



Guida per l'utente dei piani di scalabilità

AWS Auto Scaling



AWS Auto Scaling: Guida per l'utente dei piani di scalabilità

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

I marchi e l'immagine commerciale di Amazon non possono essere utilizzati in relazione a prodotti o servizi che non siano di Amazon, in una qualsiasi modalità che possa causare confusione tra i clienti o in una qualsiasi modalità che denigri o discrediti Amazon. Tutti gli altri marchi non di proprietà di Amazon sono di proprietà delle rispettive aziende, che possono o meno essere associate, collegate o sponsorizzate da Amazon.

Table of Contents

Cos'è un piano di dimensionamento?	1
Risorse supportate	1
Caratteristiche e vantaggi del piano di dimensionamento	1
Come iniziare	2
Come funzionano i piani di dimensionamento	2
Disponibilità regionale	3
Prezzi	3
Come funzionano i piani di dimensionamento	5
Best practice	8
Altre considerazioni	8
Evitare l'errore ActiveWithProblems	10
Nozioni di base	11
Fase 1: individuare le risorse scalabili	12
Prerequisiti	12
Aggiungi il gruppo Auto Scaling al nuovo piano di dimensionamento	12
Ottieni maggiori informazioni su come individuare le tue risorse scalabili	14
Fase 2: specificare la strategia di dimensionamento	16
Fase 3: configurare le impostazioni avanzate (facoltativo)	19
Impostazioni generali	19
Impostazioni dimensionamento dinamico	22
Impostazioni dimensionamento predittivo	22
Fase 4: creare il piano di dimensionamento	24
(Facoltativo) Visualizzare le informazioni di dimensionamento per una risorsa	24
Fase 5: Pulizia	27
Eliminare il gruppo Auto Scaling	28
Passaggio 6: fasi successive	28
Sicurezza	30
Endpoint VPC (AWS PrivateLink)	30
Crea un endpoint VPC dell'interfaccia per i piani di dimensionamento	31
Crea una policy dell'endpoint VPC per i piani di dimensionamento	31
Migrazione degli endpoint	32
Protezione dei dati	33
Identity and Access Management	34
Controllo accessi	34

Come funzionano i piani di dimensionamento con IAM	35
Ruoli collegati ai servizi	38
Esempi di policy basate su identità	40
Convalida della conformità	46
Sicurezza dell'infrastruttura	48
Quote	49
Informazioni correlate	50
Cronologia dei documenti	53
.....	lvi

Cos'è un piano di dimensionamento?

Utilizza un piano di dimensionamento per configurare la scalabilità automatica per risorse scalabili correlate o associate in pochi minuti. Ad esempio, è possibile utilizzare i tag per raggruppare risorse in categorie come produzione, test o sviluppo. Quindi, è possibile cercare e impostare piani di dimensionamento per risorse scalabili appartenenti a ciascuna categoria. Oppure, se la tua infrastruttura cloud lo include AWS CloudFormation, puoi definire modelli di stack da utilizzare per creare raccolte di risorse. Quindi, creare un piano di dimensionamento per le risorse scalabili appartenenti a ogni stack.

Risorse supportate

AWS Auto Scaling supporta l'uso di piani di scalabilità per i seguenti servizi e risorse:

- Amazon Aurora: consente di aumentare o diminuire il numero di repliche di lettura Aurora assegnate a un cluster database Aurora.
- Amazon EC2 Auto Scaling: consente di avviare o terminare le istanze EC2 aumentando o diminuendo la capacità desiderata di un gruppo Auto Scaling.
- Amazon Elastic Container Service: consente di aumentare o diminuire il numero di attività desiderato in Amazon ECS.
- Amazon DynamoDB: consente di aumentare o diminuire la capacità di lettura e scrittura assegnata alla tabella Dynamo DB o a un indice secondario globale.
- Parco istanze Spot: avvia o termina le istanze EC2 aumentando o diminuendo la capacità target di un parco istanze Spot.

Caratteristiche e vantaggi del piano di dimensionamento

I piani di dimensionamento offrono le seguenti caratteristiche e vantaggi:

- Rilevamento delle risorse: AWS Auto Scaling fornisce l'individuazione automatica delle risorse per aiutare a trovare risorse scalabili nell'applicazione.
- Dimensionamento dinamico: i piani di dimensionamento utilizzano i servizi Amazon EC2 Auto Scaling e Application Auto Scaling per regolare la capacità delle risorse scalabili per gestire le modifiche del traffico o del carico di lavoro. I parametri di dimensionamento dinamico possono essere parametri di utilizzo o throughput standard o personalizzati.

- Consigli di dimensionamento integrati: AWS Auto Scaling fornisce strategie di dimensionamento con suggerimenti che è possibile utilizzare per ottimizzare le prestazioni, i costi o un equilibrio tra i due.
- Dimensionamento predittivo: i piani di dimensionamento supportano anche il dimensionamento predittivo per i gruppi Auto Scaling. Ciò aiuta a scalare la capacità Amazon EC2 più velocemente quando si verificano regolarmente picchi.

Important

Se si utilizzano piani di dimensionamento solo per il dimensionamento predittivo, si consiglia vivamente di utilizzare invece le policy di dimensionamento predittivo direttamente sui gruppi con dimensionamento automatico. Questa opzione introdotta di recente offre maggiori funzionalità, come l'utilizzo di aggregazioni di parametri per crearne di nuovi e personalizzati o conservare i dati cronologici dei parametri nelle distribuzioni blu/verdi. Per ulteriori informazioni, consulta [Dimensionamento predittivo per Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

Come iniziare

Utilizza le risorse seguenti per creare e utilizzare un piano di dimensionamento:

- [Come funzionano i piani di dimensionamento](#)
- [Best practice per i piani di dimensionamento](#)
- [Nozioni di base sui piani di dimensionamento](#)

Come funzionano i piani di dimensionamento

È possibile creare, accedere e gestire i piani di dimensionamento utilizzando una qualsiasi delle seguenti interfacce:

- AWS Management Console: fornisce un'interfaccia Web da utilizzare per l'accesso ai piani di dimensionamento. Se ti sei registrato a un Account AWS, puoi accedere ai tuoi piani di scalabilità accedendo a AWS Management Console, utilizzando la casella di ricerca sulla barra di navigazione per cercare e quindi AWS Auto Scaling selezionando. AWS Auto Scaling

- **AWS Command Line Interface (AWS CLI)** — Fornisce comandi per un ampio set di Servizi AWS ed è supportato su Windows, macOS e Linux. Per iniziare, consulta la [\(Guida per l'utente di AWS Command Line Interface\)](#). Per ulteriori informazioni, consulta [piani di dimensionamento automatico](#) nella Documentazione di riferimento sui comandi AWS CLI .
- **AWS Tools for Windows PowerShell**— Fornisce comandi per un'ampia gamma di AWS prodotti per coloro che eseguono script nell' PowerShell ambiente. Per iniziare, consulta la [Guida per l'utente di AWS Tools for Windows PowerShell](#). Per ulteriori informazioni, consulta la [Documentazione di riferimento per Cmdlet AWS Tools for PowerShell](#).
- **AWS SDK**: forniscono operazioni API specifiche per la lingua e si occupano di molti dettagli di connessione, come il calcolo delle firme, la gestione dei tentativi di richiesta e la gestione degli errori. Per ulteriori informazioni, consulta [SDK di AWS](#).
- **API HTTPS**: forniscono operazioni API di basso livello accessibili tramite richieste HTTPS. Per ulteriori informazioni, consulta la [Documentazione di riferimento delle API di AWS Auto Scaling](#).
- **AWS CloudFormation**— Supporta la creazione di piani di scalabilità utilizzando modelli. CloudFormation Per ulteriori informazioni, consulta il [AWS::AutoScalingPlans::ScalingPlan](#) riferimento nella Guida per l'AWS CloudFormation utente.

Disponibilità regionale

L' AWS Auto Scaling API è disponibile in diverse regioni Regioni AWS e fornisce un endpoint per ciascuna di queste regioni. Per un elenco di tutte le regioni e gli endpoint in cui l'API è attualmente disponibile, consulta la sezione [AWS Auto Scaling Riferimenti generali di AWSEndpoint e ARN](#) per Amazon Web Services in

Prezzi

Tutte le funzionalità del piano di dimensionamento sono abilitate per l'uso. Le funzionalità vengono fornite senza costi aggiuntivi oltre ai costi del servizio CloudWatch e Cloud AWS alle altre risorse utilizzate.

Note

La funzionalità di scalabilità predittiva si basa sull' CloudWatch [GetMetricData](#) operazione di raccolta di dati metrici storici per la previsione della capacità, il che comporta costi. Tuttavia,

se abiliti la scalabilità predittiva con una policy di Amazon EC2 Auto Scaling anziché con un piano di scalabilità, non ci sono costi per le chiamate verso `GetMetricData`

Come funzionano i piani di dimensionamento

AWS Auto Scaling ti permette di utilizzare piani di dimensionamento per configurare un insieme di istruzioni per il dimensionamento delle proprie risorse. Se si lavora con AWS CloudFormation o si aggiungono tag a risorse scalabili, è possibile impostare piani di dimensionamento per diversi set di risorse, per applicazione. La console di AWS Auto Scaling fornisce suggerimenti per le strategie di dimensionamento personalizzate per ciascuna risorsa. Dopo aver creato i piani di dimensionamento, unisce i metodi di dimensionamento dinamici e predittivi a supporto della strategia di dimensionamento.

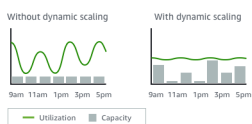
Cos'è una strategia di dimensionamento?

La strategia di dimensionamento indica a AWS Auto Scaling come ottimizzare l'utilizzo delle risorse nel piano di dimensionamento. È possibile ottimizzare la disponibilità, il costo o un equilibrio di entrambi. In alternativa, puoi creare strategie personalizzate in funzione dei parametri e delle soglie definiti. È possibile impostare delle strategie separate per ciascuna risorsa o per tipo di risorsa.



Che cos'è il dimensionamento dinamico?

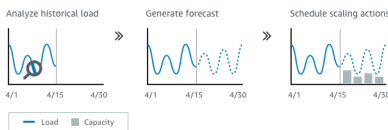
Il dimensionamento dinamico crea policy di dimensionamento di monitoraggio obiettivo per le risorse nel piano di dimensionamento. Tali policy di dimensionamento regolano la capacità di risorse in risposta alle variazioni di utilizzo delle risorse. L'obiettivo è fornire capacità sufficiente per mantenere l'utilizzo delle risorse entro il valore obiettivo specificato dalla strategia di dimensionamento. Questa operazione può essere paragonata al modo in cui il termostato regola la temperatura di una casa. Tu selezioni la temperatura, il termostato si occupa del resto.



Ad esempio, è possibile configurare il piano di dimensionamento e mantenere il numero di operazioni che il Amazon Elastic Container Service (Amazon ECS) esegue al 75% di CPU. Quando l'utilizzo della CPU del servizio supera il 75% (ossia più del 75% della CPU riservata per il servizio in uso), questo attiva la policy di dimensionamento e aggiunge un'altra attività al servizio che possa aiutarti con il picco di carico.

Che cos'è il dimensionamento predittivo?


Il dimensionamento predittivo usa il machine learning per analizzare il carico di lavoro storico di ogni risorsa ed eseguire regolarmente previsioni di carico futuro. Questo è simile a come funzionano le previsioni meteorologiche. Utilizzando la previsione, il dimensionamento predittivo genera operazioni programmate di dimensionamento per assicurare che le capacità delle risorse siano disponibili prima che la tua applicazione ne abbia bisogno. Come nel dimensionamento dinamico, il dimensionamento predittivo lavora per mantenere l'utilizzo al valore target specificato dalla strategia di dimensionamento.



Ad esempio, puoi abilitare il dimensionamento predittivo e configurare la tua strategia di dimensionamento per mantenere l'utilizzo medio della CPU del gruppo Auto Scaling al 50%. Considera una previsione in base alla quale i picchi di traffico si verificano ogni giorno alle 8 del mattino. Il piano di dimensionamento crea le operazioni di dimensionamento programmate future per assicurare che il gruppo Auto Scaling sia pronto a gestire il traffico in anticipo. Questo aiuta a mantenere le prestazioni dell'applicazione costanti, con l'obiettivo di avere sempre la capacità richiesta per mantenere l'utilizzo delle risorse il più possibile vicino al 50% in qualsiasi momento.

Di seguito sono riportati i concetti chiave per comprendere il dimensionamento predittivo:

- **Caricamento delle previsioni:** AWS Auto Scaling analizza fino a 14 giorni di storico per un determinato parametro di carico e prevede la domanda futura per i due giorni successivi. Questi dati sono disponibili in intervalli di un'ora e vengono aggiornati quotidianamente.
- **Operazioni di dimensionamento programmate:** AWS Auto Scaling pianifica le operazioni di dimensionamento che aumentano e diminuiscono in modo proattivo la capacità delle risorse in base alle previsioni di carico. Al momento previsto, AWS Auto Scaling aggiorna la capacità minima della risorsa con il valore specificato dall'operazione di dimensionamento programmata. L'obiettivo è mantenere l'utilizzo delle risorse entro il valore obiettivo specificato dalla strategia di dimensionamento. Se la tua applicazione richiede una capacità superiore rispetto a quella prevista, il dimensionamento dinamico è disponibile per aggiungere ulteriore capacità.
- **Comportamento di capacità massima:** i limiti di capacità minima e massima per la scalabilità automatica si applicano a ciascuna risorsa. Tuttavia, è possibile controllare se la tua applicazione è in grado di aumentare la capacità oltre quella massima quando la capacità di previsione è superiore alla massima.

