

Gestione delle prestazioni: White paper sulle procedure ottimali

Sommario

[Introduzione](#)

[Premesse](#)

[Fattori di successo critici](#)

[Indicatori per la gestione delle prestazioni](#)

[Flusso del processo di gestione delle prestazioni](#)

[Sviluppo di un concetto operativo di gestione della rete](#)

[Misura prestazioni](#)

[Esecuzione di un'analisi proattiva degli errori](#)

[Indicatori gestione prestazioni](#)

[Documentazione degli obiettivi aziendali di gestione della rete](#)

[Documentazione degli accordi sui livelli di servizio](#)

[Creare un elenco di variabili per la baseline](#)

[Analizzare le analisi delle baseline e delle tendenze](#)

[Documentazione di una metodologia di analisi di simulazione](#)

[Documentazione della metodologia utilizzata per aumentare le prestazioni della rete](#)

[Riepilogo](#)

[Informazioni correlate](#)

[Introduzione](#)

La gestione delle prestazioni implica l'ottimizzazione dei tempi di risposta dei servizi di rete e la gestione della coerenza e della qualità dei servizi di rete singoli e complessivi. Il servizio più importante è la necessità di misurare il tempo di risposta dell'utente/applicazione. Per la maggior parte degli utenti, i tempi di risposta rappresentano il fattore critico per il successo delle prestazioni. Questa variabile determina la percezione del successo della rete sia da parte degli utenti che degli amministratori delle applicazioni.

[Premesse](#)

La pianificazione della capacità è il processo mediante il quale si determinano i requisiti per le risorse di rete future al fine di evitare un impatto sulle prestazioni o sulla disponibilità delle applicazioni business-critical. Per quanto riguarda la pianificazione della capacità, la linea di base della rete (CPU, memoria, buffer, ottetti in/out, ecc.) può influire sul tempo di risposta. Pertanto, tenere presente che i problemi di prestazioni spesso sono correlati alla capacità. Nelle reti si tratta in genere di larghezza di banda e dati che devono attendere in coda prima di poter essere trasmessi tramite la rete. Nelle applicazioni vocali, questo tempo di attesa influisce quasi sicuramente sugli utenti, in quanto fattori quali il ritardo e la variazione influiscono sulla qualità

della chiamata vocale.

Un altro problema importante che complica la gestione delle prestazioni è che, sebbene l'elevata disponibilità della rete sia un fattore mission-critical sia per le reti di grandi aziende che per quelle di provider di servizi, la tendenza è quella di cercare vantaggi economici a breve termine a rischio di (spesso imprevisi) maggiori costi nel lungo periodo. Durante ogni ciclo di budget, gli amministratori di rete e il personale addetto all'implementazione dei progetti faticano a trovare un equilibrio tra prestazioni e rapidità di implementazione. Inoltre, gli amministratori di rete devono affrontare sfide che includono il rapido sviluppo dei prodotti per soddisfare le finestre di mercato limitate, tecnologie complesse, consolidamento aziendale, mercati concorrenti, downtime non programmati, mancanza di esperienza e strumenti spesso insufficienti.

Alla luce di queste sfide, come si inseriscono le prestazioni nel framework di gestione della rete? La funzione principale di un sistema ideale di gestione della rete è ottimizzare le capacità operative di una rete. Una volta accettato questo come obiettivo finale per la gestione della rete, l'obiettivo principale della gestione della rete è mantenere le prestazioni della rete al massimo.

Un sistema ideale di gestione della rete include le seguenti operazioni principali:

- Informa l'operatore dell'imminente deterioramento delle prestazioni.
- Fornisce semplici alternative di routing e soluzioni alternative in caso di deterioramento o guasto delle prestazioni.
- Fornisce gli strumenti per individuare le cause del deterioramento o del guasto delle prestazioni.
- Funge da stazione principale per la resilienza e la sopravvivenza della rete.
- Comunica le prestazioni in tempo reale.

In base a questa definizione di sistema ideale, la gestione delle prestazioni diventa essenziale per la gestione della rete. Questi problemi di gestione delle prestazioni sono critici:

- Prestazioni utente
- Prestazioni delle applicazioni
- Pianificazione della capacità
- Gestione proattiva degli errori

È importante notare che con le applicazioni più recenti, come voce e video, le prestazioni sono la variabile chiave per il successo e se non è possibile ottenere prestazioni coerenti, il servizio viene considerato di basso valore e non funziona. In altri casi, gli utenti soffrono semplicemente di prestazioni variabili con timeout intermittenti delle applicazioni che riducono la produttività e la soddisfazione degli utenti.

In questo documento vengono descritti i problemi più critici relativi alla gestione delle prestazioni, tra cui i fattori critici per il successo, gli indicatori prestazionali e una mappa di processo di alto livello per la gestione delle prestazioni. Descrive inoltre i concetti di disponibilità, tempo di risposta, accuratezza, utilizzo e pianificazione della capacità e include una breve discussione sul ruolo dell'analisi proattiva degli errori nella gestione delle prestazioni e nel sistema ideale di gestione della rete.

Fattori di successo critici

I fattori critici per il successo identificano i requisiti per le procedure ottimali di implementazione. Per essere considerato un fattore di successo critico, un processo o una procedura deve

migliorare la disponibilità oppure l'assenza della procedura deve diminuire la disponibilità. Inoltre, il fattore di successo critico deve essere misurabile in modo che l'organizzazione possa determinare l'entità del successo.

Nota: per informazioni dettagliate, vedere [Indicatori di gestione delle prestazioni](#).

Questi sono i fattori critici per il successo della gestione delle prestazioni:

- Raccogliere una linea di base per i dati della rete e delle applicazioni.
- Eseguire un'analisi di simulazione sulla rete e sulle applicazioni.
- Generare rapporti sulle eccezioni per i problemi di capacità.
- Determinare il sovraccarico di gestione della rete per tutti i servizi di gestione della rete proposti o potenziali.
- Analizzare le informazioni sulla capacità.
- Esaminare periodicamente le informazioni sulla capacità per la rete e le applicazioni, nonché per la base e le eccezioni.
- disporre di procedure di upgrade o tuning per gestire i problemi di capacità sia in modo reattivo che a lungo termine.

[Indicatori per la gestione delle prestazioni](#)

Gli indicatori di prestazioni forniscono il meccanismo mediante il quale un'organizzazione può misurare i fattori critici di successo. Gli indicatori di prestazioni per la pianificazione delle prestazioni includono:

- Documentare gli obiettivi aziendali di gestione della rete. Potrebbe trattarsi di un concetto formale di operazioni per la gestione della rete o di una dichiarazione meno formale delle caratteristiche e degli obiettivi richiesti.
- Definizione di obiettivi di livello di servizio dettagliati e misurabili.
- Fornire la documentazione relativa agli accordi sui livelli di servizio con grafici o diagrammi che illustrino le modalità di esecuzione o meno di tali accordi nel tempo.
- Raccogliere un elenco delle variabili per la baseline, ad esempio l'intervallo di polling, il sovraccarico di gestione della rete sostenuto, le possibili soglie di trigger, se la variabile viene utilizzata come trigger per una trap e l'analisi delle tendenze utilizzata per ciascuna variabile.
- Organizzare una riunione periodica per esaminare l'analisi della baseline e delle tendenze.
- Documentare una metodologia di analisi di simulazione. Ciò dovrebbe comprendere la modellizzazione e la verifica, ove applicabile.
- Quando le soglie vengono superate, sviluppare la documentazione sulla metodologia utilizzata per aumentare le risorse di rete. Un elemento da documentare è la linea temporale richiesta per inserire una larghezza di banda WAN aggiuntiva e una tabella dei costi.

