CLOUDFLARE

WHITE PAPER

Resilienza della rete e dei servizi Cloudflare

Indice

Panoramica

Internet è basato su sistemi imperfetti progettati per dare priorità ai tempi di attività rispetto a tutto il resto. Protocolli come TCP e BGP si basano sui principi dei sistemi distribuiti: prevedere i guasti e gestirli, e Internet è stato progettato di conseguenza. Tuttavia, i punti di guasto esistono ancora e possono causare un impatto. I provider di rete devono essere in grado di pianificare i guasti, rilevarli quando si verificano e mitigare l'impatto subito dagli utenti.

A causa della natura in tempo reale di molte applicazioni su Internet, le reti devono non solo coprire la maggior parte possibile di Internet, ma devono anche rispondere e mitigare l'impatto in tempo reale. I tempi di inattività nell'ordine di minuti non sono accettabili: i clienti si aspettano che la risoluzione avvenga in pochi secondi.

Cloudflare fornisce servizi di infrastruttura di rete critici per supportare il modo in cui le organizzazioni proteggono e comunicano su Internet. Abbiamo sviluppato la nostra rete e i servizi che fornisce per mantenere il massimo livello di eccellenza operativa. Cloudflare elabora in media oltre 84 milioni di richieste HTTP e 61 milioni di query DNS al secondo, fornendo servizi a milioni di proprietà Internet e utenti.

In che modo Cloudflare fornisce un servizio affidabile su questa scala, date le caratteristiche imprevedibili di Internet?

La risposta viene dall'architettura della rete Cloudflare, che opera con capacità di resilienza progettate per funzionare in modo indipendente e resistere allo spettro delle interruzioni. Le nostre capacità di elaborazione, rete e storage, insieme ai nostri processi operativi, sono progettate per rendere Cloudflare affidabile quanto il segnale di linea della rete telefonica tradizionale. Cloudflare offre il metaforico "cloud tone" per i servizi di rete e sicurezza su cui i clienti fanno affidamento, nonostante le condizioni di Internet in un dato momento.

Questo white paper approfondisce le sfide dell'operare in tempo reale su Internet e come Cloudflare è posizionata in modo univoco per risolverle.

Vivere in un mondo imperfetto

Internet è un luogo imperfetto, eppure le organizzazioni ne hanno bisogno per costruire la propria attività e gestire organizzazioni che collegano utenti, dati e dispositivi distribuiti alle applicazioni nel cloud. Qualsiasi interruzione del servizio ha gravi implicazioni. Nel corso degli anni, Cloudflare ha segnalato una serie di importanti interruzioni dei servizi Internet in tutto il mondo, che vanno dagli incidenti agli attacchi DDoS alle calamità naturali e altro ancora.

Internet è una rete diversificata che è stata realizzata come un collettivo vagamente accoppiato di migliaia di reti partecipanti di varie dimensioni e capacità. Queste reti hanno diversi livelli di servizio, con alcune che operano al meglio delle loro possibilità a titolo di buona volontà, e altre che operano come parte dei loro servizi commerciali. Attraverso accordi reciproci per lo scambio di traffico, che può o meno essere destinato agli host sulla propria rete, Internet funziona come un tessuto globale grazie alla benevola partecipazione dei punti di scambio Internet e dei fornitori di servizi regionali e locali.

Tuttavia, con tale diversità arriva un certo grado di imprevedibilità. Il percorso intrapreso da un determinato pacchetto si basa su una sequenza di stime e tentativi per la consegna. In assenza di dati in tempo reale o di modelli predittivi sulle condizioni di rete effettive in un dato momento, un pacchetto in genere effettua il suo salto successivo utilizzando percorsi predefiniti o il percorso presunto più breve. Nessuna di queste opzioni tiene conto dello stato del servizio di rete nelle sue decisioni di routing. Ciò significa che i pacchetti sono spesso soggetti a una serie di condizioni debilitanti:

- Al livello più elementare, le reti possono riscontrare problemi temporanei e brownout che riducono il throughput e causano la perdita di pacchetti.
- Man mano che le reti si saturano, la congestione danneggia le prestazioni e l'esperienza dell'utente.
- Sebbene questi tipi di rallentamenti possano verificarsi in normali condizioni operative, un numero crescente di interruzioni è causato intenzionalmente da persone ostili, che consumano le risorse disponibili in modo ostile.