 Note

Ora è possibile utilizzare le policy di dimensionamento predittivo dei gruppi Auto Scaling. Per ulteriori informazioni, consulta [Dimensionamento predittivo per Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

Best practice per i piani di dimensionamento

Le seguenti best practice consentono di sfruttare al massimo i piani di dimensionamento:

- Quando si crea un modello o una configurazione di avvio, abilitare il monitoraggio dettagliato per ottenere i dati dei parametri CloudWatch per le istanze EC2 a una frequenza di un minuto per garantire una risposta più rapida alle modifiche del carico. Il dimensionamento sui parametri a intervalli di 5 minuti potrebbe rallentare il tempo di risposta e causare il dimensionamento su dati di parametro obsoleti. Per impostazione predefinita, le istanze EC2 sono abilitate per il monitoraggio base, il che significa che i dati dei parametri per le istanze sono disponibili a intervalli di 5 minuti. Per un costo aggiuntivo, è possibile abilitare il monitoraggio dettagliato per ottenere i dati dei parametri per le istanze con una frequenza di un minuto. Per ulteriori informazioni, consulta la sezione relativa alla [configurazione del monitoraggio per le istanze di Auto Scaling](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.
- È consigliabile abilitare anche i parametri del gruppo Auto Scaling. In caso contrario, i dati della capacità effettiva non vengono mostrati nei grafici di previsione della capacità disponibili al completamento della procedura guidata per la creazione del piano di dimensionamento. Per ulteriori informazioni, consulta [Monitoraggio dei parametri di CloudWatch per le istanze e i gruppi Auto Scaling](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.
- Controlla il tipo di istanza utilizzato dal gruppo Auto Scaling e fai attenzione a usare un tipo di istanza con prestazioni burstable. Le istanze Amazon EC2 espandibili, ossia le istanze T3 e T2, sono progettate per offrire un livello base di prestazioni della CPU con la possibilità di raggiungere un livello più alto quando richiesto dal carico di lavoro. A seconda dell'utilizzo di destinazione specificato dal piano di dimensionamento, è possibile correre il rischio di superare la linea di base e quindi terminare i crediti di CPU, che limitano le prestazioni. Per ulteriori informazioni, consulta [Crediti CPU e prestazioni di base per istanze espandibili](#). Per configurare queste istanze come unlimited, consulta [Utilizzo di un gruppo Auto Scaling per avviare un'istanza espandibile come illimitata](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Altre considerazioni

Note

Esiste una versione più recente del dimensionamento predittivo, rilasciata a maggio 2021. Alcune funzionalità introdotte in questa versione non sono disponibili nei piani di

scalabilità e per accedere a tali funzionalità è necessario utilizzare una politica di scalabilità predittiva impostata direttamente nel gruppo con dimensionamento automatico. Per ulteriori informazioni, consulta [Dimensionamento predittivo per Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

Tieni presente le seguenti considerazioni aggiuntive:

- Il dimensionamento predittivo impiega previsioni di carico per pianificare la capacità in futuro. La qualità delle previsioni varia in base alla ciclicità del carico e all'applicabilità del modello di previsione qualificato. Il dimensionamento predittivo può essere eseguito in modalità di sola previsione per valutare la qualità delle previsioni e le operazioni di dimensionamento create dalle previsioni. È possibile impostare la modalità del dimensionamento predittivo su Forecast only (Solo previsione) quando si crea il piano di dimensionamento, quindi modificarlo in Forecast and scale (Previsione e dimensionamento) al termine della valutazione della qualità della previsione. Per ulteriori informazioni, consulta [Impostazioni dimensionamento predittivo](#) e [Monitoraggio e valutazione delle previsioni](#).
- Se si sceglie di specificare parametri diversi per il dimensionamento predittivo, è necessario accertarsi che il parametro di dimensionamento e quello di caricamento siano fortemente correlati. Il valore del parametro deve aumentare e diminuire in proporzione al numero di istanze nel gruppo Auto Scaling. In questo modo, i dati dei parametri possono essere utilizzati per scalare proporzionalmente il numero di istanze. Ad esempio, il parametro di carico equivale al conteggio totale di richieste e il parametro di dimensionamento equivale all'utilizzo medio della CPU. Se il conteggio totale delle richieste aumenta del 50%, l'utilizzo medio della CPU deve anche aumentare del 50%, purché la capacità rimanga invariata.
- Prima di creare un piano di dimensionamento, è necessario eliminare le operazioni di dimensionamento pianificate in precedenza non più necessarie mediante l'accesso alle console in cui sono state create. AWS Auto Scaling non crea un'operazione di dimensionamento predittivo che si sovrappone a un'operazione di dimensionamento pianificata esistente.
- Le tue impostazioni personalizzate per la capacità minima e massima, insieme ad altre impostazioni utilizzate per il dimensionamento dinamico, vengono visualizzate in altre console. Tuttavia, è consigliabile che, dopo la creazione di un piano di dimensionamento, queste impostazioni non vengano modificate da altre console, perché il piano di dimensionamento non riceve aggiornamenti da altre console.
- Il piano di dimensionamento può contenere risorse provenienti da più servizi, ma ogni risorsa può trovarsi in un solo piano di dimensionamento alla volta.

Evitare l'errore ActiveWithProblems

Un errore "ActiveWithProblems" può verificarsi quando viene creato o vengono aggiunte risorse a un piano di dimensionamento. L'errore si verifica quando il piano di dimensionamento è attivo, ma non è stato possibile applicare la configurazione di dimensionamento di una o più risorse.

In genere, ciò accade perché una risorsa dispone già di una policy di dimensionamento o un gruppo Auto Scaling non soddisfa i requisiti minimi per il dimensionamento predittivo.

Se una delle risorse dispone già di policy di dimensionamento da varie console dei servizi, per impostazione predefinita AWS Auto Scaling non sovrascrive queste altre policy di dimensionamento o ne crea di nuove. È possibile eliminare facoltativamente le policy di dimensionamento esistenti e sostituirle con policy di dimensionamento di monitoraggio del target create dalla console AWS Auto Scaling. Questa operazione viene effettuata abilitando l'impostazione Replace external scaling policies (Sostituisci policy di dimensionamento esterne) per ogni risorsa che dispone di policy di dimensionamento da sovrascrivere.

Con il dimensionamento predittivo, si consiglia di attendere 24 ore dopo la creazione di un nuovo gruppo Auto Scaling per configurare il dimensionamento predittivo. Devono essere presenti almeno 24 ore di dati storici per generare la previsione iniziale. Se il gruppo dispone di meno di 24 ore di dati storici ed è abilitato il dimensionamento predittivo, il piano di dimensionamento non sarà in grado di generare una previsione fino al prossimo periodo di previsione dopo che il gruppo ha raccolto la quantità di dati richiesta. Tuttavia, è anche possibile modificare e salvare il piano di dimensionamento per riavviare il processo di previsione non appena sono disponibili i dati di 24 ore.

Nozioni di base sui piani di dimensionamento

Prima di creare un piano di dimensionamento da usare con l'applicazione, esamina accuratamente l'applicazione quando viene eseguita in Cloud AWS. Prendi nota di quanto segue:

- Se si dispone di regole di dimensionamento esistenti create da altre console. Puoi sostituire le policy di dimensionamento esistenti oppure mantenerle (senza che ti sia consentito apportare modifiche ai relativi valori) quando crei il piano di dimensionamento.
- L'utilizzo di destinazione che ha senso per ogni risorsa scalabile nell'applicazione in base alla risorsa nel suo complesso. Ad esempio, esamina la quantità di CPU che le istanze EC2 in un gruppo Auto Scaling dovrebbero utilizzare rispetto alla CPU disponibile. Oppure, per un servizio come DynamoDB che utilizza un modello di throughput con provisioning, esamina la quantità di attività di lettura e scrittura che una tabella o un indice dovrebbe utilizzare rispetto alla velocità effettiva disponibile; in altre parole, il rapporto tra la capacità utilizzata e la capacità fornita. È possibile modificare l'utilizzo di destinazione in qualsiasi momento dopo aver creato il piano di dimensionamento.
- Quanto tempo occorre per avviare e configurare un server. Queste informazioni ti aiuteranno a configurare una finestra per il riscaldamento di ciascuna istanza EC2 dopo l'avvio, per assicurarti di non avviare un nuovo server mentre il precedente è ancora in fase di avvio.
- Se la cronologia dei parametri è sufficientemente lunga per essere utilizzata con il dimensionamento predittivo (se si utilizzano gruppi Auto Scaling appena creati). In generale, avere 14 giorni completi di dati storici si traduce in previsioni più accurate. Il periodo minimo è di 24 ore.

Più comprendi gli aspetti della tua applicazione, più efficace potrai rendere il tuo piano di dimensionamento.

Le seguenti attività ti consentono di acquisire familiarità con i piani di dimensionamento. Creerai un piano di dimensionamento per un singolo gruppo Auto Scaling e abiliterai il dimensionamento predittivo e il dimensionamento dinamico.

Processi

- [Fase 1: individuare le risorse scalabili](#)
- [Fase 2: specificare la strategia di dimensionamento](#)
- [Fase 3: configurare le impostazioni avanzate \(facoltativo\)](#)
- [Fase 4: creare il piano di dimensionamento](#)

- [Fase 5: Pulizia](#)
- [Passaggio 6: fasi successive](#)

Fase 1: individuare le risorse scalabili

Questa sezione include un'introduzione pratica alla creazione di piani di dimensionamento nella console AWS Auto Scaling. Se si tratta del primo piano di dimensionamento, ti consigliamo di iniziare creando un piano di dimensionamento di esempio utilizzando un gruppo Amazon EC2 Auto Scaling.

Prerequisiti

Per utilizzare un piano di dimensionamento, crea un gruppo Auto Scaling. Avvia almeno un'istanza Amazon EC2 nel gruppo Auto Scaling. Per maggiori informazioni, consulta [Nozioni di base su Amazon EC2 Auto Scaling](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

Utilizza un gruppo Auto Scaling con i parametri CloudWatch abilitati per avere i dati di capacità sui grafici disponibili al completamento del percorso guidati di Creazione di un piano di dimensionamento. Per ulteriori informazioni, consulta [Abilita parametri del gruppo Auto Scaling](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

Genera un carico per alcuni giorni o più per avere i dati dei parametri di CloudWatch disponibili per la funzionalità di dimensionamento predittivo, se possibile.

Verifica di disporre delle autorizzazioni necessarie per l'utilizzo dei piani di dimensionamento. Per ulteriori informazioni, consulta [Identity and Access Management per piani di dimensionamento](#).

Aggiungi il gruppo Auto Scaling al nuovo piano di dimensionamento

Quando crei un piano di dimensionamento dalla console, ti aiuta a trovare le risorse scalabili come primo passo. Prima di iniziare, verifica che siano soddisfatti i seguenti requisiti

- Hai creato un gruppo Auto Scaling e avviato almeno un'istanza EC2, come descritto nella sezione precedente.
- Il gruppo Auto Scaling creato esiste da almeno 24 ore.

Per iniziare, crea un piano di dimensionamento.

1. Apri la console AWS Auto Scaling alla pagina <https://console.aws.amazon.com/autoscaling/>.

2. Sulla barra di navigazione nella parte superiore della schermata, seleziona la stessa Regione utilizzata durante la creazione del gruppo Auto Scaling.
3. Nella pagina di benvenuto, scegliere Get started (Inizia).

Management Tools

AWS Auto Scaling

helps you quickly and easily scale multiple resources

AWS Auto Scaling enables you to quickly discover all of the scalable resources underlying your application and set up application scaling in minutes using built-in scaling recommendations.

Create a scaling plan

Optimize your application in just a few steps

[Get started](#)

Pricing

AWS Auto Scaling is free.

AWS Auto Scaling is enabled by Amazon CloudWatch and carries no additional fees. Service fees for your application resources and Amazon CloudWatch apply.

How it works

4. Nella pagina Find scalable resources (Trova risorse scalabili), esegui una delle seguenti operazioni:
 - Scegli Search by CloudFormation stack (Cerca per stack CloudFormation) e seleziona lo stack AWS CloudFormation da utilizzare.
 - Scegli Search by tag (Cerca per tag). Quindi, per ciascun tag, scegli una chiave di tag da Key (Chiave) e valori di tag da Value (Valore). Per aggiungere tag, scegliere Add another row (Aggiungi un'altra riga). Per rimuovere i tag, scegliere Remove (Rimuovi).
 - Scegli Choose EC2 Auto Scaling groups (Scegli gruppi EC2 di Auto Scaling), quindi scegli uno o più gruppi Auto Scaling.

Note

Per un tutorial introduttivo, scegli Choose EC2 Auto Scaling groups (Scegli gruppi EC2 Auto Scaling), quindi scegli il gruppo Auto Scaling creato.

Choose a method

Search by CloudFormation stack
Search for resources provisioned by an AWS CloudFormation stack.

Search by tag
Search for resources by tags applied to them.

Choose EC2 Auto Scaling groups
Choose one or more Auto Scaling groups to include in your scaling plan.

Choose Auto Scaling groups [Info](#)

Auto Scaling groups

Choose Auto Scaling groups ▼

my-auto-scaling-group X

5. Scegli Next (Successivo) per continuare con il processo di creazione del piano di dimensionamento.

Ottieni maggiori informazioni su come individuare le tue risorse scalabili

Se è già stato creato un piano di dimensionamento di esempio e si desidera crearne altri, consulta i seguenti scenari per l'utilizzo di uno stack di CloudFormation o di un set di tag. È possibile utilizzare questa sezione per decidere se scegliere l'opzione Search by CloudFormation stack (Cerca per stack di CloudFormation) o Search by tag (Cerca per tag) per individuare le risorse scalabili quando si utilizza la console per creare il piano di dimensionamento.