[Flusso del processo di gestione delle prestazioni](#)

Questi passaggi forniscono un flusso di processo di alto livello per la gestione delle prestazioni:

1. [Sviluppo di un concetto operativo di gestione della rete](#)[Definite le feature richieste: Obiettivi di servizi, scalabilità e disponibilità](#)[Definizione degli obiettivi di disponibilità e gestione della rete](#)[Definizione di SLA e metriche delle prestazioni](#)[Definizione degli SLA](#)

2. [Misura prestazioni](#)[Raccolta dei dati di base della rete](#)[Disponibilità delle misure](#)[Misura tempo di risposta](#)[Precisione misura](#)[Utilizzo misure](#)[Pianificazione della capacità](#)
3. [Esecuzione di un'analisi proattiva degli errori](#)[Utilizzo delle soglie per la gestione proattiva degli errori](#)[Implementazione della gestione della rete](#)[Metriche delle operazioni di rete](#)

Sviluppo di un concetto operativo di gestione della rete

Prima di definire le variabili dettagliate relative alle prestazioni e alla capacità di una rete, è necessario esaminare il concetto generale di funzionamento per la gestione della rete all'interno dell'organizzazione. Definendo questo concetto generale, si crea una base aziendale sulla quale è possibile creare definizioni precise delle funzionalità desiderate nella rete. Se non si sviluppa un concetto operativo per la gestione della rete, è possibile che la mancanza di obiettivi cambi continuamente a causa delle richieste dei clienti.

In genere, il concetto di gestione di rete delle operazioni viene prodotto come primo passaggio nella fase di definizione del sistema del programma di gestione di rete. Lo scopo è descrivere le caratteristiche generali desiderate del sistema da un punto di vista operativo. Questo documento viene utilizzato per coordinare gli obiettivi aziendali generali (non quantitativi) di operazioni di rete, progettazione, progettazione, altre business unit e utenti finali. Questo documento ha lo scopo di definire la pianificazione operativa a lungo termine per la gestione e il funzionamento della rete. Vengono inoltre fornite indicazioni per lo sviluppo di tutta la documentazione relativa alle definizioni successive, ad esempio gli accordi sui livelli di servizio. Questa serie iniziale di definizioni non può ovviamente concentrarsi troppo strettamente sulla gestione di problemi specifici della rete, ma su quelle voci che sottolineano l'importanza per l'organizzazione complessiva e in relazione ai costi che devono essere gestiti. Alcuni obiettivi sono:

- Individuare le caratteristiche essenziali per un uso efficiente dell'infrastruttura di rete.
- Identificare i servizi/le applicazioni supportati dalla rete.
- Avvio della gestione completa dei servizi.
- Avviare metriche basate sulle prestazioni per migliorare il servizio complessivo.
- Raccolta e distribuzione delle informazioni sulla gestione delle prestazioni.
- Supportare la valutazione strategica della rete con il feedback degli utenti.

In altre parole, il concetto operativo di gestione della rete deve concentrarsi sugli obiettivi organizzativi generali e sulla filosofia aziendale per raggiungere tali obiettivi. Gli ingredienti principali sono le definizioni di livello più alto della missione, gli obiettivi della missione, gli obiettivi del sistema, il coinvolgimento dell'organizzazione e la filosofia operativa generale.

In qualità di gestore di rete, è possibile unificare le aspettative di prestazioni spesso incoerenti degli utenti. Ad esempio, se il requisito principale per la rete è il trasferimento di file di grandi dimensioni da una posizione a un'altra, si desidera concentrarsi su un throughput elevato e meno sui tempi di risposta degli utenti interattivi. Fai attenzione a non limitare la tua visione delle prestazioni a meno che tu non consideri una varietà di problemi. Ad esempio, quando si esegue il test di una rete, esaminare i livelli di carico utilizzati. Il carico si basa spesso su pacchetti molto piccoli e sul throughput su pacchetti molto grandi. Uno di questi test delle prestazioni può produrre un'immagine molto positiva, ma in base al carico del traffico di rete, i test potrebbero non offrire un'immagine reale delle prestazioni. Studiare le prestazioni della rete nel maggior numero possibile di condizioni di carico di lavoro e documentare le prestazioni.

Inoltre, mentre molte organizzazioni di gestione della rete dispongono di tecniche di allarme efficaci per notificare ai tecnici i guasti dei dispositivi, è molto più difficile definire e implementare un processo di valutazione delle prestazioni delle applicazioni end-to-end. Pertanto, mentre il

centro operativo di rete (NOC) può rispondere rapidamente a un router o a uno switch disattivato, le condizioni di rete che potrebbero compromettere le prestazioni della rete e influire sulla percezione dell'utente potrebbero passare inosservate fino a quando la percezione non diventa negativa. Per quanto difficile, questo secondo processo può fornire enormi vantaggi sia all'organizzazione aziendale che alla gestione della rete.

Infine, accertarsi di non creare aspettative irrealistiche sulle prestazioni di rete. Le aspettative irrealistiche vengono in genere create quando si fraintendono i dettagli dei protocolli di rete o delle applicazioni. Spesso le prestazioni insoddisfacenti non sono imputabili alla rete, ma piuttosto alla progettazione inadeguata delle applicazioni. L'unico modo per documentare e misurare le prestazioni delle applicazioni consiste nel disporre di una base delle prestazioni della rete prima dell'installazione dell'applicazione.

[Definite le feature richieste: Obiettivi di servizi, scalabilità e disponibilità](#)

La prima fase della gestione delle prestazioni, della pianificazione continua della capacità e della progettazione della rete consiste nel definire le funzionalità e/o i servizi richiesti. Questo passaggio richiede la comprensione delle applicazioni, dei flussi di traffico di base, del numero di utenti e siti e dei servizi di rete richiesti. Il primo utilizzo di queste informazioni consiste nel determinare la criticità dell'applicazione rispetto agli obiettivi organizzativi. È inoltre possibile applicare queste informazioni per creare una knowledge base da utilizzare nella progettazione logica in modo da comprendere i requisiti relativi a larghezza di banda, interfaccia, connettività, configurazione e dispositivi fisici. Questo passaggio iniziale consente agli architetti di rete di creare un modello della rete.

Creare obiettivi di scalabilità della soluzione per aiutare i tecnici di rete a progettare reti che soddisfino i futuri requisiti di crescita e garantire che i progetti proposti non risentano di limitazioni delle risorse dovute alla crescita o all'estensione della rete. I vincoli delle risorse possono includere:

- Traffico complessivo
- Volume
- Numero di cicli di lavorazione
- Numero di circuiti virtuali
- Conteggi router adiacenti
- Domini di trasmissione
- Velocità effettiva dispositivo
- Capacità dei supporti

Il planner della rete dovrebbero determinare la durata richiesta della progettazione, le estensioni o i siti previsti necessari per tutta la durata della progettazione, il volume di nuovi utenti e il volume di traffico o le modifiche previste. Questo piano contribuisce a garantire che la soluzione proposta soddisfi i requisiti di crescita per tutta la durata prevista del progetto.