In Cloudflare, la nostra missione è contribuire a realizzare un Internet migliore. A tal fine, creiamo un'infrastruttura per rendere Internet più veloce, più affidabile e più sicuro. Ciò è reso possibile dai servizi da e attraverso la rete Cloudflare, una delle reti più grandi al mondo che opera sui nostri server bare metal (senza virtualizzazione), una dorsale privata e un'enorme impronta globale raggiungibile, in media, entro 50 ms dal 95% della popolazione mondiale.[1]

I clienti spesso vogliono capire come questo venga fatto. In che modo Cloudflare progetta servizi su tale scala, senza compromessi tra sicurezza e prestazioni e con l'affidabilità di cui le organizzazioni hanno bisogno? La risposta viene da una serie di obiettivi North Star, che stabiliscono come realizziamo la nostra infrastruttura per rimanere performante nonostante le forze contrarie, sia esterne che interne.

Come Cloudflare progetta tenendo conto della resilienza

Molte organizzazioni si concentrano sul miglioramento della disponibilità cercando di reagire meglio o più rapidamente ai guasti. Questo richiede di investire e testare costantemente le loro capacità e processi di failover. Sebbene questi siano obiettivi validi, Cloudflare la pensa in modo diverso: i nostri team di ingegneri della resilienza dedicano enormi sforzi per ridurre gli scenari in cui è necessaria una risposta di disaster recovery.

L'ingegneria della resilienza di Cloudflare inizia con una semplice premessa: **come realizzare un'infrastruttura critica che rimanga operativa supponendo che si verifichino guasti?**

Quando i guasti si verificano inevitabilmente, i servizi resilienti di Cloudflare rilevano e isolano i guasti in modo che non influiscano sulla disponibilità dei servizi. I guasti vengono risolti fuori banda dalla nostra erogazione di servizi. Ci impegniamo a essere indipendenti dai guasti in tutta la nostra flotta.

Per comprendere i principi della resilienza, è utile definire prima i concetti chiave alla base dei percorsi di traffico con Cloudflare e gli obiettivi di progettazione per i sistemi chiave. Al livello più astratto, l'architettura di Cloudflare può essere suddivisa in due segmenti: **il piano di controllo e il piano dati.** Ognuno ha una resilienza unica e un assetto di fronte alle calamità.

Resilienza del piano di controllo

Il piano di controllo fornisce l'interfaccia di gestione che stabilisce la fonte di verità per la configurazione dei servizi di rete e di sicurezza nell'ambiente del cliente. Il piano di controllo stesso non elabora il traffico (che è il ruolo del piano dati). Indica al piano dati quali criteri applicare e gestisce le configurazioni tra diversi datacenter.

I servizi del piano di controllo di Cloudflare sono generalmente distribuiti in una topografia centralizzata più tradizionale su tre datacenter logicamente correlati ma indipendenti in un'area geografica primaria (ad esempio, negli Stati Uniti). Questi tre datacenter vengono replicati con capacità equivalente in un'area secondaria (ad esempio, UE). I servizi del piano di controllo sono progettati per essere resilienti e mantenere un'erogazione coerente dei servizi in caso di guasto di un singolo datacenter nella sede principale. La perdita di ulteriori datacenter nella sede principale attiverebbe un failover verso i datacenter dell'altra area geografica (ad esempio, in Europa rispetto agli Stati Uniti).

In linea con l'attenzione di Cloudflare sulla resilienza prima del ripristino, continuiamo a investire per rafforzare il nostro stato di resilienza.

Ad esempio:

- Utilizziamo sempre più spesso le due aree geografiche del piano di controllo in una configurazione active-active, che aumenta contemporaneamente la nostra capacità/reattività, nonché la nostra tolleranza ai guasti. Di conseguenza, possiamo resistere a più tipi di guasti senza interruzioni dei servizi o necessità di failover.
- Stiamo anche aumentando la granularità del modo in cui spostiamo i servizi tra le varie sedi del piano di controllo, consentendoci di rispondere con maggiore precisione ai problemi dell'infrastruttura locale.

