



Guida per l'utente

Dimensionamento automatico Amazon EC2



Dimensionamento automatico Amazon EC2: Guida per l'utente

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

I marchi e il trade dress di Amazon non possono essere utilizzati in relazione a prodotti o servizi che non siano di Amazon, in qualsiasi modo che possa causare confusione tra i clienti o in qualsiasi modo 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 Amazon EC2 Auto Scaling?	1
Caratteristiche di Amazon EC2 Auto Scaling	1
Prezzi per Amazon EC2 Auto Scaling	3
Inizia a usare	4
Utilizzo dei gruppi Auto Scaling	4
Vantaggi di Auto Scaling	5
Esempio: copertura delle necessità variabili	5
Esempio: architettura per applicazioni Web	7
Esempio: distribuzione di istanze tra le zone di disponibilità	9
Ciclo di vita dell'istanza	12
Aumento orizzontale	13
Istanze in servizio	13
Riduzione orizzontale	14
Come distaccare un'istanza	15
Come allegare un'istanza	15
Hook del ciclo di vita	15
Come entrare e uscire dallo stato di standby	16
Quote	16
Limitazione delle richieste per l'API Amazon EC2 Auto Scaling	19
Tassi di terminazione EC2	19
Altri servizi	19
Configurazione	20
Preparazione all'utilizzo di Amazon EC2	20
Preparazione all'utilizzo della AWS CLI	20
Inizia a usare	21
Tutorial: crea il tuo primo gruppo Auto Scaling	22
Preparazione della spiegazione passo per passo	22
Fase 1: creazione di un modello di avvio	23
Fase 2: creazione di un gruppo con dimensionamento automatico con una singola istanza ...	24
Fase 3: verifica del gruppo con dimensionamento automatico	25
Fase 4: termina un'istanza nel gruppo con dimensionamento automatico	26
Fase 5: fasi successive	27
Fase 6: pulizia	27
Tutorial: Configurazione di un'applicazione con dimensionamento e bilanciamento del carico	28

Prerequisiti	30
Fase 1: configurazione di un modello di avvio o una configurazione di avvio	31
Fase 2: creazione di un gruppo con dimensionamento automatico	35
Fase 3: come verificare che il load balancer sia allegato	36
Fase 4: Fasi successive	37
Fase 5: rimozione	37
Risorse correlate	39
Modelli di avvio	40
Autorizzazioni per lavorare con i modelli di lancio	41
Operazioni API supportate dai modelli di avvio	41
Creare un modello di avvio per un gruppo con dimensionamento automatico	41
Crea il tuo modello di avvio (console)	42
Modifica le impostazioni dell'interfaccia di rete di default (console)	45
Modifica la configurazione di archiviazione (console)	47
Creare un modello di avvio da un'istanza esistente (console)	51
Risorse correlate	51
Limitazioni	51
Crea un modello di avvio utilizzando le impostazioni avanzate	52
Impostazioni richieste	52
Impostazioni avanzate	53
Richiedi istanze spot	57
Blocchi di capacità per ML	59
Migra i tuoi gruppi di Auto Scaling ai modelli di avvio	64
Passo 1: ricerca dei gruppi con dimensionamento automatico che utilizzano configurazioni di avvio	65
Passo 2: copia di una configurazione di avvio in un modello di avvio	67
Passo 3: aggiornare un gruppo con dimensionamento automatico per utilizzare un modello di avvio	68
Passo 4: sostituzione delle istanze	69
Informazioni aggiuntive	70
Migra gli CloudFormation stack verso i modelli di avvio	70
Trova gruppi con dimensionamento automatico che utilizzano una configurazione di avvio	71
Aggiornare uno stack per utilizzare un modello di avvio	71
Comprendi l'aggiornamento dei comportamenti delle risorse stack	75
Monitora la migrazione	76
Riferimento della mappatura di avvio	77

AWS CLI esempi per lavorare con i modelli di lancio	78
Esempio di utilizzo	79
Creare un modello di avvio di base	80
Specificare dei tag che etichettano le istanze all'avvio	81
Specificare un ruolo IAM da passare alle istanze	81
Assegnare un indirizzo IP pubblico	81
Specificare uno script di dati utente che configura le istanze all'avvio	82
Specificare di una mappatura dei dispositivi a blocchi di un'AMI	82
Specificare gli host dedicati per portare licenze software da fornitori esterni	82
Specificare un'interfaccia di rete esistente	83
Creare più interfacce di rete	83
Gestire i modelli di avvio	84
Aggiornare un gruppo con dimensionamento automatico per utilizzare un modello di avvio ...	87
Usa i parametri di Systems Manager anziché gli ID AMI	88
Crea un modello di avvio che specifichi un parametro per l'AMI	88
Verifica che un modello di avvio ottenga l'ID AMI corretto	93
Risorse correlate	94
Limitazioni	94
Configurazioni di avvio	96
Creazione di una configurazione di avvio	96
Creazione di una configurazione di avvio	97
Configurazione di IMDS	100
Creazione di una configurazione di avvio utilizzando un'istanza EC2	102
Modifica una configurazione di avvio	107
Gruppi Auto Scaling	109
Creare gruppi con dimensionamento automatico utilizzando modelli di avvio	111
Creare un gruppo utilizzando un modello di avvio	111
Creare un gruppo utilizzando la procedura guidata di EC2	115
Utilizzare più tipi di istanze e opzioni di acquisto	119
Creare gruppi con dimensionamento automatico utilizzando configurazioni di avvio	165
Creare un gruppo utilizzando una configurazione di avvio	166
Creare un gruppo utilizzando un'istanza EC2	170
Aggiornamento di un gruppo con dimensionamento automatico	176
Aggiornamento di istanze di dimensionamento automatico	177
Assegnazione di tag a gruppi e istanze	178
Limitazioni all'uso e alla denominazione dei tag	179

Ciclo di vita dell'assegnazione di tag all'istanza EC2	180
Assegnazione dei tag ai gruppi Auto Scaling	180
Eliminazione dei tag	184
Tag di sicurezza	185
Controllo dell'accesso ai tag	186
Usa i tag per filtrare i gruppi Auto Scaling	187
Policy di manutenzione delle istanze	190
Panoramica	191
Impostare una policy di manutenzione delle istanze per il tuo gruppo	199
Hook del ciclo di vita	204
Disponibilità dell'hook del ciclo di vita	205
Considerazioni e limitazioni	205
Risorse correlate	207
Come funzionano gli hook del ciclo di vita	208
Preparazione all'aggiunta di un hook del ciclo di vita	210
Recupero dello stato del ciclo di vita di destinazione	218
Aggiungere un hook del ciclo di vita	220
Completare un'operazione del ciclo di vita	224
Tutorial: configurazione dei dati utente per recuperare lo stato del ciclo di vita di destinazione tramite i metadati dell'istanza	226
Tutorial : Configurazione di un hook del ciclo di vita che richiama una funzione Lambda	235
Warm pool	245
Concetti principali	246
Prerequisiti	248
Aggiornamento di istanze in un pool caldo	249
Risorse correlate	250
Limitazioni	250
Utilizzare un hook del ciclo di vita	251
Creazione di un pool caldo per un gruppo con dimensionamento automatico	256
Visualizzare lo stato del controllo dell'integrità	257
AWS CLI esempi di utilizzo di piscine calde	261
Scollega le istanze	263
Considerazioni sul distacco delle istanze	264
Considerazioni per il collegamento delle istanze	265
Sposta un'istanza in un gruppo diverso utilizzando detach e attach	266
Rimuovere temporaneamente le istanze	271

Come funziona lo stato di standby	271
Considerazioni	272
Stato di integrità di un'istanza in standby	273
Rimuovi temporaneamente un'istanza impostandola in standby	271
Eliminazione dell'infrastruttura Auto Scaling	277
Elimina il gruppo con scalabilità automatica	278
(Facoltativo) Elimina la configurazione di avvio	279
(Facoltativo) Elimina il modello di avvio	279
(Facoltativo) Elimina il load balancer e i gruppi di destinazione	280
(Facoltativo) Eliminare gli allarmi CloudWatch	281
AWS Esempi SDK per lavorare con i gruppi di Auto Scaling	282
Creazione di un gruppo con dimensionamento automatico	282
Aggiornamento di un gruppo con dimensionamento automatico	298
Descrivi un gruppo di Auto Scaling	308
Eliminazione di un gruppo con dimensionamento automatico	323
Ricicla le tue istanze	336
Aggiornamento istanza	337
Come funziona l'aggiornamento di un'istanza	337
Informazioni sui valori predefiniti	343
Avvio dell'aggiornamento di un'istanza	346
Monitora l'aggiornamento di un'istanza	359
Annullamento dell'aggiornamento di un'istanza	362
Annullamento delle modifiche tramite un rollback	363
Utilizzo dell'opzione Ignora la corrispondenza	368
Aggiunta di checkpoint	378
Durata massima dell'istanza	384
Considerazioni	384
Impostazione della durata massima dell'istanza	385
Limitazioni	386
Dimensionamento del gruppo	388
Scegli il tuo metodo di dimensionamento	389
Impostare i limiti di dimensionamento	390
Impostazione della preparazione dell'istanza di default	392
Considerazioni sulle prestazioni di dimensionamento	393
Scegli il tempo di riscaldamento predefinito dell'istanza	394
Abilitazione della preparazione di default dell'istanza per un gruppo	395

Verificare il riscaldamento predefinito dell'istanza per un gruppo	397
Trova le politiche di scalabilità con un tempo di riscaldamento dell'istanza impostato in precedenza	397
Cancela la preparazione dell'istanza precedentemente impostata per una policy di dimensionamento	399
Dimensionamento manuale	399
Modifica la capacità desiderata del gruppo con dimensionamento automatico	400
Terminare un'istanza nel gruppo con dimensionamento automatico (AWS CLI)	404
Dimensionamento programmato	405
Come funziona il dimensionamento programmato	406
Pianificazioni ricorrenti	406
Time zone (Fuso orario)	407
Considerazioni	408
Creazione di un'operazione pianificata	408
Visualizza i dettagli delle azioni pianificate	410
Verificare un'attività di dimensionamento	411
Eliminazione di un'operazione pianificata	412
Limitazioni	412
Dimensionamento dinamico	413
Come funziona il dimensionamento dinamico	414
Più policy di dimensionamento	415
Policy di dimensionamento con monitoraggio degli obiettivi	417
Policy di dimensionamento per fasi e di dimensionamento semplice	431
Tempo di raffreddamento per il dimensionamento	449
Scalabilità basata su Amazon SQS	452
Verificare un'attività di dimensionamento	460
Disabilitare una policy di dimensionamento	462
Eliminazione di una policy di dimensionamento	465
AWS CLI esempi di politiche di scalabilità	468
Dimensionamento predittivo	471
Funzionamento del dimensionamento predittivo	472
Crea una politica di scalabilità predittiva	475
Valutazione delle policy di dimensionamento predittivo	484
Sovrascrivere la previsione	493
Utilizzare parametri personalizzato	498
Verificare la terminazione dell'istanza	510

Scenari di policy di terminazione	510
Configura le politiche di terminazione	515
Creare una policy di terminazione personalizzata con Lambda	521
Utilizzare la protezione di riduzione orizzontale dell'istanza	527
Progettazione per la chiusura graduale delle istanze	532
Sospensione e ripresa dei processi	536
Tipi di processi	536
Considerazioni	537
Sospendere i processi	538
Riprendere i processi	539
In che modo i processi sospesi influiscono sugli altri processi	540
Monitoraggio	544
Controlli dell'integrità	546
Informazioni sui controlli dell'integrità	547
Come visualizzare il motivo dei fallimenti dei controlli dell'integrità	555
Impostazione del periodo di tolleranza del controllo dell'integrità	556
Monitora con AWS Health Dashboard	559
Monitoraggio dei parametri CloudWatch	560
Visualizzare i grafici di monitoraggio nella console Dimensionamento automatico Amazon EC2	561
Parametri CloudWatch per Dimensionamento automatico Amazon EC2	566
Configurare il monitoraggio per le istanze Auto Scaling	574
Registra le chiamate API con AWS CloudTrail	577
Informazioni su Amazon EC2 Auto Scaling in CloudTrail	577
Comprendere delle voci di file di log di Amazon EC2 Auto Scaling	578
Risorse correlate	580
Opzioni di notifica Amazon SNS	580
Amazon SNS e Amazon EC2 Auto Scaling	581
Utilizzare altri Servizi	588
Ribilanciamento della capacità	588
Panoramica	589
Comportamento di ribilanciamento della capacità	590
Considerazioni	591
Abilitazione del ribilanciamento della capacità (console)	593
Abilitare il ribilanciamento della capacità. (AWS CLI)	594
Risorse correlate	599

Limitazioni	599
Prenotazioni della capacità	600
Fase 1: creazione di prenotazioni della capacità	601
Fase 2: creazione di un gruppo di prenotazione della capacità	603
Fase 3: creazione di un modello di avvio	605
Fase 4: creazione di un gruppo con dimensionamento automatico	607
Risorse correlate	609
AWS CloudShell	609
AWS CloudFormation	610
Auto Scaling e modelli di Amazon EC2 AWS CloudFormation	610
Scopri di più su AWS CloudFormation	611
Compute Optimizer	611
Limitazioni	612
Risultati	612
Visualizzare le raccomandazioni	613
Considerazioni sulla valutazione dei suggerimenti	614
Sistema di bilanciamento del carico elastico	615
Tipi di sistema di bilanciamento del carico elastico	616
Preparati a collegare un sistema di bilanciamento del carico	617
Come collegare un sistema di bilanciamento del carico	620
Configurazione di un sistema di bilanciamento del carico dalla console Dimensionamento automatico Amazon EC2	624
Verifica lo stato di collegamento	625
Aggiungere zone di disponibilità	627
AWS CLI esempi di utilizzo di Elastic Load Balancing	630
VPC Lattice	638
Prepararsi a collegare un gruppo target	640
Collega un gruppo target VPC Lattice	644
Verifica lo stato di collegamento	649
EventBridge	650
Riferimento eventi Dimensionamento automatico Amazon EC2	650
Esempi di eventi e modelli relativi al pool caldo	660
Crea EventBridge regole	665
Amazon VPC	670
VPC di default	671
VPC non di default	672

Considerazioni sulla scelta di sottoreti VPC	672
Indirizzamento IP in un VPC	673
Interfacce di rete in un VPC	673
Tenancy di posizionamento delle istanze	674
AWS Outposts	674
Più risorse per l'approfondimento della conoscenza di VPC	674
Sicurezza	676
Sicurezza dell'infrastruttura	676
Risorse correlate	677
Resilienza	677
Risorse correlate	679
Protezione dei dati	679
Utilizzalo AWS KMS keys per crittografare i volumi Amazon EBS	680
Risorse correlate	681
AWS KMS politica chiave da utilizzare con volumi crittografati	681
Identity and Access Management	687
Controllo accessi	688
Funzionamento di Dimensionamento automatico Amazon EC2 con IAM	689
Autorizzazioni API	699
Policy gestite	700
Ruoli collegati ai servizi	704
Esempi di policy basate su identità	713
Prevenzione del problema "confused deputy" tra servizi	722
Supporto modello di avvio	724
Ruoli IAM per le applicazioni in esecuzione sulle istanze Amazon EC2	732
Convalida della conformità	736
Conformità PCI DSS	737
Utilizzare gli endpoint VPC per la connettività privata	737
Creazione di un endpoint VPC dell'interfaccia	738
Creazione di una policy di endpoint VPC	738
Risoluzione dei problemi	740
Recuperare un messaggio di errore	740
Disattiva le attività di scalabilità	742
Altre risorse per la risoluzione dei problemi	743
Avvio di un'istanza non riuscito	744
La configurazione richiesta non è attualmente supportata.	745

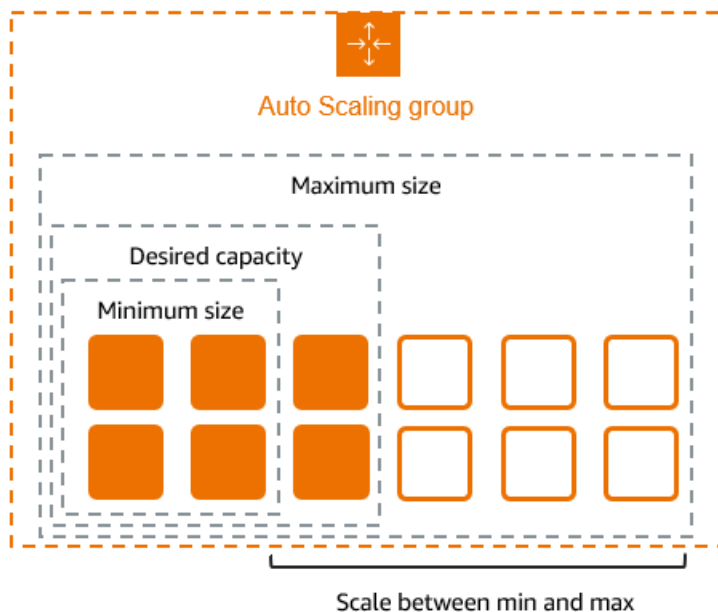
Il gruppo di sicurezza <nome del gruppo di sicurezza> non esiste. Avvio di un'istanza EC2 non riuscito.	746
La coppia di chiavi <coppia di chiavi associata all'istanza EC2> non esiste. Avvio di un'istanza EC2 non riuscito.	746
Il tipo di istanza richiesto (<tipo di istanza>) non è supportato nella zona di disponibilità richiesta (<zona di disponibilità dell'istanza>)... ..	747
Il prezzo della richiesta Spot pari a 0,015 è inferiore al prezzo minimo richiesto di gestione di una Richiesta Spot pari a 0,0735... ..	747
Nome dispositivo <device name> non valido/Caricamento del nome del dispositivo non valido. Avvio di un'istanza EC2 non riuscito.	747
Il valore (<nome associato al dispositivo di archiviazione dell'istanza>) per il parametro virtualName non è valido... Avvio di un'istanza EC2 non riuscito.	748
Mappature dei dispositivi a blocchi EBS non supportate per le AMI di instance-store.	748
I gruppi di collocamento non possono essere usati con le istanze di tipo “<instance type>”. Avvio di un'istanza EC2 non riuscito.	749
Cliente. InternalError: errore del client all'avvio.	749
Al momento non disponiamo di una sufficiente capacità di tipo di istanza nella zona di disponibilità richiesta. Avvio di un'istanza EC2 non riuscito.	750
La prenotazione richiesta non ha una capacità compatibile e disponibile sufficiente per questa richiesta. Avvio di un'istanza EC2 non riuscito.	751
La tua prenotazione Blocchi di capacità <reservation id> non è ancora attiva. Avvio di un'istanza EC2 non riuscito.	752
Non è disponibile alcuna capacità Spot che corrisponda alla tua richiesta. Avvio di un'istanza EC2 non riuscito.	752
<numero di istanze> istanza/e già in esecuzione. Avvio di un'istanza EC2 non riuscito.	753
Problemi relativi ad AMI	753
L'ID AMI <ID dell'AMI> non esiste. Avvio di un'istanza EC2 non riuscito.	754
L'AMI <ID AMI> è in sospenso e non può essere eseguita. Avvio di un'istanza EC2 non riuscito.	754
<Nome dispositivo> Nome dispositivo non valido. Avvio di un'istanza EC2 non riuscito.	754
L'architettura “arm64” del tipo di istanza specificato non corrisponde all'architettura “x86_64” dell'AMI specificata... Avvio dell'istanza EC2 non riuscito.	755
L'AMI “<ID AMI>” è disabilitata e non può essere eseguita. Avvio di un'istanza EC2 non riuscito.	756
Problemi del load balancer	757

Uno o più gruppi di destinazione non trovati. Convalida della configurazione del load balancer non riuscita.	758
Impossibile trovare Load Balancer <il tuo load balancer>. Convalida della configurazione del load balancer non riuscita.	758
Non vi è alcun load balancer ATTIVO denominato <nome del load balancer>.	
Aggiornamento della configurazione del sistema di load balancer non riuscito.	759
L'istanza EC2 <ID istanza > non è in VPC. Aggiornamento della configurazione del sistema di load balancer non riuscito.	759
Problemi dei modelli di avvio	759
È necessario utilizzare un modello di avvio completo valido (valore non valido)	759
Assenza dell'autorizzazione a utilizzare il modello di avvio (autorizzazioni insufficienti)	760
Controlli dell'integrità	762
Un'istanza è stata messa fuori servizio in risposta a un errore di controllo dell'integrità dell'istanza EC2	763
Un'istanza è stata messa fuori servizio in risposta a un riavvio pianificato EC2	763
Un'istanza è stata messa fuori servizio in risposta a un controllo dello stato EC2 che indicava che era stata terminata o interrotta	764
Un'istanza è stata messa fuori servizio in risposta a un errore di controllo dell'integrità del sistema ELB	765
Informazioni correlate	767
Cronologia dei documenti	770
.....	dcccxv

Cos'è Amazon EC2 Auto Scaling?

Amazon EC2 Auto Scaling ti aiuta ad assicurarti di avere a disposizione il numero corretto di istanze Amazon EC2 per gestire il carico dell'applicazione. È possibile creare raccolte di istanze EC2, denominate gruppi Auto Scaling. Puoi specificare il numero minimo di istanze in ogni gruppo con scalabilità automatica e Amazon EC2 Auto Scaling garantisce che il gruppo non scenda mai a dimensioni inferiori. Puoi specificare il numero massimo di istanze in ogni gruppo con scalabilità automatica e Amazon EC2 Auto Scaling garantisce che il gruppo non salga mai a dimensioni superiori. Se specifichi la capacità desiderata, sia quando crei il gruppo o in qualsiasi momento successivo, Amazon EC2 Auto Scaling garantisce che il tuo gruppo disponga di un determinato numero di istanze. Se specifichi le policy di dimensionamento, Amazon EC2 Auto Scaling può avviare o terminare istanze in base all'aumento o alla diminuzione della richiesta delle applicazioni.

Ad esempio, il seguente gruppo Auto Scaling ha una dimensione minima di quattro istanze, una capacità desiderata di sei istanze e una dimensione massima di dodici istanze. Le policy di dimensionamento definite regolano il numero di istanze entro il numero minimo e massimo, in base ai criteri specificati.