Se non si esamina la scalabilità della soluzione, è possibile che si debba implementare una serie di modifiche di progettazione reattive. Questa modifica può includere gerarchie aggiuntive, aggiornamenti di supporti o aggiornamenti hardware. Nelle organizzazioni che si affidano a cicli di budget abbastanza precisi per i principali acquisti di hardware, queste modifiche possono essere un importante ostacolo al successo generale. In termini di disponibilità, le reti possono sperimentare limitazioni impreviste delle risorse che causano periodi di non disponibilità e misure reattive.

I test di interoperabilità e interoperabilità possono essere fondamentali per il successo dell'installazione di nuove soluzioni. L'interoperabilità può riferirsi a diversi fornitori di hardware o a diverse topologie o soluzioni che devono interagire durante o dopo l'implementazione di una rete. I problemi di interoperabilità possono includere la segnalazione hardware di problemi di routing o trasporto attraverso lo stack di protocolli. I problemi di interoperabilità possono verificarsi prima, durante o dopo la migrazione di una soluzione di rete. La pianificazione dell'interoperabilità deve includere la connettività tra dispositivi diversi e i problemi di topologia che possono verificarsi durante le migrazioni.

Il confronto delle soluzioni è la pratica in cui vengono confrontati diversi progetti potenziali in relazione ad altre procedure relative ai requisiti delle soluzioni. Questa pratica aiuta a garantire che la soluzione sia la più adatta per un particolare ambiente e che i pregiudizi personali non guidino il processo di progettazione. Il confronto può includere diversi fattori quali costi, resilienza, disponibilità, rischi, interoperabilità, gestibilità, scalabilità e prestazioni. Una volta implementata la progettazione, tutte queste funzionalità possono influire in modo significativo sulla disponibilità complessiva della rete. È inoltre possibile confrontare supporti, gerarchia, ridondanza, protocolli di routing e funzionalità simili. Creare un grafico con i fattori sull'asse X e le possibili soluzioni sull'asse Y per riepilogare i confronti tra le soluzioni. Il confronto dettagliato delle soluzioni in un ambiente lab consente inoltre di analizzare in modo obiettivo nuove soluzioni e funzionalità in relazione ai diversi fattori di confronto.

Come parte del concetto operativo di gestione della rete, è essenziale definire gli obiettivi per la rete e i servizi supportati in modo che tutti gli utenti possano comprenderli. Le attività che seguono lo sviluppo del concetto operativo sono fortemente influenzate dalla qualità di tale documento.

Questi sono gli obiettivi di prestazioni standard:

- Tempo di risposta
- Utilizzo
- Velocità effettiva
- Capacità (velocità effettiva massima)

Anche se queste misurazioni possono essere inutili per una LAN semplice, possono essere molto difficili in una rete di campus a commutazione o in una rete aziendale multi-vendor. Quando si utilizza un concetto ben ponderato di piano operativo, ogni obiettivo di prestazioni viene definito in modo misurabile. Ad esempio, il tempo di risposta minimo per l'applicazione "x" è di 500 Ms o meno durante le ore di punta. Definisce le informazioni per identificare la variabile, il modo in cui misurarla e il periodo di tempo su cui l'applicazione di gestione della rete deve concentrarsi.

[Definizione degli obiettivi di disponibilità e gestione della rete](#)

Gli obiettivi di disponibilità definiscono il livello di servizio o i requisiti dei livelli di servizio per un servizio di rete. In questo modo la soluzione soddisfa i requisiti di disponibilità dei prodotti. Definire classi di servizio diverse per una particolare organizzazione e specificare i requisiti di rete per ogni classe appropriati al requisito di disponibilità. Anche diverse aree della rete potrebbero richiedere livelli diversi di disponibilità. Un obiettivo di maggiore disponibilità potrebbe richiedere un aumento della ridondanza e delle procedure di supporto. Quando si definisce un obiettivo di disponibilità per un particolare servizio di rete e si misura la disponibilità, l'organizzazione di rete può comprendere i componenti e i livelli di servizio richiesti per raggiungere gli SLA previsti.

Definire gli obiettivi di gestibilità per garantire che la gestione complessiva della rete non sia priva di funzionalità di gestione. Per impostare gli obiettivi di gestibilità, è necessario comprendere il processo di supporto e gli strumenti di gestione della rete associati all'organizzazione. Gli obiettivi

di gestibilità devono comprendere la conoscenza del modo in cui le nuove soluzioni si inseriscono nel modello di supporto e di strumento attuale con riferimenti a eventuali differenze potenziali o a nuovi requisiti. Ciò è fondamentale per la disponibilità della rete, in quanto la capacità di supportare nuove soluzioni è fondamentale per il successo dell'installazione e per soddisfare gli obiettivi di disponibilità.

Gli obiettivi di gestibilità dovrebbero rivelare tutte le informazioni MIB o degli strumenti di rete importanti necessarie per supportare una rete potenziale, la formazione necessaria per supportare il nuovo servizio di rete, i modelli di assegnazione del personale per il nuovo servizio ed eventuali altri requisiti di supporto. Spesso queste informazioni non vengono rilevate prima dell'installazione e la disponibilità complessiva risente della mancanza di risorse assegnate al supporto della nuova progettazione della rete.

[Definizione di SLA e metriche delle prestazioni](#)

SLA e metriche delle prestazioni consentono di definire e misurare le prestazioni delle nuove soluzioni di rete per garantire che soddisfino i requisiti di prestazioni. Le prestazioni della soluzione proposta possono essere misurate con strumenti di monitoraggio delle prestazioni o con un semplice ping sull'infrastruttura di rete proposta. Gli SLA delle prestazioni devono includere il volume medio previsto di traffico, il volume massimo di traffico, il tempo medio di risposta e il tempo massimo di risposta consentito. Queste informazioni possono quindi essere utilizzate in un secondo momento nella sezione di convalida della soluzione e consentono di determinare le prestazioni e la disponibilità della rete richieste.

[Definizione degli SLA](#)

Un aspetto importante della progettazione della rete è la definizione del servizio per utenti o clienti. Le aziende definiscono questi accordi sui livelli di servizio, mentre i fornitori di servizi li definiscono gestione dei livelli di servizio. La gestione dei livelli di servizio include in genere la definizione dei tipi di problemi e della loro gravità, nonché le responsabilità dell'help desk, quali il percorso di segnalazione e il tempo necessario prima dell'escalation a ogni livello di supporto, il tempo necessario per iniziare a risolvere il problema e il tempo necessario per chiudere gli obiettivi in base alla priorità. Altri fattori importanti sono il tipo di servizio fornito nell'area della pianificazione della capacità, della gestione proattiva degli errori, della notifica di gestione delle modifiche, delle soglie, dei criteri di aggiornamento e della sostituzione dell'hardware.

Quando le organizzazioni non definiscono i livelli di servizio in anticipo, diventa difficile migliorare o ottenere i requisiti delle risorse identificati in una data successiva. Diventa anche difficile capire quali risorse aggiungere per aiutare a supportare la rete. In molti casi, queste risorse vengono applicate solo dopo che sono stati rilevati problemi.