Quando scegli l'opzione Search by CloudFormation stack (Cerca per stack di CloudFormation) o Search by tag (Cerca per tag) nella fase 1 della procedura guidata di Create Scaling Plan (Crea un piano di dimensionamento) le risorse scalabili associate allo stack o al set di tag per il piano di dimensionamento diventano disponibili. Quando definisci il piano di dimensionamento, puoi specificare quali di queste risorse includere o escludere.

Individuazione di risorse scalabili utilizzando uno stack di CloudFormation

Quando si utilizza CloudFormation, lavori con gli stack per fornire le risorse. Tutte le risorse di uno stack sono definite dal modello dello stack. Il piano di dimensionamento aggiunge un livello di orchestrazione nella parte superiore dello stack che semplifica la configurazione del dimensionamento per più risorse. Senza un piano di dimensionamento, è necessario impostare il dimensionamento individualmente per ogni risorsa scalabile. Ciò significa capire l'ordine per il provisioning delle risorse e le policy di dimensionamento e cogliere le sfumature del funzionamento di queste dipendenze.

Nella console AWS Auto Scaling, puoi selezionare uno stack esistente perché sia scansionato da in cerca delle risorse configurabili per la scalabilità automatica. AWS Auto Scaling troverà solo le risorse che sono definite nello stack selezionato. Non si espande attraverso gli stack nidificati.

Affinché i servizi ECS siano individuabili in uno stack CloudFormation, la console AWS Auto Scaling deve sapere quale cluster ECS sta eseguendo il servizio. A questo scopo, i servizi ECS devono trovarsi nello stesso stack CloudFormation del cluster ECS che sta eseguendo il servizio. Altrimenti, devono far parte del cluster predefinito. Per essere identificato correttamente, il nome del servizio ECS deve inoltre essere univoco in ciascuno di questi cluster ECS.

Per ulteriori informazioni su CloudFormation, consulta [Cos'è AWS CloudFormation?](#) nella Guida per l'utente di AWS CloudFormation.

Individuazione di risorse scalabili tramite tag

I tag forniscono metadati che possono essere utilizzati per individuare le risorse scalabili correlate nella console AWS Auto Scaling, utilizzando i filtri dei tag.

Utilizza i tag per individuare una qualsiasi delle seguenti risorse:

- Cluster di database Aurora
- Gruppi Auto Scaling
- Tabelle DynamoDB e indici secondari globali

Quando esegui una ricerca in base a più tag, una risorsa viene rilevata solo se dispone di tutti i tag elencati.

Per ulteriori informazioni sull'assegnazione di tag, consulta la documentazione seguente.

- Scopri come assegnare [tag ai cluster di Aurora](#) nella Guida per l'utente di Amazon Aurora.
- Scopri come assegnare [tag ai gruppi Auto Scaling](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.
- Scopri come assegnare [tag alle risorse Dynamo DB](#) nella Guida per gli sviluppatori di Amazon DynamoDB.
- Per ulteriori informazioni sulle best practice per [l'assegnazione di tag alle risorse AWS](#), consulta la Riferimenti generali di AWS.

Fase 2: specificare la strategia di dimensionamento

Utilizzare la procedura seguente per specificare strategie di dimensionamento per le risorse individuate nel passaggio precedente.

Per ogni tipo di risorsa, AWS Auto Scaling sceglie il parametro più frequentemente utilizzato per determinare la quantità della risorsa in uso in un qualsiasi momento. Si può scegliere la strategia di dimensionamento più appropriata per ottimizzare le prestazioni dell'applicazione in base a tale parametro. Quando si abilita la funzione di dimensionamento dinamico e quella di dimensionamento predittivo, la strategia di dimensionamento è condivisa tra le due funzioni. Per ulteriori informazioni, consulta [Come funzionano i piani di dimensionamento](#).

Sono disponibili le seguenti strategie di dimensionamento:

- Optimize for availability (Ottimizza per la disponibilità) - AWS Auto Scaling dimensiona automaticamente la risorsa per mantenere l'utilizzo delle risorse al 40%. Questa opzione è utile quando l'applicazione ha esigenze di dimensionamento urgenti e imprevedibili.
- Bilancia disponibilità e costo - AWS Auto Scaling dimensiona automaticamente la risorsa per mantenere l'utilizzo delle risorse al 50%. Questa opzione aiuta a mantenere un'elevata disponibilità riducendo al contempo i costi.
- Optimize for cost (Ottimizza per il costo) - AWS Auto Scaling dimensiona automaticamente la risorsa per mantenere l'utilizzo delle risorse al 70%. Questa opzione è utile per ridurre i costi se l'applicazione è in grado di gestire una ridotta capacità del buffer quando la domanda è soggetta a modifiche impreviste.

Ad esempio, il piano di dimensionamento configura il gruppo Auto Scaling in modo da aggiungere o rimuovere istanze Amazon EC2 in base alla quantità di CPU utilizzata in media per tutte le istanze nel gruppo. Puoi scegliere se ottimizzare l'utilizzo per la disponibilità, per i costi o una combinazione dei due modificando la strategia di dimensionamento.

In alternativa, puoi configurare una strategia personalizzata se una di quelle esistenti non soddisfa le tue esigenze. Con una strategia personalizzata, è possibile modificare il valore di utilizzo target, scegliere un altro parametro o entrambe le opzioni.

⚠ Important

Per il tutorial di base, completa solo la prima fase della seguente procedura, quindi scegli Next (Successivo) per continuare.

Specificare strategie di dimensionamento

1. Nella pagina Specify scaling strategy (Specifica strategia di dimensionamento), in Scaling plan details (Dettagli del piano di dimensionamento), Name (Nome), immettere un nome per il piano di dimensionamento. Il nome del piano di dimensionamento deve essere univoco all'interno del set di piani di dimensionamento per la regione. Può avere un massimo di 128 caratteri e non deve contenere barre "|", barre "/" o due punti ":".
2. Tutte le risorse incluse sono elencate per tipo di risorsa. Per Gruppi Auto Scaling, procedi come segue:

Auto Scaling groups (1) Include in scaling plan

Specify a scaling strategy for 1 Auto Scaling group.

Scaling strategy
The strategy defines the scaling metric and target value used to scale your resources.

Optimize for availability
Keep the average CPU utilization of your Auto Scaling groups at 40% to provide high availability and ensure capacity to absorb spikes in demand.

Balance availability and cost
Keep the average CPU utilization of your Auto Scaling groups at 50% to provide optimal availability and reduce costs.

Optimize for cost
Keep the average CPU utilization of your Auto Scaling groups at 70% to ensure lower costs.

Custom
Choose your own scaling metric, target value, and other settings.

Enable predictive scaling
Support your scaling strategy by continually forecasting load and proactively scheduling capacity ahead of when you need it. [Info](#)

Enable dynamic scaling
Support your scaling strategy by creating target tracking scaling policies to monitor your scaling metric and increase or decrease capacity as you need it. [Info](#)

▶ **Configuration details**

- a. Salta questo passaggio per utilizzare la strategia di dimensionamento e i parametri di default. Per utilizzare una strategia o parametri di ridimensionamento diversi, procedi con i seguenti passaggi:
 - i. Per Strategia di dimensionamento, scegli la strategia di dimensionamento desiderata.

Per il tutorial introduttivo, assicurati di scegliere Optimize for availability (Ottimizza per disponibilità). Ciò specifica che l'utilizzo medio della CPU del gruppo Auto Scaling verrà mantenuto al 40%.

- ii. Se hai scelto Custom (Personalizza), espandi Configuration details (Dettagli di configurazione) per scegliere i parametri e il valore di destinazione desiderati.
 - Per Scaling metric (Parametro di dimensionamento), scegli il parametro di dimensionamento desiderato.
 - Per Target value (Valore di destinazione), scegli il valore di destinazione desiderato, ad esempio l'utilizzo o il throughput di destinazione durante un intervallo di un minuto.
 - Per Load metric (Parametro di carico) [solo gruppi Auto Scaling], scegli il parametro di carico desiderato da utilizzare per il dimensionamento predittivo.
 - In Replace external scaling policies (Sostituisci le policy di dimensionamento esterno), scegli se AWS Auto Scaling può eliminare le policy di dimensionamento create all'esterno del piano di dimensionamento (ad esempio da altre console) e sostituirle con nuove policy di dimensionamento di monitoraggio obiettivi create dal piano di dimensionamento.
 - b. (Facoltativo) Per impostazione predefinita, per i gruppi Auto Scaling è attivo il dimensionamento predittivo. Per disabilitare il dimensionamento predittivo per i gruppi Auto Scaling, deseleziona Enable predictive scaling (Abilita dimensionamento predittivo).
 - c. (Facoltativo) Per impostazione predefinita, il dimensionamento dinamico è abilitato per ogni tipo di risorsa. Per disabilitare il dimensionamento dinamico per un tipo di risorsa, deseleziona Enable dynamic scaling (Abilita dimensionamento dinamico).
 - d. (Facoltativo) Per impostazione predefinita, quando si specifica un'origine dell'applicazione da cui vengono individuate più risorse scalabili, tutti i tipi di risorse vengono automaticamente inclusi nel piano di dimensionamento. Per omettere un tipo di risorsa dal piano di dimensionamento, deselezionare Includi in piano di dimensionamento.
3. (Facoltativo) Per specificare una strategia di dimensionamento per un altro tipo di risorsa, ripeti i passaggi precedenti.
 4. Una volta terminato, scegli Next (Successivo) per continuare con il processo di creazione del piano di dimensionamento.

Fase 3: configurare le impostazioni avanzate (facoltativo)

Ora che hai specificato la strategia di dimensionamento da usare per ogni tipo di risorsa, puoi scegliere di personalizzare le impostazioni predefinite per ogni risorsa utilizzando la fase *Configure advanced settings* (Configura impostazioni avanzate). Per ogni tipo di risorsa, sono disponibili per la personalizzazione più gruppi di impostazioni. Nella maggior parte dei casi, tuttavia, le impostazioni predefinite devono essere più efficienti, ad eccezione dei valori per la capacità minima e la capacità massima, che possono essere attentamente regolati.

Ignora questa procedura se desideri mantenere le impostazioni predefinite. Puoi modificare queste impostazioni in qualsiasi momento modificando il piano di dimensionamento.

Important

Per il tutorial introduttivo, apportiamo qualche modifica per aggiornare la capacità massima del gruppo Auto Scaling e per abilitare il dimensionamento predittivo in modalità di sola previsione. Anche se non è necessario personalizzare tutte le impostazioni per il tutorial, esaminiamo brevemente anche le impostazioni in ciascuna sezione.


Impostazioni generali

Utilizza questa procedura per visualizzare e personalizzare le impostazioni specificate nella fase precedente, in base alle risorse. È anche possibile personalizzare la capacità minima e la capacità massima di ciascuna risorsa.

Per visualizzare e personalizzare le impostazioni generali

1. Nella pagina *Configure advanced settings* (Configura impostazioni avanzate), selezionare la freccia alla sinistra di ogni intestazione della sezione per espanderla. Per il tutorial, espandi la sezione *Auto Scaling groups* (Gruppi Auto Scaling).
2. Nella tabella visualizzata, scegli il gruppo Auto Scaling gruppo che stai utilizzando in questo tutorial.
3. Lasciare selezionata l'opzione *Include in scaling plan* (Includi in piano di dimensionamento). Se questa opzione non è selezionata, la risorsa è omessa dal piano di dimensionamento. Se non si include almeno una risorsa, il piano di dimensionamento non può essere creato.
4. Per espandere la visualizzazione e visualizzare i dettagli della sezione *Impostazioni generali*, selezionare la freccia alla sinistra dell'intestazione della sezione.

5. È possibile effettuare scelte per uno qualsiasi dei seguenti elementi. Per questo tutorial, individuare l'impostazione Maximum capacity (Capacità massima) e inserire il valore 3 al posto di quello attuale.
- Scaling strategy (Strategia di dimensionamento) - Consente di ottimizzare la disponibilità, il costo o un equilibrio di entrambi, oppure di specificare una strategia personalizzata.
 - Enable dynamic scaling (Abilita il dimensionamento dinamico) - Se questa impostazione è deselezionata, la risorsa selezionata non può essere dimensionata utilizzando una configurazione di dimensionamento del monitoraggio degli obiettivi.
 - Enable predictive scaling (Abilita il dimensionamento predittivo) - [Solo gruppi Auto Scaling] Se questa impostazione è deselezionata, il gruppo selezionato non può dimensionare utilizzando il dimensionamento predittivo.
 - Scaling metric (Parametro di dimensionamento) - Specifica il parametro di dimensionamento da utilizzare. Se si sceglie Custom (Personalizzata), è possibile specificare un parametro di carico personalizzato da utilizzare al posto dei parametri di carico predefiniti disponibili nella console. Per maggiori informazioni, consulta il prossimo argomento in questa sezione.
 - Target value (Valore obiettivo) - Specifica il valore di utilizzo obiettivo da usare.
 - Load metric (Parametro di carico) - [Solo gruppi Auto Scaling] Specifica il parametro di carico da utilizzare. Se si sceglie Custom (Personalizzata), è possibile specificare un parametro di carico personalizzato da utilizzare al posto dei parametri di carico predefiniti disponibili nella console. Per maggiori informazioni, consulta il prossimo argomento in questa sezione.
 - Minimum capacity (Capacità minima) - Specifica la capacità minima per la risorsa. AWS Auto Scaling assicura che la tua risorsa non superi mai queste dimensioni.
 - Maximum capacity (Capacità massima) - Specifica la capacità massima per la risorsa. AWS Auto Scaling assicura che la tua risorsa non superi mai queste dimensioni.