Caratteristiche di Amazon EC2 Auto Scaling

Con Amazon EC2 Auto Scaling, le istanze EC2 sono organizzate in gruppi di Auto Scaling in modo da poter essere trattate come un'unità logica ai fini della scalabilità e della gestione. I gruppi di Auto

Scaling utilizzano modelli di avvio (o configurazioni di avvio) come modelli di configurazione per le loro istanze EC2.

Le seguenti sono le caratteristiche principali di Amazon EC2 Auto Scaling:

Monitoraggio dello stato delle istanze in esecuzione

Amazon EC2 Auto Scaling monitora automaticamente lo stato e la disponibilità delle istanze utilizzando i controlli dello stato di EC2 e sostituisce le istanze terminate o danneggiate per mantenere la capacità desiderata.

Controlli dell'integrità personalizzati

Oltre ai controlli di integrità integrati, puoi definire controlli di integrità personalizzati specifici per la tua applicazione per verificare che risponda come previsto. Se un'istanza non supera il controllo di integrità personalizzato, viene sostituita automaticamente per mantenere la capacità desiderata.

Bilanciamento della capacità tra le zone di disponibilità

Puoi specificare più zone di disponibilità per il tuo gruppo Auto Scaling e Amazon EC2 Auto Scaling bilancia le istanze in modo uniforme tra le zone di disponibilità man mano che il gruppo cresce. Ciò garantisce disponibilità e resilienza elevate proteggendo le applicazioni dai guasti in un'unica posizione.

Più tipi di istanza e opzioni di acquisto

All'interno di un unico gruppo di Auto Scaling, puoi lanciare più tipi di istanze e opzioni di acquisto (istanze Spot e On-Demand), che consentono di ottimizzare i costi tramite l'utilizzo delle istanze Spot. Puoi anche usufruire degli sconti Reserved Instance e Savings Plan utilizzandoli insieme alle istanze On-Demand del gruppo.

Sostituzione automatica delle istanze spot

Se il tuo gruppo include istanze Spot, Amazon EC2 Auto Scaling può richiedere automaticamente la capacità Spot sostitutiva in caso di interruzione delle istanze Spot. Tramite Capacity Rebalancing, Amazon EC2 Auto Scaling può anche monitorare e sostituire in modo proattivo le istanze Spot che presentano un elevato rischio di interruzione.

Bilanciamento del carico

Puoi utilizzare il bilanciamento del carico e i controlli di integrità di Elastic Load Balancing per garantire una distribuzione uniforme del traffico delle applicazioni sulle istanze integre. Ogni volta che le istanze vengono avviate o terminate, Amazon EC2 Auto Scaling registra e annulla automaticamente la registrazione delle istanze dal sistema di bilanciamento del carico.

Scalabilità

Amazon EC2 Auto Scaling offre anche diversi modi per scalare i gruppi di Auto Scaling. L'utilizzo della scalabilità automatica consente di mantenere la disponibilità delle applicazioni e ridurre i costi aggiungendo capacità per gestire i picchi di carico e rimuovendo la capacità quando la domanda è inferiore. È inoltre possibile regolare manualmente le dimensioni del gruppo Auto Scaling in base alle esigenze.

Aggiornamento istanza

La funzionalità di aggiornamento delle istanze fornisce un meccanismo per aggiornare le istanze in modo continuativo quando aggiorni l'AMI o il modello di avvio. Puoi anche utilizzare un approccio graduale, noto come implementazione Canary, per testare una nuova AMI o avviare un modello su un piccolo set di istanze prima di distribuirlo all'intero gruppo.

Hook del ciclo di vita

I Lifecycle Hook sono utili per definire azioni personalizzate che vengono richiamate all'avvio di nuove istanze o prima che le istanze vengano terminate. Questa funzionalità è particolarmente utile per creare architetture basate sugli eventi, ma aiuta anche a gestire le istanze durante il loro ciclo di vita.

Support per carichi di lavoro stateful

I Lifecycle Hook offrono anche un meccanismo per mantenere lo stato persistente allo spegnimento. Per garantire la continuità delle applicazioni con stato, puoi anche utilizzare una protezione scalabile o politiche di terminazione personalizzate per evitare che le istanze con processi a esecuzione prolungata vengano terminate anticipatamente.

Per ulteriori informazioni sui vantaggi di Amazon EC2 Auto Scaling, consulta [Vantaggi di Amazon EC2 Auto Scaling](#).

Prezzi per Amazon EC2 Auto Scaling

Non ci sono costi aggiuntivi con Amazon EC2 Auto Scaling, quindi è facile provarlo e vedere come può apportare vantaggi alla tua architettura. AWS Paghi solo per le AWS risorse (ad esempio, istanze EC2, volumi EBS e CloudWatch allarmi) che utilizzi.

Inizia a usare

Per iniziare, completa il tutorial [Crea il tuo primo gruppo di Auto Scaling per creare un gruppo](#) Auto Scaling e scopri come reagisce quando un'istanza di quel gruppo termina.

Utilizzo dei gruppi Auto Scaling

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 gruppi di Auto Scaling accedendo a AWS Management Console, utilizzando la casella di ricerca sulla barra di navigazione per cercare i gruppi di Auto Scaling e quindi selezionando i gruppi 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 [Preparazione all'utilizzo della AWS CLI](#). Per ulteriori informazioni, consulta [scalabilità automatica](#) nella Documentazione di riferimento sui comandi della 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 di query:** forniscono operazioni API di basso livello accessibili tramite richieste HTTPS. L'utilizzo dell'API di query è il modo più diretto di accedere a Servizi AWS. Tuttavia, richiede che l'applicazione gestisca dettagli di basso livello, come la generazione dell'hash per firmare la richiesta e la gestione degli errori. Per ulteriori informazioni, consulta [Riferimento delle API Amazon EC2 Auto Scaling](#).
- **AWS CloudFormation**— Supporta la creazione di gruppi di Auto Scaling utilizzando CloudFormation modelli. Per ulteriori informazioni, consulta [Creazione di gruppi Auto Scaling con AWS CloudFormation](#).

Per connettersi a livello di codice a un Servizio AWS, si utilizza un endpoint. l'utente della regione.

Vantaggi di Amazon EC2 Auto Scaling

L'aggiunta di Amazon EC2 Auto Scaling all'architettura dell'applicazione è un modo per massimizzare i vantaggi del cloud. AWS Quando usi Amazon EC2 Auto Scaling, le applicazioni usufruiscono dei seguenti vantaggi:

- **Maggiore tolleranza ai guasti.** Amazon EC2 Auto Scaling può rilevare quando un'istanza non è integra, terminarla e avviarne una nuova in sostituzione. Puoi anche configurare Amazon EC2 Auto Scaling per l'utilizzo di più zone di disponibilità. Se una zona di disponibilità non è più accessibile, Amazon EC2 Auto Scaling può avviare le istanze in un'altra zona per compensare.
- **Maggiore disponibilità.** Amazon EC2 Auto Scaling può aiutarti a garantire che l'applicazione disponga sempre della giusta quantità di capacità per gestire le necessità di traffico correnti.
- **Migliore gestione dei costi.** Amazon EC2 Auto Scaling è in grado di aumentare e diminuire dinamicamente la capacità in base alle esigenze. Poiché le istanze EC2 utilizzate sono a pagamento, è possibile risparmiare denaro avviando le istanze quando sono necessarie e terminandole quando non lo sono.

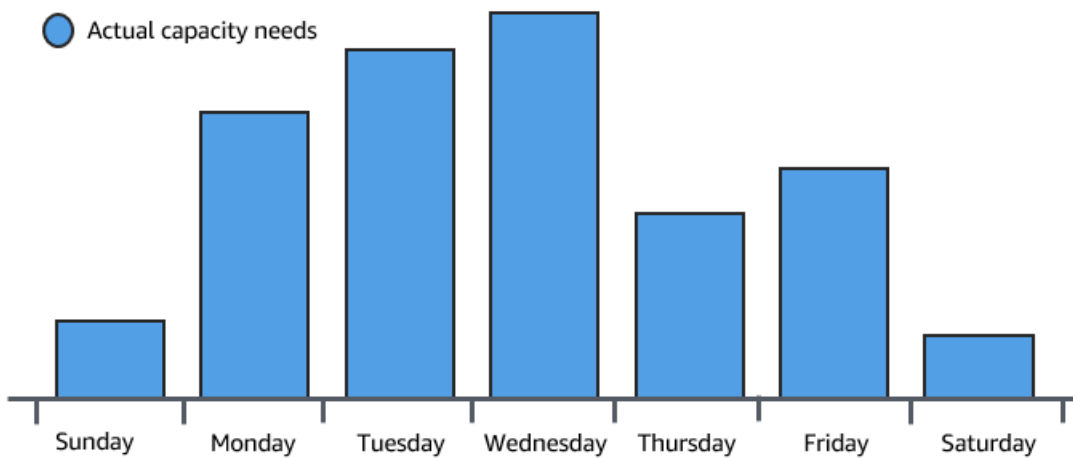
Indice

- [Esempio: copertura delle necessità variabili](#)
- [Esempio: architettura per applicazioni Web](#)
- [Esempio: distribuzione di istanze tra le zone di disponibilità](#)
 - [Distribuzione delle istanze](#)
 - [Attività di ribilanciamento](#)

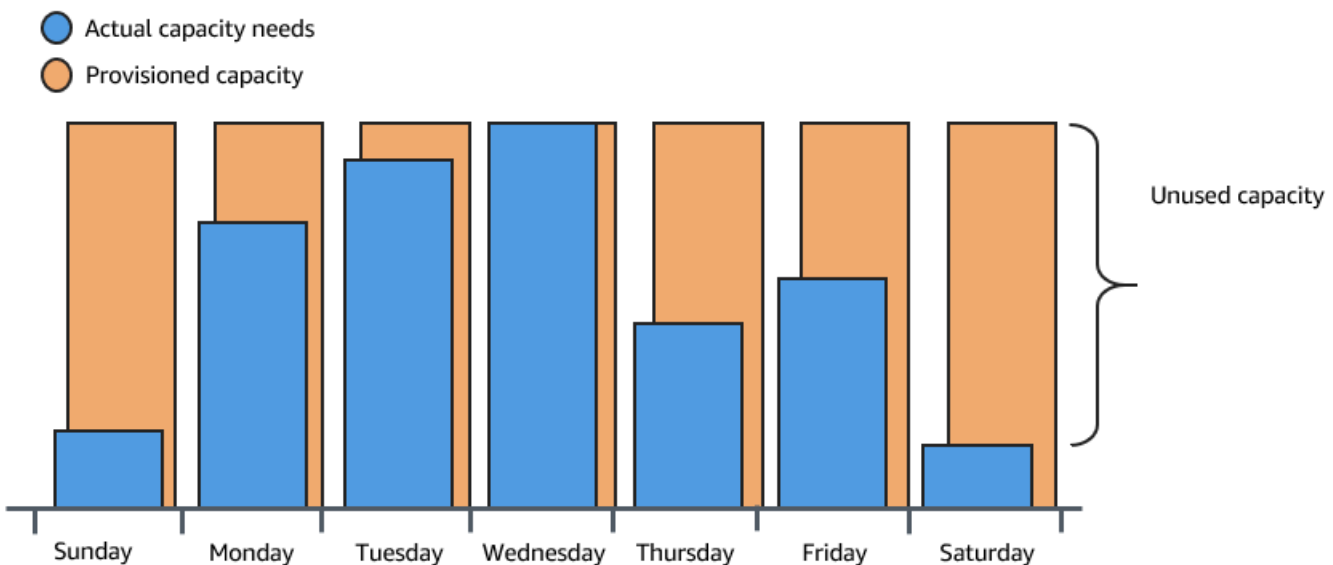
Esempio: copertura delle necessità variabili

Per dimostrare alcuni dei vantaggi di Amazon EC2 Auto Scaling, considera un'applicazione Web di base in esecuzione su AWS. Questa applicazione consente ai dipendenti di cercare sale conferenze da utilizzare per le riunioni. All'inizio e alla fine della settimana, l'utilizzo di questa applicazione è minimo. Nei giorni centrali della settimana più dipendenti programmano riunioni, perciò la richiesta relativa all'applicazione aumenta in modo significativo.

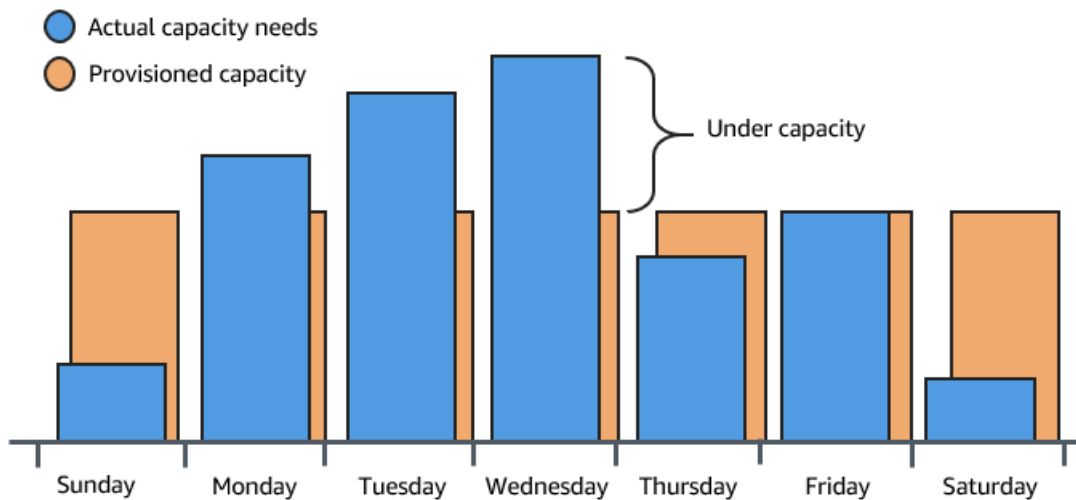
Il grafico seguente mostra la quantità di capacità dell'applicazione che viene utilizzata durante una settimana.



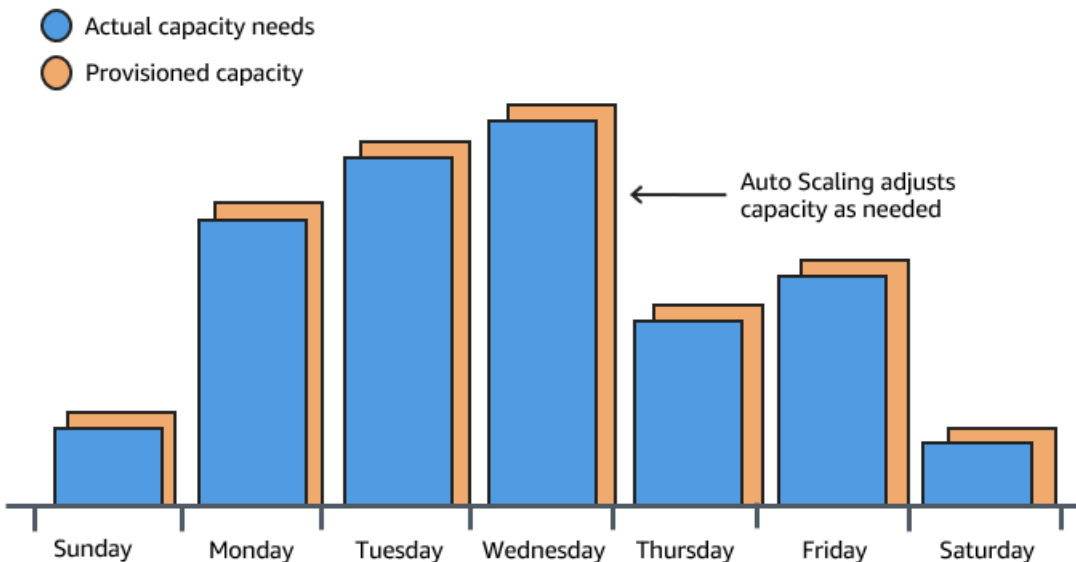
Tradizionalmente, sono disponibili due modi per pianificare le modifiche della capacità. Il primo è la possibilità di aggiungere un numero sufficiente di server, in modo che l'applicazione disponga sempre di capacità sufficiente per soddisfare le richieste. Lo svantaggio di questa opzione, tuttavia, è che ci sono giorni in cui l'applicazione non ha bisogno di molta capacità. La capacità aggiuntiva resta inutilizzata e, in pratica, aumenta il costo del mantenimento dell'applicazione in esecuzione.



La seconda opzione consiste nel disporre di capacità sufficiente per gestire la richiesta media dell'applicazione. Questa opzione è meno costosa, in quanto non è necessario acquistare apparecchiature da utilizzare solo occasionalmente. Tuttavia, si rischia di creare un'esperienza cliente insoddisfacente, quando la richiesta dell'applicazione supera la relativa capacità.



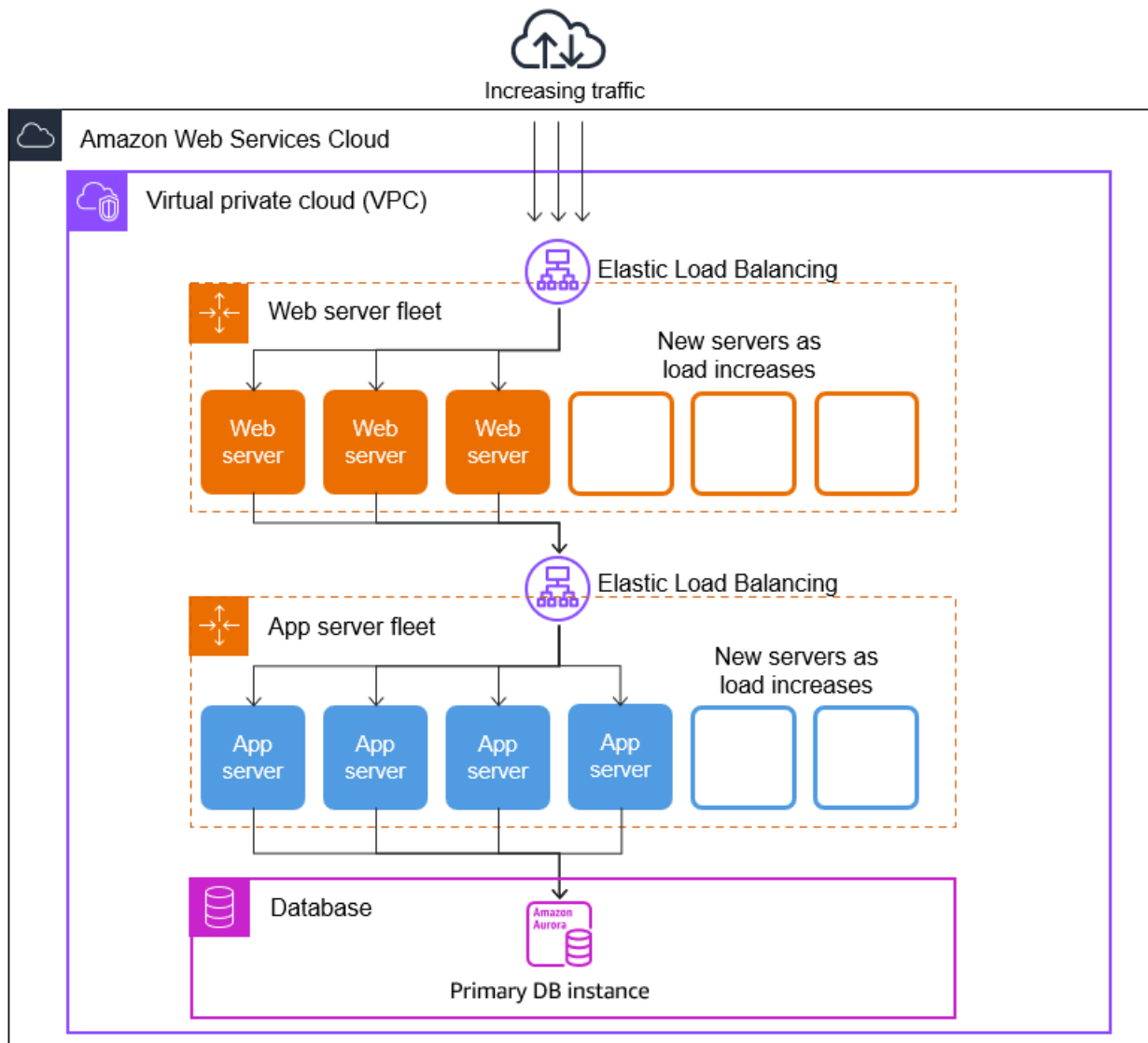
Aggiungendo Amazon EC2 Auto Scaling a questa applicazione, disponi di una terza opzione. Puoi aggiungere nuove istanze per l'applicazione solo quando necessario e terminarle quando non sono più necessarie. Poiché Amazon EC2 Auto Scaling usa le istanze EC2, dovrai pagare solo per le istanze utilizzate e quando le utilizzi. In questo modo avrai un'architettura economicamente vantaggiosa che offre la migliore esperienza cliente riducendo al minimo i costi.



Esempio: architettura per applicazioni Web

In uno scenario comune di applicazioni Web è possibile eseguire più copie della tua app contemporaneamente, per coprire il volume del traffico dei clienti. Queste copie multiple dell'applicazione sono ospitate in istanze EC2 identiche (server cloud), ciascuna delle quali gestisce le richieste dei clienti.

Amazon EC2 Auto Scaling gestisce l'avvio e la terminazione di queste istanze EC2 per tuo conto. Definisci una serie di criteri (come un CloudWatch allarme Amazon) che determinano quando il gruppo Auto Scaling avvia o chiude le istanze EC2. L'aggiunta di gruppi Auto Scaling all'architettura di rete aiuta ad aumentare la disponibilità e la tolleranza ai guasti della tua applicazione.



Puoi creare tutti i gruppi Auto Scaling di cui hai bisogno. Ad esempio, puoi creare un gruppo con scalabilità automatica per ogni livello.

Per distribuire il traffico tra le istanze nei gruppi con scalabilità automatica, puoi introdurre un load balancer all'interno dell'architettura. Per ulteriori informazioni, consulta [Sistema di bilanciamento del carico elastico](#).

