La funcionalidad del plano de control incluye sobrecarga de servicio como routing, árbol de extensión, keepalives de interfaz y administración SNMP del dispositivo. Estos requisitos del plano de control utilizan CPU, memoria, almacenamiento en búfer, colas y ancho de banda, al igual que el tráfico que atraviesa la red. Muchos de los requisitos del plano de control también son esenciales para la funcionalidad general del sistema. Si no disponen de los recursos que necesitan, la red falla.

CPU

El plano de control y el plano de datos suelen utilizar la CPU en cualquier dispositivo de red. En cuanto a la gestión de la capacidad y el rendimiento, debe asegurarse de que el dispositivo y la red disponen de suficiente CPU para funcionar en todo momento. Una CPU insuficiente a menudo puede colapsar una red porque los recursos inadecuados en un dispositivo pueden afectar a toda la red. Una CPU insuficiente también puede aumentar la latencia, ya que los datos deben esperar para ser procesados cuando no hay switching de hardware sin la CPU principal.

Placa base o E/S

La placa de interconexiones o E/S se refiere a la cantidad total de tráfico que un dispositivo puede gestionar, descrita generalmente en términos de tamaño de BUS o capacidad de placa de interconexiones. La placa de interconexiones insuficiente normalmente da lugar a paquetes perdidos, lo que puede dar lugar a retransmisiones y tráfico adicional.

Memoria

La memoria es otro recurso que tiene requisitos de plano de datos y plano de control. Se requiere memoria para información como tablas de enrutamiento, tablas ARP y otras estructuras de datos. Cuando los dispositivos se quedan sin memoria, algunas operaciones del dispositivo pueden fallar. La operación podría afectar a los procesos del plano de control o a los procesos del plano de datos, según la situación. Si los procesos del plano de control fallan, toda la red puede degradarse. Por ejemplo, esto puede suceder cuando se requiere memoria adicional para la convergencia de ruteo.

Tamaños de interfaz y de tubería

Los tamaños de la interfaz y de la tubería se refieren a la cantidad de datos que se pueden enviar simultáneamente en cualquier conexión. A menudo se denomina incorrectamente velocidad de una conexión, pero los datos realmente no viajan a diferentes velocidades de un dispositivo a otro. La velocidad de silicio y la capacidad de hardware ayudan a determinar el ancho de banda

disponible en función de los medios. Además, los mecanismos de software pueden "regular" los datos para ajustarse a las asignaciones de ancho de banda específicas de un servicio. Normalmente, esto se ve en las redes de proveedores de servicios para Frame Relay o ATM que tienen capacidades de velocidad inherentes de 1.54kpbs a 155mbps y superiores. Cuando hay limitaciones de ancho de banda, los datos se colocan en una cola de transmisión. Una cola de transmisión puede tener diferentes mecanismos de software para priorizar los datos dentro de la cola; sin embargo, cuando hay datos en la cola, debe esperar a los datos existentes antes de poder reenviar los datos fuera de la interfaz.

Colas, latencia y fluctuación

Las colas, la latencia y la fluctuación también afectan al rendimiento. Puede ajustar la cola de transmisión para que afecte al rendimiento de diferentes maneras. Por ejemplo, si la cola es grande, los datos esperan más tiempo. Cuando las colas son pequeñas, los datos se descartan. Esto se denomina taildrop y es aceptable para las aplicaciones TCP ya que los datos se retransmitirán. Sin embargo, la voz y el vídeo no funcionan bien con el descarte de cola o incluso con una latencia de cola significativa que requiere una atención especial al ancho de banda o a los tamaños de las canalizaciones. El retraso de la cola también puede ocurrir con las colas de entrada si el dispositivo no tiene recursos suficientes para reenviar inmediatamente el paquete. Esto puede deberse a la CPU, la memoria o los búferes.

Latencia describe el tiempo de procesamiento normal desde el momento en que se recibe hasta el momento en que se reenvía el paquete. Los routers y switches de datos modernos normales tienen una latencia extremadamente baja (< 1 ms) en condiciones normales sin limitaciones de recursos. Los dispositivos modernos con procesadores de señal digital para convertir y comprimir paquetes de voz analógica pueden tardar más, incluso hasta 20 ms.