Test del caos

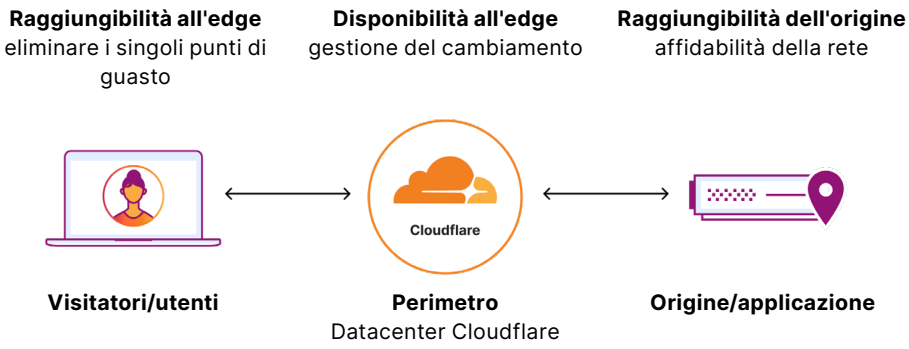
La resilienza non è un'attività da impostare e dimenticare. Anche i piani di resilienza più solidi possono rivelarsi inefficaci a causa della “deriva” del sistema: il lento, spesso impercettibile accumulo di modifiche che possono degradare i comportamenti previsti e introdurre modalità di guasto impreviste. Per affrontare in modo proattivo questo rischio, Cloudflare esegue regolarmente test di caos, sondando sistematicamente potenziali vulnerabilità legate alla deriva.

Resilienza del piano dati

Il piano dati elabora il traffico dei clienti Cloudflare in base ai criteri stabiliti dal piano di controllo. Sebbene i servizi del piano dati prendano la direzione dal piano di controllo, non dipende dal piano di controllo per funzionare. Tutti i criteri vengono mantenuti tramite Quicksilver, il nostro archivio chiave-valore distribuito a livello globale, per garantire che i servizi rimangano operativi con una configurazione nota e funzionante in caso di interruzione della comunicazione con il piano di controllo.

Anycast svolge un ruolo importante nella ridondanza dei datacenter. I datacenter Cloudflare, situati in oltre 330 città, sono autonomi in locale e tuttavia intercambiabili tra loro tramite Anycast e BGP. Questo perché ogni datacenter può elaborare in locale qualsiasi servizio, senza essere codipendente dai servizi in un altro datacenter. Con Anycast, ogni datacenter è effettivamente ridondante con gli altri. Poiché ogni datacenter partecipa ad Anycast, non è necessario indicare al client di passare a un datacenter alternativo a un altro indirizzo IP.

Che si tratti di un consumatore che visita un sito web protetto da Cloudflare, di un dipendente che accede ad app connesse a Internet o di un ufficio che si connette alla propria WAN, tutti questi scenari utilizzano BGP per trovare il datacenter Cloudflare Anycast più vicino. Se il datacenter scelto non dovesse più essere disponibile, BGP si sposterebbe automaticamente al miglior datacenter Cloudflare successivo.



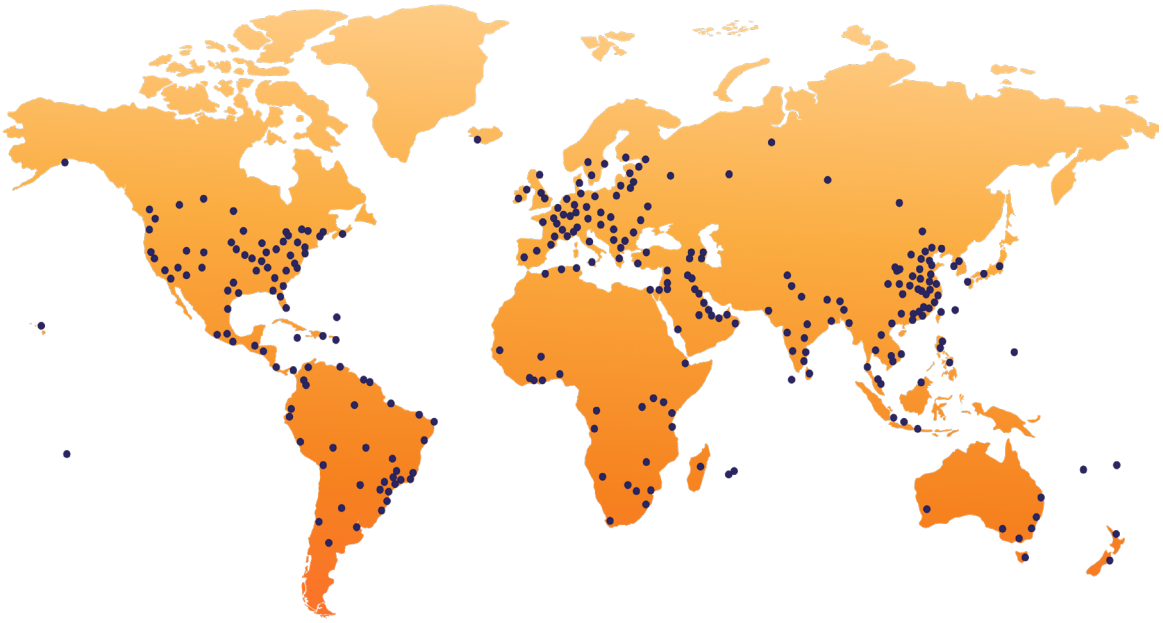
Cloudflare affronta la resilienza del piano dati risolvendo tre diversi problemi:

- Raggiungibilità all'edge: garantire che il traffico degli utenti finali possa raggiungere i datacenter ed eliminare i singoli punti di guasto all'ingresso del traffico
- Disponibilità all'edge: mantenimento della qualità del codice e dell'uptime del software attraverso una rigorosa gestione delle modifiche
- Raggiungibilità dell'origine: routing adattivo alle applicazioni dei clienti per garantire che non vi siano perdite sui percorsi in uscita

Raggiungibilità all'edge

La raggiungibilità all'edge è la capacità degli utenti finali di raggiungere la rete di Cloudflare. È probabilmente l'aspetto più importante del problema della resilienza. Se i provider di servizi Internet (ISP) o i datacenter non funzionano, la raggiungibilità all'edge viene ridotta o degradata, impedendo o rallentando gli utenti di arrivare dove devono essere su Internet. Cloudflare risolve i problemi di raggiungibilità all'edge in quattro modi principali:

1. Rete Anycast



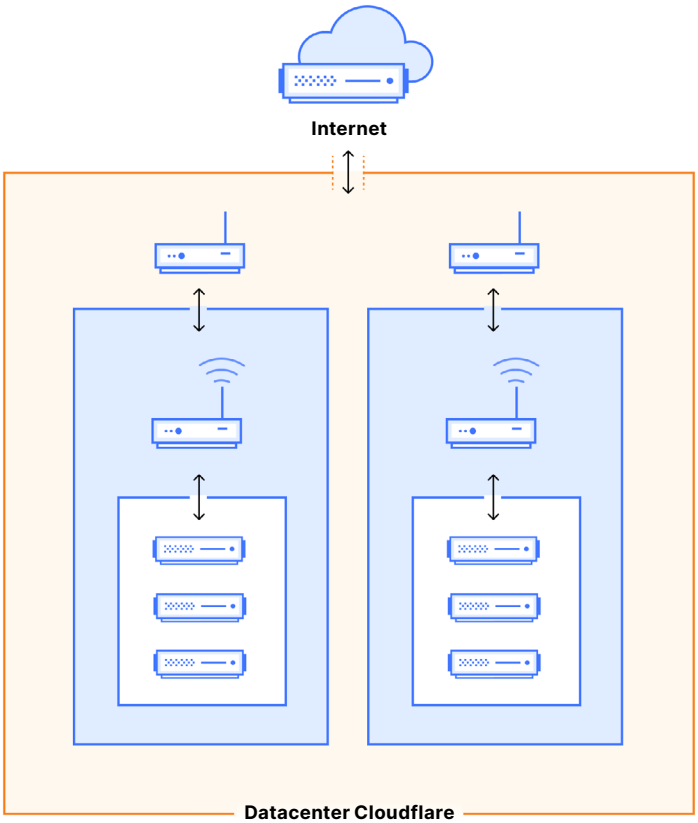
Migliori prestazioni di rete con resilienza sono integrate nella nostra architettura di rete, solidamente basata sulla tecnologia Anycast. Anycast, la superpotenza ingegneristica di Cloudflare, significa che lo spazio IP è pubblicizzato ovunque: se uno qualsiasi dei nostri punti di presenza (PoP) è offline, il traffico viene semplicemente instradato verso altre posizioni anziché essere interrotto. Il fatto che Cloudflare sia presente in così tante città in tutto il mondo e sia in peering con reti locali significa che elaboriamo il traffico dei clienti il più vicino possibile all'utente. Ad esempio, anche se una connessione del provider di servizi Internet viene disconnessa o un datacenter si disconnette dalla rete elettrica, il traffico rimane inalterato.