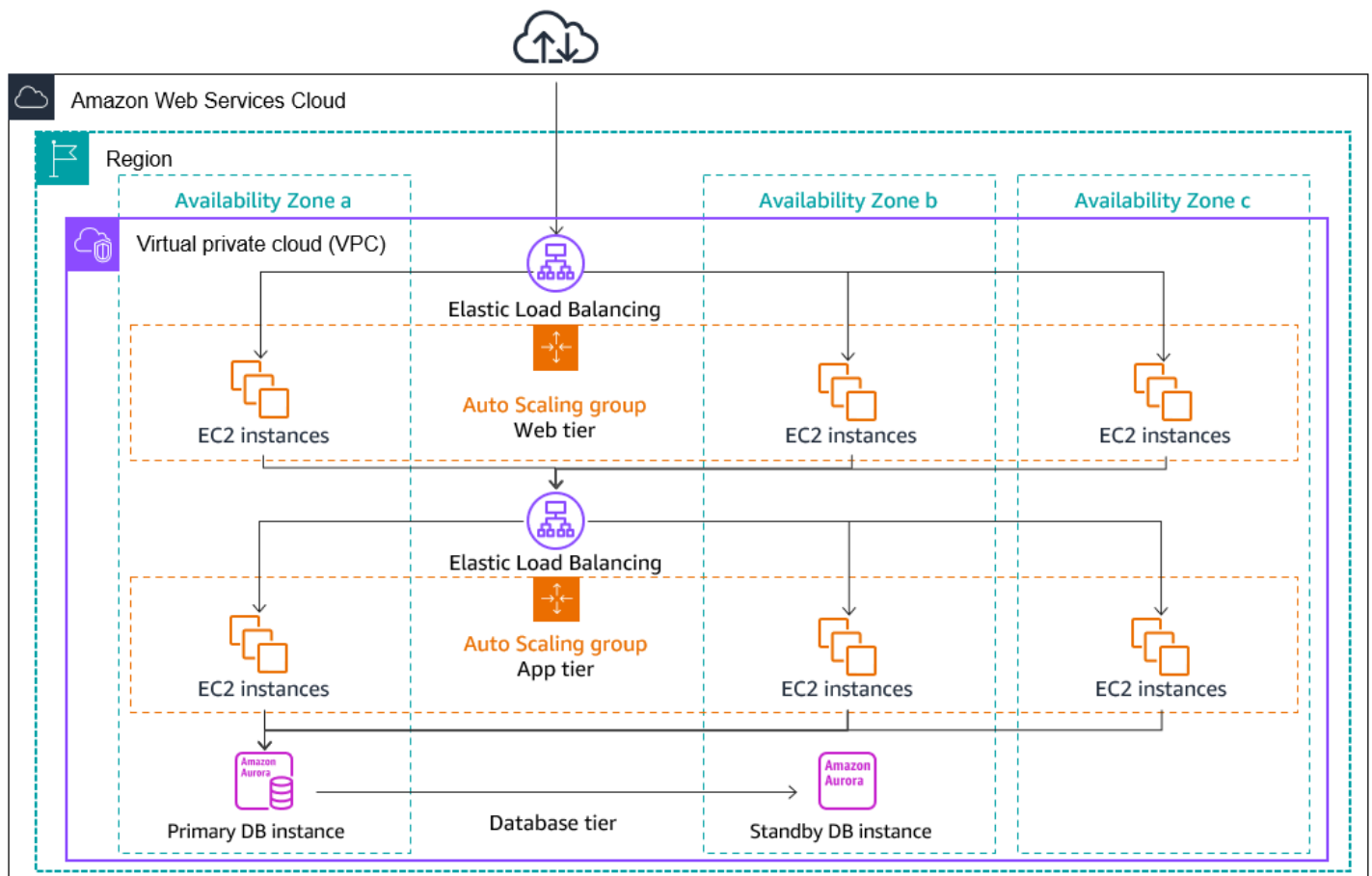
Esempio: distribuzione di istanze tra le zone di disponibilità

Le zone di disponibilità sono posizioni isolate all'interno di una determinata Regione AWS. Ogni regione ha più zone di disponibilità progettate per fornire un'elevata disponibilità per la regione. Le zone di disponibilità sono indipendenti e pertanto quando si progetta l'applicazione per l'uso di più zone se ne aumenta la disponibilità. Per ulteriori informazioni, consulta [Resilienza in Amazon EC2 Auto Scaling](#).

Una zona di disponibilità è identificata dal Regione AWS codice seguito da una lettera identificativa (ad esempio, us-east-1a). Se crei il tuo VPC e le sottoreti anziché utilizzare il VPC di default, puoi definire una o più sottoreti in ciascuna zona di disponibilità. Ogni sottorete deve risiedere totalmente all'interno di una zona di disponibilità e non può estendersi in altre zone. Per ulteriori informazioni, consulta [Come funziona Amazon VPC](#) nella Guida per l'utente di Amazon VPC.

Quando crei un gruppo con scalabilità automatica, è necessario scegliere il VPC e le sottoreti in cui distribuire il gruppo. Dimensionamento automatico Amazon EC2 crea le istanze nelle sottoreti scelte. Ogni istanza è quindi associata a una zona di disponibilità specifica scelta da Dimensionamento automatico Amazon EC2. All'avvio delle istanze, Dimensionamento automatico Amazon EC2 prova a distribuirle uniformemente tra le zone per un'elevata disponibilità e affidabilità.

L'immagine seguente mostra una panoramica dell'architettura a più livelli implementata in tre zone di disponibilità.



Distribuzione delle istanze

Dimensionamento automatico Amazon EC2 prova automaticamente a mantenere un numero equivalente di istanze in ogni zona di disponibilità abilitata. Dimensionamento automatico Amazon EC2 esegue questa operazione tentando di avviare nuove istanze nella zona di disponibilità con il minor numero di istanze. Se sono state scelte più sottoreti in una zona di disponibilità, Dimensionamento automatico Amazon EC2 seleziona in modo casuale una sottorete della zona di disponibilità. Se il tentativo non riesce, Amazon EC2 Auto Scaling tenta di avviare le istanze in un'altra zona di disponibilità finché l'operazione non va a buon fine.

In circostanze in cui una zona di disponibilità diventa non integra o non disponibile, la distribuzione delle istanze potrebbe essere non uniforme tra le zone di disponibilità. Quando la zona di disponibilità viene ripristinata, Dimensionamento automatico Amazon EC2 ribilancia automaticamente il gruppo con scalabilità automatica. Ciò avviene avviando le istanze nelle zone di disponibilità abilitate con il minor numero di istanze e terminandole in un'altra posizione.

Attività di ribilanciamento

Le attività di ribilanciamento rientrano in due categorie: ribilanciamento della zona di disponibilità e ribilanciamento della capacità.

Ribilanciamento della zona di disponibilità

Dopo determinate operazioni, il gruppo con scalabilità automatica può diventare non bilanciato tra le zone di disponibilità. Amazon EC2 Auto Scaling compensa ribilanciando le zone di disponibilità. Le seguenti operazioni possono portare alla necessità di attività di ribilanciamento:

- Puoi modificare le zone di disponibilità associate al gruppo con scalabilità automatica.
- Hai terminato o distaccato esplicitamente le istanze o hai messo le istanze in standby e il gruppo è diventato sbilanciato.
- Una zona di disponibilità che aveva in precedenza capacità insufficiente viene ripristinata e dispone quindi di capacità aggiuntiva.
- Una zona di disponibilità il cui precedente prezzo istanza Spot era superiore al tuo prezzo massimo ora ha un prezzo istanza Spot inferiore al tuo prezzo massimo.

Durante il ribilanciamento, Dimensionamento automatico Amazon EC2 lancia nuove istanze prima di terminare quelle precedenti. In questo modo, il ribilanciamento non compromette le prestazioni o la disponibilità delle applicazioni.

Poiché Dimensionamento automatico Amazon EC2 prova ad avviare nuove istanze prima di terminare quelle vecchie, se si è raggiunta o si sta per raggiungere la capacità massima specificata le attività di ribilanciamento potrebbero essere ostacolate o bloccate completamente.

Per evitare questo problema, il sistema può temporaneamente superare la capacità massima specificata di un gruppo durante un'attività di ribilanciamento. Per impostazione predefinita, ciò è possibile con un margine del 10% o un'istanza, a seconda di quale sia maggiore. Il margine viene esteso solo se il gruppo ha raggiunto o è vicino alla capacità massima e deve essere ribilanciato. L'estensione dura solo il tempo necessario per ribilanciare il gruppo, in genere pochi minuti.

In alternativa, puoi stabilire soglie per un gruppo con dimensionamento automatico utilizzando una policy di manutenzione dell'istanza e il gruppo può solo aumentare o diminuire la capacità entro tale intervallo di soglia. In questo modo, puoi controllare la velocità alla quale il gruppo si riequilibra. Per ulteriori informazioni, consulta [Policy di manutenzione delle istanze](#).

Ribilanciamento della capacità

Quando utilizzi le istanze spot, puoi abilitare il ribilanciamento della capacità nei gruppi con scalabilità automatica. Ciò consente a Dimensionamento automatico Amazon EC2 di avviare un'istanza spot ogni volta che Amazon EC2 riporta che un'istanza spot è a rischio elevato di interruzione. Dopo aver avviato una nuova istanza, ne termina quindi una vecchia. Per ulteriori informazioni, consulta [Utilizza il ribilanciamento della capacità per gestire le interruzioni spot di Amazon EC2](#).

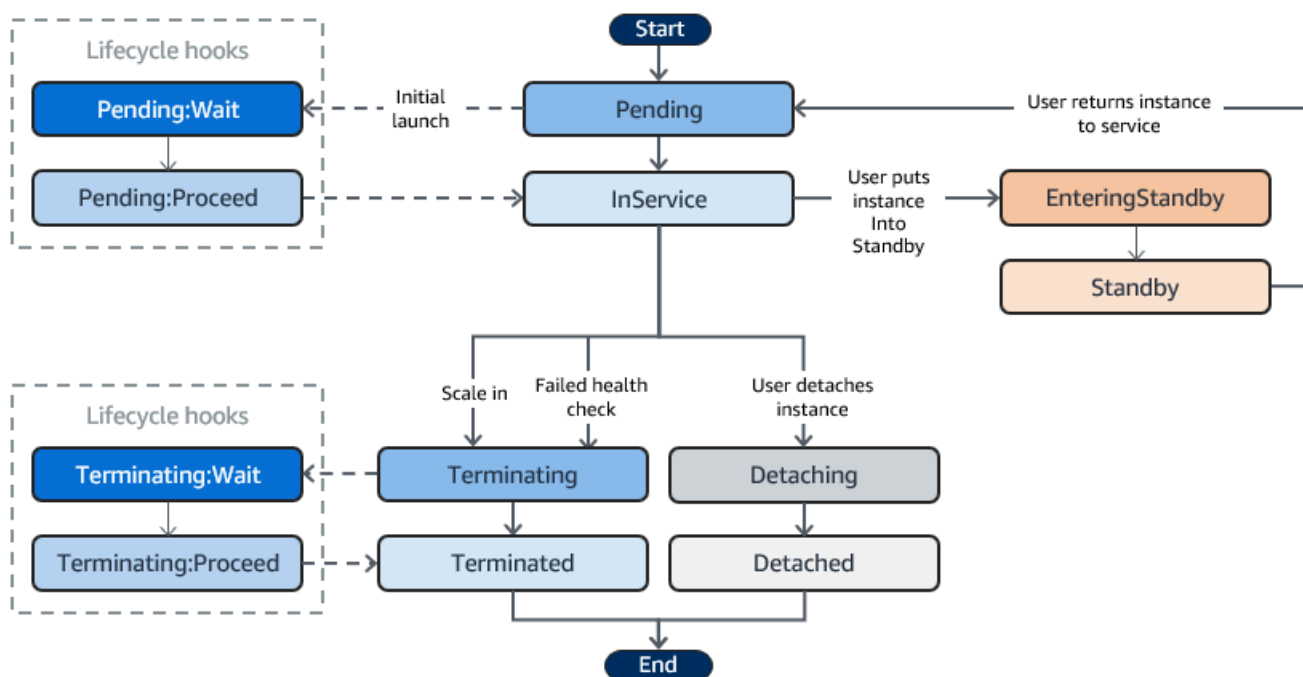
Ciclo di vita delle istanze di Amazon EC2 Auto Scaling

Le istanze EC2 in un gruppo con scalabilità automatica dispongono di un percorso, o un ciclo di vita, che differisce da quello delle altre istanze EC2. Il ciclo di vita inizia quando il gruppo con scalabilità automatica avvia un'istanza e la mette in servizio. Il ciclo di vita finisce quando l'utente termina l'istanza o il gruppo con scalabilità automatica mette l'istanza fuori servizio e la termina.

Note

I prezzi vengono calcolati in base all'utilizzo delle istanze non appena vengono avviate, incluso il tempo in cui non sono ancora in servizio.

La figura che segue illustra le transizioni tra gli stati di un'istanza nel ciclo di vita di Amazon EC2 Auto Scaling.



Aumento orizzontale

I seguenti eventi di aumento orizzontale indicano al gruppo con scalabilità automatica di avviare le istanze EC2 e allegarle al gruppo:

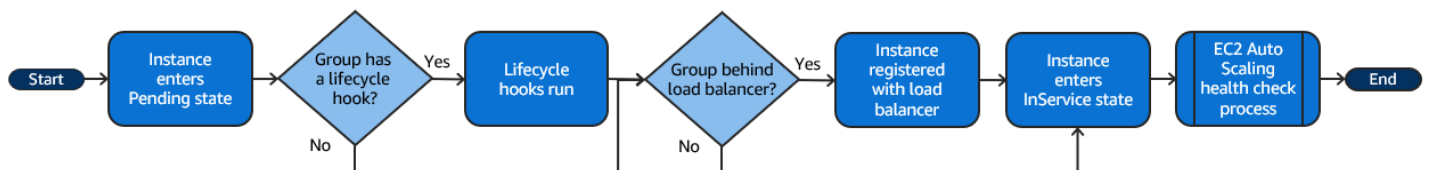
- Hai aumentato manualmente le dimensioni del gruppo. Per ulteriori informazioni, consulta [Modifica della capacità desiderata di un gruppo con dimensionamento automatico esistente](#).
- Hai creato una policy di dimensionamento per aumentare automaticamente le dimensioni del gruppo in base a un determinato aumento della richiesta. Per ulteriori informazioni, consulta [Dimensionamento dinamico per Dimensionamento automatico Amazon EC2](#).
- Hai impostato il dimensionamento con una pianificazione per aumentare le dimensioni del gruppo in un determinato momento. Per ulteriori informazioni, consulta [Dimensionamento programmato per Dimensionamento automatico Amazon EC2](#).

Quando si verifica un evento di aumento orizzontale, il gruppo con scalabilità automatica avvia il numero richiesto di istanze EC2, utilizzando il modello di avvio assegnato. Le istanze si avviano nello stato Pending. Se aggiungi un hook del ciclo di vita al gruppo con scalabilità automatica, puoi eseguire un'operazione personalizzata. Per ulteriori informazioni, consulta [Hook del ciclo di vita](#).

Quando un'istanza è completamente configurata e supera i controlli dell'integrità di Amazon EC2, viene allegata al gruppo con scalabilità automatica e acquisisce lo stato InService. L'istanza viene calcolata nell'ambito della capacità desiderata del gruppo con scalabilità automatica.

Se il gruppo con scalabilità automatica è configurato per ricevere traffico da un load balancer Elastic Load Balancing, Dimensionamento automatico Amazon EC2 registra automaticamente l'istanza con il load balancer prima di contrassegnare l'istanza come InService.

Di seguito vengono riepilogati i passaggi per registrare un'istanza con un sistema di bilanciamento del carico per un evento di scalabilità orizzontale.



Istanze in servizio

Le istanze restano in stato InService finché non si verifica uno dei seguenti eventi:

- Si verifica un evento di riduzione orizzontale e Amazon EC2 Auto Scaling sceglie di terminare l'istanza, in modo da ridurre le dimensioni del gruppo con scalabilità automatica. Per ulteriori informazioni, consulta [Verificare quali istanze Auto Scaling vengono terminate durante la riduzione orizzontale](#).
- L'istanza è stata messa in stato Standby. Per ulteriori informazioni, consulta [Come entrare e uscire dallo stato di standby](#).
- Distacchi l'istanza dal gruppo con scalabilità automatica. Per ulteriori informazioni, consulta [Scollegare o collegare istanze](#).
- L'istanza non supera un numero richiesto di controlli dell'integrità, perciò viene rimossa dal gruppo con scalabilità automatica, terminata e sostituita. Per ulteriori informazioni, consulta [Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico](#).

Riduzione orizzontale

I seguenti eventi di riduzione orizzontale indicano al gruppo con scalabilità automatica di distaccare le istanze EC2 dal gruppo e di terminarle:

- Hai ridotto manualmente le dimensioni del gruppo. Per ulteriori informazioni, consulta [Modifica della capacità desiderata di un gruppo con dimensionamento automatico esistente](#).
- Hai creato una policy di dimensionamento per diminuire automaticamente le dimensioni del gruppo in base a una determinata diminuzione della richiesta. Per ulteriori informazioni, consulta [Dimensionamento dinamico per Dimensionamento automatico Amazon EC2](#).
- Hai impostato il dimensionamento con una pianificazione per diminuire le dimensioni del gruppo in un determinato momento. Per ulteriori informazioni, consulta [Dimensionamento programmato per Dimensionamento automatico Amazon EC2](#).

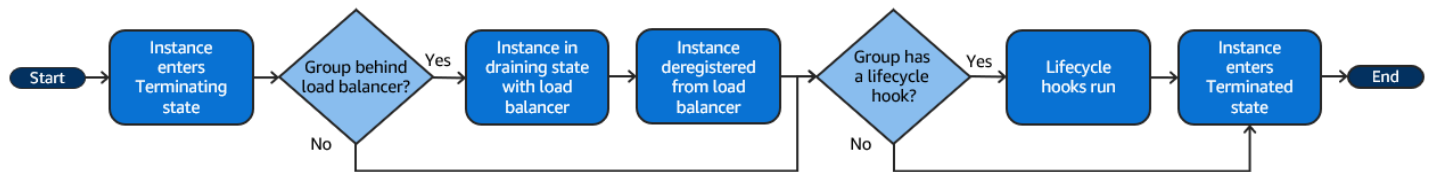
È importante creare un evento di riduzione orizzontale corrispondente a ogni evento di aumento orizzontale che crei. In questo modo è possibile garantire che le risorse assegnate alla tua applicazione corrispondano il più possibile alla richiesta per tali risorse.

Quando si verifica un evento di riduzione orizzontale, il gruppo con scalabilità automatica distacca una o più istanze. Il gruppo con scalabilità automatica utilizza la policy di terminazione per determinare quali istanze terminare. Le istanze in fase di distacco dal gruppo con dimensionamento automatico vanno in stato `Terminating` e non possono essere rimesse in servizio.

Se il gruppo con dimensionamento automatico è configurato per ricevere traffico da un load balancer Elastic Load Balancing, Dimensionamento automatico Amazon EC2 annulla automaticamente la registrazione dell'istanza in fase di terminazione dal sistema di bilanciamento del carico. L'annullamento della registrazione dell'istanza garantisce che tutte le nuove richieste vengano reindirizzate ad altre istanze del gruppo di destinazione del load balancer mentre le connessioni all'istanza esistenti possono continuare fino alla scadenza del ritardo di cancellazione.

Se aggiungi un hook del ciclo di vita al gruppo con dimensionamento automatico, puoi eseguire un'operazione personalizzata su un'istanza di terminazione. Per ulteriori informazioni, consulta [Hook del ciclo di vita](#). Infine, l'istanza è completamente terminata e in stato Terminated.

Di seguito vengono riepilogati i passaggi per annullare la registrazione di un'istanza con un sistema di bilanciamento del carico per un evento di scalabilità.



Come distaccare un'istanza

È possibile distaccare un'istanza dal gruppo con scalabilità automatica. Dopo essere stata distaccata, puoi gestire l'istanza separatamente dal gruppo con scalabilità automatica o allegarla a un altro gruppo.

Per ulteriori informazioni, consulta [Scollegare o collegare istanze](#).

Come allegare un'istanza

Puoi collegare un'istanza EC2 in esecuzione che soddisfa determinati criteri del gruppo con scalabilità automatica. Dopo essere stata allegata, l'istanza viene gestita come parte del gruppo con scalabilità automatica.

Per ulteriori informazioni, consulta [Scollegare o collegare istanze](#).

Hook del ciclo di vita

Puoi aggiungere un hook del ciclo di vita al gruppo con scalabilità automatica, in modo che sia possibile eseguire operazioni personalizzate all'avvio o alla terminazione delle istanze.

Quando Amazon EC2 Auto Scaling risponde a un evento di aumento orizzontale delle dimensioni, avvia una o più istanze. Le istanze si avviano nello stato `Pending`. Se hai aggiunto un hook del ciclo di vita `autoscaling:EC2_INSTANCE_LAUNCHING` al gruppo con scalabilità automatica, le istanze passano dallo stato `Pending` allo stato `Pending:Wait`. Al termine dell'operazione del ciclo di vita le istanze passano allo stato `Pending:Proceed`. Quando le istanze sono completamente configurate, vengono alleggiate al gruppo con scalabilità automatica ed entrano nello stato `InService`.

Quando Amazon EC2 Auto Scaling risponde a un evento di riduzione orizzontale delle dimensioni, termina una o più istanze. Le istanze vengono distaccate dal gruppo con scalabilità automatica e vanno in stato `Terminating`. Se hai aggiunto un hook del ciclo di vita `autoscaling:EC2_INSTANCE_TERMINATING` al gruppo con scalabilità automatica, le istanze passano dallo stato `Terminating` allo stato `Terminating:Wait`. Al termine dell'operazione del ciclo di vita le istanze passano allo stato `Terminating:Proceed`. Una volta completata la terminazione delle istanze, queste vanno in stato `Terminated`.

Per ulteriori informazioni, consulta [Hook del ciclo di vita di Amazon EC2 Auto Scaling](#).

Come entrare e uscire dallo stato di standby

Puoi mettere qualsiasi istanza in stato `InService` in stato di `Standby`. In questo modo è possibile eliminare l'istanza dal servizio, risolvere i problemi o modificarla, quindi riportarla in servizio.

Le istanze in stato di `Standby` continueranno a essere gestite dal gruppo con scalabilità automatica. Tuttavia, non saranno parte attiva dell'applicazione finché non le avrai riportate in servizio.

Per ulteriori informazioni, consulta [Rimuovere temporaneamente le istanze dal gruppo con scalabilità automatica](#).

Quote per Amazon EC2 Auto Scaling

Your Account AWS ha delle quote predefinite, precedentemente denominate limiti, per ogni servizio. AWS Salvo diversa indicazione, ogni quota si applica a una regione specifica. Se per alcune quote è possibile richiedere aumenti, altre quote non possono essere modificate.