[Misura prestazioni](#)

Performance Management è un termine generico che include la configurazione e la misurazione di aree prestazionali distinte. In questa sezione vengono descritti i seguenti sei concetti di gestione delle prestazioni:

- [Raccolta dei dati di base della rete](#)
- [Disponibilità delle misure](#)
- [Misura tempo di risposta](#)
- [Precisione misura](#)

- [Utilizzo misure](#)
- [Pianificazione della capacità](#)

[Raccolta dei dati di base della rete](#)

La maggior parte delle Intranet aziendali dispone di larghezza di banda sufficiente. Tuttavia, senza dati adeguati, potrebbe non essere possibile escludere la congestione della rete come fattore che contribuisce a prestazioni inadeguate delle applicazioni. Uno degli indizi della congestione o degli errori è se le prestazioni scarse sono intermittenti o dipendenti dall'ora del giorno. Un esempio di questa condizione è quando le prestazioni sono adeguate a tarda sera, ma molto lente al mattino e durante le ore di punta.

Dopo aver definito il concetto di gestione della rete delle operazioni e i dati di implementazione necessari, è necessario raccogliere questi dati nel tempo. Questo tipo di raccolta costituisce la base per la linea di base della rete.

Eseguire una base della rete corrente prima dell'installazione di una nuova soluzione (modifica dell'applicazione o del sistema operativo IOS) e dopo l'installazione per misurare le aspettative impostate per la nuova soluzione. Questa base di riferimento consente di determinare se la soluzione soddisfa gli obiettivi di prestazioni e disponibilità e la capacità di benchmark. Un tipico report di base su router/switch include problemi di capacità relativi a CPU, memoria, gestione dei buffer, utilizzo di collegamenti/supporti e throughput. È possibile includere anche altri tipi di dati di previsione, in base agli obiettivi definiti nel concetto di operazioni. Ad esempio, una baseline della disponibilità dimostra una maggiore stabilità/disponibilità dell'ambiente di rete. Eseguire un confronto di base tra gli ambienti vecchi e nuovi per verificare i requisiti della soluzione.

Un'altra baseline specializzata è la baseline dell'applicazione, utile quando si valutano i requisiti di rete delle applicazioni. Queste informazioni possono essere utilizzate per la fatturazione e/o la definizione del budget nel ciclo di aggiornamento. Le baseline dell'applicazione possono inoltre essere importanti nell'area della disponibilità dell'applicazione in relazione ai servizi preferiti o alle qualità del servizio per applicazione. Le informazioni di base dell'applicazione consistono principalmente nella larghezza di banda utilizzata dalle applicazioni per periodo di tempo. Alcune applicazioni di gestione della rete possono inoltre offrire prestazioni di base. Anche un'analisi stratificata del tipo di traffico (Telnet o FTP) è importante per la pianificazione. In alcune organizzazioni, le aree più critiche della rete con risorse limitate vengono monitorate per i relatori più esperti. Gli amministratori di rete possono utilizzare queste informazioni per definire il budget, pianificare o ottimizzare la rete. Quando si esegue il tuning della rete, è possibile modificare i parametri di qualità del servizio o della coda per il servizio o l'applicazione di rete.

[Disponibilità delle misure](#)

Una delle metriche principali utilizzate dai gestori di rete è la disponibilità. La disponibilità è la misura del tempo per cui un sistema di rete o un'applicazione è disponibile per un utente. Dal punto di vista della rete, la disponibilità rappresenta l'affidabilità dei singoli componenti di una rete.

Ad esempio, per misurare la disponibilità, è possibile coordinare le chiamate telefoniche all'help desk con le statistiche raccolte dai dispositivi gestiti. Tuttavia, gli strumenti di disponibilità non sono in grado di determinare tutte le cause dell'errore.

La ridondanza di rete è un altro fattore da considerare quando si misura la disponibilità. La perdita di ridondanza indica una riduzione del livello di servizio piuttosto che un errore totale della rete. Il risultato potrebbe essere un tempo di risposta più lento e una perdita di dati dovuta alla perdita di

pacchetti. È inoltre possibile che i risultati vengano visualizzati in altre aree di misurazione delle prestazioni, quali l'utilizzo e il tempo di risposta.

Infine, se si esegue il rispetto di un contratto di servizio, è necessario tenere conto delle interruzioni pianificate. Tali interruzioni possono essere il risultato di spostamenti, aggiunte e modifiche, arresti degli impianti o altri eventi che non si desidera vengano segnalati. Non si tratta solo di un compito difficile, ma potrebbe anche essere un compito manuale.

Misura tempo di risposta

Il tempo di risposta di rete è il tempo necessario al traffico per spostarsi tra due punti. I tempi di risposta più lenti del normale, rilevati tramite un confronto di base o che superano una soglia, possono indicare una congestione o un errore di rete.

I tempi di risposta sono la misura migliore dell'utilizzo della rete da parte del cliente e consentono di misurare l'efficacia della rete. Indipendentemente dalla causa della lentezza nella risposta, gli utenti sono frustrati dal ritardo del traffico. Nelle reti distribuite, molti fattori influiscono sul tempo di risposta, ad esempio:

- Congestione della rete
- Indirizzamento verso destinazione inferiore a quello desiderato (o nessuna route)
- Dispositivi di rete sottoalimentati
- Guasti di rete, ad esempio durante le trasmissioni
- Errori di disturbo o CRC

Nelle reti che utilizzano le code QoS, la misurazione del tempo di risposta è importante per determinare se i tipi di traffico corretti si spostano nella rete nel modo previsto. Ad esempio, quando si implementa il traffico vocale su reti IP, i pacchetti voce devono essere consegnati puntualmente e a velocità costante per mantenere una buona qualità vocale. È possibile generare traffico classificato come traffico vocale per misurare il tempo di risposta del traffico come appare agli utenti.

È possibile misurare il tempo di risposta per contribuire a risolvere i conflitti tra gli application server e i gestori di rete. Gli amministratori di rete sono spesso ritenuti colpevoli quando un'applicazione o un server appare lento. L'amministratore di rete deve dimostrare che il problema non è la rete. La raccolta dei dati relativi ai tempi di risposta costituisce un mezzo indiscutibile per dimostrare o smentire che la rete è la causa dei problemi dell'applicazione.

Ove possibile, è consigliabile misurare il tempo di risposta così come viene visualizzato agli utenti. L'utente percepisce la risposta come il tempo che intercorre tra la pressione di Invio o il clic su un pulsante e la visualizzazione della schermata. Il tempo trascorso include il tempo necessario per l'elaborazione del traffico da parte di ciascun dispositivo di rete, della workstation utente e del server di destinazione.

Sfortunatamente, misurazioni a questo livello sono quasi impossibili a causa del numero di utenti e della mancanza di strumenti. Inoltre, incorporando i tempi di risposta di utenti e server, si ottiene uno scarso valore quando si determina la crescita futura della rete o si risolvono problemi di rete.

È possibile utilizzare i dispositivi di rete e i server per misurare il tempo di risposta. È inoltre possibile utilizzare strumenti quali ICMP per misurare le transazioni, anche se non tiene conto dei ritardi introdotti in un sistema mentre gli strati superiori lo elaborano. Questo approccio risolve il problema della conoscenza delle prestazioni di rete.