 Note

Quando utilizzi il dimensionamento predittivo, hai la possibilità di scegliere un comportamento diverso per la capacità massima in base alla capacità di previsione. Questa impostazione si trova nella sezione Predictive scaling settings (Impostazioni dimensionamento predittivo).

Parametri personalizzati

AWS Auto Scaling fornisce i parametri più comunemente utilizzati per il dimensionamento automatico. Tuttavia, in base alle esigenze, è preferibile ottenere i dati provenienti da parametri differenti anziché i parametri nella console. Amazon CloudWatch offre molti parametri diversi tra cui scegliere. CloudWatch consente inoltre di pubblicare i propri parametri.

Per specificare un parametro personalizzato CloudWatch, utilizza JSON. Prima di seguire queste istruzioni, ti consigliamo di familiarizzare con la [Guida per l'utente di Amazon CloudWatch](#).

Per specificare un parametro personalizzato, è necessario creare un payload in formato JSON utilizzando un set di parametri richiesti da un modello. Aggiungi i valori per ciascun parametro da CloudWatch. Forniamo il modello nell'ambito delle opzioni personalizzate per Scaling metric (Parametro di dimensionamento) e Load metric (Parametro di carico) nelle impostazioni avanzate del piano di dimensionamento.

JSON rappresenta i dati in due modi:

- Un oggetto, una raccolta non ordinata di coppie nome-valore. Un oggetto viene definito all'interno di parentesi graffe sinistra ({) e destra (}). Ogni coppia nome-valore inizia con il nome, seguita dai due punti e dal valore. Le coppie nome-valore sono separate da virgole.
- Una matrice, che è un insieme ordinato di valori. Una matrice viene definita all'interno di parentesi quadre sinistra ([) e destra (]). Gli elementi nella matrice sono separati da virgole.

Ecco un esempio del modello JSON con valori di esempio per ogni parametro:

```
{
  "MetricName": "MyBackendCPU",
  "Namespace": "MyNamespace",
  "Dimensions": [
    {
      "Name": "MyOptionalMetricDimensionName",
      "Value": "MyOptionalMetricDimensionValue"
    }
  ],
  "Statistic": "Sum"
}
```

Per ulteriori informazioni, consulta [Specifiche di parametri di dimensionamento personalizzati](#) e [Specifiche di parametri di caricamento personalizzati](#) nella Documentazione di riferimento dell'API AWS Auto Scaling.

Impostazioni dimensionamento dinamico

Utilizza questa procedura per visualizzare e personalizzare le impostazioni per la policy di dimensionamento di monitoraggio dei target creata da AWS Auto Scaling.

Per visualizzare e personalizzare le impostazioni per il dimensionamento dinamico

1. Per espandere la visualizzazione e visualizzare i dettagli della sezione Impostazioni dimensionamento dinamico, selezionare la freccia alla sinistra dell'intestazione della sezione.
2. È possibile effettuare scelte per i seguenti elementi. Tuttavia, le impostazioni predefinite vanno bene per questo tutorial.
 - Replace external scaling policies (Sostituisci le policy di dimensionamento esterne) - Se questa impostazione è disattivata, conserva le policy di dimensionamento esistenti create dall'esterno del piano di dimensionamento e non ne crea di nuove.
 - Disable scale-in (Disabilita dimensionamento orizzontale) - Se questa impostazione è deselezionata, la scalabilità automatica per ridurre la capacità corrente della risorsa è consentita quando il parametro specificato è inferiore al valore obiettivo.
 - Cooldown (Raffreddamento) - Crea tempi di raffreddamento per il dimensionamento orizzontale. Il tempo di raffreddamento è il tempo che la policy di dimensionamento attende prima di rendere effettiva una precedente attività di dimensionamento. Per ulteriori informazioni, consulta [Tempo di raffreddamento](#) nella Guida per l'utente di Application Auto Scaling. (Questa impostazione non viene utilizzata se la risorsa è un gruppo Auto Scaling.)
 - Instance warmup (Riscaldamento dell'istanza) - [Solo gruppi Auto Scaling] Controlla il tempo necessario perché un'istanza appena avviata inizi a contribuire ai parametri di CloudWatch. Per ulteriori informazioni, consulta [Riscaldamento delle istanze](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

Impostazioni dimensionamento predittivo

Se la risorsa è un gruppo Auto Scaling, utilizza questa procedura per visualizzare e personalizzare le impostazioni utilizzate da AWS Auto Scaling per il dimensionamento predittivo.

Per visualizzare e personalizzare le impostazioni per il dimensionamento predittivo

1. Per espandere la visualizzazione e visualizzare i dettagli della sezione Impostazioni dimensionamento predittivo, selezionare la freccia alla sinistra dell'intestazione della sezione.
2. È possibile effettuare scelte per i seguenti elementi. Per questo tutorial, modificare Predictive scaling mode (Modalità dimensionamento predittivo) in Forecast only (Sola previsione).
 - Predictive scaling mode (Modalità di dimensionamento predittivo) - Specifica la modalità di dimensionamento. L'impostazione predefinita è Forecast and scale (Previsione e dimensionamento). Se imposti il valore su Forecast only (Solo previsione), il piano di dimensionamento effettua la previsione sulla capacità futura ma non applica le operazioni di dimensionamento.
 - Pre-launch instances (Pre-avvia le istanze) - Regola le operazioni di dimensionamento programmate per l'esecuzione prioritaria in caso di dimensionamento orizzontale. Ad esempio, la previsione dichiara di aggiungere capacità alle 10:00 e il tempo relativo al buffer è 5 minuti (300 secondi). Il runtime dell'operazione di dimensionamento corrispondente è quindi 9:55. Questa funzione è utile per i gruppi Auto Scaling, dove può richiedere alcuni minuti dal momento in cui viene avviata un'istanza fino a quando non entra in servizio. Il tempo effettivo può variare in base a diversi fattori, ad esempio la dimensione dell'istanza e la presenza o meno di script di avvio da completare. Il valore predefinito è 300 secondi.
 - Max Capacity Behavior (Comportamento capacità massima) - Controlla se la risorsa selezionata può aumentare oltre la capacità massima quando la capacità di previsione è prossima o superiore alla capacità massima specificata al momento. L'opzione predefinita è Enforce the maximum capacity setting (Applica l'impostazione per la capacità massima).
 - Enforce the maximum capacity setting (Applica l'impostazione per la capacità massima) - AWS Auto Scaling non può dimensionare una capacità delle risorse superiore alla capacità massima. La capacità massima è applicata come un limite insuperabile.
 - Set the maximum capacity to equal forecast capacity (Imposta la capacità massima pari alla capacità prevista) - AWS Auto Scaling può dimensionare la capacità delle risorse a un livello superiore alla capacità massima affinché sia pari ma non superiore alla capacità prevista.
 - Increase maximum capacity above forecast capacity (Aumenta la capacità massima oltre la capacità prevista) - AWS Auto Scaling può dimensionare la capacità delle risorse a un livello superiore alla capacità massima di un valore di buffer specificato. L'intenzione è quella di assegnare alla policy di dimensionamento del monitoraggio di target capacità aggiuntiva in caso di traffico imprevisto.

- **Max capacity behavior buffer (Buffer di comportamento di capacità massima)** - Se hai scelto **Increase maximum capacity above forecast capacity (Aumenta la capacità massima oltre la capacità prevista)**, seleziona la dimensione del buffer di capacità da utilizzare quando la capacità prevista è prossima o superiore alla capacità massima. Il valore è specificato come una percentuale relativa alla capacità di previsione. Ad esempio, con un buffer del 10%, se la capacità di previsione è 50 e quella massima è 40, la capacità massima effettiva è 55.
3. Quando la definizione delle impostazioni personalizzate è terminata, seleziona **Next (Successivo)**.

Note

Per annullare qualsiasi modifica, seleziona le risorse e scegli **Revert to original (Ripristina originale)**. In questo modo viene ripristinato l'ultimo stato noto delle risorse selezionate all'interno del piano di dimensionamento.

Fase 4: creare il piano di dimensionamento

Nella pagina **Review and create (Rivedi e crea)** rivedi i dettagli del piano di dimensionamento e scegli **Create scaling plan (Crea piano di dimensionamento)**. Si viene reindirizzati a una pagina che mostra lo stato del piano di dimensionamento. Il completamento della creazione del piano di dimensionamento può richiedere un momento affinché le risorse vengano aggiornate.

Con il dimensionamento predittivo, **AWS Auto Scaling** analizza la cronologia di uno specifico parametro di carico relativa agli ultimi 14 giorni (è necessario un periodo minimo di 24 ore) e genera così una previsione per i due giorni successivi. Pianifica quindi operazioni di dimensionamento programmate per regolare la capacità delle risorse e soddisfare la richiesta prevista per ogni ora del periodo di previsione.

Dopo la creazione del piano di dimensionamento, visualizzare i dettagli del piano di dimensionamento scegliendo il nome dalla schermata **Piani di dimensionamento**.

(Facoltativo) Visualizzare le informazioni di dimensionamento per una risorsa

Utilizzare questa procedura per visualizzare le informazioni di dimensionamento create per una risorsa.

I dati vengono presentati nei seguenti modi:

- Grafici che mostrano i dati della cronologia di parametri recenti da CloudWatch.
- Grafici di dimensionamento predittivo che mostrano previsioni di carico e capacità basati sui dati provenienti da AWS Auto Scaling.
- Una tabella che elenca tutte le operazioni di dimensionamento predittivo programmate per la risorsa.

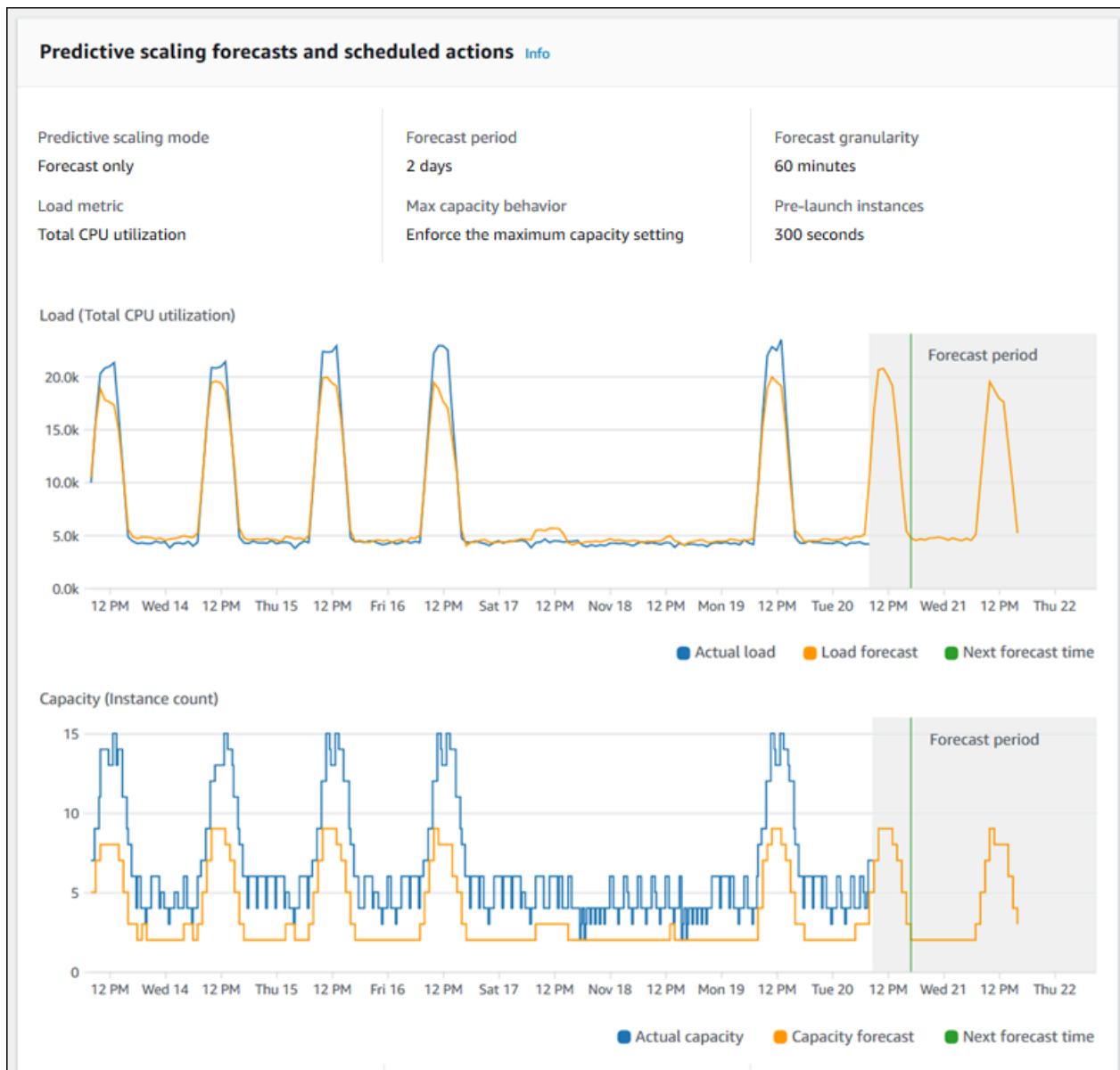
Per visualizzare le informazioni di dimensionamento per una risorsa

1. Apri la console AWS Auto Scaling alla pagina <https://console.aws.amazon.com/autoscaling/>.
2. Nella pagina Scaling plans (Piani di dimensionamento), selezionare il piano di dimensionamento.
3. Nella pagina Scaling plan details (Dettagli del piano di dimensionamento), scegliere la risorsa da visualizzare.