Per visualizzare le quote per Amazon EC2 Auto Scaling, apri la [Console di Service Quotas](#). Nel pannello di navigazione, scegli Servizi AWS e seleziona Dimensionamento automatico Amazon EC2.

Per richiedere un aumento delle quote, consultare [Richiesta di aumento delle quote](#) nella Guida per l'utente di Service Quotas. Se la quota non è ancora disponibile in Service Quotas, utilizza il

[modulo dei limiti di Auto Scaling](#). Gli aumenti di quota sono vincolati alla regione per la quale sono stati richiesti.

Tutte le richieste vengono inviate a AWS Support. È possibile tenere traccia del caso di richiesta nella console AWS Support.

Risorse Amazon EC2 Auto Scaling

Hai Account AWS le seguenti quote relative al numero di gruppi di Auto Scaling e alle configurazioni di avvio che puoi creare.

Risorsa	Quota predefinita
Gruppi Auto Scaling per regione	500
Configurazioni di avvio per regione	200

Configurazione del gruppo con scalabilità automatica

Hai Account AWS le seguenti quote relative alla configurazione dei gruppi di Auto Scaling. Non possono essere modificati.

Risorsa	Quota
Policy di dimensionamento per gruppo con scalabilità automatica	50
Operazioni programmate per gruppo con scalabilità automatica	125
Regolazioni di fase per policy di dimensionamento a fasi	20
Hook del ciclo di vita per gruppo con scalabilità automatica	50
Argomenti SNS per gruppo con scalabilità automatica	10
Classic Load Balancer per gruppo con scalabilità automatica	50
Gruppi di destinazione Elastic Load Balancer per gruppo con dimensionamento automatico	50

Risorsa	Quota
Gruppi di destinazione di VPC Lattice per gruppo con dimensionamento automatico	5

Operazioni delle API del gruppo con scalabilità automatica

Amazon EC2 Auto Scaling fornisce operazioni API per apportare modifiche ai gruppi Auto Scaling in batch. Di seguito sono riportati i limiti API sul numero massimo di elementi (numero massimo di membri dell'array) consentiti in una singola operazione. Non possono essere modificati.

Operazione	Numero massimo di membri dell'array
AttachInstances	20 ID istanza
AttachLoadBalancers	10 load balancer
AttachLoadBalancerTargetGroups	10 gruppi di destinazioni
BatchDeleteScheduledAction	50 operazioni pianificate
BatchPutScheduledUpdateGroupAction	50 operazioni pianificate
DetachInstances	20 ID istanza
DetachLoadBalancers	10 load balancer
DetachLoadBalancerTargetGroups	10 gruppi di destinazioni
EnterStandby	20 ID istanza
ExitStandby	20 ID istanza
SetInstanceProtection	50 ID istanza

Limitazione delle richieste per l'API Amazon EC2 Auto Scaling

Le richieste API di Amazon EC2 Auto Scaling vengono limitate utilizzando uno schema token bucket per mantenere la larghezza di banda del servizio. Per ulteriori informazioni, consulta la [percentuale di richieste API](#) nel riferimento alle API di Amazon EC2 Auto Scaling.

Tassi di terminazione EC2

Dimensionamento automatico Amazon EC2 determina in modo dinamico il numero di operazioni di terminazione dell'istanza EC2 che può eseguire in un momento in cui il gruppo con dimensionamento automatico viene ridotto. Ciò significa che potresti notare variazioni nel numero di istanze terminate contemporaneamente tra i gruppi con dimensionamento automatico. Queste variazioni sono causate da considerazioni esterne, ad esempio se Dimensionamento automatico Amazon EC2 deve annullare la registrazione delle istanze con un sistema di bilanciamento del carico.

Altri servizi

Le quote per altri servizi, come Amazon EC2 e Amazon VPC, possono influire sui gruppi di Auto Scaling. Puoi utilizzarle Service Quotas per aggiornare le quote per le istanze EC2 e altre risorse del tuo Account AWS. Nella Service Quotas console, puoi visualizzare tutte le quote di servizio disponibili e richiederne un aumento. Per ulteriori informazioni, consulta [Richiesta di un aumento di quota](#) nel Manuale dell'utente di Service Quotas .

Per le quote specifiche dei modelli di avvio, consulta [Restrizioni dei modelli di avvio](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Impostare Amazon EC2 Auto Scaling

Prima di iniziare a utilizzare Amazon EC2 Auto Scaling, completa le seguenti attività.

Processi

- [Preparazione all'utilizzo di Amazon EC2](#)
- [Preparazione all'utilizzo della AWS CLI](#)

Preparazione all'utilizzo di Amazon EC2

Se non hai mai utilizzato Amazon EC2 in precedenza, completa le operazioni descritte nella documentazione di Amazon EC2. Per ulteriori informazioni, consulta [Configurazione con Amazon EC2](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux o [Configurazione con Amazon EC2](#) nella Guida per l'utente utente Amazon EC2 per le istanze Windows.

Preparazione all'utilizzo della AWS CLI

Puoi utilizzare gli strumenti a riga di comando AWS per inviare comandi alla riga di comando del tuo sistema al fine di eseguire Dimensionamento automatico Amazon EC2 e altre attività AWS.

Per utilizzare la AWS Command Line Interface (AWS CLI), scarica, installa e configura la versione 1 o 2 della AWS CLI. La stessa funzionalità di Dimensionamento automatico Amazon EC2 è disponibile nelle versioni 1 e 2. Per installare la versione 1 della AWS CLI, consulta la sezione [Installazione, aggiornamento e disinstallazione della AWS CLI](#) nella Guida per l'utente della versione 1 della AWS CLI. Per installare la versione 2 della AWS CLI, consulta la sezione [Installazione o aggiornamento della versione più recente della AWS CLI](#) nella Guida per l'utente della versione 2 della AWS CLI.

AWS CloudShell ti permette di saltare l'installazione della AWS CLI nel tuo ambiente di sviluppo e di utilizzarlo invece nella AWS Management Console. Oltre a evitare l'installazione, non è necessario configurare le credenziali e non è necessario specificare una regione. La tua sessione di AWS Management Console fornisce questo contesto alla AWS CLI. Puoi usare AWS CloudShell nelle Regioni AWS supportate. Per ulteriori informazioni, consulta [Crea gruppi di Auto Scaling dalla riga di comando usando AWS CloudShell](#).

Per ulteriori informazioni, consulta [scalabilità automatica](#) nella Documentazione di riferimento sui comandi della AWS CLI.

Nozioni di base su Amazon EC2 Auto Scaling

Per iniziare a usare Amazon EC2 Auto Scaling, puoi seguire i tutorial che ti introducono al servizio.

Argomenti

- [Tutorial: crea il tuo primo gruppo Auto Scaling](#)
- [Tutorial: Configurazione di un'applicazione con dimensionamento e bilanciamento del carico](#)

Per ulteriori tutorial incentrati su strumenti specifici per la gestione del ciclo di vita delle istanze in un gruppo di Auto Scaling, consulta i seguenti argomenti:

- [Tutorial : Configurazione di un hook del ciclo di vita che richiama una funzione Lambda](#). Questo tutorial mostra come usare Amazon per EventBridge creare regole che richiamano le funzioni Lambda in base agli eventi che accadono alle istanze del tuo gruppo Auto Scaling.
- [Tutorial: configurazione dei dati utente per recuperare lo stato del ciclo di vita di destinazione tramite i metadati dell'istanza](#). In questo tutorial viene illustrato come utilizzare il servizio di metadati di istanza (IMDS) per richiamare un'azione dall'interno dell'istanza stessa.

Prima di creare un gruppo con dimensionamento automatico da usare con l'applicazione, esamina accuratamente l'applicazione quando viene eseguita nell' Cloud AWS. Considera i seguenti aspetti:

- Su quante zone di disponibilità deve essere distribuito il gruppo con dimensionamento automatico.
- Quali risorse esistenti, ad esempio gruppi di sicurezza o Amazon Machine Image (AMI), è possibile usare.
- Se desideri eseguire il dimensionamento per aumentare o diminuire la capacità oppure solo per assicurarti che un determinato numero di server siano sempre in esecuzione. Ricorda che Amazon EC2 Auto Scaling può fare entrambe le cose contemporaneamente.
- Quali parametri sono più importanti per le prestazioni dell'applicazione.
- Quanto tempo è necessario per avviare e fornire un server.

Più comprendi gli aspetti della tua applicazione, più efficace potrai rendere l'architettura Auto Scaling.

Tutorial: crea il tuo primo gruppo Auto Scaling

Questo tutorial fornisce un'introduzione pratica ad Amazon EC2 Auto Scaling tramite AWS Management Console. Creerai un modello di lancio che definisce le tue istanze EC2 e un gruppo di Auto Scaling contenente una singola istanza. Dopo aver avviato il gruppo Auto Scaling, chiuderai l'istanza e verificherai che sia stata rimossa dal servizio e sostituita. Per mantenere un numero costante di istanze, Amazon EC2 Auto Scaling rileva e risponde automaticamente ai controlli di integrità e raggiungibilità di Amazon EC2.

[Quando ti registri AWS, puoi iniziare a usare Amazon EC2 Auto Scaling gratuitamente utilizzando AWS il piano gratuito.](#) È possibile utilizzare il piano gratuito per avviare e utilizzare un'istanza t2.micro gratuitamente per 12 mesi (nelle regioni in cui t2.micro non è disponibile, è possibile utilizzare un'istanza t3.micro sotto il piano gratuito). Se avvii un'istanza che non rientra nel piano gratuito, ti verranno addebitati i costi standard di utilizzo di Amazon EC2 per l'istanza. Per ulteriori informazioni, consulta [Prezzi di Amazon EC2](#).

Attività

- [Preparazione della spiegazione passo per passo](#)
- [Fase 1: creazione di un modello di avvio](#)
- [Fase 2: creazione di un gruppo con dimensionamento automatico con una singola istanza](#)
- [Fase 3: verifica del gruppo con dimensionamento automatico](#)
- [Fase 4: termina un'istanza nel gruppo con dimensionamento automatico](#)
- [Fase 5: fasi successive](#)
- [Fase 6: pulizia](#)

Preparazione della spiegazione passo per passo

Questa spiegazione passo per passo presuppone che si abbia familiarità con l'avvio di istanze EC2 e che siano già stati creati una coppia di chiavi e un gruppo di sicurezza. Per ulteriori informazioni, consulta [Impostazioni con Amazon EC2](#) nella Guida per l'utente per istanze Linux di Amazon EC2.

Per iniziare a usare Amazon EC2 Auto Scaling, puoi utilizzare il VPC predefinito per il tuo Account AWS. Il VPC di default include una sottorete pubblica di default in ogni zona di disponibilità e un gateway Internet allegato al VPC. Puoi visualizzare i VPC nella pagina [Your VPCs \(I tuoi VPC\)](#) della console Amazon Virtual Private Cloud (Amazon VPC).

Fase 1: creazione di un modello di avvio

In questo passaggio, crei un modello di lancio che specifica il tipo di istanza EC2 che Amazon EC2 Auto Scaling crea per te. Vengono incluse informazioni quali l'ID Amazon Machine Image (AMI) da utilizzare, il tipo di istanza, le coppie di chiavi e i gruppi di sicurezza.

Per creare un modello di avvio

1. Apri la console Amazon EC2 e vai alla pagina [Launch templates](#).
2. Sulla barra di navigazione, seleziona una Regione AWS. Il modello di avvio e il gruppo con scalabilità automatica che crei sono vincolate alla regione specificata.
3. Scegli Crea modello di avvio.
4. In Launch template name (Nome del modello di) avvio, inserisci **my-template-for-auto-scaling**.
5. In Auto Scaling guidance (Guida Auto Scaling), seleziona la casella di controllo.
6. In Applicazioni e Immagini OS (Amazon machine image), scegli una versione di Amazon Linux 2 (HVM) dall'elenco Quick Start (Avvio rapido). L'AMI serve come modello di configurazione di base per le istanze.
7. In Instance type (Tipo di istanza), scegli una configurazione hardware per le istanze compatibile con l'AMI che hai specificato.
8. (Facoltativo) In Key pair name (Nome coppia di chiavi), scegli una coppia di chiavi esistente. Puoi utilizzare le coppie di chiavi per connetterti a un'istanza Amazon EC2 instance con SSH. La connessione a un'istanza non è inclusa nell'ambito di questo tutorial. Pertanto, non è necessario che specifichi una coppia di chiavi, a meno che tu non intenda connetterti all'istanza usando SSH.
9. Per Impostazioni di rete, amplia Configurazione avanzata di rete e esegui queste operazioni:
 - a. Per configurare un'interfaccia di rete primaria, selezionare Aggiungi interfaccia di rete.
 - b. Per l'assegnazione automatica dell'IP pubblico, specifica se l'istanza riceve un indirizzo IPv4 pubblico. Per impostazione predefinita, Amazon EC2 assegna un indirizzo IPv4 pubblico se l'istanza EC2 viene avviata in una sottorete predefinita o se l'istanza viene avviata in una sottorete configurata per assegnare automaticamente un indirizzo IPv4 pubblico. Se non hai bisogno di connetterti alla tua istanza, scegli Disabilita.
 - c. Per Security group ID, scegli un gruppo di sicurezza nello stesso VPC che intendi utilizzare come VPC per il tuo gruppo Auto Scaling. Se non specifichi un gruppo di sicurezza, l'istanza viene automaticamente associata al gruppo di sicurezza predefinito per il VPC.

- d. Per Elimina alla terminazione, scegli Sì per eliminare l'interfaccia di rete quando l'istanza viene eliminata.
10. Scegli Crea modello di avvio.
 11. Nella pagina di conferma, scegli Crea gruppo con dimensionamento automatico).

Fase 2: creazione di un gruppo con dimensionamento automatico con una singola istanza

Utilizza la procedura seguente per continuare da dove avevi interrotto dopo aver creato un modello di avvio.

Creazione di un gruppo con scalabilità automatica

1. Nella pagina Scegli il modello o la configurazione di avvio, in Nome per il gruppo con dimensionamento automatico, inserisci **my-first-asg**.
2. Seleziona Successivo.

Viene visualizzata la pagina Scegli le opzioni di avvio dell'istanza, che consente di scegliere le impostazioni di rete VPC che desideri vengano utilizzate dal gruppo Auto Scaling e offre opzioni per l'avvio di istanze On-Demand e Spot.

3. Nella sezione Rete, mantieni VPC impostato sul VPC predefinito scelto Regione AWS o seleziona il tuo VPC. Il VPC di default viene configurato automaticamente per fornire la connettività Internet all'istanza. Questo VPC include una sottorete pubblica in ciascuna zona di disponibilità nella regione.
4. In Availability Zones and subnets (Zone di disponibilità e sottoreti), scegli una sottorete per ogni zona di disponibilità che desideri includere. Per un'elevata disponibilità, utilizza sottoreti in più zone di disponibilità. Per ulteriori informazioni, consulta [Considerazioni sulla scelta di sottoreti VPC](#).
5. Nella sezione Instance type requirements (Requisiti del tipo di istanza), utilizza l'impostazione di default per semplificare questo passaggio. (Non sovrascrivere il modello di avvio.) Per questo tutorial, avvierai solo un'istanza on demand utilizzando il tipo di istanza specificato nel modello di avvio.
6. Mantieni il resto delle impostazioni di default per questo tutorial e scegli Skip to review (Passa alla verifica).

Note

La dimensione iniziale del gruppo è determinata dalla sua capacità desiderata. Il valore di default è istanza 1.

7. In Verifica, verifica le informazioni per il gruppo, quindi scegli Crea gruppo con dimensionamento automatico.

Fase 3: verifica del gruppo con dimensionamento automatico

Dopo aver creato un gruppo con dimensionamento automatico è possibile verificare che questo abbia avviato un'istanza EC2.

Tip

Nella seguente procedura, vengono esaminate le sezioni Cronologia attività e Istanze del gruppo con dimensionamento automatico. In entrambi, le colonne con nome dovrebbero essere già visualizzate. Per visualizzare le colonne nascoste o modificare il numero di righe visualizzate, scegli l'icona a forma di ingranaggio nell'angolo in alto a destra di ogni sezione per aprire le preferenze modali, aggiorna le impostazioni secondo necessità e scegli Confirm (Conferma).

Come verificare che il gruppo con dimensionamento automatico ha avviato un'istanza EC2

1. Apri la [pagina Gruppi con dimensionamento automatico](#) nella console Amazon EC2.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico appena creato.

Si aprirà un riquadro diviso nella parte inferiore della pagina Gruppi con dimensionamento automatico. La prima scheda disponibile è la scheda Dettagli che mostra le informazioni sul gruppo con dimensionamento automatico.

3. Scegli la seconda scheda, Activity (Attività). In Cronologia attività, è possibile visualizzare lo stato di avanzamento delle attività associate al gruppo con dimensionamento automatico. La colonna Status (Stato) mostra lo stato corrente della tua istanza. Durante l'avvio dell'istanza, la colonna di stato mostra `Not yet in service`. Una volta avviata l'istanza, lo stato diventa `Successful`. Per visualizzare l'attuale stato di un'istanza, puoi utilizzare anche il pulsante di aggiornamento.

4. Nella scheda Instance management (Gestione istanze) in Instances (Istanze), è possibile visualizzare lo stato dell'istanza.
5. Verifica che l'istanza sia stata avviata correttamente. L'avvio di un'istanza richiede pochi minuti.
 - La colonna Lifecycle (Ciclo di vita) mostra lo stato dell'istanza. Inizialmente, l'istanza è in stato Pending. Lo stato di un'istanza pronta a ricevere traffico è InService.
 - La colonna Health status mostra il risultato dei controlli di integrità di Amazon EC2 Auto Scaling sulla tua istanza.

Fase 4: termina un'istanza nel gruppo con dimensionamento automatico

Puoi usare queste fasi per avere ulteriori informazioni su come funziona Amazon EC2 Auto Scaling, in particolare su come avviare nuove istanze quando è necessario. La dimensione minima per il gruppo con dimensionamento automatico creato in questo tutorial è un'istanza. Pertanto, se chiudi l'istanza in esecuzione, Amazon EC2 Auto Scaling dovrà avviare una nuova istanza per sostituirla.

1. Apri la [pagina Gruppi con dimensionamento automatico](#) nella console Amazon EC2.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.
3. Nella scheda Instance management (Gestione istanze) seleziona l'ID dell'istanza in Instances (Istanze).

Questo ti porta alla pagina Instances (Istanze) della console Amazon EC2, dove è possibile terminare l'istanza.

4. Seleziona Actions (Operazioni), Instance State (Stato istanza), Terminate (Termina). Quando viene richiesta la conferma, scegli Yes, Terminate (Sì, termina).
5. Nel pannello di navigazione, nella sezione Dimensionamento automatico, seleziona Gruppi con dimensionamento automatico. Seleziona il gruppo con dimensionamento automatico, quindi scegli la scheda Attività.

Quando si termina un'istanza dalla pagina Istanze, sono necessari uno o due minuti dopo la chiusura dell'istanza prima che venga avviata una nuova istanza. Nella cronologia delle attività, quando inizia l'attività di dimensionamento, viene mostrata una voce per la terminazione della prima istanza e una voce per l'avvio di una nuova istanza. Usa il pulsante di aggiornamento finché non vedi le nuove voci.

6. Nella scheda Instance management (Gestione istanze), solo la nuova istanza viene visualizzata nella sezione Instances (Istanze).

7. Nel pannello di navigazione, in Instances (Istanze), scegli Instances (Istanze). Questa pagina mostra sia l'istanza terminata che la nuova istanza in esecuzione.

Fase 5: fasi successive

Vai al passaggio successivo se desideri eliminare l'infrastruttura di base che hai appena creato. In caso contrario, puoi utilizzare questa infrastruttura come base e provare uno o più delle seguenti operazioni:

- Connettiti all'istanza Linux utilizzando Session Manager o SSH. Per ulteriori informazioni, consulta [Connessione all'istanza Linux tramite Session Manager](#) e [Connessione all'istanza Linux da Linux o macOS tramite SSH](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.
- Configura una notifica Amazon SNS per avvisare l'utente ogni volta che il gruppo con scalabilità automatica avvia o termina istanze. Per ulteriori informazioni, consulta [Opzioni di notifica Amazon SNS](#).
- Dimensiona manualmente il gruppo con scalabilità automatica per testare la notifica SNS. Per ulteriori informazioni, consulta [Modifica la capacità desiderata del gruppo con dimensionamento automatico](#).

Puoi anche iniziare ad acquisire familiarità con i concetti di dimensionamento automatico leggendo informazioni su [Policy di dimensionamento con monitoraggio degli obiettivi](#). Se il carico dell'applicazione cambia, il gruppo con scalabilità automatica può aumentare orizzontalmente (aggiungere istanze) e ridursi orizzontalmente (eseguire meno istanze) in maniera automatica regolando la capacità desiderata del gruppo tra i limiti di capacità minima e massima. Per ulteriori informazioni sull'impostazione dei limiti, consulta [Impostare i limiti di dimensionamento per il gruppo con dimensionamento automatico](#).

Fase 6: pulizia

Puoi eliminare la tua infrastruttura di scaling o eliminare solo il tuo gruppo Auto Scaling e mantenere il modello di lancio per utilizzarlo in seguito.

Se viene avviata un'istanza che non rientra nel [AWS Free Tier \(Piano gratuito AWS\)](#), dovrai terminare le istanze per evitare costi aggiuntivi. Quando si termina l'istanza, verranno eliminati anche i dati ad essa associati.

Come eliminare il gruppo con dimensionamento automatico

1. Apri la [pagina Gruppi con dimensionamento automatico](#) nella console Amazon EC2.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico (my-first-asg).
3. Scegli Elimina.
4. Quando viene richiesta la conferma, digita **delete** per confermare l'eliminazione del gruppo con dimensionamento automatico specificato e scegli Elimina.

Un'icona di caricamento nella colonna Nome indica che il gruppo con dimensionamento automatico è in fase di eliminazione. Al termine dell'eliminazione, le colonne Desired (Desiderate), Min e Max mostreranno le istanze 0 per il gruppo con dimensionamento automatico. Sono necessari alcuni minuti per terminare l'istanza ed eliminare il gruppo. Aggiorna l'elenco per visualizzare lo stato corrente.

Se desideri mantenere il modello di avvio, salta la procedura seguente.

