

Gestione della capacità e delle prestazioni: white paper sulle best practice

Sommario

[Introduzione](#)

[Panoramica sulla gestione della capacità e delle prestazioni](#)

[CPU](#)

[Backplane o I/O](#)

[Memoria](#)

[Interfaccia e dimensioni dei tubi](#)

[Accodamento, latenza e jitter](#)

[Velocità e distanza](#)

[Caratteristiche dell'applicazione](#)

[Procedure ottimali per la gestione di capacità e prestazioni](#)

[Gestione del livello del servizio](#)

[Analisi di simulazione della rete e delle applicazioni](#)

[Baseline e trend](#)

[Gestione delle eccezioni](#)

[Gestione QoS](#)

[Raccolta e reporting delle informazioni sulla capacità](#)

[Determinate le vostre esigenze](#)

[Definisci processo](#)

[Definizione aree capacità](#)

[Definizione delle variabili di capacità](#)

[Interpretare i dati](#)

[Informazioni correlate](#)

Introduzione

L'elevata disponibilità della rete è un requisito mission-critical all'interno delle reti di grandi aziende e provider di servizi. I responsabili di rete devono affrontare sfide sempre più complesse per garantire una maggiore disponibilità, inclusi tempi di inattività non pianificati, mancanza di esperienza, strumenti insufficienti, tecnologie complesse, consolidamento aziendale e mercati concorrenti. La gestione della capacità e delle prestazioni consente ai responsabili della rete di raggiungere nuovi obiettivi aziendali a livello mondiale e di ottenere disponibilità e prestazioni di rete coerenti.

In questo documento vengono esaminati gli argomenti seguenti:

- Problemi generali di capacità e prestazioni, compresi i rischi e i potenziali problemi di capacità all'interno delle reti.
- Procedure ottimali per la gestione della capacità e delle prestazioni, tra cui analisi di simulazione, baseline, trend, gestione delle eccezioni e gestione QoS.

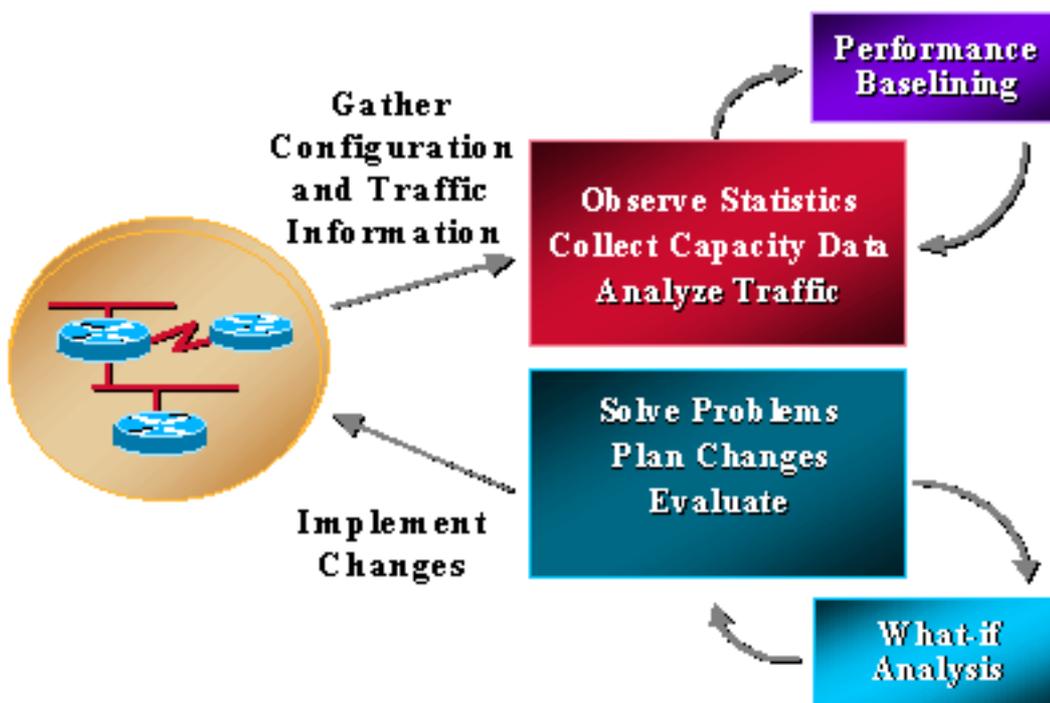
- Come sviluppare una strategia di pianificazione della capacità, che includa tecniche, strumenti, variabili MIB e soglie comuni utilizzati nella pianificazione della capacità.

Panoramica sulla gestione della capacità e delle prestazioni

La pianificazione della capacità è il processo di determinazione delle risorse di rete necessarie per impedire un impatto sulle prestazioni o sulla disponibilità delle applicazioni business-critical. La gestione delle prestazioni è la pratica di gestire i tempi di risposta, la coerenza e la qualità dei servizi di rete per i singoli servizi e per i servizi complessivi.