Monitoraggio e valutazione delle previsioni

Una volta che il piano di dimensionamento è in esecuzione, è possibile monitorare la previsione del carico, la previsione della capacità e le operazioni di dimensionamento per esaminare le prestazioni del dimensionamento predittivo. Tutti questi dati sono disponibili nella console AWS Auto Scaling per tutti i gruppi Auto Scaling abilitati al dimensionamento predittivo. Ricorda che il piano di dimensionamento richiede almeno 24 ore di dati storici del carico per effettuare la previsione iniziale.

Nel seguente esempio, il lato sinistro di ogni grafico mostra un modello storico. Il lato destro mostra la previsione generata dal piano di dimensionamento per il periodo di previsione. Vengono tracciati entrambi i valori, attuali e previsti (in blu e in arancione).



AWS Auto Scaling apprende i dati automaticamente. In primo luogo, effettua una previsione del carico. Quindi, il calcolo della previsione della capacità determina il numero minimo di istanze necessarie per supportare l'applicazione. In base alla previsione della capacità, AWS Auto Scaling pianifica le operazioni di dimensionamento che dimensionano il gruppo Auto Scaling prima delle modifiche del carico previste. Se il dimensionamento dinamico è abilitato (scelta consigliata), il gruppo Auto Scaling può dimensionare orizzontalmente capacità aggiuntiva (o rimuovere capacità) in base all'utilizzo corrente del gruppo di istanze.

Quando si valuta l'efficacia del dimensionamento predittivo, monitora quanto corrispondono i valori effettivi e quelli di previsione nel corso del tempo. Quando crei un piano di dimensionamento, AWS Auto Scaling offre grafici in base ai dati effettivi più recenti. Fornisce inoltre una previsione

iniziale per le successive 48 ore. Tuttavia al momento della creazione del piano di dimensionamento sono disponibili pochissimi dati di previsione per confrontare i dati effettivi. Attendi che il piano di dimensionamento abbia ottenuto i valori di previsione per alcuni periodi prima di confrontare i valori di previsione storici con i valori effettivi. Dopo alcuni giorni di previsioni giornaliere, avrai un campione più ampio di valori di previsione da confrontare con i valori effettivi.

Per i modelli che ricorrono su base giornaliera, l'intervallo di tempo tra la creazione del piano di dimensionamento e la valutazione dell'efficacia della previsione può essere anche di pochi giorni. Tuttavia, questo lasso di tempo non è sufficiente per valutare la previsione in base a un cambiamento di modello recente. Ad esempio, supponiamo che stai osservando le previsioni per un gruppo Auto Scaling che ha iniziato una nuova campagna di marketing la scorsa settimana. La campagna aumenta in modo significativo il traffico Web per gli stessi due giorni ogni settimana. In situazioni come questa, consigliamo di attendere che il gruppo raccolga per una o due settimane intere i nuovi dati prima di valutare l'efficacia della previsione. Lo stesso suggerimento vale per un gruppo Auto Scaling appena creato che ha iniziato a raccogliere solo i dati dei parametri.

Se i valori effettivi e di previsione non corrispondono dopo averli monitorati per un periodo di tempo appropriato, è necessario considerare anche la scelta del parametro di carico. Per essere efficace, il parametro di carico deve rappresentare una misura affidabile e accurata del carico totale in tutte le istanze nel gruppo Auto Scaling. Il parametro di carico è fondamentale per il dimensionamento predittivo. Se scegli un parametro di carico non ottimale, ciò potrebbe impedire al dimensionamento predittivo di effettuare accurate previsioni di carico e di capacità e la pianificazione delle corrette modifiche alla capacità per il gruppo Auto Scaling.

Fase 5: Pulizia

Dopo aver completato il tutorial sulle nozioni di base, puoi scegliere di mantenere il tuo piano di dimensionamento. Tuttavia, se non utilizzi questo piano di dimensionamento attivamente, potresti considerare di eliminarlo, in modo che l'account non vada incontro a spese superflue.

L'eliminazione di un piano di dimensionamento comporta l'eliminazione delle policy di dimensionamento con monitoraggio degli obiettivi, i relativi allarmi CloudWatch associati e le operazioni di dimensionamento predittivo che AWS Auto Scaling ha creato a tuo nome.

L'eliminazione di un piano di dimensionamento non elimina lo stack AWS CloudFormation, il gruppo Auto Scaling o altre risorse scalabili.

Per eliminare un piano di dimensionamento

1. Apri la console AWS Auto Scaling alla pagina <https://console.aws.amazon.com/autoscaling/>.
2. Nella pagina Scaling plans (Piani di dimensionamento), selezionare il piano di dimensionamento creato per questo tutorial e selezionare Delete (Elimina).
3. Quando viene richiesta la conferma, seleziona Delete (Elimina).

Una volta eliminato il piano di dimensionamento, le tue risorse non sono in grado di ripristinare la propria capacità originale. Ad esempio, se il tuo gruppo Auto Scaling viene dimensionato a 10 istanze quando elimini il piano di dimensionamento, rimarrà dimensionato a 10 istanze una volta eliminato il piano di dimensionamento. È possibile aggiornare la capacità di risorse specifiche mediante l'accesso alla console per ciascun servizio.

Eliminare il gruppo Auto Scaling

Per evitare che l'account accumuli costi Amazon EC2, è necessario eliminare anche il gruppo Auto Scaling creato per questo tutorial.

Per istruzioni dettagliate, consulta [Eliminazione del gruppo Auto Scaling](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

Passaggio 6: fasi successive

Ora che hai familiarizzato con i piani di dimensionamento e alcune delle sue caratteristiche, prova a creare il tuo modello di piano di dimensionamento utilizzando AWS CloudFormation.

Un modello AWS CloudFormation è un file di testo in formato JSON o YAML che descrive l'infrastruttura Amazon Web Services necessaria per l'esecuzione di un'applicazione o di un servizio insieme a tutte le interconnessioni tra i componenti dell'infrastruttura. Con AWS CloudFormation, è possibile distribuire e gestire una raccolta di risorse associata come stack. AWS CloudFormation è disponibile senza alcun costo aggiuntivo, è a pagamento solo per le risorse AWS necessarie per eseguire le applicazioni. Le risorse sono costituite da qualsiasi risorsa AWS definita all'interno del modello. Per ulteriori informazioni, consulta [Concetti di AWS CloudFormation](#) nella Guida per l'utente di AWS CloudFormation.

Nella Guida per l'utente di AWS CloudFormation è riportato un modello semplice per iniziare. Il modello di esempio è disponibile come tale nella sezione [AWS::AutoScalingPlans: :ScalingPlan](#) della documentazione di riferimento del modello AWS CloudFormation. Il modello di esempio crea un piano

di dimensionamento per un singolo gruppo Auto Scaling e consente il dimensionamento predittivo e il dimensionamento dinamico.

Per ulteriori informazioni, consulta [Nozioni di base su AWS CloudFormation](#) nella Guida per l'utente di AWS CloudFormation.

Sicurezza del piano di dimensionamento

La sicurezza del cloud AWS è la massima priorità. In qualità di AWS cliente, puoi beneficiare di un data center e di un'architettura di rete progettati per soddisfare i requisiti delle organizzazioni più sensibili alla sicurezza.

La sicurezza è una responsabilità condivisa tra AWS te e te. Il [modello di responsabilità condivisa](#) descrive questo aspetto come sicurezza del cloud e sicurezza nel cloud:

- **Sicurezza del cloud:** AWS è responsabile della protezione dell'infrastruttura che gestisce AWS i servizi nel AWS cloud. AWS ti fornisce anche servizi che puoi utilizzare in modo sicuro. Per maggiori informazioni sui programmi di conformità applicabili AWS Auto Scaling, consulta la sezione [AWS Servizi rientranti nell'ambito dei programmi di conformità](#) ().
- **Sicurezza nel cloud:** la tua responsabilità è determinata dal AWS servizio che utilizzi. Inoltre, sei responsabile anche di altri fattori, tra cui la riservatezza dei dati, i requisiti dell'azienda e le leggi e le normative applicabili.

Questa documentazione aiuta a comprendere come applicare il modello di responsabilità condivisa quando si utilizzano piani di dimensionamento e aiuta a comprendere come gestirne l'accesso.

Argomenti

- [Piani di dimensionamento ed endpoint VPC di interfaccia \(AWS PrivateLink\)](#)
- [Protezione dei dati](#)
- [Identity and Access Management per piani di dimensionamento](#)
- [Convalida della conformità](#)
- [Sicurezza dell'infrastruttura](#)

Piani di dimensionamento ed endpoint VPC di interfaccia (AWS PrivateLink)

Puoi migliorare il livello di sicurezza del tuo VPC AWS Auto Scaling configurando l'uso di un endpoint VPC di interfaccia. Gli endpoint di interfaccia sono alimentati da AWS PrivateLink, una tecnologia che consente di accedere in modo privato alle AWS Auto Scaling API limitando tutto il traffico di rete tra il VPC e la rete. AWS Auto Scaling AWS Con gli endpoint di interfaccia, non è necessario nemmeno un gateway Internet, un dispositivo NAT o un gateway privato virtuale.

La configurazione non è obbligatoria AWS PrivateLink, ma è consigliata. [Per ulteriori informazioni sugli AWS PrivateLink endpoint VPC, consulta Cos'è? AWS PrivateLink](#) nella Guida.AWS PrivateLink

Argomenti

- [Crea un endpoint VPC dell'interfaccia per i piani di dimensionamento](#)
- [Crea una policy dell'endpoint VPC per i piani di dimensionamento](#)
- [Migrazione degli endpoint](#)

Crea un endpoint VPC dell'interfaccia per i piani di dimensionamento

Crea un endpoint per i piani AWS Auto Scaling di scalabilità utilizzando il seguente nome di servizio:

```
com.amazonaws.region.autoscaling-plans
```

Per ulteriori informazioni, consulta [Accedere a un AWS servizio utilizzando un endpoint VPC di interfaccia nella Guida.AWS PrivateLink](#)

Non è necessario modificare altre impostazioni. AWS Auto Scaling Le altre chiamate API Servizi AWS utilizzano endpoint di servizio o endpoint VPC con interfaccia privata, a seconda di quale dei due siano in uso.

Crea una policy dell'endpoint VPC per i piani di dimensionamento

Puoi allegare una policy al tuo endpoint VPC per controllare l'accesso all'API. AWS Auto Scaling La policy specifica:

- Il principale che può eseguire operazioni.
- Le operazioni che possono essere eseguite.
- La risorsa su cui è possibile eseguire le operazioni.

Nell'esempio seguente viene illustrata una policy di endpoint VPC che nega a chiunque l'autorizzazione per eliminare un piano di dimensionamento tramite l'endpoint. Inoltre, la policy di esempio concede a chiunque l'autorizzazione per eseguire tutte le altre operazioni.

```
{
  "Statement": [
    {
      "Action": "*",
```

```
    "Effect": "Allow",
    "Resource": "*",
    "Principal": "*"
  },
  {
    "Action": "autoscaling-plans:DeleteScalingPlan",
    "Effect": "Deny",
    "Resource": "*",
    "Principal": "*"
  }
]
```

Per ulteriori informazioni, consulta [Policy dell'endpoint VPC](#) nella Guida di AWS PrivateLink .

Migrazione degli endpoint

Il 22 novembre 2019, abbiamo introdotto `autoscaling-plans.region.amazonaws.com` come nuovo nome host ed endpoint DNS predefiniti per le chiamate all'API. AWS Auto Scaling Il nuovo endpoint è compatibile con l'ultima versione di and SDK. AWS CLI Se non l'hai già fatto, installa la versione più recente AWS CLI e gli SDK per utilizzare il nuovo endpoint. Per aggiornare il AWS CLI, consulta [Installazione dell' AWS CLI utilizzo di pip](#) nella Guida per l'AWS Command Line Interface utente. Per informazioni sugli AWS SDK, consulta [Tools for Amazon Web Services](#).

Important

Per motivi di compatibilità con le versioni precedenti, l'`autoscaling.region.amazonaws.com` endpoint esistente continuerà a essere supportato per le chiamate all'API. AWS Auto Scaling Per configurare l'endpoint `autoscaling.region.amazonaws.com` come endpoint VPC di interfaccia privato, consulta [Amazon EC2 Auto Scaling ed endpoint VPC di interfaccia](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

Endpoint da chiamare quando si utilizza la CLI o l'API AWS Auto Scaling

Nella versione corrente di AWS Auto Scaling, le chiamate all' AWS Auto Scaling API vanno automaticamente all'`autoscaling-plans.region.amazonaws.com` endpoint anziché `autoscaling.region.amazonaws.com`

È possibile richiamare il nuovo endpoint nell'interfaccia a riga di comando (CLI) utilizzando il parametro seguente con ogni comando per specificare l'endpoint: `--endpoint-url https://autoscaling-plans.region.amazonaws.com`.