Per eliminare il modello di avvio

1. Apri la pagina [Configurazioni di avvio](#) della console Amazon EC2.
2. Seleziona il modello di avvio (my-template-for-auto-scaling).
3. Scegliere Actions (Operazioni), Delete template (Elimina modello).
4. Quando viene richiesta la conferma, digitare **Delete** per confermare l'eliminazione del modello di avvio specificato e scegliere Delete (Elimina).

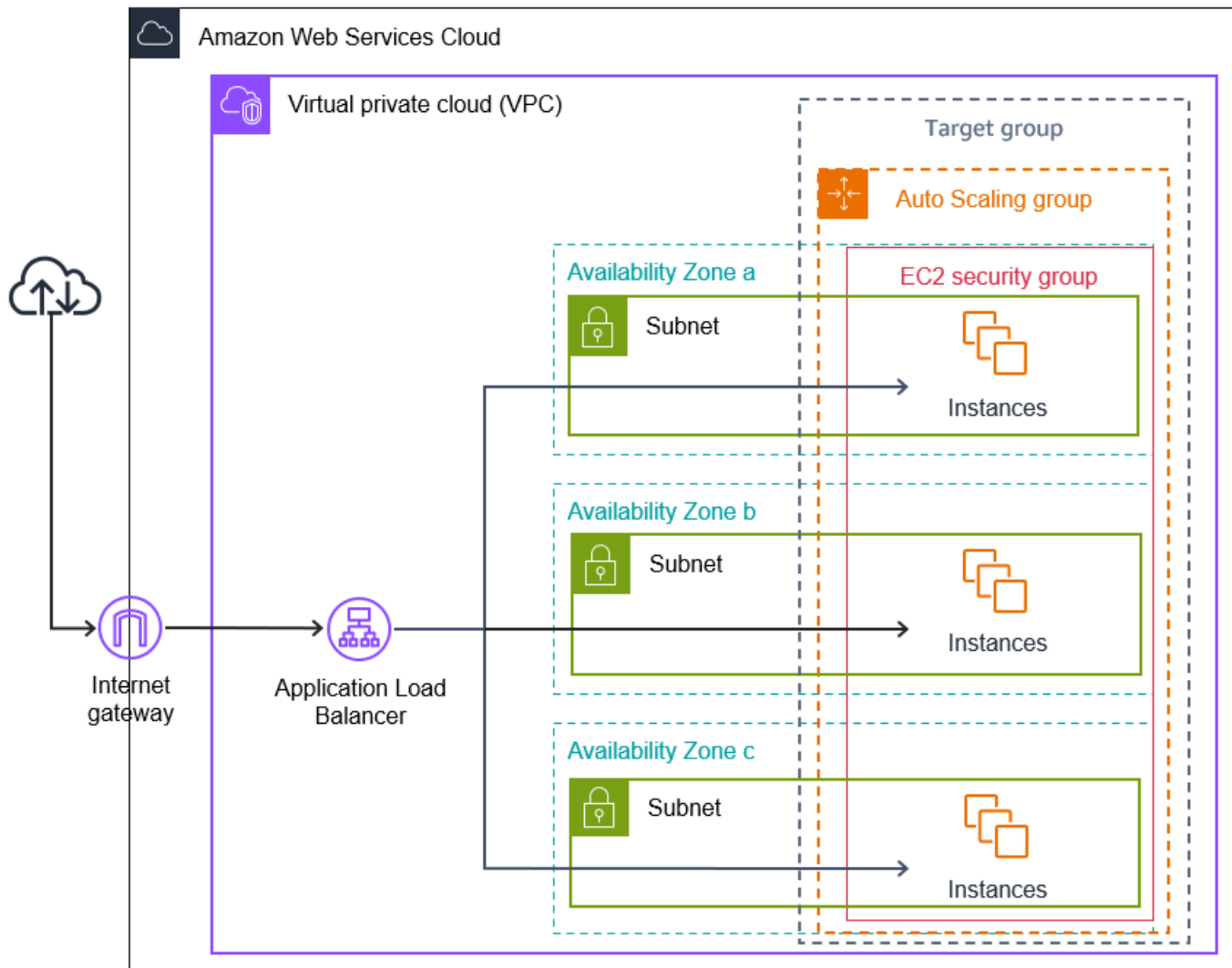
Tutorial: Configurazione di un'applicazione con dimensionamento e bilanciamento del carico

Important

Prima di esplorare questo tutorial, ti consigliamo di leggere prima il seguente tutorial introduttivo: [Crea il tuo primo gruppo di Auto Scaling](#).

La registrazione del gruppo Auto Scaling con un load balancer Elastic Load Balancer ti consente di configurare un'applicazione con bilanciamento del carico. Elastic Load Balancing funziona con Amazon EC2 Auto Scaling per distribuire il traffico in entrata tra le istanze Amazon EC2 integre. Ciò aumenta la scalabilità e la disponibilità dell'applicazione. È possibile abilitare Elastic Load Balancing all'interno di più zone di disponibilità per aumentare la tolleranza ai guasti delle applicazioni.

In questo tutorial vengono illustrate le fasi di base per l'impostazione di un'applicazione con bilanciamento del carico quando viene creato il gruppo Auto Scaling. Una volta terminato, l'architettura dovrebbe apparire simile al diagramma riportato di seguito:



Elastic Load Balancing supporta diversi tipi di bilanciamento del carico. Per questo tutorial, è consigliabile utilizzare un Application Load Balancer.

Per ulteriori informazioni sull'introduzione di un load balancer nell'architettura, vedi [Utilizzare sistema di bilanciamento del carico elastico per distribuire il traffico tra le istanze nel gruppo con dimensionamento automatico.](#)

Attività

- [Prerequisiti](#)
- [Fase 1: configurazione di un modello di avvio o una configurazione di avvio](#)
- [Fase 2: creazione di un gruppo con dimensionamento automatico](#)
- [Fase 3: come verificare che il load balancer sia allegato](#)
- [Fase 4: Fasi successive](#)
- [Fase 5: rimozione](#)
- [Risorse correlate](#)

Prerequisiti

- Un load balancer e un gruppo di destinazione. Assicurati di scegliere le stesse zone di disponibilità per il load balancer che intendi abilitare per il tuo gruppo Auto Scaling. Per ulteriori informazioni, consulta [Nozioni di base dell'Elastic Load Balancing](#) nella Guida per l'utente di 'Elastic Load Balancing.
- Un gruppo di sicurezza per il modello di avvio o la configurazione di avvio. Il gruppo di sicurezza deve permettere l'accesso dal load balancer sulla porta del listener (in genere la porta 80 per il traffico HTTP) e sulla porta che desideri che il bilanciamento del carico Elastic Load Balancing utilizzi per i controlli di integrità. Per ulteriori informazioni, consulta la documentazione relativa:
 - [Gruppi di sicurezza di destinazione](#) nella Guida per l'utente di Application Load Balancer
 - [Gruppi di sicurezza di destinazione](#) nella Guida per l'utente di Network Load Balancer

Se le istanze dispongono di indirizzi IP pubblici, puoi anche scegliere di permettere il traffico SSH, se è necessario che tu ti connetta alle istanze.

- (Facoltativo) Un ruolo IAM che consente all'applicazione di accedere a. AWS
- (Facoltativo) Un'Amazon Machine Image (AMI) come modello di origine per le tue istanze Amazon EC2. Per crearne una ora, avvia un'istanza. Specifica il ruolo IAM (se ne hai creato uno) ed eventuali script di configurazione di cui hai bisogno come dati utente. Connessione e personalizzazione dell'istanza. Ad esempio, puoi installare software e applicazioni, copiare i dati e allegare volumi EBS aggiuntivi. Testa l'applicazione sull'istanza per assicurarsi che quest'ultima

sia configurata correttamente. Salva questa configurazione aggiornata come AMI personalizzata. Puoi terminare l'istanza, se in seguito non ne avrai bisogno. Le istanze avviate da questa nuova AMI personalizzata includono le personalizzazioni apportate durante la creazione dell'AMI stessa.

- Un virtual private cloud (VPC). Questo tutorial fa riferimento al VPC di default, ma puoi anche utilizzare il tuo. In quest'ultimo caso, assicurati che il VPC abbia una sottorete mappata a ciascuna zona di disponibilità della Regione in cui stai lavorando. Per creare il load balancer, ti serviranno almeno due sottoreti pubbliche. Inoltre, dovrai avere a tua disposizione due sottoreti private o due sottoreti pubbliche per creare il gruppo Auto Scaling e registrarlo con il load balancer.

Fase 1: configurazione di un modello di avvio o una configurazione di avvio

Per questo tutorial, utilizza un modello di avvio o una configurazione di avvio.

Argomenti

- [Selezione o crea un modello di lancio](#)
- [Creazione o selezione di una configurazione di avvio](#)

Selezione o crea un modello di lancio

Se disponi già di un modello di avvio che desideri utilizzare, selezionalo utilizzando la procedura seguente.

Come selezionare un modello di avvio esistente

1. Apri la pagina [Configurazioni di avvio](#) della console Amazon EC2.
2. Sulla barra di navigazione nella parte superiore dello schermo, scegli la Regione in cui è stato creato il load balancer.
3. Seleziona un modello di avvio.
4. Scegli Operazioni, Crea un gruppo con dimensionamento automatico.

In alternativa, per creare un nuovo modello di avvio, utilizza la procedura seguente.

Per creare un modello di avvio

1. Apri la pagina [Configurazioni di avvio](#) della console Amazon EC2.

2. Sulla barra di navigazione nella parte superiore dello schermo, scegli la Regione in cui è stato creato il load balancer.
3. Scegli Crea modello di avvio.
4. Immetti un nome e una descrizione per la versione iniziale del modello di avvio.
5. Per Immagini di applicazioni e sistema operativo (Amazon Machine Image), scegliere l'ID dell'AMI per le istanze. Puoi cercare tra tutte le AMI disponibili o selezionarne una dall'elenco Recent (Recenti) o Quick Start (Avvio rapido). Se l'AMI desiderata non viene visualizzata, puoi scegliere Browse more AMIs (Sfoggia altre AMI) per sfogliare il catalogo completo di AMI.
6. In Instance type (Tipo di istanza), seleziona una configurazione hardware per le istanze compatibile con l'AMI che hai specificato.
7. (Facoltativo) In Key pair name (login) (Nome coppia di chiavi (Accesso)) inserisci il nome della coppia di chiavi da utilizzare durante il collegamento alle istanze.
8. Per Impostazioni di rete, amplia Configurazione avanzata di rete e esegui queste operazioni:
 - a. Per configurare un'interfaccia di rete primaria, selezionare Aggiungi interfaccia di rete.
 - b. Per l'assegnazione automatica dell'IP pubblico, specifica se le istanze ricevono indirizzi IPv4 pubblici. Per impostazione predefinita, Amazon EC2 assegna un indirizzo IPv4 pubblico se l'istanza EC2 viene avviata in una sottorete predefinita o se l'istanza viene avviata in una sottorete configurata per assegnare automaticamente un indirizzo IPv4 pubblico. Se non hai bisogno di connetterti alle tue istanze, puoi scegliere Disattiva per impedire alle istanze del tuo gruppo di ricevere traffico direttamente da Internet. In questo caso, le istanze riceveranno traffico solo dal load balancer.
 - c. In Security group ID (ID gruppo di sicurezza), specifica un gruppo di sicurezza per le istanze dallo stesso VPC del load balancer.
 - d. In Delete on termination (Elimina al termine), scegli Yes (Sì). Questo elimina l'interfaccia di rete quando il gruppo Auto Scaling si riduce e termina l'istanza a cui è allegata l'interfaccia di rete.
9. (Facoltativo) Per distribuire in modo sicuro le credenziali alle istanze, in Advanced details (Dettagli avanzati), IAM instance profile (Profilo dell'istanza IAM), inserisci l'Amazon Resource Name (ARN) del ruolo IAM.
10. (Facoltativo) Per specificare i dati utente o uno script di configurazione per le istanze, incollali in Advanced details (Dettagli avanzati), User data (Dati utente).
11. Scegli Crea modello di avvio.
12. Nella pagina di conferma, scegli Crea gruppo con dimensionamento automatico).

Creazione o selezione di una configurazione di avvio

Note

Sconsigliamo vivamente di utilizzare le configurazioni di avvio in nuove applicazioni perché si tratta di una funzionalità legacy senza investimenti pianificati. Inoltre, i nuovi account creati a partire dal 1° giugno 2023 non avranno la possibilità di creare nuove configurazioni di lancio tramite la console. Per ulteriori informazioni, consulta [Configurazioni di avvio](#).

Come selezionare una configurazione di avvio esistente

1. Apri la pagina [Launch configurations \(Configurazioni di avvio\)](#) della console Amazon EC2.
2. Sulla barra di navigazione nella parte superiore dello schermo, scegli la Regione in cui è stato creato il sistema di bilanciamento del carico.
3. Seleziona una configurazione di avvio.
4. Scegli Operazioni, Crea un gruppo con dimensionamento automatico.

In alternativa, per creare una nuova configurazione di avvio, utilizza la procedura seguente.

Per creare una configurazione di avvio


1. Apri la pagina [Launch configurations \(Configurazioni di avvio\)](#) della console Amazon EC2. Quando viene richiesta la conferma, scegli Visualizza configurazioni di avvio per confermare che desideri visualizzare la pagina delle configurazioni di avvio.
2. Sulla barra di navigazione nella parte superiore dello schermo, scegli la Regione in cui è stato creato il sistema di bilanciamento del carico.
3. Scegli Create launch configuration (Crea una configurazione di avvio) e inserisci un nome per la configurazione di avvio.
4. In Amazon Machine Image (AMI), inserisci l'ID dell'AMI per le tue istanze come criterio di ricerca.
5. In Instance type (Tipo di istanza), seleziona una configurazione hardware per l'istanza.
6. In Additional configuration (Configurazione aggiuntiva), presta particolare attenzione ai campi seguenti:
 - a. (Facoltativo) Per distribuire in modo sicuro le credenziali per la tua istanza EC2, in IAM instance profile (Profilo dell'istanza IAM), seleziona il tuo ruolo IAM. Per ulteriori

informazioni, consulta [Ruoli IAM per le applicazioni in esecuzione sulle istanze Amazon EC2](#).

- b. (Facoltativo) Per specificare i dati utente o uno script di configurazione per le istanze, incollali in Advanced details (Dettagli avanzati), User data (Dati utente).
 - c. (Facoltativo) In Advanced details (Dettagli avanzati), IP address type (Tipo di indirizzo IP), mantieni il valore di default. Quando crei il gruppo Auto Scaling, puoi assegnare un indirizzo IP pubblico alle istanze del gruppo utilizzando le sottoreti per le quali è abilitato l'attributo di indirizzamento IP pubblico, ad esempio le sottoreti predefinite nel VPC di default. In alternativa, se non devi connetterti alle istanze, puoi scegliere Do not assign a public IP address to any instances (Non assegnare un indirizzo IP pubblico a nessuna istanza) per impedire alle istanze del gruppo di ricevere traffico direttamente da Internet. In questo caso, le istanze riceveranno traffico solo dal load balancer.
7. In Gruppi di sicurezza, scegli un gruppo di sicurezza esistente dallo stesso VPC del load balancer. Se lasci Create new security group (Crea nuovo gruppo di sicurezza) selezionato, viene configurata una regola SSH predefinita per le istanze Amazon EC2 che eseguono Linux. Viene configurato un ruolo RDP di default per le istanze Amazon EC2 che eseguono Windows.
 8. In Key pair (login) (Coppia di chiavi (login)), scegli un'opzione in Key pair options (Opzioni di coppia di chiavi).

Se hai già configurato una coppia di chiavi dell'istanza Amazon EC2 instance, è possibile sceglierla in questa pagina.

Se invece non hai ancora una coppia di chiavi dell'istanza Amazon EC2, scegli Create a new key pair (Crea una nuova coppia di chiavi) e assegna un nome riconoscibile. Scegli Download key pair (Scarica coppia di chiavi) per scaricare la coppia di chiavi sul computer.

 Important

Se hai bisogno di connetterti alle istanze, non scegliere Proceed without a key pair (Procedi senza una coppia di chiavi).

9. Selezionare la casella di controllo di conferma, quindi scegliere Create launch configuration (Crea configurazione di avvio).
10. Seleziona la casella di controllo accanto al nome della nuova configurazione di avvio e scegli Operazioni, Gruppo con dimensionamento automatico.

Fase 2: creazione di un gruppo con dimensionamento automatico

Utilizza la procedura seguente per continuare dal punto in cui hai interrotto dopo aver selezionato o creato il modello o la configurazione di avvio.

Creazione di un gruppo con dimensionamento automatico

1. Nella pagina Scegli il modello o la configurazione di avvio, in Nome gruppo con dimensionamento automatico, inserisci un nome per il gruppo con dimensionamento automatico.
2. [Solo modello di avvio] In Versione del modello di avvio, scegli se per l'aumento orizzontale il gruppo con dimensionamento automatico debba usare la versione di default, quella più recente o una versione specifica del modello di avvio.
3. Seleziona Successivo.

Viene visualizzata la pagina Scegli le opzioni di avvio dell'istanza che consente di scegliere le impostazioni di rete VPC che si desidera utilizzare il gruppo con dimensionamento automatico e offre opzioni per l'avvio di istanze on demand e istanze Spot (se scegli un modello di avvio).

4. Nella sezione Network (Rete), per VPC, scegli il VPC utilizzato per il load balancer. Se sceglierai il VPC di default, questo verrà configurato automaticamente per fornire la connettività Internet alle istanze. Questo VPC include una sottorete pubblica in ciascuna zona di disponibilità nella regione.
5. Per Availability Zones and subnets (Zone di disponibilità e sottoreti), scegli una o più sottoreti ciascuna zona di disponibilità che desideri includere, in base alle zone di disponibilità in cui si trova il load balancer. Per ulteriori informazioni, consulta [Considerazioni sulla scelta di sottoreti VPC](#).
6. [Solo modello di avvio] Nella sezione Instance type requirements (Requisiti del tipo di istanza), utilizza l'impostazione predefinita per semplificare questo passaggio. (Non sovrascrivere il modello di avvio.) Per questo tutorial, avvierai solo le istanze on demand utilizzando il tipo di istanza specificato nel modello di avvio.
7. Scegli Next (Avanti) per accedere alla pagina Configure advanced options (Configura opzioni avanzate).
8. Per collegare il gruppo a un sistema di bilanciamento del carico esistente, nella sezione Load balancing (Bilanciamento del carico), scegli Attach to an existing load balancer (Collega a un load balancer esistente). È possibile scegliere Choose from your load balancer target groups (Scegli tra i gruppi di destinazione del load balancer) o Choose from Classic Load Balancers (Scegli tra Classic Load Balancer). È quindi possibile scegliere il nome di un gruppo di

destinazione per Application Load Balancer o Network Load Balancer creato oppure scegliere il nome di un Classic Load Balancer.

9. (Facoltativo) Per utilizzare i controlli d'integrità dell'Elastic Load Balancing per Health checks (Controlli dell'integrità), scegli ELB in Health check type (Tipo di controllo dell'integrità).
10. Al termine della configurazione del gruppo Auto Scaling, scegli Skip to review (Vai alla verifica).
11. Nella pagina Review (Verifica), riesamina i dettagli del gruppo Auto Scaling. Per apportare delle modifiche, scegli Edit (Modifica). Al termine, scegli Create Auto Scaling group (Crea gruppo Auto Scaling).

Dopo aver creato il gruppo Auto Scaling con il load balancer allegato, quest'ultimo registra automaticamente le nuove istanze, una volta che queste saranno online. A questo punto è disponibile solo un'istanza, quindi non c'è molto da registrare. Tuttavia, ora puoi aggiungere altre istanze aggiornando la capacità desiderata del gruppo. Per step-by-step istruzioni, consulta [Modifica la capacità desiderata del gruppo con dimensionamento automatico](#).

Fase 3: come verificare che il load balancer sia allegato

Come verificare che il load balancer sia allegato

1. Dalla [pagina Gruppi con dimensionamento automatico](#) della console Amazon EC2 seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.
2. Nella scheda Details (Dettagli), il campo Load balancing (Bilanciamento del carico), mostra tutti i gruppi di destinazione del load balancer o del Classic Load Balancer allegati.
3. Puoi verificare che le istanze siano state avviate correttamente nella scheda Activity (Attività), in Activity history (Cronologia delle attività). La colonna Status (Stato) indica se il gruppo Auto Scaling ha avviato le istanze. Se le istanze non vengono avviate, puoi trovare idee per la risoluzione dei problemi comuni di avvio delle istanze in [Risoluzione dei problemi di Amazon EC2 Auto Scaling](#).
4. Nella scheda Instance management (Gestione istanze) in Instances (Istanze), è possibile verificare che le istanze siano pronte a ricevere traffico. Inizialmente, lo stato delle istanze è Pending. Lo stato di un'istanza pronta a ricevere traffico è InService. La colonna Health status (Stato) mostra il risultato dei controlli dell'integrità sulle istanze Amazon EC2 Auto Scaling. Sebbene un'istanza possa essere contrassegnata come integra, il load balancer invierà traffico solo alle istanze che superano i controlli di integrità relativi.
5. Verifica che le istanze siano registrate con il load balancer. Apri la [pagina Target groups \(Gruppi di destinazione\)](#) nella console Amazon EC2. Seleziona il tuo gruppo di destinazione e poi scegli

la scheda Targets (Destinazioni). Se lo stato delle istanze è `initial`, è probabilmente perché sono ancora in fase di registrazione, o sono ancora sottoposti al controllo dell'integrità. Quando lo stato delle istanze è `healthy`, queste sono pronte per l'uso.

Fase 4: Fasi successive

Ora che hai completato questo tutorial, puoi scoprire di più su:

- Dimensionamento automatico Amazon EC2 determina se un'istanza è integra in base allo stato dei controlli dell'integrità utilizzati dal gruppo con dimensionamento automatico. Se abiliti i controlli dello stato del bilanciamento del carico e un'istanza non supera i controlli di integrità, il gruppo Auto Scaling considera l'istanza non integra e la sostituisce. Per ulteriori informazioni, consulta [Controlli dell'integrità](#).
- Puoi espandere la tua applicazione a una zona di disponibilità aggiuntiva nella stessa Regione per aumentare la tolleranza ai guasti in caso di interruzione di un servizio. Per ulteriori informazioni, consulta [Aggiungere zone di disponibilità](#).
- Puoi configurare il gruppo Auto Scaling affinché utilizzi una policy di dimensionamento con monitoraggio degli obiettivi. Ciò aumenta o diminuisce automaticamente il numero di istanze man mano la domanda varia. Ciò consente al gruppo di gestire le modifiche nel traffico ricevuto dall'applicazione. Per ulteriori informazioni, consulta [Policy di dimensionamento con monitoraggio degli obiettivi](#).