La fluctuación describe la brecha entre paquetes para las aplicaciones de transmisión, incluidas las de voz y vídeo. Si los paquetes llegan en diferentes momentos con diferentes intervalos entre paquetes, la fluctuación es alta y la calidad de voz se degrada. La fluctuación es principalmente un factor del retraso de la cola.

Velocidad y distancia

La velocidad y la distancia también influyen en el rendimiento de la red. Las redes de datos tienen una velocidad de reenvío de datos uniforme basada en la velocidad de la luz. Esto es aproximadamente 100 millas por milisegundo. Si una organización ejecuta una aplicación cliente-servidor internacionalmente, puede esperar un retraso de reenvío de paquetes correspondiente. La velocidad y la distancia pueden ser factores fundamentales en el rendimiento de las aplicaciones cuando éstas no están optimizadas para el rendimiento de la red.

Características de aplicación

Las características de las aplicaciones son la última área que afecta a la capacidad y al rendimiento. Problemas como el tamaño reducido de las ventanas, las señales de mantenimiento de las aplicaciones y la cantidad de datos enviados a través de la red con respecto a lo que se necesita pueden afectar al rendimiento de una aplicación en muchos entornos, especialmente las WAN.

Prácticas recomendadas de administración de capacidad y

rendimiento

En esta sección se analizan en detalle las cinco prácticas recomendadas principales de gestión de la capacidad y el rendimiento:

- [Administración de nivel de servicio](#)
- [Análisis de hipótesis de aplicaciones y redes](#)
- [Definición de líneas de base y tendencias](#)
- [Gestión de excepciones](#)
- [Gestión de QoS](#)

Administración de nivel de servicio

La gestión del nivel de servicio define y regula otros procesos de gestión de capacidad y rendimiento necesarios. Los administradores de red son conscientes de que necesitan planificar su capacidad, pero se enfrentan a limitaciones presupuestarias y de personal que impiden una solución completa. La gestión de nivel de servicio es una metodología probada que ayuda con los problemas de recursos mediante la definición de un entregable y la creación de responsabilidad bidireccional para un servicio vinculado a ese entregable. Puede lograr esto de dos maneras:

- Cree un acuerdo de nivel de servicio entre los usuarios y la organización de red para un servicio que incluya gestión de capacidad y rendimiento. El servicio incluiría informes y recomendaciones para mantener la calidad del servicio. Sin embargo, los usuarios deben estar preparados para financiar el servicio y cualquier actualización necesaria.
- La organización de la red define su capacidad y el servicio de gestión del rendimiento y, a continuación, intenta obtener financiación para ese servicio y las actualizaciones caso por caso.

En cualquier caso, la organización de la red debe empezar por definir un servicio de gestión del rendimiento y planificación de la capacidad que incluya los aspectos del servicio que pueden proporcionar actualmente y los que están previstos en el futuro. Un servicio completo incluiría un análisis hipotético de los cambios de la red y de las aplicaciones, la definición de líneas de base y tendencias para las variables de rendimiento definidas, la gestión de excepciones para las variables de capacidad y rendimiento definidas y la gestión de QoS.

Análisis de hipótesis de aplicaciones y redes

Realizar un análisis de hipótesis de aplicaciones y redes para determinar el resultado de un cambio planificado. Sin un análisis hipotético, las organizaciones asumen riesgos significativos para cambiar el éxito y la disponibilidad general de la red. En muchos casos, los cambios en la red han provocado un colapso congestivo que ha provocado muchas horas de inactividad en la producción. Además, una cantidad asombrosa de introducciones de aplicaciones fallan y causan impacto en otros usuarios y aplicaciones. Estos fallos continúan produciéndose en muchas organizaciones de red, pero se pueden evitar completamente con unas cuantas herramientas y algunos pasos de planificación adicionales.