Nota: i problemi di prestazioni sono in genere correlati alla capacità. Le applicazioni sono più lente perché la larghezza di banda e i dati devono attendere in coda prima di essere trasmessi attraverso la rete. Nelle applicazioni vocali, problemi come il ritardo e l'instabilità influiscono direttamente sulla qualità della chiamata vocale.

La maggior parte delle organizzazioni già raccoglie alcune informazioni relative alla capacità e lavora in modo coerente per risolvere i problemi, pianificare le modifiche e implementare nuove funzionalità di capacità e prestazioni. Tuttavia, le organizzazioni non eseguono regolarmente analisi di tendenza e di simulazione. L'analisi di simulazione è il processo di determinazione dell'effetto di una modifica di rete. Trending è il processo che consente di definire le linee di base per i problemi di capacità e prestazioni della rete e di esaminare le linee di base per le tendenze della rete per comprendere i requisiti di aggiornamento futuri. La gestione della capacità e delle prestazioni deve includere anche la gestione delle eccezioni, in cui i problemi vengono identificati e risolti prima che gli utenti vengano contattati, e la gestione QoS, in cui gli amministratori di rete pianificano, gestiscono e identificano i singoli problemi di prestazioni dei servizi. L'immagine seguente mostra i processi di gestione della capacità e delle prestazioni.



Anche la gestione della capacità e delle prestazioni presenta dei limiti, in genere correlati alla CPU e alla memoria. Di seguito sono indicati i potenziali motivi di preoccupazione:

- CPU
- Backplane o I/O

- Memoria e buffer
- Interfaccia e dimensioni dei tubi
- Accodamento, latenza e jitter
- Velocità e distanza
- Caratteristiche dell'applicazione

Alcuni riferimenti alla pianificazione della capacità e alla gestione delle prestazioni citano anche qualcosa chiamato "piano dati" e "piano di controllo". Il piano dati è semplicemente un problema di capacità e prestazioni relativo ai dati che attraversano la rete, mentre il piano di controllo indica le risorse necessarie per mantenere il corretto funzionamento del piano dati. La funzionalità del control plane include il sovraccarico del servizio, ad esempio routing, spanning tree, keep-alive dell'interfaccia e gestione SNMP del dispositivo. Questi requisiti del control plane utilizzano CPU, memoria, buffering, code e larghezza di banda proprio come il traffico che attraversa la rete. Molti dei requisiti dei piani di controllo sono inoltre essenziali per la funzionalità complessiva del sistema. Se non dispongono delle risorse necessarie, la rete si interrompe.

CPU

La CPU viene in genere utilizzata sia dal control plane che dal data plane su qualsiasi dispositivo di rete. Nella gestione della capacità e delle prestazioni, è necessario verificare che il dispositivo e la rete dispongano di CPU sufficiente per funzionare sempre. Una CPU insufficiente può spesso comprimere una rete perché risorse inadeguate su un dispositivo possono influire sull'intera rete. Una CPU insufficiente può anche aumentare la latenza, in quanto i dati devono attendere di essere elaborati quando non è presente alcuna commutazione hardware senza la CPU principale.

Backplane o I/O

Il backplane o I/O si riferisce alla quantità totale di traffico che un dispositivo è in grado di gestire, generalmente descritta in termini di dimensioni del BUS o capacità del backplane. Un backplane insufficiente provoca in genere la perdita di pacchetti, che possono comportare ritrasmissioni e traffico aggiuntivo.

Memoria

La memoria è un'altra risorsa che ha i requisiti del piano dati e del piano di controllo. La memoria è necessaria per informazioni quali tabelle di routing, tabelle ARP e altre strutture di dati. Quando la memoria dei dispositivi si esaurisce, alcune operazioni sul dispositivo possono non riuscire. L'operazione può influire sui processi del piano di controllo o del piano dati, a seconda della situazione. Se i processi del control plane non riescono, l'intera rete può deteriorarsi. Ad esempio, questo può accadere quando è necessaria memoria aggiuntiva per la convergenza del routing.

Interfaccia e dimensioni dei tubi

Le dimensioni dell'interfaccia e della pipe si riferiscono alla quantità di dati che possono essere inviati contemporaneamente su una singola connessione. Spesso viene erroneamente definito come la velocità di una connessione, ma in realtà i dati non viaggiano a velocità diverse da un dispositivo all'altro. La velocità del silicio e le funzionalità hardware consentono di determinare la larghezza di banda disponibile in base al supporto. Inoltre, i meccanismi software possono "limitare" i dati per conformarsi alle allocazioni di larghezza di banda specifiche per un servizio. Ciò si verifica in genere nelle reti dei provider di servizi per frame relay o ATM che dispongono intrinsecamente di funzionalità di velocità da 1,54 kbps a 155 mbs e superiori. In caso di limitazioni

della larghezza di banda, i dati vengono accodati in una coda di trasmissione. Una coda di trasmissione può disporre di diversi meccanismi software per assegnare la priorità ai dati all'interno della coda; tuttavia, quando nella coda sono presenti dati, è necessario attendere i dati esistenti prima di inoltrarli all'esterno dell'interfaccia.