2. Tutti i servizi Cloudflare sono erogati da tutti i punti di presenza

Oltre ad Anycast, i datacenter Cloudflare sono progettati in modo da elaborare il traffico in locale, senza essere codipendenti dalle catene proxy verso altre risorse di elaborazione. Eseguiamo quasi ogni servizio su ogni macchina. Ciò significa che un datacenter può essere portato offline senza alcun impatto sul cliente. Le macchine all'interno del datacenter sono sostituibili l'una con l'altra perché ce ne sono centinaia che eseguono lo stesso servizio che sono in grado di intervenire, nel caso in cui una si guasti.

3. Punti di presenza in più colocation

La progettazione e le posizioni dei datacenter di Cloudflare riflettono le esigenze dei nostri clienti. Una rete Anycast ci consente di aggiungere/rimuovere datacenter a piacimento, ma dobbiamo monitorare costantemente le prestazioni dei clienti. Abbiamo adattato le nostre topologie di datacenter per consentire a sezioni di capacità di elaborazione (colocation) di guastarsi indipendentemente l'una dall'altra. Questi datacenter (con più colocation) sono chiamati punti di presenza in più colocation o posizioni MCP.



Queste posizioni separano la connettività rivolta a Internet dalla connettività di elaborazione interna per consentire di portare offline individualmente le strutture di colocation. Ciò significa che anche se si verifica un problema con una colocation, l'intero PoP in un'area geografica può rimanere online, fornendo tempi di attività e prestazioni maggiori a un cliente. Questo tipo di datacenter rimuove anche i singoli punti di guasto con dispositivi ridondanti nel livello rivolto a Internet: se un router rivolto a Internet (edge) si guasta, l'altro router all'edge può prendere il traffico e garantire che la posizione rimanga operativa.

Questo modello operativo aiuta le sedi MCP a evitare lo spostamento del traffico a meno che non sia assolutamente necessario e aumenta ulteriormente i tempi di attività dei clienti di Cloudflare.

4. Cloudflare Traffic Manager

I datacenter MCP collaborano per formare la più ampia rete Cloudflare. Questa rete sfrutta Anycast per garantire che il traffico dei clienti venga servito. Anycast è migliorato con una gestione deterministica del traffico per garantire che le richieste dei clienti vengano servite dove possiamo servirle, con le migliori prestazioni possibili. Questo sistema di gestione del traffico, Traffic Manager, funziona sondando continuamente la rete di Cloudflare e spostando automaticamente il traffico lontano dai datacenter che subiscono un sovraccarico della CPU. Ciò impedisce la congestione nei datacenter ad alto traffico; invece, il traffico viene instradato in modo intelligente verso un altro datacenter in grado di gestirlo.

Disponibilità all'edge

La disponibilità all'edge si riferisce alla capacità di Cloudflare di elaborare il traffico una volta che raggiunge la nostra rete. Quando le modifiche agli strumenti di rete o al software determinano modifiche non intenzionali, la disponibilità può diminuire e influire sull'esperienza dell'utente. Per evitare che si verifichino incidenti derivanti dalla modifica del codice, Cloudflare ha investito molto nei seguenti controlli di distribuzione:

1. Funnel di distribuzione

Quando si distribuisce il software, garantire la qualità del codice inizia monitorando e limitando la capacità di sviluppatori e clienti di introdurre modifiche nell'ecosistema. Cloudflare limita il numero di modi in cui chiunque può introdurre cambiamenti nella nostra infrastruttura in modo da poter monitorare da vicino ogni cambiamento e garantire che superino una serie di test prima di essere distribuiti in produzione.

2. Gestione delle modifiche del raggio di esplosione

Un altro modo in cui Cloudflare supporta la disponibilità all'edge durante la distribuzione consiste nel limitare le distribuzioni ai datacenter o ai gruppi di test prima di implementarli ampiamente. Ci riferiamo a questo come gestione del raggio di esplosione.