Normalmente, se necesitan algunos procesos nuevos para realizar un análisis condicional de calidad. El primer paso es identificar los niveles de riesgo para todos los cambios y exigir un análisis de hipótesis más profundo para los cambios de mayor riesgo. El nivel de riesgo puede ser un campo obligatorio para todos los envíos de cambios. Los cambios de mayor nivel de riesgo requerirían un análisis hipotético definido del cambio. Un análisis de hipótesis de red determina el

efecto de los cambios de red en la utilización de la red y los problemas de recursos del plano de control de la red. Un análisis hipotético de la aplicación determinaría el éxito de la aplicación del proyecto, los requisitos de ancho de banda y cualquier problema de recursos de red. Las tablas siguientes son ejemplos de asignación de nivel de riesgo y requisitos de prueba correspondientes:

Nivel de riesgos	Definición	Cambiar recomendaciones de planificación
1	<ul style="list-style-type: none"> • Gran impacto potencial en un gran número de usuarios (más de 500) o servicios empresariales críticos debido a la introducción de nuevos productos, software, topologías o funciones. • El cambio implica un tiempo de inactividad de red esperado. 	<ul style="list-style-type: none"> • Validar laboratorio de nueva solución. La validación de laboratorio incluye pruebas y validaciones documentadas de soluciones y análisis hipotéticos que muestran el impacto en la infraestructura existente. Recomendamos soluciones piloto. Las nuevas soluciones requieren la cumplimentación de un documento de soporte de operaciones. • Realice la revisión del diseño de Cisco NSA. • Crear plan de retirada. • Crear plan de implementación.

		<ul style="list-style-type: none"> • Crear proceso de cambio.
2	<ul style="list-style-type: none"> • Gran impacto potencial en un gran número de usuarios (más de 500) o servicios empresariales críticos debido a un gran aumento del tráfico o los usuarios, cambios en la red troncal o cambios en el routing. • El cambio puede requerir algún tiempo de inactividad. 	<ul style="list-style-type: none"> • Realizar análisis hipotéticos para determinar el impacto en el entorno existente (se debe realizar en un entorno de laboratorio). • Pruebe y revise los cambios de routing para comprobar su funcionalidad. • Crear plan de retirada. • Realice una revisión del diseño para obtener información sobre los principales cambios de routing o de red troncal. • Crear plan de implementación. • Crear proceso de cambio.
3	<ul style="list-style-type: none"> • Impacto potencial medio en un número menor de usuarios o servicios empresariales debido a cualquier cambio no estándar. • Incluye un nuevo producto, software, topología, adición de funciones para nuevos usuarios, aumento del tráfico o topología no 	<ul style="list-style-type: none"> • Realizar análisis de ingeniería de la nueva solución (puede requerir validación de laboratorio). • Crear plan de implementación.

	<p>estándar.</p> <ul style="list-style-type: none"> • El cambio puede requerir algún tiempo de inactividad. 	<p>n.</p> <ul style="list-style-type: none"> • Crear proceso de cambio.
4	<ul style="list-style-type: none"> • Menor impacto potencial en el servicio o el usuario. • Incluye la adición de nuevos módulos de red de plantillas estándar, como switches/concentradores de servidor o de creación en routers. • Incluye la instalación de nuevos sitios WAN o servicios de acceso adicionales de eficacia probada. • Todos los cambios del nivel de riesgo 3 se han probado técnicamente en el entorno de producción. • El cambio puede requerir algún tiempo de inactividad. 	<ul style="list-style-type: none"> • Crear plan de implementación. • Crear proceso de cambio.
5	<ul style="list-style-type: none"> • Ningún impacto en el usuario o el servicio. • Incluye la adición de usuarios individuales a la red y cambios de configuración estándar, como contraseña, banner, SNMP u otros parámetros de configuración estándar. • Sin tiempo de inactividad. 	<ul style="list-style-type: none"> • El proceso de cambio es opcional.

Una vez que haya definido dónde necesita el análisis de hipótesis, puede definir el servicio.

Puede realizar un análisis de hipótesis de red con herramientas de modelado o con un laboratorio que imite el entorno de producción. Las herramientas de modelado están limitadas por el grado de comprensión que tenga la aplicación de los problemas de recursos de dispositivos y, dado que la mayoría de los cambios de red son nuevos dispositivos, es posible que la aplicación no comprenda el efecto del cambio. El mejor método es crear alguna representación de la red de producción en un laboratorio y probar el software, la función, el hardware o la configuración deseados bajo carga mediante generadores de tráfico. La fuga de rutas (u otra información de control) de la red de producción al laboratorio también mejora el entorno del laboratorio. Pruebe los requisitos de recursos adicionales con diferentes tipos de tráfico, incluido el tráfico SNMP, de difusión, multidifusión, cifrado o comprimido. Con todas estas metodologías diferentes, analice los requisitos de recursos de dispositivos durante situaciones de tensión potenciales, como la

convergencia de rutas, la inestabilidad de enlaces y los reinicios de dispositivos. Los problemas de utilización de recursos incluyen áreas de recursos de capacidad normal como CPU, memoria, utilización de backplane, búferes y colas.