Accodamento, latenza e jitter

Anche le operazioni di accodamento, latenza e variazione influiscono sulle prestazioni. È possibile ottimizzare la coda di trasmissione per influire sulle prestazioni in diversi modi. Ad esempio, se la coda è grande, i dati attendono più a lungo. Quando le code sono piccole, i dati vengono eliminati. Questa operazione è detta "taildrop" ed è accettabile per le applicazioni TCP in quanto i dati verranno ritrasmessi. Tuttavia, la voce e il video non funzionano bene con la perdita della coda o anche una latenza significativa della coda che richiede un'attenzione particolare alla larghezza di banda o alle dimensioni dei tubi. Il ritardo della coda può verificarsi anche con le code di input se il dispositivo non dispone di risorse sufficienti per inoltrare immediatamente il pacchetto. Ciò può essere dovuto alla CPU, alla memoria o ai buffer.

La latenza descrive il tempo di elaborazione normale dal momento della ricezione a quello dell'inoltro del pacchetto. I moderni switch di dati e i router normali hanno una latenza estremamente bassa (< 1 ms) in condizioni normali senza vincoli di risorse. I dispositivi moderni dotati di processori di segnale digitale per convertire e comprimere i pacchetti voce analogici possono richiedere più tempo, anche fino a 20 ms.

Jitter descrive il gap tra pacchetti per le applicazioni di streaming, inclusi voce e video. Se i pacchetti arrivano in momenti diversi con intervalli diversi tra i pacchetti, l'instabilità è elevata e la qualità della voce peggiora. La variazione è principalmente un fattore di ritardo di accodamento.

Velocità e distanza

Velocità e distanza sono fattori che influiscono sulle prestazioni della rete. Le reti di dati hanno una velocità di inoltro dei dati coerente basata sulla velocità della luce. Si tratta di circa 100 miglia al millisecondo. Se un'organizzazione esegue un'applicazione client-server a livello internazionale, può prevedere un ritardo di inoltro dei pacchetti corrispondente. La velocità e la distanza possono essere un fattore fondamentale per le prestazioni delle applicazioni quando queste non sono ottimizzate per le prestazioni della rete.

Caratteristiche dell'applicazione

Le caratteristiche delle applicazioni sono l'ultima area che influisce sulla capacità e sulle prestazioni. Problemi quali le dimensioni ridotte delle finestre, i pacchetti keepalive delle applicazioni e la quantità di dati inviati sulla rete rispetto a quanto richiesto possono influire sulle prestazioni di un'applicazione in molti ambienti, in particolare le WAN.

Procedure ottimali per la gestione di capacità e prestazioni

In questa sezione vengono descritte in dettaglio le cinque procedure ottimali principali per la gestione della capacità e delle prestazioni:

- [Gestione del livello del servizio](#)
- [Analisi di simulazione della rete e delle applicazioni](#)

- [Baseline e trend](#)
- [Gestione delle eccezioni](#)
- [Gestione QoS](#)

Gestione del livello del servizio

La gestione dei livelli di servizio definisce e regola altri processi di gestione della capacità e delle prestazioni richiesti. I responsabili di rete sono consapevoli di aver bisogno di una pianificazione della capacità, ma devono far fronte a vincoli di budget e di personale che impediscono di ottenere una soluzione completa. La gestione dei livelli di servizio è una metodologia collaudata che aiuta a risolvere i problemi relativi alle risorse definendo un servizio e creando una responsabilità bidirezionale per un servizio associato a tale servizio. A tale scopo, è possibile procedere in due modi:

- Creare un accordo sui livelli di servizio tra gli utenti e l'organizzazione di rete per un servizio che includa la gestione della capacità e delle prestazioni. Il servizio include report e suggerimenti per mantenere la qualità del servizio. Tuttavia, gli utenti devono essere pronti a finanziare il servizio e gli eventuali aggiornamenti necessari.
- L'organizzazione di rete definisce il proprio servizio di gestione della capacità e delle prestazioni e quindi tenta di ottenere i fondi necessari per tale servizio e di eseguire aggiornamenti caso per caso.

In ogni caso, l'organizzazione della rete dovrebbe iniziare definendo un servizio di pianificazione della capacità e di gestione delle prestazioni che includa quali aspetti del servizio possono attualmente fornire e quali sono le previsioni per il futuro. Un servizio completo potrebbe includere un'analisi di simulazione per le modifiche alla rete e alle applicazioni, la definizione di baseline e trend per le variabili di prestazioni definite, la gestione delle eccezioni per le variabili di capacità e prestazioni definite e la gestione QoS.