Sebbene non sia consigliato, è inoltre possibile richiamare il vecchio endpoint nell'interfaccia a riga di comando (CLI) utilizzando il parametro seguente per specificare l'endpoint: `--endpoint-url https://autoscaling.region.amazonaws.com`.

Per i vari SDK utilizzati per richiamare le API, consultare la documentazione relativa all'SDK di interesse per informazioni su come indirizzare le richieste a un endpoint specifico. Per ulteriori informazioni, consulta [Strumenti per Amazon Web Services](#).

Protezione dei dati

Il modello di [responsabilità AWS condivisa modello](#) di si applica alla protezione dei dati in AWS Auto Scaling. Come descritto in questo modello, AWS è responsabile della protezione dell'infrastruttura globale che gestisce tutti i Cloud AWS. L'utente è responsabile del controllo dei contenuti ospitati su questa infrastruttura. Inoltre, sei responsabile della configurazione della protezione e delle attività di gestione per i Servizi AWS che utilizzi. Per ulteriori informazioni sulla privacy dei dati, vedi [Domande frequenti sulla privacy dei dati](#). Per informazioni sulla protezione dei dati in Europa, consulta il post del blog [AWS Shared Responsibility Model and GDPR](#) nel Blog sulla sicurezza AWS .

Ai fini della protezione dei dati, consigliamo di proteggere Account AWS le credenziali e configurare i singoli utenti con AWS IAM Identity Center or AWS Identity and Access Management (IAM). In tal modo, a ogni utente verranno assegnate solo le autorizzazioni necessarie per svolgere i suoi compiti. Ti suggeriamo, inoltre, di proteggere i dati nei seguenti modi:

- Utilizza l'autenticazione a più fattori (MFA) con ogni account.
- Usa SSL/TLS per comunicare con le risorse. AWS È richiesto TLS 1.2 ed è consigliato TLS 1.3.
- Configura l'API e la registrazione delle attività degli utenti con. AWS CloudTrail
- Utilizza soluzioni di AWS crittografia, insieme a tutti i controlli di sicurezza predefiniti all'interno Servizi AWS.
- Utilizza i servizi di sicurezza gestiti avanzati, come Amazon Macie, che aiutano a individuare e proteggere i dati sensibili archiviati in Amazon S3.

- Se hai bisogno di moduli crittografici convalidati FIPS 140-2 per l'accesso AWS tramite un'interfaccia a riga di comando o un'API, utilizza un endpoint FIPS. Per ulteriori informazioni sugli endpoint FIPS disponibili, consulta il [Federal Information Processing Standard \(FIPS\) 140-2](#).

Ti consigliamo vivamente di non inserire mai informazioni riservate o sensibili, ad esempio gli indirizzi e-mail dei clienti, nei tag o nei campi di testo in formato libero, ad esempio nel campo Nome. Ciò include quando lavori o Servizi AWS utilizzi la console, l'API AWS Auto Scaling o gli SDK. AWS CLI AWS I dati inseriti nei tag o nei campi di testo in formato libero utilizzati per i nomi possono essere utilizzati per la fatturazione o i log di diagnostica. Quando fornisci un URL a un server esterno, ti suggeriamo vivamente di non includere informazioni sulle credenziali nell'URL per convalidare la tua richiesta al server.

Identity and Access Management per piani di dimensionamento

AWS Identity and Access Management (IAM) è uno strumento Servizio AWS che aiuta un amministratore a controllare in modo sicuro l'accesso alle AWS risorse. Gli amministratori IAM controllano chi può essere autenticato (effettuato l'accesso) e autorizzato (disporre delle autorizzazioni) a utilizzare le risorse. AWS Auto Scaling IAM è uno Servizio AWS strumento che puoi utilizzare senza costi aggiuntivi.

Per la documentazione IAM completa, consulta la [Guida per l'utente IAM](#).

Controllo accessi

Per autenticare le richieste, è necessario disporre di credenziali valide, ma a meno che non si disponga delle autorizzazioni non è possibile creare o accedere ai piani di dimensionamento. Ad esempio, è necessario disporre delle autorizzazioni per creare piani di dimensionamento, configurare il dimensionamento predittivo e così via.

Le sezioni seguenti forniscono i dettagli su come un amministratore IAM può utilizzare IAM per proteggere i piani di dimensionamento, controllando chi può lavorarci.

Argomenti

- [Come funzionano i piani di dimensionamento con IAM](#)
- [Ruolo collegato al servizio di dimensionamento predittivo](#)
- [Esempi di policy basate su identità per piani di dimensionamento](#)

Come funzionano i piani di dimensionamento con IAM

Prima di utilizzare IAM per gestire chi può creare, accedere e gestire i piani di AWS Auto Scaling scalabilità, è necessario comprendere quali funzionalità IAM sono disponibili per l'uso con i piani di scalabilità.

Argomenti

- [Policy basate su identità](#)
- [Policy basate su risorse](#)
- [Liste di controllo degli accessi \(ACL\)](#)
- [Autorizzazione basata su tag](#)
- [Ruoli IAM](#)

Policy basate su identità

Con le policy IAM basate su identità, puoi specificare operazioni e risorse consentite o rifiutate, nonché le condizioni in base alle quali le operazioni sono consentite o rifiutate. I piani di dimensionamento supportano operazioni, risorse e chiavi di condizione specifiche. Per informazioni su tutti gli elementi utilizzati in una policy JSON, consulta [Documentazione di riferimento degli elementi delle policy JSON IAM](#) nella Guida per l'utente IAM.

Azioni

Gli amministratori possono utilizzare le policy AWS JSON per specificare chi ha accesso a cosa. Cioè, quale principale può eseguire azioni su quali risorse, e in quali condizioni.

L'elemento `Action` di una policy JSON descrive le azioni che è possibile utilizzare per consentire o negare l'accesso a una policy. Le azioni politiche in genere hanno lo stesso nome dell'operazione AWS API associata. Ci sono alcune eccezioni, ad esempio le azioni di sola autorizzazione che non hanno un'operazione API corrispondente. Esistono anche alcune operazioni che richiedono più operazioni in una policy. Queste operazioni aggiuntive sono denominate operazioni dipendenti.

Includi le operazioni in una policy per concedere le autorizzazioni a eseguire l'operazione associata.

Le operazioni del piano di dimensionamento nelle istruzioni della policy IAM utilizzano il seguente prefisso prima dell'operazione: `autoscaling-plans:`. Le istruzioni della policy devono includere un elemento `Action` o `NotAction`. I piani di dimensionamento hanno un proprio set di operazioni che descrivono le attività che è possibile eseguire con questo servizio.

Per specificare più operazioni in una singola istruzione, separale con virgole, come illustrato nell'esempio seguente.

```
"Action": [  
    "autoscaling-plans:DescribeScalingPlans",  
    "autoscaling-plans:DescribeScalingPlanResources"
```

Puoi specificare più operazioni tramite caratteri jolly (*). Ad esempio, per specificare tutte le operazioni che iniziano con la parola `Describe`, includi la seguente operazione.

```
"Action": "autoscaling-plans:Describe*"
```

Per visualizzare un elenco completo delle operazioni del piano di dimensionamento che possono essere utilizzate nelle istruzioni della policy, consulta [Operazioni, risorse e chiavi di condizione per AWS Auto Scaling](#) in Service Authorization Reference.

Risorse

L'elemento `Resource` specifica l'oggetto o gli oggetti ai quali si applica l'operazione.

I piani di dimensionamento non includono risorse definite dal servizio che possano essere utilizzate come elemento `Resource` di una dichiarazione di policy IAM. Pertanto, non vi sono Amazon Resource Name (ARN) che è possibile utilizzare in una policy IAM. Per controllare l'accesso alle operazioni del piano di dimensionamento, utilizza sempre un * (asterisco) come risorsa durante la scrittura di una policy IAM.

Chiavi di condizione

L'elemento `Condition` (o blocco `Condition`) consente di specificare le condizioni in cui una dichiarazione è attiva. Ad esempio, potresti decidere che una policy venga applicata solo dopo una data specifica. Per esprimere le condizioni, utilizza chiavi di condizione predefinite.

I piani di dimensionamento non forniscono chiavi di condizione specifiche del servizio, ma supportano l'utilizzo di alcune chiavi di condizione globali. Per visualizzare tutte le chiavi di condizione AWS globali, consulta le [chiavi di contesto delle condizioni AWS globali](#) nella Guida per l'utente IAM.

L'elemento `Condition` è facoltativo.

Esempi

Per visualizzare esempi di policy basate su identità per i piani di dimensionamento, consulta [Esempi di policy basate su identità per piani di dimensionamento](#).

Policy basate su risorse

Altri servizi Amazon Web Services, come ad esempio Amazon Simple Storage Service, supportano policy di autorizzazioni basate su risorse. Ad esempio, è possibile allegare una policy di autorizzazione a un bucket S3 per gestire le autorizzazioni di accesso a quel bucket.

I piani di dimensionamento non supportano le policy basate su risorse.

Liste di controllo degli accessi (ACL)

I piani di dimensionamento non supportano le liste di controllo accessi (ACL).

Autorizzazione basata su tag

Non è possibile applicare i tag ai piani di dimensionamento. Non dispongono nemmeno di risorse definite dal servizio che possono essere taggate. Pertanto, non supportano il controllo dell'accesso basato sui tag su una risorsa.

I piani di dimensionamento possono contenere risorse su cui è possibile applicare tag, come i gruppi Auto Scaling, che supportano il controllo dell'accesso in base ai tag. Per ulteriori informazioni, consulta la documentazione per quel Servizio AWS.

Ruoli IAM

Un [ruolo IAM](#) è un'entità all'interno dell' Account AWS che dispone di autorizzazioni specifiche.

Utilizzo di credenziali temporanee

Puoi utilizzare credenziali temporanee per effettuare l'accesso utilizzando la federazione, per assumere un ruolo IAM o per assumere un ruolo multi-account. È possibile ottenere credenziali di sicurezza temporanee chiamando operazioni AWS STS API come [AssumeRoleo GetFederationToken](#).

I piani di dimensionamento supportano l'utilizzo di credenziali temporanee.

Ruoli collegati ai servizi per piani di dimensionamento

AWS Auto Scaling utilizza ruoli collegati ai servizi per le autorizzazioni necessarie per chiamare altri AWS servizi per conto dell'utente. I ruoli collegati ai servizi semplificano la configurazione dei piani di dimensionamento perché permette di evitare l'aggiunta manuale delle autorizzazioni necessarie. Per ulteriori informazioni, consulta [Utilizzo di ruoli collegati a servizi](#) nella Guida per l'utente di IAM .

AWS Auto Scaling utilizza alcuni tipi di ruoli collegati ai servizi per chiamare altri utenti per conto dell'utente quando si utilizza un Servizi AWS piano di scalabilità:

- Ruolo collegato al servizio di scalabilità predittiva: consente AWS Auto Scaling di accedere ai dati metrici storici da CloudWatch. Consente inoltre la creazione di azioni pianificate per i gruppi Auto Scaling in base a una previsione di carico e alla previsione della capacità. Per ulteriori informazioni, consulta [Ruolo collegato al servizio di dimensionamento predittivo](#).
- Ruolo collegato al servizio Amazon EC2 Auto Scaling: AWS Auto Scaling consente di accedere e gestire le politiche di scalabilità di tracciamento degli obiettivi per i gruppi Auto Scaling. Per ulteriori informazioni, consulta [Ruoli collegati ai servizi per Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento.
- Ruolo collegato al servizio Application Auto Scaling: consente di accedere e gestire AWS Auto Scaling le politiche di scalabilità del tracciamento degli obiettivi per altre risorse scalabili. Esiste un ruolo collegato ai servizi per ciascun servizio. Per ulteriori informazioni, consulta [Ruoli collegati ai servizi per Application Auto Scaling](#) nella Guida per l'utente di Application Auto Scaling.

È possibile utilizzare la procedura seguente per stabilire se l'account dispone già di un ruolo collegato al servizio.

Per stabilire se esiste già un ruolo collegato al servizio

1. Aprire la console IAM all'indirizzo <https://console.aws.amazon.com/iam/>.
2. Nel riquadro di navigazione, seleziona Ruoli.
3. Cerca nell'elenco `AWSServiceRole` per trovare i ruoli collegati al servizio presenti nel tuo account. Cerca il nome del ruolo collegato al servizio da controllare.

Ruoli di servizio

AWS Auto Scaling non ha ruoli di servizio per i piani di scalabilità.

Ruolo collegato al servizio di dimensionamento predittivo

AWS Auto Scaling utilizza ruoli collegati ai servizi per le autorizzazioni necessarie per chiamare altri utenti per conto dell'utente quando si utilizza un AWS piano di scalabilità. Per ulteriori informazioni, consulta [Ruoli collegati ai servizi per piani di dimensionamento](#).

Le seguenti sezioni descrivono come creare e gestire i ruoli collegati al servizio per il dimensionamento predittivo. Per iniziare, configura le autorizzazioni per consentire a un'entità IAM

(ad esempio un utente, un gruppo o un ruolo) di creare, modificare o eliminare un ruolo collegato ai servizi.

Autorizzazioni concesse dal ruolo collegato ai servizi

AWS Auto Scaling utilizza il ruolo collegato al servizio denominato `AWSServiceRoleForAutoScalingPlans_EC2 AutoScaling` per richiamare le seguenti azioni per tuo conto quando abiliti la scalabilità predittiva:

- `cloudwatch:GetMetricData`
- `autoscaling:DescribeAutoScalingGroups`
- `autoscaling:DescribeScheduledActions`
- `autoscaling:BatchPutScheduledUpdateGroupAction`
- `autoscaling:BatchDeleteScheduledAction`

`AWSServiceRoleForAutoScalingPlans_EC2 AutoScaling` si fida che il `autoscaling-plans.amazonaws.com` servizio assuma il ruolo.