Las nuevas aplicaciones también deben realizar un análisis hipotético para determinar el éxito de la aplicación y los requisitos de ancho de banda. Normalmente, este análisis se realiza en un entorno de laboratorio utilizando un analizador de protocolos y un simulador de retraso de WAN para comprender el efecto de la distancia. Solo necesita un PC, un concentrador, un dispositivo de retraso WAN y un router de laboratorio conectados a la red de producción. Puede simular el ancho de banda en el laboratorio acelerando el tráfico mediante el modelado de tráfico genérico o la limitación de velocidad en el router de prueba. El administrador de red puede trabajar junto con el grupo de aplicaciones para comprender los requisitos de ancho de banda, los problemas de ventanas y los posibles problemas de rendimiento de la aplicación en entornos LAN y WAN.

Realice un análisis de hipótesis antes de implementar cualquier aplicación empresarial. Si no lo hace, el grupo de aplicaciones culpa a la red por su bajo rendimiento. Si de alguna manera puede requerir un análisis hipotético de la aplicación para nuevas implementaciones a través del proceso de gestión de cambios, puede ayudar a evitar implementaciones sin éxito y comprender mejor los aumentos repentinos en el consumo de ancho de banda para los requisitos tanto de cliente-servidor como de lotes.

Línea de base y tendencias

Las líneas de base y las tendencias permiten a los administradores de red planificar y completar las actualizaciones de red antes de que un problema de capacidad provoque tiempos de inactividad de la red o problemas de rendimiento. Comparar la utilización de recursos durante períodos de tiempo sucesivos o destilar información a lo largo del tiempo en una base de datos y permitir a los planificadores consultar los parámetros de utilización de recursos para la última hora, día, semana, mes y año. En cualquier caso, alguien debe revisar la información de manera semanal, quincenal o mensual. El problema con la línea de base y las tendencias es que se requiere una cantidad abrumadora de información para revisar en redes grandes.

Puede resolver este problema de varias maneras:

- Cree una gran capacidad y realice el switching en el entorno LAN para que la capacidad no sea un problema.
- Divida la información de tendencias en grupos y concéntrese en áreas críticas o de alta disponibilidad de la red, como sitios WAN críticos o LAN de Data Center.
- Los mecanismos de presentación de informes pueden resaltar las áreas que se encuentran por encima de un determinado umbral y requieren especial atención. Si implementa primero las áreas de disponibilidad críticas, puede reducir significativamente la cantidad de información necesaria para la revisión.

Con todos los métodos anteriores, aún debe revisar la información periódicamente. La evaluación y las tendencias son un esfuerzo proactivo y, si la organización solo dispone de recursos para ofrecer soporte reactivo, los usuarios no leerán los informes.

Muchas soluciones de administración de redes proporcionan información y gráficos sobre las variables de recursos de capacidad. Desafortunadamente, la mayoría de las personas solo utilizan estas herramientas para dar soporte reactivo a un problema existente; esto contradice el objetivo de establecer una referencia y establecer tendencias. Dos herramientas eficaces para proporcionar información sobre la tendencia de capacidad de las redes de Cisco son el producto

Concord Network Health y los productos INS EnterprisePRO. En muchos casos, las organizaciones de red ejecutan lenguajes de scripting sencillos para recopilar información de capacidad. A continuación, se incluyen algunos informes de ejemplo que se recopilaron a través del script para la utilización de enlaces, la utilización de CPU y el rendimiento de ping. Otras variables de recursos que pueden ser importantes para la tendencia incluyen la memoria, la profundidad de la cola, el volumen de difusión, el búfer, la notificación de congestión de Frame Relay y la utilización de la placa de interconexiones. Consulte esta tabla para obtener información sobre el uso de links y el uso de CPU:

Utilización de enlaces

Recurso	Dirección	Segmento	Utilización media (%)	Utilización máxima (%)
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