Analisi di simulazione della rete e delle applicazioni

Eseguire un'analisi di simulazione della rete e dell'applicazione per determinare il risultato di una modifica pianificata. Senza un'analisi di simulazione, le organizzazioni corrono rischi significativi per modificare il successo e la disponibilità complessiva della rete. In molti casi, i cambiamenti della rete hanno provocato un collasso congestizio che ha causato molte ore di inattività della produzione. Inoltre, una quantità impressionante di introduzioni di applicazioni non riesce e causa impatto ad altri utenti e applicazioni. Questi guasti continuano in molte organizzazioni di rete, ma sono completamente prevenibili con alcuni strumenti e alcune fasi di pianificazione aggiuntive.

In genere sono necessari alcuni nuovi processi per eseguire un'analisi di simulazione di qualità. Il primo passo consiste nell'identificare i livelli di rischio per tutte le modifiche e richiedere un'analisi di simulazione più approfondita per le modifiche che comportano maggiori rischi. Il livello di rischio può essere un campo obbligatorio per tutti gli invii di modifiche. Modifiche di livello di rischio più elevate richiederebbero quindi un'analisi di simulazione definita della modifica. Un'analisi di simulazione della rete determina l'effetto delle modifiche di rete sull'utilizzo della rete e i problemi relativi alle risorse del control plane di rete. Un'analisi di simulazione dell'applicazione determinerebbe il successo dell'applicazione del progetto, i requisiti di larghezza di banda e qualsiasi problema relativo alle risorse di rete. Le tabelle seguenti sono esempi di assegnazione del livello di rischio e dei corrispondenti requisiti di prova:

Liv	Definizione	Suggerimenti per la
-----	-------------	---------------------

ello di rischio		pianificazione delle modifiche
1	<ul style="list-style-type: none"> • Impatto potenziale elevato su un numero elevato di utenti (oltre 500) o servizi business critical a causa dell'introduzione di nuovi prodotti, software, topologie o funzionalità. • La modifica comporta il tempo di inattività previsto della rete. 	<ul style="list-style-type: none"> • Convalida laboratorio nuova soluzione. La convalida in laboratorio include test e convalida documentati della soluzione e analisi di simulazione che mostrano l'impatto sull'infrastruttura esistente. Consigliamo progetti pilota per le soluzioni. Le nuove soluzioni richiedono il completamento di un documento di supporto alle operazioni. • Esegue la revisione del progetto Cisco NSA. • Creare un piano di ripristino. • Creare il piano di implementazione. • Crea processo di modifica.
2	<ul style="list-style-type: none"> • Elevato impatto potenziale 	<ul style="list-style-type: none"> • Eseguire

	<p>su un numero elevato di utenti (oltre 500) o servizi business critical a causa di un notevole aumento del traffico o degli utenti, modifiche alla backbone o modifiche al routing.</p> <ul style="list-style-type: none"> • La modifica potrebbe richiedere tempi di inattività. 	<p>un'analisi di simulazione per determinare l'impatto sull'ambiente esistente (da eseguire in un ambiente lab).</p> <ul style="list-style-type: none"> • Verificare e rivedere le modifiche del ciclo per verificarne la funzionalità. • Creare un piano di ripristino. • Esegue l'analisi della progettazione per le principali modifiche di instradamento o backbone. • Creare il piano di implementazione. • Crea processo di modifica.
3	<ul style="list-style-type: none"> • Impatto potenziale medio su un numero inferiore di utenti o servizi aziendali a causa di modifiche non standard. • Include nuovo prodotto, software, topologia, aggiunta di funzionalità o nuovi utenti, aumento del traffico o topologia non standard. • La modifica potrebbe richiedere tempi di inattività. 	<ul style="list-style-type: none"> • Eseguire l'analisi tecnica della nuova soluzione (potrebbe essere necessaria la convalida in laboratorio). • Creare il piano di implementazione. • Crea processo di modifica.
4	<ul style="list-style-type: none"> • Riduzione dell'impatto 	<ul style="list-style-type: none"> • Creare il piano

	<p>potenziale su servizi e utenti.</p> <ul style="list-style-type: none"> • Include l'aggiunta di nuovi moduli di rete modello standard, ad esempio la creazione o gli switch/hub server sui router. • Include la creazione di nuovi siti WAN o servizi di accesso collaudati aggiuntivi. • Tutte le modifiche al livello di rischio 3 sono state dimostrate tecnicamente nell'ambiente di produzione. • La modifica potrebbe richiedere tempi di inattività. 	<p>di implementazione.</p> <ul style="list-style-type: none"> • Crea processo di modifica.
5	<ul style="list-style-type: none"> • Nessun impatto sugli utenti o sui servizi. • Include l'aggiunta di singoli utenti alla rete e le modifiche alla configurazione standard come password, banner, SNMP o altri parametri di configurazione standard. • Nessun tempo di inattività. 	<ul style="list-style-type: none"> • Processo di modifica facoltativo.