Fase 5: rimozione

Dopo aver creato le risorse per questo tutorial, è consigliabile considerare di effettuare la pulizia per evitare addebiti superflui.

Come eliminare il gruppo con dimensionamento automatico

1. Apri la [pagina Gruppi con dimensionamento automatico](#) nella console Amazon EC2.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.
3. Scegli Elimina.
4. Quando viene richiesta la conferma, digita **delete** per confermare l'eliminazione del gruppo con dimensionamento automatico specificato e scegli Elimina.

Un'icona di caricamento nella colonna Nome indica che il gruppo con dimensionamento automatico è in fase di eliminazione. Al termine dell'eliminazione, le colonne Desired (Desiderate), Min e Max mostreranno le istanze 0 per il gruppo con dimensionamento automatico. Sono necessari alcuni minuti per terminare l'istanza ed eliminare il gruppo. Aggiorna l'elenco per visualizzare lo stato corrente.

Se desideri mantenere il modello di avvio, salta la procedura seguente.

Per eliminare il modello di avvio

1. Apri la pagina [Configurazioni di avvio](#) della console Amazon EC2.
2. Seleziona il modello di avvio.
3. Scegliere Actions (Operazioni), Delete template (Elimina modello).
4. Quando viene richiesta la conferma, digitare **Delete** per confermare l'eliminazione del modello di avvio specificato e scegliere Delete (Elimina).

Se desideri mantenere la configurazione di avvio, salta la procedura seguente.

Per eliminare la configurazione di avvio

1. Apri la pagina [Launch configurations \(Configurazioni di avvio\)](#) della console Amazon EC2.
2. Seleziona la configurazione di avvio.
3. Scegli Actions (Operazioni), Delete launch configuration (Elimina configurazione di avvio).
4. Quando viene richiesta la conferma, seleziona Delete (Elimina).

Se desideri mantenere il load balancer per uso futuro, salta la procedura seguente.

Come eliminare il load balancer

1. Apri la [pagina bilanciatori del carico](#) nella console Amazon EC2.
2. Seleziona il load balancer e scegli Actions (Operazioni), Delete (Elimina).
3. Quando viene richiesta la conferma, seleziona Sì, elimina.

Come eliminare i gruppi di destinazione

1. Apri la [pagina Target groups \(Gruppi di destinazione\)](#) nella console Amazon EC2.
2. Selezionare il gruppo di destinazione e scegliere Operazioni, Elimina.
3. Quando viene richiesta la conferma, seleziona Sì, elimina.

Risorse correlate

Con AWS CloudFormation, puoi creare ed eseguire il provisioning AWS delle implementazioni dell'infrastruttura in modo prevedibile e ripetuto, utilizzando file modello per creare ed eliminare una raccolta di risorse insieme come una singola unità (uno stack). Per ulteriori informazioni, consulta la [AWS CloudFormation Guida per l'utente](#).

Per una procedura dettagliata che utilizza un modello di stack per effettuare il provisioning di un gruppo con dimensionamento automatico e di un Application Load Balancer, consulta [Spiegazione passo per passo: creazione di un'applicazione con dimensionamento e sistema di bilanciamento del carico](#) nella Guida per l'utente di AWS CloudFormation . Utilizza questa procedura e questo modello di esempio come punto di partenza per creare modelli simili adatti alle tue esigenze.

Modelli di avvio

Un modello di avvio è simile a una [configurazione di avvio](#), in quanto specifica le informazioni di configurazione dell'istanza. Sono inclusi l'ID dell'Amazon Machine Image (AMI), il tipo di istanza, una coppia di chiavi, i gruppi di sicurezza e altri parametri che utilizzi per avviare le istanze EC2. Tuttavia, la definizione di un modello di avvio invece di una configurazione di avvio permette di avere più versioni di un modello.

Con il controllo delle versioni dei modelli di avvio, è possibile creare un sottoinsieme del set completo di parametri. Quindi, è possibile riutilizzarlo per creare altre versioni dello stesso modello di avvio. Ad esempio, è possibile creare un modello di avvio che definisce una configurazione di base senza un AMI o uno script di dati utente. Dopo aver creato il modello di avvio, è possibile creare una nuova versione e aggiungere l'AMI e i dati utente con la versione più recente dell'applicazione per l'esecuzione dei test. Ne conseguono due versioni del modello di avvio. L'archiviazione di una configurazione di base consente di mantenere i parametri di configurazione generali richiesti. Puoi creare una nuova versione del modello di avvio dalla configurazione di base in qualsiasi momento tu desideri farlo. Inoltre, quando non ti servono più, puoi eliminare le versioni utilizzate per testare l'applicazione.

Noi ti suggeriamo di utilizzare i modelli di avvio per accertarti di accedere alle caratteristiche e ai miglioramenti più recenti. Quando si utilizzano configurazioni di avvio, non tutte le caratteristiche di Amazon EC2 Auto Scaling sono disponibili. Ad esempio, non è possibile creare un gruppo con dimensionamento automatico che avvii istanze Spot e on demand oppure che specifichi più tipi di istanza. Per configurare queste caratteristiche, sarà necessario utilizzare un modello di avvio. Per ulteriori informazioni, consulta [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#).

Con i modelli di avvio, puoi anche utilizzare le funzionalità più recenti di Amazon EC2. Ciò include i parametri di Systems Manager (ID AMI), l'attuale generazione di volumi EBS con capacità di IOPS allocata (io2), l'assegnazione dei tag del volume EBS, le istanze T2 in modalità illimitata, le prenotazioni della capacità, i blocchi di capacità e gli host dedicati, solo per citarne alcuni.

Quando crei un modello di avvio, tutti i parametri sono facoltativi. Tuttavia, se un modello di avvio non specifica un AMI, non potrai aggiungerlo durante la creazione del gruppo con dimensionamento automatico. Se specifichi un AMI ma nessun tipo di istanza, potrai aggiungere uno o più tipi di istanza durante la creazione del gruppo con dimensionamento automatico.

Indice

- [Autorizzazioni per lavorare con i modelli di lancio](#)

- [Operazioni API supportate dai modelli di avvio](#)
- [Creare un modello di avvio per un gruppo con dimensionamento automatico](#)
- [Crea un modello di avvio utilizzando le impostazioni avanzate](#)
- [Migra i tuoi gruppi di Auto Scaling ai modelli di avvio](#)
- [Migra AWS CloudFormation gli stack ai modelli di lancio](#)
- [Esempi di creazione e gestione di modelli di avvio con \(\) AWS Command Line InterfaceAWS CLI](#)
- [Usa AWS Systems Manager i parametri anziché gli ID AMI nei modelli di avvio](#)

Autorizzazioni per lavorare con i modelli di lancio

Le procedure in questa sezione presuppongono che tu disponga già delle autorizzazioni necessarie per creare i modelli di avvio. Per informazioni sulla concessione di autorizzazioni da parte di un amministratore, consulta [Controllare l'accesso ai modelli di avvio con le autorizzazioni IAM](#) nella Manuale utente di Amazon EC2 per istanze Linux.

Tieni presente che se non disponi di autorizzazioni sufficienti per l'utilizzo e la creazione di risorse specificate in un modello di avvio, riceverai un errore corrispondente quando tenti di specificarlo per un gruppo con dimensionamento automatico. Per ulteriori informazioni, consulta [Risoluzione dei problemi di Amazon EC2 Auto Scaling: modelli di avvio](#).

Per esempi di policy IAM che consentono di chiamare le operazioni RunInstances API e CreateAutoScalingGroupUpdateAutoScalingGroup, con un modello di avvio, consulta [Supporto modello di avvio](#).

Operazioni API supportate dai modelli di avvio

Per un elenco delle operazioni API supportate dai modelli di avvio, consulta le [azioni di Amazon EC2](#) nella [Documentazione di riferimento per le API di EC2](#).

Creare un modello di avvio per un gruppo con dimensionamento automatico

Prima di poter creare un gruppo con dimensionamento automatico utilizzando un modello di avvio, è necessario crearne uno che includa i parametri necessari per avviare un'istanza, ad esempio l'ID dell'Amazon Machine Image (AMI) e un tipo di istanza.

Per creare nuovi modelli di avvio, completa le seguenti procedure.

Indice

- [Crea il tuo modello di avvio \(console\)](#)
- [Modifica le impostazioni dell'interfaccia di rete di default \(console\)](#)
- [Modifica la configurazione di archiviazione \(console\)](#)
- [Creare un modello di avvio da un'istanza esistente \(console\)](#)
- [Risorse correlate](#)
- [Limitazioni](#)

Important

I parametri del modello di avvio non vengono convalidati completamente al momento della creazione. Se si specificano valori errati per i parametri o se non si utilizzano combinazioni di parametri supportate, non è possibile avviare istanze utilizzando questo modello di avvio. Assicurati di specificare i valori corretti per i parametri e di utilizzare le combinazioni di parametri supportate. Ad esempio, per avviare istanze con un'AMI AWS Graviton o Graviton2 basata su Arm, è necessario specificare un tipo di istanza compatibile con Arm. Per ulteriori informazioni, consulta [Restrizioni del modello di avvio](#) nella Manuale per l'utente di Amazon EC2 per le istanze Linux).

Crea il tuo modello di avvio (console)

Di seguito viene descritto come configurare un modello di avvio di base:

- Specifica l'Amazon Machine Image (AMI) da cui avviare le istanze.
- Scegli un tipo di istanza compatibile con l'AMI specificata.
- Specifica la coppia di chiavi da utilizzare durante il collegamento alle istanze, utilizzando ad esempio SSH.
- Aggiungi uno o più gruppi di sicurezza per consentire l'accesso di rete alle istanze.
- Specifica altri eventuali volumi da collegare a ogni istanza.
- Aggiungi tag personalizzati (coppie chiave-valore) alle istanze e ai volumi.

Per creare un modello di avvio

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nel pannello di navigazione, in Istanze, scegli Modelli di avvio.
3. Scegli Crea modello di avvio. Immetti un nome e una descrizione per la versione iniziale del modello di avvio.
4. (Facoltativo) Nella sezione Guida al dimensionamento automatico, seleziona la casella di controllo affinché Amazon EC2 fornisca indicazioni utili per creare un modello da utilizzare con Dimensionamento automatico Amazon EC2.
5. In Contenuto del modello di avvio, compila ogni campo obbligatorio e tutti i campi facoltativi come richiesto.
 - a. Application and OS Images (Amazon Machine Image) (Immagini di applicazione e SO (Amazon Machine Image): (obbligatorio) scegli l'ID dell'AMI per le tue istanze. Puoi cercare tutte le AMI disponibili o selezionarne una dall'elenco Recent (Recenti) o Quick Start (Avvio rapido). Se l'AMI necessaria non viene visualizzata, scegli Browse more AMIs (Sfoglialtre AMI) per sfogliare il catalogo completo di AMI.

Per scegliere un'AMI personalizzata, devi prima creare l'AMI da un'istanza personalizzata. Per ulteriori informazioni, consulta [Creare un'AMI](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

- b. Per Tipo istanza, scegli un singolo tipo di istanza compatibile con l'AMI specificata.

Per utilizzare la selezione del tipo di istanza basata su attributi, scegli Avanzato, Specifica gli attributi del tipo di istanza, quindi specifica le seguenti opzioni:

- Numero di vCPU: inserire il numero minimo e massimo di vCPU. Per indicare l'assenza di limite, inserisci un valore minimo uguale a 0 e lascia vuoto il campo del valore massimo.
- Quantità di memoria (MiB): inserire la quantità minima e massima di memoria, in MiB. Per indicare l'assenza di limite, inserisci un valore minimo uguale a 0 e lascia vuoto il campo del valore massimo.
- Espandi Attributi del tipo di istanza facoltativi e scegli Aggiungi attributo per limitare ulteriormente i tipi di istanze utilizzabili per soddisfare la capacità desiderata. Per informazioni su ciascun attributo, consulta [InstanceRequirementsRequest](#) in riferimento alle API di Amazon EC2.
- Tipi di istanza risultanti: è possibile visualizzare i tipi di istanza che corrispondono ai requisiti di calcolo specificati, come ad esempio vCPU, memoria e archiviazione.

- Per escludere i tipi di istanza, scegli Aggiungi attributo. Dall'elenco Attributo, scegli Tipi di istanza esclusi. Dall'elenco Valore attributo, seleziona i tipi di istanza da escludere.
- c. Coppia di chiavi (accesso): per Nome coppia di chiavi, scegli una coppia di chiavi esistente oppure scegli Crea nuova coppia di chiavi per crearne una nuova. Per ulteriori informazioni, consulta [Coppie di chiavi Amazon EC2 e istanze Linux](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.
- d. Network settings (Impostazioni di rete): per Firewall (security groups) (Firewall (gruppi di sicurezza)), utilizza uno o più gruppi di sicurezza oppure lascia vuoto il campo per configurare uno o più gruppi di sicurezza come parte dell'interfaccia di rete. Per ulteriori informazioni, consulta [Gruppi di sicurezza Amazon EC2 per le istanze Linux](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Se non specifichi alcun gruppo di sicurezza nel modello di avvio, Amazon EC2 utilizzerà il gruppo di sicurezza di default per il VPC in cui il gruppo con dimensionamento automatico avvierà le istanze. Per impostazione predefinita, questo gruppo di sicurezza non permette il traffico in ingresso da reti esterne. Per ulteriori informazioni, consulta [Gruppi di sicurezza di default per i VPC](#) nella Guida per l'utente di Amazon VPC.

- e. Esegui una di queste operazioni:
 - Modifica le impostazioni dell'interfaccia di rete di default. Ad esempio, è possibile abilitare o disabilitare la funzione di indirizzamento IPv4 pubblica, che sostituisce l'impostazione di assegnazione automatica degli indirizzi IPv4 pubblici nella sottorete. Per ulteriori informazioni, consulta [Modifica le impostazioni dell'interfaccia di rete di default \(console\)](#).
 - Se desideri mantenere le impostazioni dell'interfaccia di rete di default, salta questo passaggio.
- f. Esegui una di queste operazioni:
 - Modifica la configurazione dell'archiviazione. Per ulteriori informazioni, consulta [Modifica la configurazione di archiviazione \(console\)](#).
 - Se desideri mantenere la configurazione di archiviazione di default, salta questo passaggio.
- g. Per Tag risorse, specifica i tag fornendo combinazioni di chiavi e valori. Se si specificano i tag delle istanze nel modello di avvio e si sceglie di propagare i tag del gruppo con dimensionamento automatico alle relative istanze, tutti i tag vengono uniti. Se viene specificata la stessa chiave tag per un tag nel modello di avvio e un tag nel gruppo con dimensionamento automatico, il valore del tag del gruppo ha la precedenza.

6. (Facoltativo) Configura le impostazioni avanzate. Ad esempio, puoi scegliere un ruolo IAM che l'applicazione può utilizzare quando accede ad altre risorse AWS o specificare i dati utente dell'istanza che possono essere utilizzati per eseguire processi di configurazione automatizzati comuni dopo l'avvio di un'istanza. Per ulteriori informazioni, consulta [Crea un modello di avvio utilizzando le impostazioni avanzate](#).
7. Quando è tutto pronto per creare il modello di avvio, scegli Create launch template (Crea modello di avvio).
8. Per creare un gruppo con dimensionamento automatico, scegli Crea un gruppo con dimensionamento automatico dalla pagina di conferma.

Modifica le impostazioni dell'interfaccia di rete di default (console)

Le interfacce di rete forniscono connettività ad altre risorse nel tuo VPC e in Internet. Per ulteriori informazioni, consulta [Fornisci connettività di rete per le istanze Auto Scaling utilizzando Amazon VPC](#).

In questa sezione viene illustrato come modificare le impostazioni dell'interfaccia di rete di default. In questo modo puoi definire, ad esempio, se desideri assegnare un indirizzo IPv4 pubblico a ciascuna istanza invece di impostare gli indirizzi IPv4 pubblici ad assegnazione automatica nella sottorete.

Considerazioni e limitazioni

Quando modifichi le impostazioni dell'interfaccia di rete di default, tieni presente le seguenti considerazioni e limitazioni:

- Dovrai configurare il gruppo di sicurezza nell'ambito dell'interfaccia di rete e non nella sezione Security groups (Gruppi di sicurezza) del modello. Non è possibile specificare i gruppi di sicurezza in entrambe le posizioni.
- Non puoi assegnare indirizzi IP privati secondari, i cosiddetti indirizzi IP secondari, ad una interfaccia di rete.
- Se specifichi un ID di interfaccia di rete esistente, puoi avviare una sola istanza. A tale scopo, è necessario utilizzare AWS CLI o un SDK per creare il gruppo Auto Scaling. Quando crei il gruppo, devi specificare la zona di disponibilità, ma non l'ID della sottorete. Inoltre, puoi specificare un'interfaccia di rete esistente solo se ha un indice dispositivo pari a 0.
- Se specifichi più di un'interfaccia di rete, non potrai assegnare automaticamente un indirizzo IPv4 pubblico. Non è possibile, inoltre, specificare gli indici dei dispositivi duplicati tra le interfacce di

rete. Nota che entrambe le interfacce di rete, primaria e secondaria, risiederanno nella stessa sottorete.

- Quando un'istanza viene avviata, viene allocato automaticamente un indirizzo privato a ciascuna interfaccia di rete. L'indirizzo proviene dall'intervallo CIDR della sottorete in cui viene avviata l'istanza. Per ulteriori informazioni sulla specifica dei blocchi CIDR (o degli intervalli di indirizzi IP) per il VPC o la sottorete, consulta la [Guida per l'utente di Amazon VPC](#).

Modifica delle impostazioni dell'interfaccia di rete di default

1. In Network settings (Impostazioni di rete), espandi Advanced network configuration (Configurazione di rete avanzata).
2. Scegli Add network interface (Aggiungi interfaccia di rete) per configurare l'interfaccia di rete primaria, facendo attenzione ai seguenti campi:
 - a. Indice dispositivo: mantieni il valore di default, 0, per applicare le modifiche all'interfaccia di rete primaria (eth0).
 - b. Interfaccia di rete: mantieni il valore di default, New interface (Nuova interfaccia), affinché Amazon EC2 Auto Scaling crei automaticamente una nuova interfaccia di rete all'avvio di un'istanza. In alternativa, puoi scegliere un'interfaccia di rete esistente disponibile con un indice di dispositivo pari a 0, ma questo limita il gruppo con dimensionamento automatico a una sola istanza.
 - c. Description (Descrizione): (facoltativo) inserisci un nome descrittivo.
 - d. Sottorete: mantieni l'impostazione di default Don't include in launch template (Non includere nel modello di avvio).

Se l'AMI specifica una sottorete per l'interfaccia di rete, si verifica un errore. Si consiglia di disattivare l'opzione Auto Scaling guidance (Guida Auto Scaling) come soluzione alternativa. Dopo aver apportato questa modifica, non verrà ricevuto alcun messaggio di errore. Tuttavia, indipendentemente da dove è specificata la sottorete, le impostazioni della sottorete del gruppo con dimensionamento automatico hanno la precedenza e non possono essere sovrascritte.

- e. Assegna automaticamente IP pubblico: cambia se l'interfaccia di rete con un indice dispositivo pari a 0 riceve un indirizzo IPv4 pubblico. Per impostazione predefinita, le istanze in una sottorete di default ricevono un indirizzo IPv4 pubblico, al contrario delle istanze in una sottorete non di default. Selezionare Abilita o Disabilita per sostituire l'impostazione predefinita della sottorete.

- f. **Gruppi di sicurezza:** scegli uno o più gruppi di sicurezza per l'interfaccia di rete. Ciascun gruppo di sicurezza deve essere configurato per il VPC nel quale il gruppo con dimensionamento automatico avvierà le istanze. Per ulteriori informazioni, consulta [Gruppi di sicurezza Amazon EC2 per le istanze Linux](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.
 - g. **Elimina al termine:** scegli Sì per eliminare l'interfaccia di rete al termine dell'istanza oppure No per mantenere l'interfaccia di rete.
 - h. **Elastic Fabric Adapter:** per supportare i casi d'uso del calcolo computerizzato ad alte prestazioni e del machine learning, cambiare l'interfaccia di rete in un'interfaccia di rete Elastic Fabric Adapter. Per ulteriori informazioni, consulta [Elastic Fabric Adapter](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.
 - i. **Indice delle schede di rete:** scegli 0 per collegare l'interfaccia di rete primaria alla scheda di rete con un indice dispositivo pari a 0. Se questa opzione non è disponibile, mantieni il valore di default, Don't include in launch template (Non includere nel modello di avvio). Il collegamento dell'interfaccia di rete a una scheda di rete specifica è disponibile solo per i tipi di istanza supportati. Per ulteriori informazioni, consulta [Schede di rete](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.
 - j. **ENA Express:** per esempio i tipi che supportano ENA Express, scegli Abilita per abilitare ENA Express o Disabilita per disabilitarlo. Per maggiori informazioni, consulta [Migliora le prestazioni di rete con ENA Express su istanze Linux](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.
 - k. **ENA Express UDP:** se abiliti ENA Express, puoi opzionalmente utilizzarlo per il traffico UDP. Scegli Abilita per abilitare ENA Express UDP o Disabilita per disabilitarlo.
3. Per aggiungere un'interfaccia di rete secondaria, seleziona **Aggiungi interfaccia di rete**.

Modifica la configurazione di archiviazione (console)

Puoi modificare la configurazione di archiviazione per le istanze avviate da AMI supportate da Amazon EBS o da AMI supportate dall'archivio dell'istanza. Puoi specificare altri volumi EBS da collegare alle istanze. L'AMI include uno o più volumi di archiviazione, compreso il volume root (Volume 1 (AMI Root)).

Modifica della configurazione dell'archiviazione

1. In **Configure storage (Configura archiviazione)**, modifica la dimensione o il tipo di volume.

Se il valore specificato per la dimensione del volume è al di fuori dei limiti del tipo di volume o inferiore alla dimensione dello snapshot, viene visualizzato un messaggio di errore. Per risolvere il problema, questo messaggio fornisce il valore minimo o massimo che il campo può accettare.