Utilización de la CPU

Recurso	Dirección de sondeo	Utilización media (%)	Utilización máxima (%)
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

Utilización de enlaces

Recurso	Dirección	AvResT (mS) 09-09-98	AvResT (mS) 09-09-98	AvResT (mS) 09-09-98	AvResT (mS) 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
ABR01	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
ASR01	10.196.178.1	667.7	976.4	655.5	948.6
ASYR01S					503.4

AZWR 01	10.177.32 .1	460.1		444.7	
BEJR 01	10.195.18 .1	1023. 7	1064. 6	1184	1021.9

Gestión de excepciones

La gestión de excepciones es una metodología valiosa para identificar y resolver problemas de capacidad y rendimiento. La idea es recibir notificaciones de infracciones del umbral de capacidad y rendimiento para investigar y solucionar inmediatamente el problema. Por ejemplo, un administrador de red podría recibir una alarma por uso elevado de la CPU en un router. El administrador de red puede iniciar sesión en el router para determinar por qué la CPU es tan alta. A continuación, puede realizar alguna configuración correctiva que reduzca la CPU o crear una lista de acceso que evite el tráfico que causa el problema, especialmente si el tráfico no parece ser crítico para la empresa.

Puede configurar la administración de excepciones para problemas más críticos simplemente usando los comandos de configuración de RMON en un router o usando herramientas más avanzadas como el administrador de nivel de servicio de Netsys junto con los datos de SNMP, RMON o Netflow. La mayoría de las herramientas de administración de redes tienen la capacidad de establecer umbrales y alarmas en las infracciones. El aspecto importante del proceso de administración de excepciones es proporcionar una notificación del problema casi en tiempo real. De lo contrario, el problema puede desaparecer antes de que alguien se percate de que se ha recibido la notificación. Esto se puede hacer dentro de un NOC si la organización tiene una supervisión consistente. De lo contrario, se recomienda la notificación del localizador.