Una volta definita la posizione in cui è necessaria l'analisi di simulazione, è possibile definire il servizio.

È possibile eseguire un'analisi di simulazione della rete con strumenti di modellazione o un'esercitazione che simuli l'ambiente di produzione. Gli strumenti di modellazione sono limitati dalla capacità dell'applicazione di comprendere i problemi relativi alle risorse del dispositivo e, poiché la maggior parte delle modifiche di rete è costituita da dispositivi nuovi, l'applicazione potrebbe non comprendere l'effetto della modifica. Il metodo migliore è quello di creare una rappresentazione della rete di produzione in un laboratorio e di testare il software, la funzionalità, l'hardware o la configurazione desiderata sotto carico utilizzando i generatori di traffico. Le perdite di dati (o altre informazioni di controllo) dalla rete di produzione alle apparecchiature di laboratorio contribuiscono a migliorare l'ambiente di emulazione. Verificare i requisiti di risorse aggiuntive con diversi tipi di traffico, tra cui SNMP, broadcast, multicast, crittografato o compresso. Con tutte queste diverse metodologie, analizzare i requisiti di risorse del dispositivo durante le potenziali situazioni di stress, quali la convergenza delle route, il link flapping e il riavvio del dispositivo. I problemi relativi all'utilizzo delle risorse includono le aree delle risorse con capacità normale, quali CPU, memoria, utilizzo del backplane, buffer e accodamento.

Le nuove applicazioni dovrebbero inoltre eseguire un'analisi di simulazione per determinare il successo dell'applicazione e i requisiti di larghezza di banda. In genere, questa analisi viene eseguita in un ambiente lab utilizzando un analizzatore di protocolli e un simulatore di ritardo WAN per comprendere l'effetto della distanza. È sufficiente un PC, un hub, un dispositivo di ritardo WAN

e un router lab collegati alla rete di produzione. È possibile simulare la larghezza di banda nel lab limitando il traffico utilizzando il traffic shaping generico o la limitazione della velocità sul router di test. L'amministratore di rete può lavorare in collaborazione con il gruppo di applicazioni per comprendere i requisiti di larghezza di banda, i problemi di windowing e i potenziali problemi di prestazioni dell'applicazione in ambienti LAN e WAN.

Eseguire un'analisi di simulazione dell'applicazione prima di distribuire qualsiasi applicazione aziendale. In caso contrario, il gruppo di applicazioni attribuisce alla rete la colpa delle prestazioni insoddisfacenti. Se è possibile richiedere un'analisi di simulazione dell'applicazione per nuove distribuzioni tramite il processo di gestione delle modifiche, è possibile evitare distribuzioni non riuscite e comprendere meglio gli aumenti improvvisi dell'utilizzo della larghezza di banda per i requisiti sia client-server che batch.

Baseline e trend

La definizione dei livelli di base e delle tendenze consente agli amministratori di pianificare e completare gli aggiornamenti della rete prima che un problema di capacità causi tempi di inattività o problemi di prestazioni. Confrontare l'utilizzo delle risorse durante periodi di tempo successivi o distillare le informazioni inattive nel tempo in un database e consentire ai responsabili della pianificazione di visualizzare i parametri di utilizzo delle risorse per l'ultima ora, giorno, settimana, mese e anno. In entrambi i casi, qualcuno deve rivedere le informazioni su base settimanale, bisettimanale o mensile. Il problema con l'associazione alla baseline e i trend è che è necessaria una quantità enorme di informazioni da esaminare nelle reti di grandi dimensioni.

Per risolvere il problema, procedere in diversi modi:

- Ampia capacità e possibilità di commutazione all'interno dell'ambiente LAN, in modo che la capacità non costituisca un problema.
- Dividere le informazioni sulle tendenze in gruppi e concentrarsi sull'alta disponibilità o su aree critiche della rete, come siti WAN critici o LAN di data center.
- I meccanismi di notifica possono evidenziare le aree che superano una determinata soglia per ricevere un'attenzione particolare. Se si implementano prima le aree di disponibilità critiche, è possibile ridurre in modo significativo la quantità di informazioni necessarie per la revisione.

Con tutti i metodi precedenti, è comunque necessario esaminare le informazioni periodicamente. L'approvazione e l'analisi dei trend sono attività proattive e se l'organizzazione dispone solo di risorse per il supporto reattivo, gli utenti non potranno leggere i report.