A un livello semplicistico, è possibile temporizzare la risposta ai ping dalla stazione di gestione di rete ai punti chiave della rete, come un'interfaccia mainframe, il punto finale della connessione di un provider di servizi o gli indirizzi IP degli utenti chiave, per misurare il tempo di risposta. Il problema di questo metodo è che non riflette in modo accurato la percezione del tempo di risposta tra il computer di destinazione e quello utilizzato dall'utente. Raccoglie semplicemente informazioni e genera report sui tempi di risposta dalla prospettiva della stazione di gestione di rete. Questo metodo maschera anche i problemi dei tempi di risposta hop-by-hop attraverso la rete.

Un'alternativa al polling basato sul server consiste nella distribuzione dell'impegno più vicino all'origine e alla destinazione che si desidera simulare per la misurazione. Usare i poler di gestione della rete distribuita e implementare la funzionalità di Cisco IOS Service Assurance Agent (SAA). L'appliance ASA può essere abilitata sui router per misurare il tempo di risposta tra un router e un dispositivo di destinazione, ad esempio un server o un altro router. È inoltre possibile specificare una porta TCP o UDP, che forza l'inoltro e il indirizzamento del traffico nello stesso modo in cui viene simulato.

Con l'integrazione di voce, video e dati su reti multiservice, i clienti implementano la definizione di priorità QoS nella propria rete. Le misurazioni ICMP o UDP non riflettono accuratamente i tempi di risposta, in quanto le diverse applicazioni ricevono priorità diverse. Inoltre, con la commutazione di tag, il routing del traffico potrebbe variare in base al tipo di applicazione contenuta in un pacchetto specifico. Pertanto, un ping ICMP potrebbe ricevere priorità diverse nel modo in cui ciascun router lo gestisce e potrebbe ricevere percorsi diversi e meno efficienti.

In questo caso, l'unico modo per misurare il tempo di risposta è generare un traffico simile alla particolare applicazione o tecnologia che interessa. In questo modo, i dispositivi di rete gestiscono il traffico come farebbero per il traffico reale. Questo livello può essere raggiunto con l'appliance ASA o tramite l'utilizzo di sonde compatibili con le applicazioni di terze parti.

[Precisione misura](#)

L'accuratezza è la misura del traffico di interfaccia che non genera errori e può essere espressa in termini di percentuale che confronta la frequenza di riuscita con la frequenza totale dei pacchetti in un periodo di tempo. È necessario prima misurare il tasso di errore. Ad esempio, se due pacchetti su 100 generano un errore, la percentuale di errore sarà pari al 2% e la percentuale di accuratezza al 98%.

Con le tecnologie di rete precedenti, in particolare nell'area estesa, un certo livello di errori era accettabile. Tuttavia, con le reti ad alta velocità e i servizi WAN attuali, la trasmissione è molto più accurata e le percentuali di errore sono vicine allo zero a meno che non ci sia un problema reale. Alcune cause comuni degli errori dell'interfaccia includono:

- Cablaggio non conforme alle specifiche
- Interferenza elettrica
- Hardware o software difettoso

Utilizzate una percentuale di precisione ridotta per avviare un'analisi più approfondita. È possibile che alcuni errori vengano rilevati da un'interfaccia specifica e che l'errore sia accettabile. In questo caso, è necessario regolare la soglia di accuratezza per questa interfaccia in modo da riflettere i punti in cui il tasso di errore è inaccettabile. Il tasso di errore inaccettabile potrebbe essere stato segnalato in una baseline precedente.

Le variabili descritte in questa tabella vengono utilizzate nelle formule relative alla precisione e alla

frequenza degli errori.

Notazione	Descrizione
$\Delta seErrori$	Il delta (o differenza) tra due cicli di polling che raccolgono l'oggetto snmp ifInErrors , che rappresenta il numero di pacchetti in entrata con un errore.
$\Delta ifInUcastPkts$	Delta tra due cicli di polling che raccolgono l'oggetto snmp ifInUcastPkts , che rappresenta il numero di pacchetti unicast in entrata.
$\Delta ifInNUcastPkts$	Delta tra i due cicli di polling che raccolgono l'oggetto snmp ifInNUcastPkts , che rappresenta il numero di pacchetti non unicast in entrata (multicast e broadcast).

La formula per il tasso di errore è generalmente espressa come percentuale:

$$\text{Frequenza errori} = (\Delta ifInErrors) * 100$$

—

$$(\Delta ifPacchettiTrasmissione + (\Delta ifInPacchettiTrasmissione))$$

Si noti che gli errori in uscita non vengono considerati nelle formule relative alla frequenza degli errori e alla precisione. Infatti, un dispositivo non deve mai inserire intenzionalmente pacchetti con errori sulla rete e la frequenza degli errori dell'interfaccia in uscita non deve mai aumentare. Pertanto, il traffico in entrata e gli errori sono le uniche misure di interesse per gli errori e l'accuratezza dell'interfaccia.

La formula per la precisione prende il tasso di errore e lo sottrae da 100 (di nuovo, sotto forma di percentuale):

$$\text{Precisione} = 100 - (\Delta ifInErrors) * 100$$

—

$$(\Delta ifPktsInUcast + (\Delta ifInNUcastPkts))$$

Queste formule riflettono l'errore e la precisione in termini di contatori generici dell'interfaccia MIB II (RFC 2233). Il risultato viene espresso in termini di una percentuale che confronta gli errori con il totale dei pacchetti visualizzati e inviati. La frequenza di errore risultante viene sottratta da 100, che produce la frequenza di accuratezza. Una percentuale di precisione del 100% è perfetta.

Poiché le variabili MIB II vengono memorizzate come contatori, è necessario eseguire due cicli di polling e calcolare la differenza tra i due (da cui il Delta utilizzato nell'equazione).

[Utilizzo misure](#)

Utilizzo misura l'utilizzo di una particolare risorsa nel tempo. La misura è generalmente espressa sotto forma di percentuale in cui l'uso di una risorsa è confrontato con la sua capacità operativa

massima. Tramite le misure di utilizzo è possibile identificare la congestione (o potenziale congestione) in tutta la rete. È inoltre possibile identificare le risorse sottoutilizzate.

L'utilizzo è la misura principale per determinare quanto sono piene le pipe di rete (collegamenti). Misurare la CPU, l'interfaccia, le code e altre misurazioni della capacità relative al sistema per determinare in che misura vengono consumate le risorse del sistema di rete.

Un utilizzo elevato non è necessariamente negativo. Un basso utilizzo potrebbe indicare flussi di traffico in posizioni impreviste. Con il sovrautilizzo delle linee, gli effetti possono diventare significativi. Il sovrautilizzo si verifica quando il traffico in coda per passare su un'interfaccia è superiore a quello che può gestire. Salti improvvisi nell'utilizzo delle risorse possono indicare una condizione di errore.