El siguiente ejemplo de configuración proporciona una notificación de umbral ascendente y descendente para la CPU del router a un archivo de registro que se puede revisar de manera uniforme. Puede configurar comandos RMON similares para infracciones de umbral de utilización de links críticos u otros umbrales 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
```

Gestión de QoS

La gestión de la calidad del servicio implica la creación y supervisión de clases de tráfico específicas dentro de la red. Un tráfico proporciona un rendimiento más uniforme para grupos de aplicaciones específicos (definidos dentro de las clases de tráfico). Los parámetros de modelado de tráfico proporcionan una flexibilidad significativa en la priorización y el modelado de tráfico para clases específicas de tráfico. Entre estas funciones se incluyen capacidades como la tasa de acceso comprometida (CAR), la detección temprana aleatoria ponderada (WRED) y las colas ponderadas equitativas basadas en clases. Normalmente, las clases de tráfico se crean en función de los SLA de rendimiento para aplicaciones empresariales más críticas y requisitos de aplicaciones específicos, como la voz. El tráfico no crítico o no empresarial también se controlaría de forma que no pueda afectar a las aplicaciones y los servicios de mayor prioridad.

La creación de clases de tráfico requiere una comprensión básica de la utilización de la red, los

requisitos específicos de las aplicaciones y las prioridades de las aplicaciones empresariales. Los requisitos de las aplicaciones incluyen el conocimiento de los tamaños de los paquetes, los problemas de tiempo de espera, los requisitos de fluctuación, los requisitos de ráfaga, los requisitos de lotes y los problemas de rendimiento general. Con este conocimiento, los administradores de red pueden crear planes y configuraciones de modelado de tráfico que proporcionan un rendimiento de aplicaciones más uniforme en una variedad de topologías LAN/WAN.

Por ejemplo, una organización tiene una conexión ATM de 10 megabits entre dos sitios principales. El link a veces se congestiona debido a las grandes transferencias de archivos, lo que causa una degradación del rendimiento para el procesamiento de transacciones en línea y una calidad de voz deficiente o inutilizable.

La organización configuró cuatro clases de tráfico diferentes. Se asignó la máxima prioridad a la voz y se le permitió mantener dicha prioridad aunque se propagara por encima de la tasa de volumen de tráfico estimada. A la clase de aplicación crítica se le dio la siguiente prioridad más alta, pero no se le permitió reventar sobre el tamaño total del link menos los requisitos de ancho de banda de voz estimados. Cuando se rompe, se cae. El tráfico de transferencia de archivos simplemente recibió una prioridad más baja y el resto del tráfico encajó en algún lugar del medio.

La organización ahora tiene que realizar la gestión de QoS en este enlace para determinar la cantidad de tráfico que está tomando cada clase y medir el rendimiento dentro de cada clase. Si la organización no logra hacer esto, puede ocurrir la inanición para algunas clases o puede que los SLA de rendimiento no se cumplan dentro de una clase particular.

La gestión de las configuraciones de QoS sigue siendo una tarea difícil debido a la falta de herramientas. Un método consiste en utilizar el Administrador de rendimiento de Internet (IPM) de Cisco para enviar tráfico diferente a través del link que cae en cada una de las clases de tráfico. A continuación, podría supervisar el rendimiento de cada clase e IPM proporciona análisis de tendencias, en tiempo real y salto a salto para identificar las áreas problemáticas. Otros pueden seguir dependiendo de un método más manual como investigar la colocación en cola y los paquetes descartados dentro de cada clase de tráfico según las estadísticas de la interfaz. En algunas organizaciones, estos datos pueden recopilarse a través de SNMP o analizarse en una base de datos para obtener líneas de base y tendencias. También existen algunas herramientas en el mercado que envían tipos de tráfico específicos a través de la red para determinar el rendimiento de un servicio o aplicación en particular.

Recopilación e informes de información de capacidad

La reunión y presentación de información sobre la capacidad debe estar vinculada a las tres esferas recomendadas de la gestión de la capacidad:

- Análisis de hipótesis, centrado en el cambio de la red y en cómo afecta el cambio al entorno
- Definición de líneas de base y tendencias
- Gestión de excepciones

Dentro de cada una de estas áreas, desarrollar un plan de recolección de información. En el caso de los análisis condicionales de la red o la aplicación, se necesitan herramientas para imitar el entorno de red y comprender el efecto del cambio en relación con los posibles problemas de recursos dentro del plano de control del dispositivo o el plano de datos. En el caso de las líneas de base y las tendencias, necesitará instantáneas de los dispositivos y enlaces que muestren el uso actual de los recursos. A continuación, revise los datos a lo largo del tiempo para comprender

los posibles requisitos de actualización. Esto permite a los administradores de red planificar correctamente las actualizaciones antes de que surjan problemas de capacidad o rendimiento. Cuando surgen problemas, necesita la administración de excepciones para avisar a los administradores de red para que puedan ajustar la red o solucionar el problema.

Este proceso se puede dividir en los siguientes pasos:

1. Determine sus necesidades.
2. Defina un proceso.
3. Definir áreas de capacidad.
4. Defina las variables de capacidad.
5. Interpretar los datos.

Determine sus necesidades

El desarrollo de un plan de gestión de la capacidad y del rendimiento requiere conocer la información que necesita y el propósito de dicha información. Dividir el plan en tres áreas necesarias: una para cada análisis hipotético, establecimiento de referencias/tendencias y gestión de excepciones. En cada una de estas áreas, descubra qué recursos y herramientas están disponibles y qué se necesita. Muchas organizaciones fracasan con las implementaciones de herramientas porque tienen en cuenta la tecnología y las características de las herramientas, pero no tienen en cuenta a las personas y los conocimientos necesarios para administrar las herramientas. Incluya a las personas y los conocimientos necesarios en su plan, así como mejoras en los procesos. Estas personas pueden incluir administradores de sistemas para gestionar las estaciones de gestión de red, administradores de bases de datos para ayudar con la administración de bases de datos, administradores formados para utilizar y supervisar las herramientas y administradores de redes de nivel superior para determinar políticas, umbrales y requisitos de recopilación de información.

Definir un proceso

También es necesario un proceso para garantizar que la herramienta se utiliza de forma correcta y coherente. Es posible que necesite mejoras en los procesos para definir qué deben hacer los administradores de red cuando se produzca una infracción de umbrales o qué proceso seguir para establecer una línea de base, crear tendencias y actualizar la red. Una vez que haya determinado los requisitos y los recursos para una correcta planificación de la capacidad, puede considerar la metodología. Muchas organizaciones eligen externalizar este tipo de funcionalidad a una organización de servicios de red como INS o crear la experiencia interna porque consideran el servicio una competencia principal.

Definir áreas de capacidad

El plan de planificación de la capacidad también debería incluir una definición de las esferas de capacidad. Estas son áreas de la red que pueden compartir una estrategia de planificación de capacidad común: por ejemplo, la LAN corporativa, las oficinas exteriores de WAN, los sitios WAN críticos y el acceso telefónico. La definición de diferentes áreas es útil por varias razones:

- Diferentes áreas pueden tener diferentes umbrales. Por ejemplo, el ancho de banda LAN es mucho más barato que el ancho de banda WAN, por lo que los umbrales de utilización deben ser inferiores.

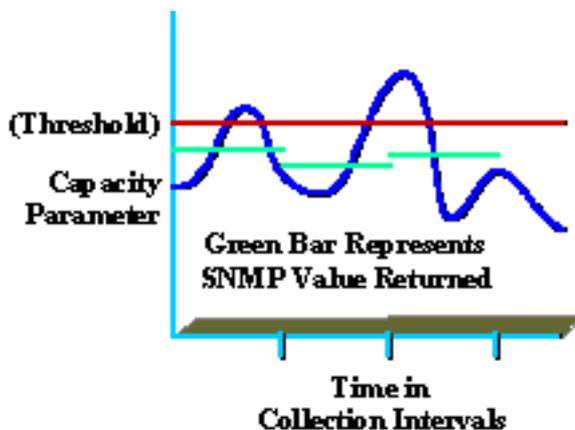
- Diferentes áreas pueden requerir el monitoreo de diferentes variables MIB. Por ejemplo, los contadores FECN y BECN en Frame Relay son críticos para entender los problemas de capacidad de Frame Relay.
- La actualización de algunas áreas de la red puede resultar más difícil o requerir más tiempo. Por ejemplo, los circuitos internacionales pueden tener plazos de entrega mucho más largos y necesitan un nivel superior de planificación correspondiente.

Definir las variables de capacidad

La siguiente área importante es la definición de las variables a monitorear y los valores de umbral que requieren acción. La definición de las variables de capacidad depende en gran medida de los dispositivos y medios utilizados en la red. En general, los parámetros como la utilización de CPU, memoria y enlaces son valiosos. Sin embargo, otras áreas pueden ser importantes para tecnologías o requisitos específicos. Estos pueden incluir profundidades de cola, rendimiento, notificación de congestión de Frame Relay, utilización de backplane, utilización de buffer, estadísticas de NetFlow, volumen de broadcast y datos RMON. Tenga en cuenta sus planes a largo plazo, pero comience con solo unas pocas áreas clave para garantizar el éxito.

Interpretar los datos

La comprensión de los datos recopilados también es clave para proporcionar un servicio de alta calidad. Por ejemplo, muchas organizaciones no entienden completamente los niveles de utilización media y máxima. El siguiente diagrama muestra un pico de parámetro de capacidad basado en un intervalo de recopilación SNMP de 5 minutos (mostrado en verde).



Aunque el valor notificado sea inferior al umbral (mostrado en rojo), los picos pueden producirse dentro del intervalo de recopilación que están por encima del valor del umbral (mostrado en azul). Esto es significativo porque durante el intervalo de recopilación, la organización puede estar experimentando valores máximos que afectan al rendimiento o la capacidad de la red. Tenga cuidado de seleccionar un intervalo de recopilación significativo que sea útil y que no cause una sobrecarga excesiva.

Otro ejemplo es la utilización media. Si los empleados solo están en la oficina de ocho a cinco, pero el uso medio es 7X24, la información puede ser engañosa.

Información Relacionada

- [Soporte Técnico - Cisco Systems](#)

Acerca de esta traducción

Cisco ha traducido este documento combinando la traducción automática y los recursos humanos a fin de ofrecer a nuestros usuarios en todo el mundo contenido en su propio idioma.

Tenga en cuenta que incluso la mejor traducción automática podría no ser tan precisa como la proporcionada por un traductor profesional.

Cisco Systems, Inc. no asume ninguna responsabilidad por la precisión de estas traducciones y recomienda remitirse siempre al documento original escrito en inglés (insertar vínculo URL).