Molte soluzioni di gestione della rete forniscono informazioni e grafici sulle variabili delle risorse di capacità. Sfortunatamente, la maggior parte delle persone utilizza questi strumenti solo per fornire un supporto reattivo a un problema esistente; in questo modo viene vanificato lo scopo dell'utilizzo delle baseline e dei trend. Due strumenti efficaci nel fornire informazioni sulle tendenze della capacità per le reti Cisco sono il prodotto Concord Network Health e i prodotti INS EnterprisePRO. In molti casi, le organizzazioni di rete utilizzano semplici linguaggi di script per raccogliere informazioni sulla capacità. Di seguito sono riportati alcuni report di esempio raccolti tramite Script per l'utilizzo dei collegamenti, l'utilizzo della CPU e le prestazioni del ping. Altre variabili di risorsa che possono essere importanti per la tendenza sono la memoria, la profondità della coda, il volume di trasmissione, il buffer, la notifica di congestione del frame relay e l'utilizzo del backplane. Per informazioni sull'utilizzo dei collegamenti e della CPU, fare riferimento alla tabella seguente:

Utilizzo collegamento

Risorsa	Indirizzo	Segmento	Utilizzo medio (%)	Utilizzo massimo (%)
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
PACR01S 3/1	10.2.5.2	384 Kbps	51.1	98.4

Utilizzo CPU

Risorsa	Indirizzo di polling	Utilizzo medio (%)	Utilizzo massimo (%)
FSTR01	10.28.142.1	60.4	80
NERT 06	10.170.2.1	47	86
NORR01	10.73.200.1	47	99
RTCR01	10.49.136.1	42	98

Utilizzo collegamento

Risorsa	Indirizzo	AvResT (mS) 09-09-98	AvResT (mS) 09-09-98	AvResT (mS) 09-09-98	AvResT (mS) 10-01-98
ADR 01	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
APRILE 01	10.190.54.1	490.7	883.4	485.2	892.5
ASAR 01	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
AZURT 01	10.177.32.1	460.1		444.7	
BEJR01	10.195.18.1	1023.7	1064.6	1184	1021.9

Gestione delle eccezioni

La gestione delle eccezioni è una metodologia valida per identificare e risolvere i problemi di capacità e prestazioni. L'idea è quella di ricevere una notifica delle violazioni delle soglie di capacità e prestazioni al fine di indagare e risolvere immediatamente il problema. Ad esempio, un amministratore di rete potrebbe ricevere un avviso relativo a un utilizzo elevato della CPU su un router. L'amministratore di rete può accedere al router per determinare il motivo per cui la CPU è

così alta. Può quindi eseguire alcune configurazioni correttive che riducono la CPU o creare un elenco degli accessi per impedire il traffico che causa il problema, soprattutto se il traffico non sembra essere di importanza critica per l'azienda.

È possibile configurare la gestione delle eccezioni per problemi più critici in modo semplice utilizzando i comandi di configurazione RMON su un router o strumenti più avanzati, ad esempio Netsys Service Level Manager in combinazione con i dati SNMP, RMON o Netflow. La maggior parte degli strumenti di gestione della rete è in grado di impostare soglie e allarmi in caso di violazioni. L'aspetto importante del processo di gestione delle eccezioni è quello di fornire una notifica quasi in tempo reale del problema. In caso contrario, il problema potrebbe risolversi prima che qualcuno si accorga di aver ricevuto la notifica. Ciò può essere fatto all'interno di un NOC se l'organizzazione ha un monitoraggio costante. In caso contrario, si consiglia di inviare una notifica al cercapersone.