Quando un'interfaccia è congestionata, il dispositivo di rete deve archiviare il pacchetto in una coda o ignorarlo. Se un router tenta di archiviare un pacchetto in una coda completa, il pacchetto viene scartato. I pacchetti ignorati si verificano quando il traffico viene inoltrato da un'interfaccia veloce a un'interfaccia più lenta. Ciò è indicato nella formula $Q = u / (1-u)$ dove u è l'utilizzo e Q è la profondità media della coda (traffico casuale presunto). Pertanto, livelli di utilizzo elevati nei collegamenti determinano una profondità media della coda elevata, che è una latenza prevedibile se si conoscono le dimensioni del pacchetto. Alcuni fornitori di report di rete indicano che è possibile ordinare meno larghezza di banda e pagare meno per la rete WAN. Tuttavia, quando si eseguono collegamenti WAN con un utilizzo del 95%, appaiono implicazioni di latenza. Inoltre, con la migrazione delle reti al VoIP, gli amministratori di rete potrebbero dover modificare le proprie policy ed eseguire i collegamenti WAN con un utilizzo di circa il 50%.

Quando un pacchetto viene scartato, il protocollo di livello superiore potrebbe forzare una ritrasmissione del pacchetto. Se vengono scartati più pacchetti, potrebbe verificarsi un traffico eccessivo di nuovi tentativi. Questo tipo di reazione può determinare l'esecuzione di backup su dispositivi situati a valle della linea. Per risolvere questo problema, è possibile impostare diversi gradi di soglia.

La misura principale utilizzata per l'utilizzo della rete è l'utilizzo dell'interfaccia. Utilizzare le formule descritte in questa tabella a seconda che la connessione misurata sia half-duplex o full-duplex:

Notazione	Descrizione
Δ seln Ottetti	Il delta (o differenza) tra due cicli di polling che raccolgono l'oggetto snmp ifInOctets , che rappresenta il conteggio degli ottetti in entrata del traffico.
Δ lfOut tOctets	Delta tra due cicli di polling che raccolgono l'oggetto snmp ifOutOctets che rappresenta il conteggio degli ottetti in uscita del traffico.
ifSpeed	La velocità dell'interfaccia indicata nell'oggetto ifSpeed del protocollo snmp. Si noti che ifSpeed potrebbe non riflettere accuratamente la velocità di un'interfaccia WAN.

Le connessioni LAN condivise tendono ad essere half-duplex principalmente perché il rilevamento delle contese richiede che un dispositivo ascolti prima di trasmettere. Le connessioni WAN sono in genere full duplex perché la connessione è point-to-point; entrambi i dispositivi possono trasmettere e ricevere contemporaneamente perché sanno che esiste solo un altro dispositivo che

condivide la connessione.

Poiché le variabili MIB II vengono memorizzate come contatori, è necessario eseguire due cicli di polling e calcolare la differenza tra i due (da cui il Delta utilizzato nell'equazione).

Per i supporti half-duplex, utilizzare questa formula per l'utilizzo dell'interfaccia:

$$(\Delta \text{InOctets} + \Delta \text{OutOctets}) * 8 * 100$$

—

$$(\text{numero di secondi in } \square) * \text{ifSpeed}$$

Per i supporti full-duplex, il calcolo dell'utilizzo è più complesso. Ad esempio, con una connessione seriale T-1 completa, la velocità della linea è di 1,544 Mbps. Ciò significa che un'interfaccia T-1 può ricevere e trasmettere 1,544 Mbps per una larghezza di banda combinata possibile di 3,088 Mbps.

Quando si calcola la larghezza di banda dell'interfaccia per le connessioni full-duplex, è possibile usare questa formula in cui si prendono i valori più grandi **in** e **out** e si genera una percentuale di utilizzo:

$$\max(\Delta \text{InOctets}, \Delta \text{OutOctets}) * 8 * 100$$

—

$$(\text{numero di secondi in } \square) * \text{ifSpeed}$$

Tuttavia, questo metodo nasconde l'utilizzo della direzione con il valore minore e fornisce risultati meno accurati. Un metodo più accurato consiste nel misurare separatamente l'utilizzo degli input e l'utilizzo degli output, ad esempio:

$$\text{Utilizzo input} = \Delta \text{InOctets} * 8 * 100$$

—

$$(\text{numero di secondi in } \square) * \text{ifSpeed}$$

E

$$\text{Utilizzo output} = \Delta \text{OutOctets} * 8 * 100$$

—

$$(\text{numero di secondi in } \square) * \text{ifSpeed}$$

Sebbene queste formule siano in qualche modo semplificate, non prendono in considerazione il sovraccarico associato a un particolare protocollo. Sono disponibili formule più precise per gestire gli aspetti specifici di ciascun protocollo. Ad esempio, la RFC 1757 contiene formule di utilizzo Ethernet che prendono in considerazione il sovraccarico dei pacchetti. Tuttavia, il team per l'elevata disponibilità ha riscontrato che le formule generali qui presentate possono essere utilizzate in modo affidabile sia sulle interfacce LAN che WAN nella maggior parte dei casi.

[Pianificazione della capacità](#)

Come accennato in precedenza, la pianificazione della capacità è il processo in cui si determinano i probabili requisiti futuri delle risorse di rete per evitare un impatto sulle prestazioni o sulla disponibilità delle applicazioni business-critical. Per ulteriori informazioni, fare riferimento al documento sulla [gestione della capacità e delle prestazioni: White paper sulle procedure ottimali](#) per informazioni più dettagliate su questo argomento.

[Esecuzione di un'analisi proattiva degli errori](#)

L'analisi proattiva degli errori è essenziale per la gestione delle prestazioni. Lo stesso tipo di dati raccolti per la gestione delle prestazioni può essere utilizzato per l'analisi proattiva degli errori. Tuttavia, i tempi e l'utilizzo di questi dati sono diversi tra la gestione proattiva degli errori e la gestione delle prestazioni.

La gestione proattiva degli errori è il modo in cui il sistema di gestione di rete ideale può raggiungere gli obiettivi stabiliti. La relazione con la gestione delle prestazioni passa attraverso la baseline e le variabili di dati utilizzate. La gestione proattiva degli errori integra eventi personalizzati, un motore di correlazione degli eventi, l'emissione di ticket per i problemi e l'analisi statistica dei dati di base al fine di collegare la gestione degli errori, delle prestazioni e delle modifiche in un sistema di gestione di rete ideale ed efficace.

Se il polling dei dati sulle prestazioni viene in genere eseguito ogni 10, 15 o anche 30 minuti, il riconoscimento di una condizione di errore deve avvenire a un intervallo di tempo molto più breve. Un metodo per la gestione proattiva degli errori consiste nell'uso di gruppi di eventi e allarmi RMON. È possibile impostare soglie sui dispositivi che non vengono sottoposte a polling da dispositivi esterni, in modo da ridurre notevolmente le soglie. Un altro metodo, non descritto in questo documento, è l'utilizzo di un sistema di gestione distribuita che consente il polling a livello locale con l'aggregazione di dati a un responsabile dei responsabili.

[Utilizzo delle soglie per la gestione proattiva degli errori](#)

La soglia è il processo in cui si definiscono i punti di interesse in flussi di dati specifici e si generano eventi quando vengono attivate le soglie. Utilizzare i dati sulle prestazioni di rete per impostare tali soglie.