Creazione di un ruolo collegato ai servizi (automatica)

Non è necessario creare manualmente il `AutoScaling` ruolo `AWSServiceRoleForAutoScalingPlans_EC2`. AWS crea questo ruolo per te quando crei un piano di scalabilità nel tuo account e abiliti la scalabilità predittiva.

AWS Per creare un ruolo collegato ai servizi per tuo conto, devi disporre delle autorizzazioni richieste. Per ulteriori informazioni, consulta [Autorizzazioni del ruolo collegato ai servizi](#) nella Guida per l'utente di IAM.

Creazione di un ruolo collegato ai servizi (manuale)

Per creare manualmente un ruolo collegato ai servizi, è possibile utilizzare la console IAM, la CLI o l'API IAM. Per ulteriori informazioni, consulta [Creazione di un ruolo collegato ai servizi](#) nella Guida per l'utente di IAM.

Come creare un ruolo collegato ai servizi (AWS CLI)

Utilizza il seguente comando [create-service-linked-role](#) CLI per creare il ruolo collegato al servizio.

```
aws iam create-service-linked-role --aws-service-name autoscaling-plans.amazonaws.com
```

Modifica del ruolo collegato ai servizi

Puoi modificare la descrizione di `AWSServiceRoleForAutoScalingPlans_EC2 AutoScaling` utilizzando IAM. Per ulteriori informazioni, consulta [Modifica di un ruolo collegato ai servizi](#) nella Guida per l'utente di IAM.

Eliminazione del ruolo collegato ai servizi

Se non hai più bisogno di utilizzare i piani di scalabilità, ti consigliamo di eliminare `AWSServiceRoleForAutoScalingPlans_EC2 AutoScaling`.

È possibile eliminare un ruolo collegato al servizio solo dopo aver eliminato tutti i piani di dimensionamento nel proprio Account AWS con il dimensionamento predittivo abilitato. Questo impedisce di rimuovere involontariamente l'autorizzazione ad accedere ai piani di dimensionamento.

Per eliminare i ruoli collegati ai servizi, puoi utilizzare la console IAM, la CLI o l'API IAM. Per ulteriori informazioni, consulta [Eliminazione del ruolo collegato al servizio](#) nella Guida per l'utente di IAM.

Dopo aver eliminato il ruolo `AWSServiceRoleForAutoScalingPlans_EC2 AutoScaling` collegato al servizio, AWS Auto Scaling crea nuovamente il ruolo se crei un piano di scalabilità con la scalabilità predittiva abilitata.

Regioni supportate

AWS Auto Scaling supporta l'utilizzo di ruoli collegati ai servizi in tutti i piani di scalabilità disponibili. Regioni AWS Per informazioni sulla disponibilità regionale dei piani di scalabilità, consulta [Endpoint e quote AWS Auto Scaling](#) nella Riferimenti generali di AWS.

Esempi di policy basate su identità per piani di dimensionamento

Per impostazione predefinita, un nuovo utente IAM non ha le autorizzazioni per svolgere alcuna operazione. Un amministratore IAM deve creare e assegnare policy IAM che diano un'autorizzazione di identità IAM (ad esempio un utente o un ruolo) per lavorare con i piani di dimensionamento.

Per informazioni su come creare una policy IAM utilizzando questi documenti di policy JSON di esempio, consulta [Creazione di policy nella scheda JSON](#) nella Guida per l'utente di IAM.

Argomenti

- [Best practice per le policy](#)
- [Consentire agli utenti di creare piani di dimensionamento](#)

- [Consentire agli utenti di abilitare il dimensionamento predittivo](#)
- [Autorizzazioni aggiuntive richieste](#)
- [Autorizzazioni necessarie per creare un ruolo collegato al servizio](#)

Best practice per le policy

Le politiche basate sull'identità determinano se qualcuno può creare, accedere o eliminare risorse nel tuo account. AWS Auto Scaling Queste operazioni possono comportare costi aggiuntivi per l' Account AWS. Quando crei o modifichi policy basate su identità, segui queste linee guida e raccomandazioni:

- Inizia con le policy AWS gestite e passa alle autorizzazioni con privilegi minimi: per iniziare a concedere autorizzazioni a utenti e carichi di lavoro, utilizza le politiche gestite che concedono le autorizzazioni per molti casi d'uso comuni. AWS Sono disponibili nel tuo. Account AWS Ti consigliamo di ridurre ulteriormente le autorizzazioni definendo politiche gestite dai AWS clienti specifiche per i tuoi casi d'uso. Per ulteriori informazioni, consulta [Policy gestite da AWS](#) o [Policy gestite da AWS per le funzioni dei processi](#) nella Guida per l'utente IAM.
- Applica le autorizzazioni con privilegi minimi: quando imposti le autorizzazioni con le policy IAM, concedi solo le autorizzazioni richieste per eseguire un'attività. Puoi farlo definendo le azioni che possono essere intraprese su risorse specifiche in condizioni specifiche, note anche come autorizzazioni con privilegi minimi. Per ulteriori informazioni sull'utilizzo di IAM per applicare le autorizzazioni, consulta [Policy e autorizzazioni in IAM](#) nella Guida per l'utente di IAM.
- Condizioni d'uso nelle policy IAM per limitare ulteriormente l'accesso: per limitare l'accesso a operazioni e risorse puoi aggiungere una condizione alle tue policy. Ad esempio, è possibile scrivere una condizione di policy per specificare che tutte le richieste devono essere inviate utilizzando SSL. Puoi anche utilizzare le condizioni per concedere l'accesso alle azioni del servizio se vengono utilizzate tramite uno specifico Servizio AWS, ad esempio AWS CloudFormation. Per ulteriori informazioni, consulta la sezione [Elementi delle policy JSON di IAM: condizione](#) nella Guida per l'utente di IAM.
- Utilizzo di IAM Access Analyzer per convalidare le policy IAM e garantire autorizzazioni sicure e funzionali: IAM Access Analyzer convalida le policy nuove ed esistenti in modo che aderiscano alla sintassi della policy IAM (JSON) e alle best practice di IAM. IAM Access Analyzer offre oltre 100 controlli delle policy e consigli utili per creare policy sicure e funzionali. Per ulteriori informazioni, consulta [Convalida delle policy per IAM Access Analyzer](#) nella Guida per l'utente di IAM.
- Richiedi l'autenticazione a più fattori (MFA): se hai uno scenario che richiede utenti IAM o un utente root nel Account AWS tuo, attiva l'MFA per una maggiore sicurezza. Per richiedere la MFA

quando vengono chiamate le operazioni API, aggiungi le condizioni MFA alle policy. Per ulteriori informazioni, consulta [Configurazione dell'accesso alle API protetto con MFA](#) nella Guida per l'utente di IAM.

Per maggiori informazioni sulle best practice in IAM, consulta [Best practice di sicurezza in IAM](#) nella Guida per l'utente di IAM.

Consentire agli utenti di creare piani di dimensionamento

Di seguito viene illustrato un esempio di policy basata sull'identità che concede le autorizzazioni per creare piani di dimensionamento.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "autoscaling-plans:*",
        "cloudwatch:PutMetricAlarm",
        "cloudwatch>DeleteAlarms",
        "cloudwatch:DescribeAlarms",
        "cloudformation:ListStackResources"
      ],
      "Resource": "*"
    }
  ]
}
```

Per utilizzare un piano di dimensionamento, l'utente finale deve disporre di autorizzazioni aggiuntive che gli consentano di lavorare con risorse specifiche nel proprio account. Queste autorizzazioni sono elencate in [Autorizzazioni aggiuntive richieste](#).

Ogni utente della console necessita inoltre di autorizzazioni che gli consentano di scoprire le risorse scalabili nel proprio account e di visualizzare grafici dei CloudWatch dati metrici dalla console. AWS Auto Scaling II set aggiuntivo di autorizzazioni necessarie per lavorare con la console è elencato di seguito: AWS Auto Scaling

- `cloudformation:ListStacks`: per elencare stack.
- `tag:GetTagKeys`: per trovare risorse scalabili che contengono determinate chiavi tag.

- `tag:GetTagValues`: per trovare le risorse che contengono determinati valori di tag.
- `autoscaling:DescribeTags`: per trovare gruppi Auto Scaling che contengono determinati tag.
- `cloudwatch:GetMetricData`: per visualizzare i dati nei grafici dei parametri.

Consentire agli utenti di abilitare il dimensionamento predittivo

Di seguito viene illustrato un esempio di policy basata sull'identità che concede le autorizzazioni per abilitare il dimensionamento predittivo. Queste autorizzazioni estendono le funzionalità dei piani di dimensionamento impostati per il dimensionamento dei gruppi Auto Scaling.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "cloudwatch:GetMetricData",
        "autoscaling:DescribeAutoScalingGroups",
        "autoscaling:DescribeScheduledActions",
        "autoscaling:BatchPutScheduledUpdateGroupAction",
        "autoscaling:BatchDeleteScheduledAction"
      ],
      "Resource": "*"
    }
  ]
}
```

Autorizzazioni aggiuntive richieste

Per configurare correttamente i piani di dimensionamento, agli utenti finali devono essere concesse le autorizzazioni per ogni servizio di destinazione per il quale configureranno il dimensionamento. Per concedere le autorizzazioni minime necessarie per lavorare con i servizi di destinazione, leggi le informazioni in questa sezione e specifica le azioni pertinenti nell'elemento `Action` di una dichiarazione di policy IAM.

Gruppi Auto Scaling

Per aggiungere gruppi Auto Scaling a un piano di dimensionamento, gli utenti devono disporre delle seguenti autorizzazioni da Amazon EC2 Auto Scaling:

- `autoscaling:UpdateAutoScalingGroup`
- `autoscaling:DescribeAutoScalingGroups`
- `autoscaling:PutScalingPolicy`
- `autoscaling:DescribePolicies`
- `autoscaling>DeletePolicy`

Servizi ECS

Per aggiungere servizi ECS a un piano di dimensionamento, gli utenti devono disporre delle seguenti autorizzazioni da Amazon ECS e Application Auto Scaling:

- `ecs:DescribeServices`
- `ecs:UpdateService`
- `application-autoscaling:RegisterScalableTarget`
- `application-autoscaling:DescribeScalableTargets`
- `application-autoscaling:DeregisterScalableTarget`
- `application-autoscaling:PutScalingPolicy`
- `application-autoscaling:DescribeScalingPolicies`
- `application-autoscaling>DeleteScalingPolicy`

parco istanze spot

Per aggiungere serie di istanze Spot a un piano di dimensionamento, gli utenti devono disporre delle seguenti autorizzazioni da Amazon EC2 e Application Auto Scaling:

- `ec2:DescribeSpotFleetRequests`
- `ec2:ModifySpotFleetRequest`
- `application-autoscaling:RegisterScalableTarget`
- `application-autoscaling:DescribeScalableTargets`
- `application-autoscaling:DeregisterScalableTarget`
- `application-autoscaling:PutScalingPolicy`
- `application-autoscaling:DescribeScalingPolicies`

- `application-autoscaling:DeleteScalingPolicy`

Tabelle o indici globali DynamoDB

Per aggiungere tabelle o indici globali DynamoDB a un piano di dimensionamento, gli utenti devono disporre delle seguenti autorizzazioni da DynamoDB e Application Auto Scaling:

- `dynamodb:DescribeTable`
- `dynamodb:UpdateTable`
- `application-autoscaling:RegisterScalableTarget`
- `application-autoscaling:DescribeScalableTargets`
- `application-autoscaling:DeregisterScalableTarget`
- `application-autoscaling:PutScalingPolicy`
- `application-autoscaling:DescribeScalingPolicies`
- `application-autoscaling>DeleteScalingPolicy`

Cluster di database Aurora

Per aggiungere cluster Aurora DB a un piano di dimensionamento, gli utenti devono disporre delle seguenti autorizzazioni da Amazon Aurora e Application Auto Scaling:

- `rds:AddTagsToResource`
- `rds>CreateDBInstance`
- `rds>DeleteDBInstance`
- `rds:DescribeDBClusters`
- `rds:DescribeDBInstances`
- `application-autoscaling:RegisterScalableTarget`
- `application-autoscaling:DescribeScalableTargets`
- `application-autoscaling:DeregisterScalableTarget`
- `application-autoscaling:PutScalingPolicy`
- `application-autoscaling:DescribeScalingPolicies`
- `application-autoscaling>DeleteScalingPolicy`

Autorizzazioni necessarie per creare un ruolo collegato al servizio

AWS Auto Scaling richiede le autorizzazioni per creare un ruolo collegato al servizio la prima volta che un utente Account AWS crea un piano di scalabilità con la scalabilità predittiva abilitata. Se il ruolo collegato al servizio non esiste già, lo crea nel tuo account. AWS Auto Scaling Il ruolo collegato al servizio concede le autorizzazioni in AWS Auto Scaling modo che possa chiamare altri servizi per tuo conto.