Appaiono solo i volumi associati a un'AMI supportata da Amazon EBS. Per visualizzare informazioni sulla configurazione di archiviazione per un'istanza avviata da un'AMI supportata dall'archivio dell'istanza, scegli Show details (Mostra dettagli) dalla sezione Instance store volumes (Volumi dell'archivio istanza).

Per specificare tutti i parametri del volume EBS, passa alla vista Advanced (Avanzate) nell'angolo in alto a destra.

2. Per le opzioni avanzate, espandi il volume che desideri modificare e configuralo come segue:
 - a. Tipo di archiviazione: il tipo di volume (EBS o temporaneo) da associare all'istanza. Il tipo di volume dell'archivio istanza (temporaneo) è disponibile solo se si seleziona un tipo di istanza che lo supporta. Per ulteriori informazioni, consulta [i volumi Amazon EBS](#) nella Amazon EBS User Guide e Amazon [EC2 instance store nella Amazon EC2](#) User Guide for Linux Instances.
 - b. Dispositivo: seleziona dall'elenco di nomi dei dispositivi disponibili per il volume.
 - c. Snapshot: selezionare lo snapshot da cui creare il volume. È inoltre possibile cercare snapshot pubblici e condivisi disponibili digitando il testo nel campo Snapshot.
 - d. Dimensioni: per i volumi EBS, è possibile specificare una dimensione di archiviazione. Se è stata selezionata un'AMI e l'istanza è idonea per il piano gratuito, per rientrare nel piano gratuito è necessario mantenersi al di sotto dei 30 GiB di archiviazione totale. Per ulteriori informazioni, consulta [Vincoli sulle dimensioni e sulla configurazione di un volume EBS nella Amazon EBS](#) User Guide.
 - e. Tipo di volume: per i volumi EBS, scegli il tipo di volume. Per ulteriori informazioni, consulta [tipi di volume di Amazon EBS](#) nella Guida per l'utente di Amazon EBS.
 - f. IOPS: se è stato selezionato il tipo di volume SSD con capacità di IOPS allocata (io1 e io2) o SSD per uso generale (gp3), è possibile inserire il numero di operazioni I/O per secondo (IOPS) che il volume può supportare. Ciò è necessario per i volumi io1, io2 e gp3. Non è supportato per gp2, st1, sc1 o volumi standard.
 - g. Delete on termination (Elimina alla terminazione): per i volumi EBS, scegli Yes (Sì) per eliminare il volume quando l'istanza viene terminata oppure No per conservare il volume.

- h. Crittografato: se il tipo di istanza supporta la crittografia EBS, scegliere Sì per abilitare la crittografia per il volume. Se hai abilitato la crittografia per impostazione predefinita in questa regione allora la crittografia è abilitata. Per ulteriori informazioni, consulta [Amazon EBS encryption e Enable encryption by default](#) nella Amazon EBS User Guide.

L'effetto di default impostando questo parametro varia con la scelta del volume di origine, come descritto nella tabella sottostante. In tutti i casi, è necessario disporre dell'autorizzazione per utilizzare quanto specificato. AWS KMS key

Risultati della crittografia

Se il parametro Encrypted è impostato su...	Inoltre, se l'origine del volume è...	Allora lo stato predefinito della crittografia è...	Note
No	Nuovo volume (vuoto)	Non crittografato*	N/D
	Snapshot non crittografato di tua proprietà	Non crittografato*	
	Snapshot crittografato di tua proprietà	Crittografato dalla stessa chiave	
	Snapshot non crittografato condiviso con te	Non crittografato*	
	Snapshot crittografato condiviso con te	Crittografato con chiave KMS predefinita	

Se il parametro Encrypted è impostato su...	Inoltre, se l'origine del volume è...	Allora lo stato predefinito della crittografia è...	Note
Sì	Nuovo volume	Crittografato con chiave KMS predefinita	Per usare una chiave KMS non di default, specifica un valore per il parametro KMS key (Chiave KMS).
	Snapshot non crittografato di tua proprietà	Crittografato con chiave KMS predefinita	
	Snapshot crittografato di tua proprietà	Crittografato dalla stessa chiave	
	Snapshot non crittografato condiviso con te	Crittografato con chiave KMS predefinita	
	Snapshot crittografato condiviso con te	Crittografato con chiave KMS predefinita	

* Se per impostazione predefinita è abilitata la crittografia, tutti i volumi appena creati (indipendentemente dall'impostazione del parametro) Encrypted (Crittografato) su Yes (Sì) vengono crittografati utilizzando la chiave KMS di default. L'impostazione di entrambi i parametri Encrypted (Crittografato) e Key (Chiave) permette di specificare una chiave KMS non di default.

- i. KMS key (Chiave KMS): se hai selezionato Yes (Sì) per Encrypted (Crittografato), allora dovrai selezionare una chiave gestita dal cliente da utilizzare per crittografare il volume. Se hai abilitato la crittografia per impostazione predefinita in questa Regione, viene selezionata automaticamente la chiave gestita dal cliente predefinita. Puoi selezionare una chiave diversa o specificare l'ARN di qualsiasi chiave gestita dal cliente creata in precedenza tramite AWS Key Management Service.

3. Per specificare altri volumi da collegare alle istanze avviate da questo modello di avvio, scegli **Aggiungi nuovo volume**.

Creare un modello di avvio da un'istanza esistente (console)

Come creare un modello di avvio da un'istanza esistente

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nel pannello di navigazione, in **Instaces (Istanze)**, scegli **Instances (Istanze)**.
3. Seleziona l'istanza e scegli **Actions (Operazioni)**, **Image and templates (Immagini e modelli)**, **Create template from instance (Crea modello dall'istanza)**.
4. Inserisci un nome e una descrizione.
5. In **Auto Scaling guidance (Guida Auto Scaling)**, seleziona la casella di controllo.
6. Regola i parametri come necessario e scegli **Create launch template (Crea modello di avvio)**.
7. Per creare un gruppo con scalabilità automatica, scegli **Create an Auto Scaling group (Crea un gruppo con scalabilità automatica)** dalla pagina di conferma.

Risorse correlate

Forniamo alcuni frammenti di template JSON e YAML che puoi usare per capire come dichiarare i modelli di lancio nei tuoi modelli di stack. AWS CloudFormation Per ulteriori informazioni, consulta le sezioni [AWS::EC2::LaunchTemplate](#) e [Crea modelli di lancio](#) con le sezioni della Guida per l'utente. AWS CloudFormation AWS CloudFormation

Per maggiori informazioni sui modelli di avvio, consulta [Avvio di un'istanza da un modello di avvio](#) nella Guida per l'utente per istanze Linux di Amazon EC2.

Limitazioni

- Sebbene sia possibile specificare una sottorete in un modello di avvio, non è necessario farlo se si utilizza il modello di avvio solo per creare gruppi con dimensionamento automatico. Non è possibile specificare la sottorete per un gruppo con dimensionamento automatico specificando la sottorete in un modello di avvio. Le sottoreti per il gruppo con dimensionamento automatico sono ricavate dalla definizione di risorse propria del gruppo.
- Per altre limitazioni sulle interfacce di rete definite dall'utente, vedi [Modifica le impostazioni dell'interfaccia di rete di default \(console\)](#).

Crea un modello di avvio utilizzando le impostazioni avanzate

Questo argomento descrive come creare un modello di avvio con impostazioni avanzate da AWS Management Console.

Per creare un modello di lancio utilizzando impostazioni avanzate

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nel riquadro di navigazione, in Istanze, scegli Launch Templates, quindi scegli Crea modello di lancio.
3. Configura il modello di lancio come descritto nei seguenti argomenti:
 - [Impostazioni richieste](#)
 - [Impostazioni avanzate](#)
4. Scegli Crea modello di avvio.

Impostazioni richieste

Quando crei un modello di lancio, devi includere le seguenti impostazioni obbligatorie.

Nome del modello di avvio

Inserisci un nome univoco che descriva il modello di lancio.

Immagini di applicazioni e sistema operativo (Amazon Machine Image)

Scegli l'Amazon Machine Image (AMI) che desideri utilizzare. Puoi cercare o cercare l'AMI che desideri utilizzare. Per una migliore efficienza di scalabilità, scegli un'AMI personalizzata completamente configurata per avviare un'istanza con il codice dell'applicazione e che richieda poche modifiche all'avvio.

Tipo di istanza

Scegli un tipo di istanza compatibile con la tua AMI. Puoi saltare l'aggiunta di un tipo di istanza al tuo modello di lancio se prevedi di utilizzare più tipi di istanze incorporati nella definizione delle risorse del gruppo Auto Scaling. Un tipo di istanza è necessario solo se non intendi creare un gruppo di istanze [miste](#).

Impostazioni avanzate

Le impostazioni avanzate sono opzionali. Se non configuri alcuna impostazione avanzata, le funzionalità specifiche non verranno aggiunte alle istanze.

Espandi la sezione Dettagli avanzati per visualizzare le impostazioni avanzate. Le sezioni seguenti descrivono le impostazioni avanzate più utili su cui concentrarsi quando si crea un modello di avvio per un gruppo di Auto Scaling. Per ulteriori informazioni, consulta [i dettagli avanzati](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Profilo dell'istanza IAM

Il profilo dell'istanza contiene il ruolo IAM che desideri utilizzare. Quando il gruppo Auto Scaling avvia un'istanza EC2, le autorizzazioni definite nel ruolo IAM associato vengono concesse alle applicazioni in esecuzione sull'istanza. Per ulteriori informazioni, consulta [Ruoli IAM per le applicazioni in esecuzione sulle istanze Amazon EC2](#).

Termination protection (Protezione da cessazione)

Se abilitata, questa funzionalità impedisce agli utenti di terminare un'istanza utilizzando la console Amazon EC2, i comandi CLI e le operazioni API. La protezione dalla terminazione fornisce un'ulteriore protezione contro la terminazione accidentale. Non impedisce ad Amazon EC2 Auto Scaling di terminare un'istanza. Per controllare quali istanze Amazon EC2 Auto Scaling può terminare, consulta [Utilizzare la protezione di riduzione orizzontale dell'istanza](#)

Monitoraggio dettagliato CloudWatch

Puoi abilitare il monitoraggio dettagliato delle tue istanze EC2 per consentire loro di inviare dati metrici ad Amazon CloudWatch a intervalli di 1 minuto. Per impostazione predefinita, le istanze EC2 inviano dati metrici a intervalli di 5 minuti. CloudWatch Vengono applicati costi aggiuntivi. Per ulteriori informazioni, consulta [Configurare il monitoraggio per le istanze Auto Scaling](#).

Specifiche del credito

Amazon EC2 offre istanze con prestazioni stabili, come T2, T3 e T3a, che consentono alle applicazioni di superare le prestazioni della CPU di base quando necessario. Per impostazione predefinita, queste istanze possono interrompersi per un periodo di tempo limitato prima che l'utilizzo della CPU venga limitato. Facoltativamente, puoi abilitare la modalità illimitata in modo che le istanze possano superare la linea di base per tutto il tempo necessario. Ciò consente alle applicazioni di mantenere elevate prestazioni della CPU quando necessario. Potrebbero essere applicati costi aggiuntivi. Per ulteriori informazioni, consulta [Usare un gruppo Auto Scaling per](#)

[avviare un'istanza dalle prestazioni espandibili come Unlimited](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Nome del gruppo di collocamento

Puoi specificare un gruppo di collocamento e utilizzare un cluster o una strategia di partizione per influenzare la posizione fisica delle istanze nel data center. AWS Per piccoli gruppi di Auto Scaling, puoi anche utilizzare la strategia di spread. Per ulteriori informazioni, consulta [Gruppi di collocamento](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Ci sono alcune considerazioni sull'utilizzo dei gruppi di posizionamento con i gruppi di Auto Scaling:

- Se viene specificato un gruppo di posizionamento sia nel modello di lancio che nel gruppo Auto Scaling, il gruppo di posizionamento per il gruppo Auto Scaling ha la precedenza. Dopo la creazione del gruppo, il gruppo di posizionamento specificato nelle impostazioni del gruppo Auto Scaling non può essere modificato.
- In AWS CloudFormation, fate attenzione se definite un gruppo di posizionamento nel modello di lancio. Amazon EC2 Auto Scaling avvierà le istanze nel gruppo di posizionamento specificato. Tuttavia, non CloudFormation riceverà segnali da quelle istanze se ne utilizzi uno [UpdatePolicy](#) con il tuo gruppo Auto Scaling (anche se questo potrebbe cambiare in futuro).

Opzione di acquisto

Puoi scegliere Richiedi istanze Spot per richiedere istanze Spot al prezzo Spot, limitato al prezzo on demand, e scegliere Personalizza per modificare le impostazioni predefinite delle istanze Spot. Per un gruppo con dimensionamento automatico, devi specificare una richiesta una tantum, senza data di fine (impostazione di default). Per ulteriori informazioni, consulta [Richiedi istanze spot per applicazioni flessibili e con tolleranza ai guasti](#). Questa impostazione può essere utile in circostanze particolari, ma in generale è meglio non specificarla e creare invece un gruppo di istanze miste. Per ulteriori informazioni, consulta [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#).

Se specifichi una richiesta di istanza spot nel modello di avvio, non puoi creare un gruppo di istanze miste. Se tenti di utilizzare un modello di avvio che richiede istanze spot con un gruppo di istanze miste, ricevi il seguente messaggio di errore: Incompatible launch template: You cannot use a launch template that is set to request Spot Instances (InstanceMarketOptions) when you configure an Auto Scaling group with a mixed instances policy. Add a different launch template to the group and try again.

Capacity Reservation

Le prenotazioni di capacità consentono di riservare la capacità per le istanze Amazon EC2 in una zona di disponibilità specifica per qualsiasi durata. Per ulteriori informazioni, consulta [Prenotazione di capacità on demand](#) nella Guida per l'utente per istanze Linux di Amazon EC2.

Puoi scegliere se avviare le istanze in:

- qualsiasi prenotazione di capacità aperta (aperta)
- una prenotazione di capacità specifica (Target by ID)
- un gruppo di prenotazioni di capacità (target per gruppo)

Per scegliere come target una prenotazione di capacità specifica, il tipo di istanza nel modello di lancio deve corrispondere al tipo di istanza della prenotazione. Quando crei il tuo gruppo Auto Scaling, utilizza la stessa zona di disponibilità della prenotazione della capacità. A seconda della scelta, Regione AWS puoi scegliere invece di scegliere come target un Capacity Block. Per ulteriori informazioni, consulta [Usa Capacity Blocks per carichi di lavoro di machine learning](#).

Per scegliere come target un gruppo di prenotazioni di capacità, consulta [Utilizza le prenotazioni della capacità on demand per prenotare la capacità in zone di disponibilità specifiche](#). Scegliendo come target un gruppo di prenotazioni di capacità, è possibile distribuire la capacità su più zone di disponibilità per migliorare la resilienza.

Tenancy

Amazon EC2 offre tre opzioni per la locazione delle istanze EC2:

- **Condiviso (condiviso):** più dispositivi Account AWS possono condividere lo stesso hardware fisico. Questa è l'opzione di locazione predefinita all'avvio di un'istanza.
- **Istanze dedicate (dedicate):** l'istanza viene eseguita su hardware single-tenant. Nessun altro AWS cliente condivide lo stesso server fisico. Per ulteriori informazioni, consulta [Istanze dedicate](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.
- **Host dedicati (host dedicato):** l'istanza viene eseguita su un server fisico dedicato al tuo utilizzo. L'utilizzo di host dedicati semplifica l'importazione di licenze personalizzate (BYOL) con requisiti hardware dedicati a EC2 e soddisfano i casi d'uso di conformità. Se scegli questa opzione, devi fornire un gruppo di risorse host per il gruppo di risorse host Tenancy. Per ulteriori informazioni, consulta [Host dedicati](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Il supporto per host dedicati è disponibile solo se si specifica un gruppo di risorse host. Non è possibile puntare a un ID host specifico o usare l'affinità di posizionamento dell'host.

- Se tenti di utilizzare un modello di avvio che specifica un ID host, ricevi il seguente messaggio di errore: `Incompatible launch template: Tenancy host ID is not supported for Auto Scaling`.
- Se tenti di utilizzare un modello di lancio che specifica l'affinità di posizionamento dell'host, ricevi il seguente messaggio di errore: `Incompatible launch template: Auto Scaling does not support host placement affinity`.

Tenancy Host Resource Group

Con AWS License Manager, puoi importare le tue licenze AWS e gestirle centralmente. Un gruppo di risorse host è un gruppo di host dedicati collegati a una configurazione di licenza specifica del License Manager. I gruppi di risorse host consentono di avviare facilmente le istanze EC2 su host dedicati che soddisfano le esigenze di licenza software. Non è necessario allocare manualmente gli host dedicati in anticipo. Vengono creati automaticamente in base alle esigenze. Si noti che quando si associa un'AMI a una configurazione di licenza, tale AMI può essere associata solo a un gruppo di risorse host alla volta. Per ulteriori informazioni, consulta [Host resource groups AWS License Manager nella License Manager User Guide](#).

Configurazioni di licenza

Con questa impostazione, puoi specificare una configurazione di licenza per le tue istanze senza limitarne la locazione agli host dedicati. La configurazione delle licenze tiene traccia delle licenze software distribuite sulle istanze in modo da poter monitorare l'utilizzo e la conformità delle licenze. Per ulteriori informazioni, consulta [Creare una licenza autogestita](#) nella Guida per l'utente di License Manager.

Metadati accessibili

Puoi scegliere se abilitare o disabilitare l'accesso all'endpoint HTTP del servizio di metadati dell'istanza. Per impostazione predefinita, l'endpoint HTTP è abilitato. Se scegli di disabilitare l'endpoint, l'accesso ai metadati dell'istanza viene disattivato. È possibile specificare la condizione per richiedere IMDSv2 solo quando l'endpoint HTTP è abilitato. Per ulteriori informazioni, consultare [Configurazione delle opzioni di metadati dell'istanza](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Versione dei metadati

Puoi scegliere di richiedere l'uso di Instance Metadata Service Version 2 (IMDSv2) quando richiedi i metadati dell'istanza. Se non specifichi un valore, l'impostazione di default è quella di supportare sia IMDSv1 che IMDSv2. Per ulteriori informazioni, consultare [Configurazione delle opzioni di metadati dell'istanza](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Limite dell'hop di risposta del token di metadati

È possibile impostare il numero consentito di hop di rete per il token di metadati. Se non specifichi un valore, il valore di default è 1. Per ulteriori informazioni, consultare [Configurazione delle opzioni di metadati dell'istanza](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Dati utente

Puoi personalizzare e completare la configurazione delle istanze al momento dell'avvio specificando script di shell o direttive cloud-init come dati utente. I dati utente vengono eseguiti all'avvio iniziale dell'istanza, consentendoti di installare automaticamente applicazioni, dipendenze o personalizzazioni al momento dell'avvio. Per maggiori informazioni, consulta [Esecuzione di comandi sull'istanza Linux all'avvio](#) nella Guida dell'utente di Amazon EC2 per le istanze Linux.

Se hai download di grandi dimensioni o script complessi, ciò aumenta il tempo necessario affinché l'istanza sia pronta per l'uso. In tal caso, potrebbe essere necessario configurare un lifecycle hook per ritardare il raggiungimento InService dello stato di un'istanza fino al completamento del provisioning. Per ulteriori informazioni sull'aggiunta di un lifecycle hook al gruppo Auto Scaling, consulta [Hook del ciclo di vita di Amazon EC2 Auto Scaling](#)

Richiedi istanze spot per applicazioni flessibili e con tolleranza ai guasti

Nel modello di avvio, puoi facoltativamente richiedere istanze Spot senza data di fine o durata. Le istanze spot Amazon EC2 rappresentano una capacità di riserva disponibile con forti sconti, rispetto al prezzo di EC2 on demand. Le istanze spot sono una scelta economicamente vantaggiosa se si può essere flessibili riguardo all'orario di esecuzione delle applicazioni e se queste possono essere interrotte. Per ulteriori informazioni sulla creazione di un modello di avvio che richiede istanze spot, consulta [Crea un modello di avvio utilizzando le impostazioni avanzate](#).

Important


Le istanze spot vengono tipicamente utilizzate per integrare le istanze on demand. È possibile specificare le stesse impostazioni utilizzate per avviare le istanze spot nell'ambito delle impostazioni del gruppo con dimensionamento automatico. Quando specifichi le impostazioni nell'ambito del gruppo con dimensionamento automatico, puoi richiedere di avviare Istanze spot solo dopo aver avviato un certo numero di istanze on demand e quindi continuare ad avviare alcune combinazioni di istanze on demand e spot durante il dimensionamento del

gruppo. Per ulteriori informazioni, consulta [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#).

Questo argomento descrive come avviare solo le istanze spot nel gruppo con dimensionamento automatico specificando le impostazioni in un modello di avvio piuttosto che nel gruppo con dimensionamento automatico stesso. Le informazioni contenute in questo argomento si applicano anche ai gruppi con dimensionamento automatico che richiedono istanze spot con una [configurazione di avvio](#). La differenza è che una configurazione di avvio richiede un prezzo massimo, ma per i modelli di avvio il prezzo massimo è facoltativo.

Quando crei un modello di avvio per avviare solo istanze Spot, tieni a mente le seguenti considerazioni:

- prezzo istanza Spot. Si paga solo il prezzo per l'istanza Spot attuale per le istanze Spot che vengono avviate. Questi prezzi cambiano lentamente nel tempo in base ai trend a lungo termine dell'offerta e della domanda. Per ulteriori informazioni, consulta [Istanze Spot](#) e [Prezzi e risparmio](#) nella Guida per l'utente Amazon EC2 per le istanze Linux.
- Impostazione del prezzo massimo. Facoltativamente, puoi includere un prezzo massimo all'ora per le istanze spot nel modello di avvio. Se il prezzo massimo supera il prezzo dell'istanza spot attuale, il servizio spot di Amazon EC2 soddisfa immediatamente la richiesta se la capacità è disponibile. Se il prezzo delle istanze spot supera il prezzo massimo per un'istanza in esecuzione nel gruppo con dimensionamento automatico, l'istanza viene cessata.

 Warning

L'applicazione potrebbe non essere eseguita se non si riceve alcuna istanza spot, ad esempio nel caso in cui il prezzo massimo sia troppo basso. Per sfruttare le istanze spot disponibili il più a lungo possibile, imposta il prezzo massimo vicino al prezzo on demand.