Esistono diversi tipi di soglie, alcune delle quali sono più applicabili a determinati tipi di dati. Le soglie sono applicabili solo ai dati numerici e convertono quindi i dati testuali in valori numerici discreti. Anche se non si conoscono tutte le possibili stringhe di testo per un oggetto, è comunque possibile enumerare le stringhe "interessanti" e assegnare tutte le altre stringhe a un valore impostato.

Esistono due classi di soglie per le due classi di dati numerici: *continuo* e *discreto*. Le soglie continue si applicano ai dati continui o delle serie temporali, come i dati memorizzati nei contatori o negli indicatori SNMP. Le soglie discrete si applicano agli oggetti enumerati o a qualsiasi dato numerico discreto. Gli oggetti booleani sono valori enumerati con due valori: true o false. I dati discreti possono anche essere chiamati dati di eventi poiché gli eventi contrassegnano la transizione da un valore a quello successivo.

Soglie continue possono attivare eventi quando l'oggetto serie temporale supera il valore specificato della soglia. Il valore dell'oggetto può essere superiore o inferiore alla soglia. Può inoltre essere utile fissare soglie di aumento e di diminuzione separate. Questa tecnica, nota come

meccanismo di isteresi, aiuta a ridurre il numero di eventi generati da questa classe di dati. Il meccanismo di isteresi consente di ridurre il volume degli eventi generati da soglie su dati di serie temporali che variano rapidamente. Questo meccanismo può essere utilizzato con qualsiasi tecnica di soglia sui dati delle serie temporali.

Il volume degli eventi viene ridotto da un allarme generato per tenere traccia del valore di un oggetto. A questo allarme sono assegnate soglie crescenti e decrescenti. L'allarme si attiva solo quando viene superata la soglia di aumento. Una volta superata questa soglia, non viene più generato un allarme crescente fino a quando la soglia in discesa non viene superata. E lo stesso meccanismo impedisce la generazione di soglie in calo fino a quando la soglia in aumento non viene nuovamente superata. Questo meccanismo può ridurre drasticamente il volume degli eventi e non elimina le informazioni necessarie per determinare se esiste un errore.

I dati delle serie temporali possono essere rappresentati come contatori, in cui ogni nuova coordinata viene aggiunta alla somma delle coordinate precedenti, oppure come contatore, in cui i dati vengono rappresentati come una frequenza su un intervallo di tempo. Esistono due diverse forme di soglie continue applicabili a ciascun tipo di dati: *soglie continue assolute* e *soglie continue relative*. Usare soglie continue assolute con contatori e soglie continue relative con contatori.

Per determinare i valori di soglia per la rete, attenersi alla seguente procedura:

1. Selezionare gli oggetti.
2. Selezionare i dispositivi e le interfacce.
3. Determinare i valori di soglia per ciascun oggetto o tipo di oggetto/interfaccia.
4. Determinare la gravità dell'evento generato da ciascuna soglia.

È necessario un notevole sforzo per determinare le soglie da utilizzare per determinati oggetti (e per quali dispositivi e interfacce). Fortunatamente, se avete raccolto una linea di base di dati sulle prestazioni, avete già svolto una quantità significativa di quel lavoro. NSA e il programma HAS (High Availability Service) possono inoltre fornire suggerimenti utili per l'impostazione degli oggetti e la creazione di intervalli. È tuttavia necessario personalizzare questi suggerimenti per la rete in uso.

Una volta raccolti i dati sulle prestazioni per la rete, il programma HAS consiglia di raggruppare le interfacce per categorie. Ciò semplifica l'impostazione delle soglie, in quanto potrebbe essere necessario determinare le soglie per il tipo di supporto di ogni categoria anziché per ogni dispositivo e oggetto su tale dispositivo. Ad esempio, si desidera impostare soglie diverse per le reti Ethernet e FDDI. Si ritiene in genere che sia possibile eseguire le reti FDDI con un utilizzo quasi del 100% rispetto a un segmento Ethernet condiviso. Tuttavia, l'Ethernet full-duplex può essere utilizzata al 100% più o meno vicino perché non è soggetta a collisioni. È possibile impostare le soglie delle collisioni su un valore molto basso per i collegamenti full-duplex perché non si dovrebbe mai verificare una collisione.

È inoltre possibile considerare la combinazione dell'importanza dell'interfaccia e della categoria/gravità del tipo di soglia. Utilizzare questi fattori per impostare la priorità dell'evento e, di conseguenza, l'importanza dell'evento e la sua attenzione da parte del personale operativo di rete.

Il raggruppamento e la categorizzazione dei dispositivi e delle interfacce di rete non possono essere sottolineati troppo. Maggiore è la possibilità di raggruppare e suddividere in categorie, più facile è integrare gli eventi di soglia nella piattaforma di gestione di rete. Utilizzare la baseline come risorsa principale per queste informazioni. Per ulteriori informazioni, fare riferimento al documento sulla [gestione della capacità e delle prestazioni: White paper sulle procedure ottimali](#) per ulteriori informazioni.

Implementazione della gestione della rete

L'organizzazione deve disporre di un sistema di gestione della rete implementato in grado di rilevare i valori di soglia definiti e di creare rapporti sui valori per periodi di tempo specifici. Utilizzare un sistema di gestione di rete RMON in grado di archiviare i messaggi di soglia in un file di log per la revisione giornaliera o una soluzione di database più completa che consenta di ricercare le eccezioni di soglia per un determinato parametro. Le informazioni dovrebbero essere costantemente a disposizione del personale e del gestore della rete. L'implementazione della gestione della rete deve includere la capacità di rilevare arresti anomali o traceback di software/hardware, affidabilità dell'interfaccia, utilizzo della CPU, utilizzo dei collegamenti, mancati riscontri nella coda o nel buffer, volume di trasmissione, transizioni vettore e reimpostazioni dell'interfaccia.

Metriche delle operazioni di rete

Un'ultima area di gestione proattiva degli errori che si sovrappone alla gestione delle prestazioni è la metrica delle operazioni di rete. Queste metriche forniscono dati preziosi per il miglioramento del processo di gestione degli errori. Come minimo, queste metriche dovrebbero includere una descrizione dettagliata di tutti i problemi che si sono verificati durante un determinato periodo. La ripartizione dovrebbe comprendere informazioni quali:

- Numero di problemi che si verificano per priorità di chiamata
- Tempo minimo, massimo e medio di chiusura per ogni priorità
- Suddivisione dei problemi per tipo (hardware, arresto anomalo del software, configurazione, alimentazione, errore dell'utente)
- Tempo di chiusura per ogni tipo di problema
- Disponibilità per gruppo di disponibilità o contratto di servizio
- Frequenza con cui sono stati soddisfatti o non soddisfatti i requisiti SLA

L'help desk spesso dispone di un sistema di creazione di rapporti con la capacità di generare metriche o rapporti. Un altro modo per raccogliere questi dati è l'uso di uno strumento di monitoraggio della disponibilità. Le metriche globali dovrebbero essere rese disponibili su base mensile. Il miglioramento dei processi basato sulla discussione deve essere implementato per migliorare i requisiti degli accordi sui livelli di servizio non rispettati o per migliorare la gestione di alcuni tipi di problemi.

Indicatori gestione prestazioni

Gli indicatori di prestazione forniscono il meccanismo mediante il quale un'organizzazione misura i fattori critici di successo.