Affinché la creazione automatica di un ruolo riesca, gli utenti devono disporre delle autorizzazioni per l'operazione `iam:CreateServiceLinkedRole`.

```
"Action": "iam:CreateServiceLinkedRole"
```

Di seguito viene illustrato un esempio di policy basata sull'identità che concede le autorizzazioni per creare un ruolo collegato al servizio.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "iam:CreateServiceLinkedRole",
      "Resource": "arn:aws:iam::*:role/aws-service-role/autoscaling-plans.amazonaws.com/AWSServiceRoleForAutoScalingPlans_EC2AutoScaling",
      "Condition": {
        "StringLike": {
          "iam:AWSServiceName": "autoscaling-plans.amazonaws.com"
        }
      }
    }
  ]
}
```

Per ulteriori informazioni, consulta [Ruolo collegato al servizio di dimensionamento predittivo](#).

Convalida della conformità


Per sapere se un Servizio AWS programma rientra nell'ambito di specifici programmi di conformità, consulta Servizi AWS la sezione [Scope by Compliance Program Servizi AWS](#) e scegli il programma

di conformità che ti interessa. Per informazioni generali, consulta Programmi di [AWS conformità](#) [Programmi](#) di di .

È possibile scaricare report di audit di terze parti utilizzando AWS Artifact. Per ulteriori informazioni, consulta [Scaricamento dei report in AWS Artifact](#) .

La vostra responsabilità di conformità durante l'utilizzo Servizi AWS è determinata dalla sensibilità dei dati, dagli obiettivi di conformità dell'azienda e dalle leggi e dai regolamenti applicabili. AWS fornisce le seguenti risorse per contribuire alla conformità:

- [Guide introduttive su sicurezza e conformità](#): queste guide all'implementazione illustrano considerazioni sull'architettura e forniscono passaggi per implementare ambienti di base incentrati sulla AWS sicurezza e la conformità.
- [Progettazione per la sicurezza e la conformità HIPAA su Amazon Web Services](#): questo white paper descrive in che modo le aziende possono utilizzare AWS per creare applicazioni idonee all'HIPAA.

 Note

Non Servizi AWS tutte sono idonee all'HIPAA. Per ulteriori informazioni, consulta la sezione [Riferimenti sui servizi conformi ai requisiti HIPAA](#).

- [AWS Risorse per](#) la per la conformità: questa raccolta di cartelle di lavoro e guide potrebbe essere valida per il tuo settore e la tua località.
- [AWS Guide alla conformità dei clienti](#): comprendi il modello di responsabilità condivisa attraverso la lente della conformità. Le guide riassumono le migliori pratiche per la protezione Servizi AWS e mappano le linee guida per i controlli di sicurezza su più framework (tra cui il National Institute of Standards and Technology (NIST), il Payment Card Industry Security Standards Council (PCI) e l'International Organization for Standardization (ISO)).
- [Valutazione delle risorse con regole](#) nella Guida per gli AWS Config sviluppatori: il AWS Config servizio valuta la conformità delle configurazioni delle risorse alle pratiche interne, alle linee guida e alle normative del settore.
- [AWS Security Hub](#)— Ciò Servizio AWS fornisce una visione completa dello stato di sicurezza interno. AWS La Centrale di sicurezza utilizza i controlli di sicurezza per valutare le risorse AWS e verificare la conformità agli standard e alle best practice del settore della sicurezza. Per un elenco dei servizi e dei controlli supportati, consulta la pagina [Documentazione di riferimento sui controlli della Centrale di sicurezza](#).

- [AWS Audit Manager](#)— Ciò Servizio AWS consente di verificare continuamente AWS l'utilizzo per semplificare la gestione dei rischi e la conformità alle normative e agli standard di settore.

Sicurezza dell'infrastruttura

In quanto servizio gestito, AWS Auto Scaling è protetto dalla sicurezza di rete AWS globale. Per informazioni sui servizi AWS di sicurezza e su come AWS protegge l'infrastruttura, consulta [AWS Cloud Security](#). Per progettare il tuo AWS ambiente utilizzando le migliori pratiche per la sicurezza dell'infrastruttura, vedi [Infrastructure Protection](#) in Security Pillar AWS Well-Architected Framework.

Utilizzate chiamate API AWS pubblicate per accedere AWS Auto Scaling attraverso la rete. I client devono supportare quanto segue:

- Transport Layer Security (TLS). È richiesto TLS 1.2 ed è consigliato TLS 1.3.
- Suite di cifratura con Perfect Forward Secrecy (PFS), ad esempio Ephemeral Diffie-Hellman (DHE) o Elliptic Curve Ephemeral Diffie-Hellman (ECDHE). La maggior parte dei sistemi moderni, come Java 7 e versioni successive, supporta tali modalità.

Inoltre, le richieste devono essere firmate utilizzando un ID chiave di accesso e una chiave di accesso segreta associata a un principale IAM. In alternativa, è possibile utilizzare [AWS Security Token Service](#) (AWS STS) per generare le credenziali di sicurezza temporanee per sottoscrivere le richieste.

Quote per i piani di dimensionamento

L'account Account AWS dispone delle seguenti quote di servizio, in precedenza definite limiti, correlate ai piani di dimensionamento.

Per richiedere un aumento, utilizza il [modulo dei limiti di Auto Scaling](#). Assicurati di specificare il tipo di risorsa con la richiesta di aumento, ad esempio, Amazon EC2 Auto Scaling, Amazon ECS o Dynamo DB.

Quote predefinite per Regione per account

Elemento	Default (Predefinito)
Numero massimo di risorse scalabili per tipo di risorsa	Le quote variano a seconda del tipo di risorsa. Amazon DynamoDB: 3000 Gruppi Amazon EC2 Auto Scaling: 200 Tutti gli altri tipi di risorse: 500
Numero massimo di piani di dimensionamento	100
Numero massimo di istruzioni di dimensionamento per piano di dimensionamento	500
Numero massimo di configurazioni di monitoraggio obiettivi per istruzione di dimensionamento	10

Tieni a mente le quote di servizio mentre dimensioni i carichi di lavoro. Ad esempio, quando raggiungi il numero massimo di unità di capacità consentite da un servizio, la scalabilità orizzontale si interromperà. Se la domanda scende e la capacità attuale diminuisce, AWS Auto Scaling può scalare di nuovo. Per evitare di raggiungere nuovamente questo limite di quota di servizio, puoi richiedere un aumento. Ogni servizio ha le proprie quote predefinite per la capacità massima della risorsa. Per informazioni sulle quote predefinite per altri servizi Amazon Web Services, consulta [Endpoint e quote di servizio](#) nella Riferimenti generali di Amazon Web Services.

Informazioni correlate

Le seguenti risorse aggiuntive possono rivelarsi utili durante l'utilizzo di questo servizio.

Risorsa	Descrizione
Consulta la sezione Endpoint e quote AWS Auto Scaling in Riferimenti generali di AWS.	Informazioni su regioni ed endpoint AWS Auto Scaling.
AWS Auto Scaling Documentazione di riferimento delle API	La documentazione per ogni operazione API mostra la sintassi della richiesta, i parametri della richiesta, gli elementi di risposta e fornisce collegamenti ad argomenti di riferimento SDK specifici per lingua.
autoscaling-plans in Riferimento ai comandi della AWS CLI	Descrizioni dei comandi della AWS CLI che possono essere utilizzati con i piani di dimensionamento.
Guida per l'utente di AWS CloudFormation	Il riferimento AWS::AutoScalingPlans::ScalingPlan nella risorsa Guida per l'utente di AWS CloudFormation ti permette di creare, modellare e gestire i piani di dimensionamento senza operazioni manuali.
Pagina del prodotto	Pagina Web principale che include informazioni su AWS Auto Scaling.
AWS re:Post	Servizio di domande e risposte gestito da AWS che offre risposte collettive e recensite da esperti alle tue domande tecniche.
Guida per l'utente di AWS CloudTrail	Scopri come utilizzare AWS CloudTrail per le chiamate di registro effettuate all'API AWS Auto Scaling e conservale in Amazon S3. È possibile utilizzare i log di CloudTrail per determinare quali chiamate sono state effettuate, l'indirizzo

Risorsa	Descrizione
	IP di origine da cui proveniva la chiamata, chi l'ha effettuata, quando e così via.
Policy di tracciamento e dimensionamento della destinazione per dimensionamento automatico Amazon EC2 nella Guida per l'utente sul dimensionamento automatico Amazon EC2	Scopri come dimensionare i tuoi gruppi con dimensionamento automatico utilizzando le politiche di dimensionamento per il monitoraggio della destinazione.
Policy di dimensionamento per il monitoraggio della destinazione nella Guida per l'utente sul dimensionamento automatico delle applicazioni.	Scopri come dimensionare le risorse scalabili aggiuntive, come indici e tabelle Dynamo DB e servizi Amazon ECS, usando le politiche di dimensionamento del monitoraggio della destinazione.
Dimensionamento predittivo per dimensionamento automatico Amazon EC2 nella Guida per l'utente sul dimensionamento automatico Amazon EC2	Questa è la versione più recente del dimensionamento predittivo, rilasciata a maggio 2021. Alcune funzionalità introdotte in questa versione non sono disponibili nei piani di scalabilità e per accedere a tali funzionalità è necessario utilizzare una politica di scalabilità predittiva impostata direttamente nel gruppo con dimensionamento automatico.

Sono disponibili le seguenti risorse generali per fornire ulteriori informazioni su AWS.

- [Corsi e workshop](#): collegamenti a corsi basati su ruoli e di specializzazione nonché a corsi gestiti dall'utente per affinare le proprie competenze su AWS e acquisire esperienza pratica.
- [Centro sviluppatori AWS](#): esplora i tutorial, scarica gli strumenti e scopri gli eventi destinati agli sviluppatori AWS.
- [Strumenti per sviluppatori AWS](#): collegamenti a strumenti per sviluppatori, SDK, kit di strumenti IDE e strumenti a riga di comando per lo sviluppo e la gestione delle applicazioni AWS.
- [Centro risorse per le nozioni di base](#): scopri come configurare il tuo Account AWS, unisciti alla community AWS e lancia la tua prima applicazione.
- [Tutorial pratici](#): segui i tutorial dettagliati per avviare la tua prima applicazione su AWS.

- [Whitepaper AWS](#): collegamenti a un elenco completo di whitepaper tecnici AWS, relativi ad argomenti come architettura, sicurezza ed economia, creati da AWS Solutions Architect o da altri esperti tecnici.
- [AWS SupportCentro](#) : il centro in cui creare e gestire i tuoi casi AWS Support. Include inoltre link ad altre risorse utili, quali forum, domande frequenti di tipo tecnico, stato d'integrità del servizio e AWS Trusted Advisor.
- [AWS Support](#): la pagina Web principale che include le informazioni su AWS Support, un canale di assistenza rapida individuale che aiuta a creare ed eseguire applicazioni nel cloud.
- [Contatti](#) - Un punto di contatto centrale per richieste relative a fatturazione, account, eventi, uso illecito e altre questioni relative ad AWS.
- [AWS Termini di utilizzo del sito](#): informazioni dettagliate sul copyright e i marchi, l'account, la licenza, l'accesso al sito e altri argomenti.

Cronologia dei documenti per i piani di ridimensionamento

Nella seguente tabella sono descritte importanti aggiunte alla documentazione di AWS Auto Scaling. Per ricevere notifiche sugli aggiornamenti della documentazione, puoi sottoscrivere il feed RSS.

Modifica	Descrizione	Data
Nuovo capitolo sulla sicurezza	Un nuovo capitolo sulla Sicurezza nella Guida per l'utente di AWS Auto Scaling fornisce informazioni su come applicare il modello di responsabilità condivisa quando si utilizza AWS Auto Scaling. Come parte di questo aggiornamento, il capitolo della guida per l'utente "Autenticazione e controllo degli accessi" è stato sostituito da una nuova sezione più utile, Identity and Access Management per AWS Auto Scaling .	12 marzo 2020
Supporto per gli endpoint Amazon VPC	Ora è possibile stabilire una connessione privata tra il VPC e AWS Auto Scaling. Per informazioni e istruzioni sulla migrazione, consulta l'argomento Piani di dimensionamento ed endpoint VPC di interfaccia .	22 novembre 2019
Supporto per aumentare la capacità massima al di sopra di quella di previsione, oltre a modifiche alla guida	Aggiunge il supporto per console per consentire al piano di dimensionamento di aumentare la capacità	9 marzo 2019

massima al di sopra di quella di previsione di un valore di buffer specificato. Per ulteriori informazioni, consulta [Impostazioni del dimensionamento predittivo](#) nella Guida per l'utente di AWS Auto Scaling. Questa versione include anche diverse sezioni riscritte nel tutorial [Nozioni di base su AWS Auto Scaling](#).

[Dimensionamento predittivo e miglioramenti](#)

Ora puoi utilizzare il dimensionamento predittivo per dimensionare in modo proattivo i gruppi Amazon EC2 Auto Scaling. Questa release, inoltre, aggiunge il supporto per la sostituzione delle policy di dimensionamento create al di fuori del piano di dimensionamento (ad esempio da altre console) e il controllo se si abilita la funzione di dimensionamento dinamico del piano. Per ulteriori informazioni, consulta [Nozioni di base su AWS Auto Scaling](#).

20 novembre 2018

[Supporto delle impostazioni per le risorse personalizzate](#)

Aggiunto il supporto per la personalizzazione di diverse impostazioni per ogni singola risorsa o più risorse nello stesso momento. Per ulteriori informazioni, consulta [Nozioni di base su AWS Auto Scaling](#).

9 ottobre 2018

[I tag come origine dell'applicazione](#)

Questa versione aggiunge supporto per specificare un set di tag come origine di un'applicazione.

23 aprile 2018

[Nuovo servizio](#)

Versione iniziale di AWS Auto Scaling.

16 gennaio 2018

Le traduzioni sono generate tramite traduzione automatica. In caso di conflitto tra il contenuto di una traduzione e la versione originale in Inglese, quest'ultima prevarrà.