Nell'esempio di configurazione riportato di seguito viene fornita una notifica di soglia crescente e decrescente per la CPU del router in un file di log che può essere esaminato in modo coerente. È possibile impostare comandi RMON simili per violazioni critiche della soglia di utilizzo del collegamento o per altre soglie 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
```

Gestione QoS

La gestione della qualità del servizio implica la creazione e il monitoraggio di classi di traffico specifiche all'interno della rete. Un traffico fornisce prestazioni più coerenti per specifici gruppi di applicazioni (definiti all'interno delle classi di traffico). I parametri di Traffic Shaping offrono una notevole flessibilità nell'assegnazione delle priorità e nel Traffic Shaping per classi di traffico specifiche. Queste funzionalità includono funzionalità come CAR (Committed Access Rate), WRED (Weighted Random Early Detection) e le code ponderate eque basate su classi. Le classi di traffico vengono in genere create in base agli SLA sulle prestazioni per le applicazioni business-critical più complesse e ai requisiti specifici delle applicazioni, ad esempio la voce. Anche il traffico non critico o non destinato alle aziende verrebbe controllato in modo tale da non influire sulle applicazioni e sui servizi con priorità più alta.

La creazione di classi di traffico richiede una conoscenza di base dell'utilizzo della rete, dei requisiti specifici delle applicazioni e delle priorità delle applicazioni aziendali. I requisiti delle applicazioni includono la conoscenza delle dimensioni dei pacchetti, dei problemi di timeout, dei requisiti di jitter, burst, batch e dei problemi di prestazioni complessivi. Sulla base di queste conoscenze, gli amministratori di rete possono creare piani e configurazioni di traffic shaping che forniscono prestazioni più coerenti delle applicazioni su diverse topologie LAN/WAN.

Ad esempio, un'organizzazione ha una connessione ATM da 10 megabit tra due siti principali. A volte il collegamento viene congestionato da trasferimenti di file di grandi dimensioni, con conseguente riduzione delle prestazioni per l'elaborazione delle transazioni online e una qualità vocale scadente o inutilizzabile.

L'organizzazione ha stabilito quattro diverse classi di traffico. Alla voce è stata assegnata la

priorità più alta e le è stato consentito di mantenere tale priorità anche se supera la velocità stimata del volume di traffico. Alla classe di applicazioni critiche è stata assegnata la priorità successiva ma non è stato consentito di frammentare oltre le dimensioni totali del collegamento meno i requisiti stimati della larghezza di banda vocale. Quando esplode, cadrà. Al traffico di trasferimento dei file è stata semplicemente assegnata una priorità più bassa e tutto il resto del traffico viene spostato da qualche parte nel mezzo.

L'organizzazione deve ora eseguire la gestione QoS su questo collegamento per determinare la quantità di traffico che ogni classe sta assorbendo e misurare le prestazioni all'interno di ogni classe. Se l'organizzazione non riesce a eseguire questa operazione, è possibile che si verifichi la carenza di alcune classi o che gli SLA delle prestazioni non vengano soddisfatti all'interno di una determinata classe.

La gestione delle configurazioni QoS è ancora un'attività difficile a causa della mancanza di strumenti. Uno dei metodi consiste nell'utilizzare Cisco Internet Performance Manager (IPM) per inviare traffico diverso attraverso il collegamento che rientra in ciascuna classe di traffico. È quindi possibile monitorare le prestazioni per ogni classe e IPM fornisce analisi dei trend, in tempo reale e hop-by-hop per individuare le aree problematiche. Altri potrebbero ancora affidarsi a un metodo più manuale, come l'analisi della coda e dei pacchetti scartati all'interno di ciascuna classe di traffico, sulla base delle statistiche dell'interfaccia. In alcune organizzazioni, questi dati possono essere raccolti tramite SNMP o analizzati in un database per individuare le baseline e i trend. Nel mercato sono inoltre disponibili alcuni strumenti che inviano tipi di traffico specifici attraverso la rete per determinare le prestazioni di un particolare servizio o applicazione.

Raccolta e reporting delle informazioni sulla capacità

La raccolta e la comunicazione delle informazioni sulla capacità dovrebbero essere collegate alle tre aree raccomandate di gestione della capacità:

- Analisi di simulazione, incentrata sulle modifiche alla rete e sul modo in cui tali modifiche influiscono sull'ambiente
- Baseline e trend
- Gestione delle eccezioni

Sviluppare un piano di raccolta delle informazioni all'interno di ciascuna di queste aree. Nel caso di analisi di simulazione della rete o dell'applicazione, sono necessari strumenti per simulare l'ambiente di rete e per comprendere l'effetto della modifica in relazione a potenziali problemi di risorse all'interno del piano di controllo del dispositivo o del piano dati. Nel caso di baseline e trend, sono necessarie istantanee per i dispositivi e i collegamenti che mostrano l'utilizzo corrente delle risorse. È quindi possibile esaminare i dati nel tempo per comprendere i potenziali requisiti di aggiornamento. Ciò consente agli amministratori di rete di pianificare correttamente gli aggiornamenti prima che si verifichino problemi di capacità o prestazioni. Quando si verificano dei problemi, è necessario disporre di una gestione delle eccezioni per avvisare gli amministratori di rete in modo che possano ottimizzare la rete o risolvere il problema.

Questo processo può essere suddiviso nei seguenti passaggi:

1. Determinate le vostre esigenze.
2. Definire un processo.
3. Definire le aree di capacità.
4. Definire le variabili di capacità.
5. Interpretare i dati.

Determinate le vostre esigenze

Lo sviluppo di un piano di gestione della capacità e delle prestazioni richiede la comprensione delle informazioni necessarie e dello scopo di tali informazioni. Dividere il piano in tre aree obbligatorie: uno per l'analisi di simulazione, l'approvazione/l'analisi dei trend e la gestione delle eccezioni. In ognuna di queste aree, individuare le risorse e gli strumenti disponibili e le esigenze. Molte organizzazioni non riescono a distribuire gli strumenti perché prendono in considerazione la tecnologia e le funzionalità degli strumenti ma non prendono in considerazione le persone e le competenze necessarie per gestirli. Includere nel piano le persone e le competenze necessarie, nonché i miglioramenti dei processi. Tali persone possono includere amministratori di sistema per la gestione delle stazioni di gestione della rete, amministratori di database per l'amministrazione del database, amministratori esperti per l'utilizzo e il monitoraggio degli strumenti e amministratori di rete di livello superiore per la determinazione di regole, soglie e requisiti di raccolta delle informazioni.

Definisci processo

È inoltre necessario un processo per garantire che lo strumento venga utilizzato in modo corretto e coerente. Potrebbe essere necessario apportare miglioramenti al processo per definire le operazioni che gli amministratori di rete devono eseguire quando si verificano violazioni di soglia o il processo da seguire per l'approvazione, l'analisi dei trend e l'aggiornamento della rete. Una volta determinati i requisiti e le risorse per una corretta pianificazione della capacità, è possibile prendere in considerazione la metodologia. Molte organizzazioni scelgono di affidare questo tipo di funzionalità a un'organizzazione di servizi di rete, ad esempio INS, o di creare competenze interne in quanto considerano il servizio una competenza fondamentale.

Definizione aree capacità

Il piano di pianificazione della capacità dovrebbe inoltre comprendere una definizione delle aree di capacità. Questi sono i settori della rete che possono condividere una strategia comune di pianificazione della capacità: ad esempio, LAN aziendale, uffici WAN, siti WAN critici e accesso dial-in. La definizione di aree diverse è utile per diversi motivi:

- Aree diverse possono avere soglie diverse. Ad esempio, la larghezza di banda della LAN è molto più economica della larghezza di banda della WAN, quindi le soglie di utilizzo dovrebbero essere inferiori.
- Diverse aree possono richiedere il monitoraggio di diverse variabili MIB. Ad esempio, i contatori FECN e BECN in Frame Relay sono fondamentali per comprendere i problemi di capacità del frame relay.
- L'aggiornamento di alcune aree della rete può risultare più difficile o richiedere più tempo. I circuiti internazionali, ad esempio, possono avere tempi di consegna molto più lunghi e necessitano di un corrispondente livello di pianificazione più elevato.

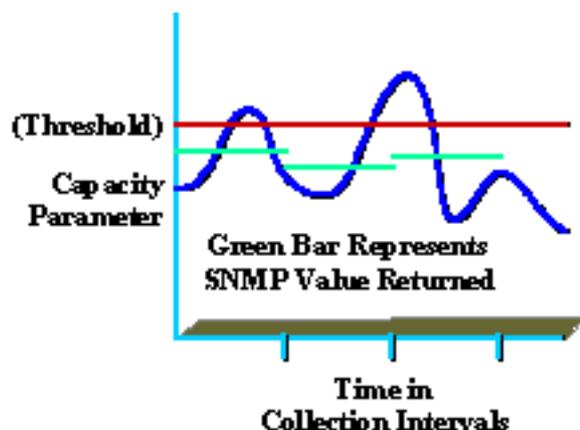
Definizione delle variabili di capacità

La successiva area importante è la definizione delle variabili da monitorare e dei valori di soglia che richiedono un'azione. La definizione delle variabili di capacità dipende in modo significativo dalle periferiche e dai supporti utilizzati nella rete. In generale, parametri quali l'utilizzo della CPU, della memoria e dei collegamenti sono importanti. Tuttavia, altre aree possono essere importanti per tecnologie o requisiti specifici. Questi possono includere profondità della coda, prestazioni,

notifica della congestione del frame relay, utilizzo del backplane, utilizzo del buffer, statistiche netflow, volume di trasmissione e dati RMON. Tenete a mente i vostri piani a lungo termine, ma iniziate con poche aree chiave per assicurarvi il successo.

Interpretare i dati

Comprendere i dati raccolti è fondamentale anche per fornire un servizio di alta qualità. Molte organizzazioni, ad esempio, non sono in grado di comprendere appieno i livelli di utilizzo massimo e medio. Il diagramma seguente mostra il picco del parametro di capacità in base a un intervallo di raccolta SNMP di 5 minuti (visualizzato in verde).



Anche se il valore segnalato era inferiore alla soglia (mostrata in rosso), i picchi possono comunque verificarsi entro l'intervallo di raccolta che sono superiori al valore di soglia (indicato in blu). Ciò è significativo perché durante l'intervallo di raccolta, l'organizzazione potrebbe riscontrare valori di picco che influiscono sulle prestazioni o sulla capacità della rete. Fare attenzione a selezionare un intervallo di raccolta significativo che sia utile e che non provochi un sovraccarico eccessivo.

Un altro esempio è l'utilizzo medio. Se i dipendenti sono in ufficio solo da otto a cinque, ma l'utilizzo medio è 24 ore su 24, 7 giorni su 7, le informazioni potrebbero essere fuorvianti.

Informazioni correlate

- [Supporto tecnico – Cisco Systems](#)

Informazioni su questa traduzione

Cisco ha tradotto questo documento utilizzando una combinazione di tecnologie automatiche e umane per offrire ai nostri utenti in tutto il mondo contenuti di supporto nella propria lingua. Si noti che anche la migliore traduzione automatica non sarà mai accurata come quella fornita da un traduttore professionista. Cisco Systems, Inc. non si assume alcuna responsabilità per l'accuratezza di queste traduzioni e consiglia di consultare sempre il documento originale in inglese (disponibile al link fornito).