Quando le modifiche alla rete vengono implementate, possono essere applicate a livello globale in pochi minuti. Contenendo le modifiche a un ambiente di distribuzione e implementando ulteriori modifiche in una cascata, possiamo monitorare gli effetti della modifica per conseguenze previste o non intenzionali, prima di interessare aree geografiche o popolazioni di utenti più grandi.

Abbiamo due modi per limitare l'impatto delle modifiche al codice:

- Limitare il numero di sedi che ricevono le modifiche; e
- Limitare il numero di utenti che ricevono le modifiche

Limitando il numero di posizioni e macchine che ricevono le modifiche, possiamo assicurarci di eseguire correttamente test A/B sul codice all'interno di una singola posizione per valutarne lo stato prima di procedere. Limitando il numero di utenti che ricevono le modifiche, possiamo prima testare a fondo le correzioni su popolazioni di utenti più piccole.

3. Distribuzione mediata dall'integrità

La distribuzione mediata dall'integrità è un sistema che valuta in modo programmatico l'idoneità di una versione in base a metriche preimpostate che forniscono un segnale “go” o “no go” in base al potenziale impatto. Questa serie di controlli automatizzati non solo può impedire l'uscita di un rilascio dannoso, ma può anche annullare un rilascio al rilevamento dell'impatto.

Ogni prodotto e servizio distribuito tramite Cloudflare deve avere un obiettivo del livello di servizio (SLO), che contiene sia una metrica che rappresenta l'integrità del prodotto, sia un obiettivo al di sotto del quale un prodotto sarebbe considerato non integro.

Gli SLO hanno tassi di consumo, o soglie di guasto accettabili. Qualsiasi servizio mediato dall'integrità fornirà SLO a un sistema automatizzato come parte dell'unione di una modifica da distribuire. In ogni ambito di distribuzione (piani gratuiti, un sottoinsieme di macchine in Ashburn, ecc.), il sistema automatizzato:

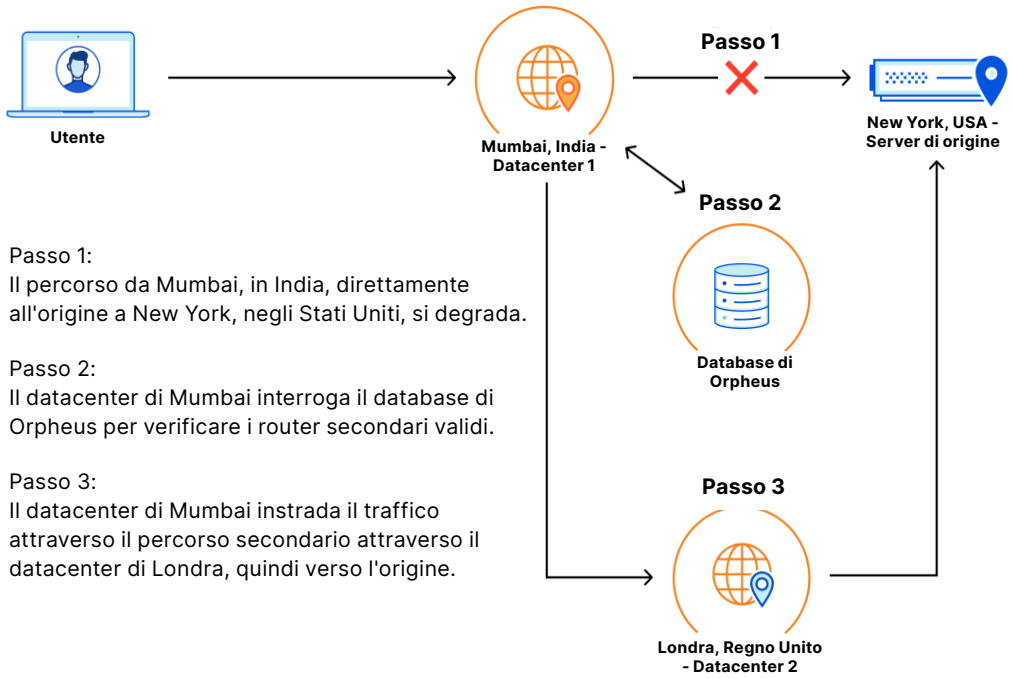
- Innanzitutto, monitorerà lo SLO del servizio per un determinato periodo di tempo per garantire che l'integrità non scenda al di sotto della soglia.
- Se lo SLO rimane entro intervalli accettabili per la durata del periodo di sospensione impostato, il sistema farà avanzare automaticamente la distribuzione a una fase più ampia.
- Tuttavia, se lo SLO viene violato, il sistema interromperà automaticamente la distribuzione ed eseguirà il rollback in modo che l'impatto venga mitigato automaticamente.