Documentazione degli obiettivi aziendali di gestione della rete

Questo documento potrebbe essere un concetto formale di operazioni per la gestione della rete o una dichiarazione meno formale di caratteristiche e obiettivi richiesti. Tuttavia, il documento deve assistere il gestore della rete nella misura in cui misura il successo.

Questo documento rappresenta la strategia di gestione della rete dell'organizzazione e deve coordinare gli obiettivi aziendali generali (non quantitativi) di operazioni di rete, progettazione, altre business unit e gli utenti finali. Questo focus consente all'organizzazione di formare le attività di pianificazione a lungo termine per la gestione e il funzionamento della rete,

che include il processo di definizione del budget. Fornisce inoltre indicazioni per l'acquisizione di strumenti e il percorso di integrazione necessario per raggiungere gli obiettivi di gestione della rete, ad esempio gli SLA.

Questo documento strategico non può concentrarsi troppo strettamente sulla gestione di problemi specifici della rete, ma su quelle questioni importanti per l'organizzazione generale, che comprendono questioni di bilancio. Ad esempio:

- Identificare un piano completo con obiettivi raggiungibili.
- Identificare ogni servizio/applicazione aziendale che richiede supporto di rete.
- Identificare le metriche basate sulle prestazioni necessarie per misurare il servizio.
- Pianificare la raccolta e la distribuzione dei dati delle metriche delle prestazioni.
- Identificare il supporto necessario per la valutazione della rete e il feedback degli utenti.
- Garantire livelli di servizio documentati, dettagliati e misurabili.

[Documentazione degli accordi sui livelli di servizio](#)

Per documentare correttamente gli SLA, è necessario definire completamente le metriche degli obiettivi dei livelli di servizio. Tale documentazione deve essere disponibile per la valutazione. Fornisce il ciclo di feedback per garantire che l'organizzazione di gestione della rete continui a misurare le variabili necessarie per mantenere il livello di contratto di assistenza.

Gli accordi sui livelli di servizio (SLA) sono documenti "attivi" in quanto il contesto aziendale e la rete sono dinamici per natura. Ciò che funziona oggi per misurare uno SLA potrebbe diventare obsoleto domani. Le operazioni di rete possono mantenere l'elevata disponibilità richiesta dall'organizzazione solo quando gli utenti stabiliscono un ciclo di feedback e agiscono su tali informazioni.

[Creare un elenco di variabili per la baseline](#)

Questo elenco include elementi quali l'intervallo di polling, il sovraccarico di gestione della rete sostenuto, le possibili soglie di attivazione, se la variabile è utilizzata come trigger per una trap e l'analisi delle tendenze utilizzata per ciascuna variabile.

Queste variabili non si limitano alle metriche necessarie per gli obiettivi dei livelli di servizio sopra menzionati. Esse devono includere almeno le seguenti variabili: stato del router, stato dello switch, informazioni di routing, dati specifici della tecnologia, utilizzo e ritardo. Queste variabili vengono sottoposte periodicamente a polling e memorizzate in un database. È quindi possibile generare rapporti in base a questi dati. Questi report possono essere utili per le operazioni di gestione della rete e per il personale di pianificazione nei modi seguenti:

- I problemi reattivi possono spesso essere risolti più rapidamente con un database cronologico.
- Questo tipo di dati è necessario per la creazione di rapporti sulle prestazioni e la pianificazione della capacità.
- Gli obiettivi dei livelli di servizio possono essere misurati in base ad esso.

[Analizzare le analisi delle baseline e delle tendenze](#)

Il personale responsabile della gestione della rete dovrebbe tenere riunioni per esaminare

periodicamente relazioni specifiche. Questo fornisce un ulteriore feedback, oltre a un approccio proattivo ai potenziali problemi della rete.

Tali riunioni dovrebbero includere personale operativo e di pianificazione. In questo modo i responsabili della pianificazione possono ricevere un'analisi operativa dei dati di base e delle tendenze. Inoltre, il personale operativo è "in loop" per alcune analisi di pianificazione.

Un altro tipo di elemento da includere in queste riunioni sono gli obiettivi dei livelli di servizio. Con l'avvicinarsi di soglie oggettive, il personale addetto alla gestione della rete può intervenire per evitare di perdere un obiettivo e, in alcuni casi, questi dati possono essere utilizzati come giustificazione di bilancio parziale. I dati possono indicare dove vanno violati gli obiettivi dei livelli di servizio se non vengono adottate misure adeguate. Inoltre, poiché tali obiettivi sono stati individuati dai servizi e dalle applicazioni aziendali, è più facile giustificarli su base finanziaria.

Effettuare queste revisioni ogni due settimane e tenere una riunione analitica più approfondita ogni sei-dodici settimane. Queste riunioni consentono di affrontare problemi sia a breve che a lungo termine.

[Documentazione di una metodologia di analisi di simulazione](#)

Un'analisi di simulazione prevede la modellazione e la verifica delle soluzioni. Prima di aggiungere una nuova soluzione alla rete (una nuova applicazione o una modifica nella versione Cisco IOS), documentare alcune delle alternative.

La documentazione per questa analisi include le domande principali, la metodologia, i set di dati e i file di configurazione. Il punto principale è che l'analisi di simulazione è un esperimento che qualcun altro dovrebbe essere in grado di ricreare con le informazioni fornite nel documento.

[Documentazione della metodologia utilizzata per aumentare le prestazioni della rete](#)

Questa documentazione include una larghezza di banda WAN aggiuntiva e una tabella dei costi che consente di aumentare la larghezza di banda per un particolare tipo di collegamento. Queste informazioni aiutano l'organizzazione a comprendere quanto tempo e denaro costa aumentare la larghezza di banda. La documentazione formale consente agli esperti di prestazioni e capacità di scoprire come e quando aumentare le prestazioni, nonché la tempistica e i costi di un'attività di questo tipo.

Esaminare periodicamente la documentazione, ad esempio nell'ambito della verifica trimestrale dei risultati, per assicurarsi che rimanga aggiornata.

[Riepilogo](#)

L'unico modo per raggiungere gli obiettivi del sistema di gestione di rete ideale è integrare attivamente i componenti della gestione delle prestazioni nel sistema. Questo obiettivo deve includere l'uso di metriche di disponibilità e tempi di risposta legate a un sistema di notifica quando le soglie vengono superate. Dovrebbe includere l'uso di una base per la pianificazione della capacità che abbia collegamenti a un modello euristico per il provisioning e la segnalazione delle eccezioni. Potrebbe disporre di un motore di simulazione o di modellazione incorporato che consente di aggiornare il modello in tempo reale e fornire un livello di pianificazione e risoluzione dei problemi tramite simulazioni software.

Anche se gran parte di questo sistema potrebbe sembrare un ideale impossibile da realizzare, ognuno dei componenti è attualmente disponibile. Inoltre, gli strumenti per integrare questi componenti esistono anche in programmi come MicroMuse. Dovremmo continuare a lavorare per questo ideale, che oggi è più realistico che mai.

[Informazioni correlate](#)

- [White paper sulle tecnologie ad alta disponibilità](#)
- [White paper sulla gestione di capacità e prestazioni](#)
- [Documentazione e supporto tecnico – Cisco Systems](#)