Questi passaggi garantiscono che l'integrità del cliente non venga influenzata dalle distribuzioni del codice e che la durata sia la più breve possibile.

Raggiungibilità dell'origine

La raggiungibilità dell'origine si riferisce alla capacità di Cloudflare di raggiungere le destinazioni, che si tratti di un'origine cliente, di un sito SaaS o dell'Internet pubblico tramite Cloudflare Gateway. Il routing di richieste dove è necessario che siano è fondamentale per gli utenti che accedono alla rete di Cloudflare. Ad esempio, Argo Smart Routing, lo strumento di Cloudflare che ottimizza le prestazioni (ovvero il Time to First Byte), sonda costantemente la rete di Cloudflare per trovare il percorso più veloce verso le origini. Argo sonda costantemente la rete di Cloudflare, trovando il percorso più veloce verso le origini.

Orpheus, la controparte di Argo, ha una filosofia simile ma una funzione diversa. Orpheus è stato creato per stabilire connessioni affidabili ai server di origine. Orpheus esamina specificamente i parametri che influiscono sulla capacità di Cloudflare di raggiungere l'origine (invece del percorso più veloce verso l'origine) e troverà percorsi che riducono al minimo la perdita di pacchetti senza influire sulle prestazioni dello stato stazionario. Ciò significa che quando si verificano problemi, il traffico viene automaticamente instradato attorno ai guasti rilevati.



Prima che Cloudflare rilasciasse Orpheus nel 2021, eravamo in grado di instradare con successo alle origini il 99,9% delle volte. Dopo aver implementato Orpheus, la nostra capacità di instradare verso le origini è salita al 99,99%. Nel prossimo anno, espanderemo Orpheus per proteggere più tipi di traffico, agire su più scenari di guasto e lavorare più velocemente per ridurre il tempo di impatto per qualsiasi utente.

Impegno per la trasparenza operativa

Anche le reti più resilienti e innovative subiranno interruzioni. Quando si verificano incidenti e i clienti vengono colpiti, Cloudflare segue una risposta di comunicazione agli incidenti che include un'indagine approfondita, un report interno sull'incidente, un report sull'incidente esterno e, se necessario, aggiornamenti di stato durante la finestra di impatto.

In alcuni casi in cui gli incidenti provocano tale impatto (o innovazione), i post mortem verranno pubblicati sul Blog di Cloudflare.

Conclusioni

Lo sforzo ingegneristico alla base della rete di Cloudflare non è poco, ma è un lavoro che svolgiamo con orgoglio per i nostri clienti. La ricompensa è la creazione di una piattaforma di rete a vantaggio dei nostri clienti e della più ampia community Internet nel suo insieme.

In definitiva, dando la priorità alla resilienza non solo come problema tecnico ma come filosofia operativa fondamentale, stiamo facendo molto di più che rafforzare le difese: stiamo costruendo attivamente un'azienda intrinsecamente predisposta per il futuro. L'approccio proattivo di test e adattamento continui significa che possiamo evolvere con agilità in base alle esigenze in continua evoluzione dei nostri clienti e del panorama dinamico di Internet stesso.

Comprendiamo che molti dei concetti rappresentati in questo documento sono incentrati su concetti di rete a cui molte aziende non sono esposte, poiché coinvolgono l'architettura interna del funzionamento di un ambiente cloud globale di classe carrier.

Se desideri un briefing approfondito per saperne di più sull'ingegneria della resilienza di Cloudflare, contatta il tuo rappresentante Cloudflare.

+44 20 3514 6970 | enterprise@cloudflare.com | www.cloudflare.com/it-it/

REV:BDES-8251.2025OCT28