- Bilanciamento tra le zone di disponibilità. Se specifichi più zone di disponibilità, Amazon EC2 con dimensionamento automatico distribuisce le richieste spot tra tutte queste zone. Se in una zona di disponibilità il prezzo massimo è troppo basso per poter soddisfare le richieste, Dimensionamento automatico Amazon EC2 verifica se le richieste siano state soddisfatte in altre zone. Se è così, Dimensionamento automatico Amazon EC2 annulla le richieste non riuscite e le ridistribuisce sulle zone di disponibilità che soddisfano le richieste. Se il prezzo in una zona di disponibilità senza richieste completate scende abbastanza da soddisfare le richieste future, Dimensionamento automatico Amazon EC2 ribilancia tutte le zone di disponibilità.

- Cessazione di istanze spot. Le istanze spot possono essere cessate in qualsiasi momento. Il servizio spot di Amazon EC2 può cessare le istanze spot nel tuo gruppo con dimensionamento automatico se la disponibilità o il prezzo delle istanze spot varia. Quando si dimensionano o si eseguono controlli dell'integrità, Dimensionamento automatico Amazon EC2 può anche cessare le istanze spot proprio come può cessare le istanze on demand. Quando un'istanza viene terminata, qualsiasi spazio di archiviazione viene eliminato.
- Mantenere la capacità desiderata. Quando un'istanza Spot viene terminata, Amazon EC2 Auto Scaling tenta di avviare un'istanza Spot per mantenere la capacità desiderata per il gruppo. Se il prezzo attuale per l'istanza Spot è inferiore al prezzo massimo, viene avviata un'istanza Spot. Se la richiesta per un'istanza Spot non ha esito positivo, continua a provare.
- Modifica del prezzo massimo. Per modificare il prezzo massimo, crea un nuovo modello di avvio o aggiorna un modello di avvio esistente con il nuovo prezzo massimo, quindi associalo al gruppo con dimensionamento automatico. Le istanze Spot esistenti rimangono in esecuzione finché il prezzo massimo specificato nel modello di avvio utilizzati per tali istanze è superiore al prezzo per l'istanza Spot corrente. Se non è stato impostato un prezzo massimo, il prezzo massimo predefinito è il prezzo on demand.

Usa Capacity Blocks per carichi di lavoro di machine learning

I Capacity Blocks ti aiutano a prenotare istanze GPU molto richieste in date future per supportare i tuoi carichi di lavoro di machine learning (ML) di breve durata.

Per una panoramica dei Capacity Blocks e del loro funzionamento, consulta [Capacity Blocks for ML](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Puoi utilizzare Capacity Blocks con i seguenti tipi di istanze EC2 e: Regioni AWS

Tipi di istanza	Regioni
p5.48xlarge	Stati Uniti orientali (Ohio), Stati Uniti orientali (Virginia settentrionale)
p4d.24xlarge	Stati Uniti orientali (Ohio), Stati Uniti occidentali (Oregon)

Per iniziare a utilizzare Capacity Blocks, devi creare una prenotazione di capacità in una zona di disponibilità specifica. I blocchi di capacità vengono forniti come prenotazioni `targeted` di capacità

in un'unica zona di disponibilità. Quando crei il modello di lancio, specifica l'ID di prenotazione e il tipo di istanza del Capacity Block. Quindi, aggiorna il tuo gruppo Auto Scaling per utilizzare il modello di avvio che hai creato e la zona di disponibilità del Capacity Block. Quando inizia la prenotazione Capacity Block, utilizza la scalabilità pianificata per avviare lo stesso numero di istanze della prenotazione Capacity Block.

Indice

- [Linee guida operative](#)
- [Specifica un blocco di capacità nel modello di avvio](#)
- [Limitazioni](#)
- [Risorse correlate](#)

Linee guida operative

Di seguito sono illustrate le linee guida operative di base e le istanze che dovresti seguire durante l'utilizzo di un blocco delle capacità con un gruppo con dimensionamento automatico.

- Ridimensiona il tuo gruppo con dimensionamento automatico fino a zero oltre 30 minuti prima dell'orario di fine della prenotazione del blocco delle capacità. Amazon EC2 interromperà tutte le istanze ancora in esecuzione 30 minuti prima della fine del blocco delle capacità.
- Ti consigliamo di utilizzare la scalabilità pianificata per la scalabilità orizzontale (aggiungere istanze) e la scalabilità orizzontale (rimuovere le istanze) negli orari di prenotazione appropriati. Per ulteriori informazioni, consulta [Dimensionamento programmato per Dimensionamento automatico Amazon EC2](#).
- Se necessario, aggiungi gli hook del ciclo di vita per eseguire un arresto graduale dell'applicazione all'interno delle istanze durante il ridimensionamento. Lascia il tempo necessario per completare l'azione del ciclo di vita prima che Amazon EC2 inizi a terminare forzatamente le istanze 30 minuti prima dell'orario di fine della prenotazione del blocco delle capacità. Per ulteriori informazioni, consulta [Hook del ciclo di vita di Amazon EC2 Auto Scaling](#).
- Assicurati che il gruppo con dimensionamento automatico punti alla versione corretta del modello di avvio per l'intera durata della prenotazione. Ti consigliamo di puntare a una versione specifica del modello di avvio anziché alla versione `$Default` o `$Latest`.

Note

Se lasci un'istanza Capacity Block in esecuzione fino alla fine della prenotazione e Amazon EC2 la recupera, le attività di scalabilità per il tuo gruppo Auto Scaling dichiarano che si trattava di "taken out of service in response to an EC2 health check that indicated it had been terminated or stopped", anche se è stata recuperata intenzionalmente alla fine del Capacity Block. Allo stesso modo, Amazon EC2 Auto Scaling tenterà di sostituire l'istanza nello stesso modo in cui fa per qualsiasi istanza che non superi un controllo di integrità. Per ulteriori informazioni, consulta [Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico](#).

Specifica un blocco di capacità nel modello di avvio

Per creare un modello di lancio destinato a un blocco di capacità specifico per il gruppo Auto Scaling, utilizzate uno dei seguenti metodi:

Console

Per specificare un blocco di capacità nel proprio modello di avvio (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nella barra di navigazione in alto, seleziona il Regione AWS punto in cui hai creato il tuo Capacity Block.
3. Nel pannello di navigazione, in Istanze, scegli Modelli di avvio.
4. Scegli Crea modello di lancio e crea il modello di lancio. Se necessario, includi l'ID dell'Amazon Machine Image (AMI), il tipo di istanza e altre impostazioni del modello di avvio.
5. Espandi la sezione Dettagli avanzati per visualizzare le impostazioni avanzate.
6. Per l'opzione di acquisto, scegli blocchi di capacità.
7. Per Prenotazione di capacità, scegli Target per ID, quindi per Prenotazione della capacità - Target per ID, scegli l'ID di prenotazione della capacità di un blocco di capacità esistente.
8. Al termine, scegliere Crea modello di avvio.

AWS CLI

Per specificare un blocco di capacità nel modello di avvio (AWS CLI)

Utilizza il seguente [create-launch-template](#) comando per creare un modello di lancio che specifichi un ID di prenotazione Capacity Block esistente. Sostituisci ciascun *placeholder input dell'utente* con le tue informazioni.

```
aws ec2 create-launch-template --launch-template-name my-template-for-capacity-block \
  --version-description AutoScalingVersion1 --region us-east-2 \
  --launch-template-data file://config.json
```

Tip

Se questo comando genera un errore, assicurati di averlo aggiornato AWS CLI localmente alla versione più recente.

Contenuto di config.json.

```
{
  "ImageId": "ami-04d5cc9b88example",
  "InstanceType": "p4d.24xlarge",
  "SecurityGroupIds": [
    "sg-903004f88example"
  ],
  "KeyName": "MyKeyPair",
  "InstanceMarketOptions": {
    "MarketType": "capacity-block"
  },
  "CapacityReservationSpecification": {
    "CapacityReservationTarget": {
      "CapacityReservationId": "cr-02168da1478b509e0"
    }
  }
}
```

Di seguito è riportato un output di esempio.

```
{
  "LaunchTemplate": {
    "LaunchTemplateId": "lt-068f72b724example",
    "LaunchTemplateName": "my-template-for-capacity-block",
    "CreateTime": "2023-10-27T15:12:44.000Z",
```

```

    "CreatedBy": "arn:aws:iam::123456789012:user/Bob",
    "DefaultVersionNumber": 1,
    "LatestVersionNumber": 1
  }
}

```

È possibile utilizzare il [describe-launch-template-versions](#) comando seguente per verificare l'ID di prenotazione Capacity Block associato al modello di lancio.

```

aws ec2 describe-launch-template-versions --launch-template-names my-template-for-capacity-block \
--region us-east-2

```

Di seguito è riportato l'output di esempio per un modello di avvio che specifica una prenotazione di blocco di capacità.

```

{
  "LaunchTemplateVersions": [
    {
      "LaunchTemplateId": "lt-068f72b724example",
      "LaunchTemplateName": "my-template-for-capacity-block",
      "VersionNumber": 1,
      "CreateTime": "2023-10-27T15:12:44.000Z",
      "CreatedBy": "arn:aws:iam::123456789012:user/Bob",
      "DefaultVersion": true,
      "LaunchTemplateData": {
        "ImageId": "ami-04d5cc9b88example",
        "InstanceType": "p5.48xlarge",
        "SecurityGroupIds": [
          "sg-903004f88example"
        ],
        "KeyName": "MyKeyPair",
        "InstanceMarketOptions": {
          "MarketType": "capacity-block"
        },
        "CapacityReservationSpecification": {
          "CapacityReservationTarget": {
            "CapacityReservationId": "cr-02168da1478b509e0"
          }
        }
      }
    }
  ]
}

```

}

Limitazioni

- Il supporto per blocchi di capacità è disponibile solo se il gruppo con dimensionamento automatico ha una configurazione compatibile. I gruppi di istanze miste e i pool caldi non sono supportati.
- Puoi scegliere come target un solo Capacity Block alla volta.

Risorse correlate

- Per i prerequisiti e i consigli per l'utilizzo delle istanze P5, consulta la Guida [introduttiva alle istanze P5 nella Guida per l'utente di Amazon EC2 per le istanze Linux](#).
- Amazon EKS supporta l'utilizzo di Capacity Blocks per supportare carichi di lavoro di apprendimento automatico (ML) di breve durata su cluster Amazon EKS. Per ulteriori informazioni, consulta [Capacity Blocks for ML](#) nella Guida per l'utente di Amazon EKS.
- Puoi utilizzare Capacity Blocks con tipi di istanze e regioni supportati. Tuttavia, le prenotazioni di capacità su richiesta offrono la flessibilità necessaria per riservare la capacità per altri tipi di istanze e regioni. Per un tutorial che mostra come utilizzare l'opzione di prenotazione della capacità su richiesta, consulta [Utilizza le prenotazioni della capacità on demand per prenotare la capacità in zone di disponibilità specifiche](#).

Migra i tuoi gruppi di Auto Scaling ai modelli di avvio

A partire dal 2023, non potrai effettuare chiamate `CreateLaunchConfiguration` con nuovi tipi di istanze Amazon EC2 rilasciati dopo il 31 dicembre 2022. Per ulteriori informazioni, consulta [Configurazioni di avvio](#).

Per migrare i gruppi di Auto Scaling dalle configurazioni di avvio ai modelli di avvio, segui i passaggi seguenti.

Important

Prima di continuare, verifica di disporre delle autorizzazioni necessarie per l'utilizzo dei modelli di avvio. Per ulteriori informazioni, consulta [Autorizzazioni per lavorare con i modelli di lancio](#).

Passo 1: ricerca dei gruppi con dimensionamento automatico che utilizzano configurazioni di avvio

Per identificare se ci sono gruppi di Auto Scaling che utilizzano ancora le configurazioni di avvio, esegui il [describe-auto-scaling-groups](#) comando seguente utilizzando. AWS CLI Sostituisci **REGION** con il tuo. Regione AWS

```
aws autoscaling describe-auto-scaling-groups --region REGION \  
--query 'AutoScalingGroups[?LaunchConfigurationName!=`null`]'
```

Di seguito è riportato un output di esempio.

```
[  
  {  
    "AutoScalingGroupName": "group-1",  
    "AutoScalingGroupARN": "arn",  
    "LaunchConfigurationName": "my-launch-config",  
    "MinSize": 1,  
    "MaxSize": 5,  
    "DesiredCapacity": 2,  
    "DefaultCooldown": 300,  
    "AvailabilityZones": [  
      "us-west-2a",  
      "us-west-2b",  
      "us-west-2c"  
    ],  
    "LoadBalancerNames": [],  
    "TargetGroupARNs": [],  
    "HealthCheckType": "EC2",  
    "HealthCheckGracePeriod": 300,  
    "Instances": [  
      {  
        "ProtectedFromScaleIn": false,  
        "AvailabilityZone": "us-west-2a",  
        "LaunchConfigurationName": "my-launch-config",  
        "InstanceId": "i-05b4f7d5be44822a6",  
        "InstanceType": "t3.micro",  
        "HealthStatus": "Healthy",  
        "LifecycleState": "InService"  
      },  
      {  
        "ProtectedFromScaleIn": false,
```

```

        "AvailabilityZone": "us-west-2b",
        "LaunchConfigurationName": "my-launch-config",
        "InstanceId": "i-0c20ac468fa3049e8",
        "InstanceType": "t3.micro",
        "HealthStatus": "Healthy",
        "LifecycleState": "InService"
    }
],
"CreatedTime": "2023-03-09T22:15:11.611Z",
"SuspendedProcesses": [],
"VPCZoneIdentifier": "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782",
"EnabledMetrics": [],
"Tags": [
    {
        "ResourceId": "group-1",
        "ResourceType": "auto-scaling-group",
        "Key": "environment",
        "Value": "production",
        "PropagateAtLaunch": true
    }
],
"TerminationPolicies": [
    "Default"
],
"NewInstancesProtectedFromScaleIn": false,
"ServiceLinkedRoleARN": "arn",
    "TrafficSources": []
},
... additional groups ...
]

```

In alternativa, per rimuovere tutto tranne i nomi dei gruppi con dimensionamento automatico con i nomi delle rispettive configurazioni di avvio e i tag nell'output, esegui il seguente comando:

```

aws autoscaling describe-auto-scaling-groups --region REGION \
  --query 'AutoScalingGroups[?LaunchConfigurationName!=`null`].{AutoScalingGroupName:
  AutoScalingGroupName, LaunchConfigurationName: LaunchConfigurationName, Tags: Tags}'

```

Di seguito viene mostrato l'output di esempio.

```
[
```

```
{
  "AutoScalingGroupName": "group-1",
  "LaunchConfigurationName": "my-launch-config",
  "Tags": [
    {
      "ResourceId": "group-1",
      "ResourceType": "auto-scaling-group",
      "Key": "environment",
      "Value": "production",
      "PropagateAtLaunch": true
    }
  ]
},
... additional groups ...
]
```

Per ulteriori informazioni sul filtraggio, vedere [Filtraggio dell' AWS CLI output nella Guida](#) per l'AWS Command Line Interface utente.

Passo 2: copia di una configurazione di avvio in un modello di avvio

È possibile copiare una configurazione di avvio in un modello di avvio utilizzando la procedura seguente. Quindi, puoi aggiungerlo al gruppo con dimensionamento automatico.

La copia di più configurazioni di avvio produce modelli di avvio con nomi identici. Per modificare il nome assegnato a un modello di avvio durante il processo di copia, è necessario copiare le configurazioni di avvio una per una.

Note

La funzionalità di copia è disponibile solo dalla console.

Come copiare una configurazione di avvio in un modello di avvio (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nel pannello sinistro di navigazione in Dimensionamento automatico, scegliere Gruppi con dimensionamento automatico.

3. Nella parte superiore della pagina, scegli Configurazioni di avvio. Quando viene richiesta la conferma, scegli Visualizza configurazioni di avvio per confermare che desideri visualizzare la pagina delle configurazioni di avvio.
4. Seleziona la configurazione di avvio che desideri copiare e scegli Copia in modello di avvio, Copia selezione. Questo imposta un nuovo modello di avvio con lo stesso nome e le stesse opzioni della configurazione di avvio selezionata.
5. In Nome del nuovo modello di avvio è possibile utilizzare il nome della configurazione di avvio (impostazione di default) oppure inserire un nuovo nome. I nomi dei modelli di avvio devono essere univoci.
6. (Facoltativo) Seleziona Crea un gruppo con dimensionamento automatico utilizzando il nuovo modello.

Se desideri completare la copia della configurazione di avvio, ignora questo passaggio. Non è necessario creare un nuovo gruppo con dimensionamento automatico.

7. Scegli Copia.

Come copiare tutte le configurazioni di avvio nei modelli di avvio (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nel riquadro di navigazione, in Dimensionamento automatico, scegliere Configurazioni di avvio.
3. Scegli Copia nel modello di avvio, Copia tutto. In questo modo vengono copiate tutte le configurazioni di avvio nell'area corrente in un nuovo modello di avvio con lo stesso nome e le stesse opzioni.
4. Scegli Copia.

Passo 3: aggiornare un gruppo con dimensionamento automatico per utilizzare un modello di avvio

Dopo aver creato un modello di avvio, sei pronto per aggiungerlo al gruppo con dimensionamento automatico.

Per aggiornare un gruppo con dimensionamento automatico per utilizzare un modello di avvio (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.

2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si apre un riquadro diviso nella parte inferiore della pagina, con le informazioni sul gruppo selezionato.

3. Nella scheda Details (Dettagli) scegli Launch configurations (Configurazioni di avvio), Edit (Modifica).
4. Scegli Switch to launch template (Passa al modello di avvio).
5. In Launch template (Modello di avvio), seleziona il modello di avvio.
6. In Version (Versione), seleziona la versione del modello di avvio in base alle tue esigenze. Dopo la creazione delle versioni di un modello di avvio, è possibile scegliere se il gruppo con dimensionamento automatico usa la versione di default o quella più recente del modello di avvio, per la scalabilità orizzontale.
7. Scegli Aggiorna.

Aggiornare un gruppo con dimensionamento automatico per utilizzare un modello di avvio (AWS CLI)

Il [update-auto-scaling-group](#) comando seguente aggiorna il gruppo Auto Scaling specificato per utilizzare la versione iniziale del modello di avvio specificato.

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg \  
--launch-template LaunchTemplateName=my-template-for-auto-scaling,Version='1'
```

Per vedere degli esempi di utilizzo di un comando CLI per aggiornare un gruppo con dimensionamento automatico e usare un modello di avvio, consulta [Aggiornare un gruppo con dimensionamento automatico per utilizzare un modello di avvio](#).

Passo 4: sostituzione delle istanze

Dopo aver sostituito la configurazione di avvio con un modello di avvio, tutte le nuove istanze utilizzeranno il nuovo modello di avvio. Le istanze esistenti non sono interessate.

Per aggiornare le istanze esistenti è possibile avviare un aggiornamento dell'istanza per sostituire le istanze nel gruppo con dimensionamento automatico, anziché sostituirne manualmente un certo numero alla volta. Per ulteriori informazioni, consulta [Usa un aggiornamento dell'istanza per aggiornare le istanze in un gruppo di Auto Scaling](#). Se il gruppo è numeroso, un aggiornamento dell'istanza può essere particolarmente utile.

In alternativa, puoi consentire il dimensionamento automatico per sostituire gradualmente le istanze esistenti con nuove istanze in base alle [policy di terminazione](#) del gruppo, oppure puoi interromperle. La terminazione manuale impone al gruppo con dimensionamento automatico di avviare nuove istanze per mantenere la capacità desiderata del gruppo. Per ulteriori informazioni, consulta [Cessazione di un'istanza](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Informazioni aggiuntive

Per ulteriori informazioni, consulta [Amazon EC2 Auto Scaling non aggiungerà più il supporto per le nuove funzionalità EC2 a Launch Configurations on the Compute Blog](#). AWS

Per un argomento che illustra come migrare gli AWS CloudFormation stack dalle configurazioni di lancio ai modelli di lancio, consulta [Migra AWS CloudFormation gli stack ai modelli di lancio](#)

Migra AWS CloudFormation gli stack ai modelli di lancio

Puoi migrare i tuoi modelli di AWS CloudFormation stack esistenti dalle configurazioni di avvio ai modelli di lancio. A tale scopo, aggiungi un modello di avvio direttamente a un modello di stack esistente, quindi associa il modello di avvio al gruppo con dimensionamento automatico nel modello di stack. Successivamente, sarà necessario utilizzare il tuo modello modificato per aggiornare lo stack.

Durante la migrazione ai modelli di lancio, questo argomento consente di risparmiare tempo fornendo istruzioni per riscrivere le configurazioni di lancio nei modelli di stack come modelli di lancio. CloudFormation Per ulteriori informazioni sulla migrazione delle configurazioni di avvio per lanciare i modelli, consulta [Migra i tuoi gruppi di Auto Scaling ai modelli di avvio](#).

Argomenti

- [Trova gruppi con dimensionamento automatico che utilizzano una configurazione di avvio](#)
- [Aggiornare uno stack per utilizzare un modello di avvio](#)
- [Comprendi l'aggiornamento dei comportamenti delle risorse stack](#)
- [Monitora la migrazione](#)
- [Riferimento della mappatura di avvio](#)

Trova gruppi con dimensionamento automatico che utilizzano una configurazione di avvio

Trova gruppi con dimensionamento automatico che utilizzano una configurazione di avvio

- Utilizzate il seguente [describe-auto-scaling-groups](#) comando per elencare i nomi dei gruppi di Auto Scaling che utilizzano configurazioni di avvio nella regione specificata. Includi l' `--filters` opzione per restringere i risultati ai gruppi associati a uno CloudFormation stack (filtrando in base alla chiave del `aws:cloudformation:stack-name` tag).

```
aws autoscaling describe-auto-scaling-groups --region REGION \  
  --filters Name=tag-key,Values=aws:cloudformation:stack-name \  
  --query 'AutoScalingGroups[?LaunchConfigurationName!  
= `null` ].AutoScalingGroupName'
```

Di seguito viene mostrato l'output di esempio.

```
[  
  "{stack-name}-group-1",  
  "{stack-name}-group-2",  
  "{stack-name}-group-3"  
]
```

Sono disponibili altri AWS CLI comandi utili per trovare i gruppi di Auto Scaling in cui migrare e filtrare l'output. [Migra i tuoi gruppi di Auto Scaling ai modelli di avvio](#)

Important

Se le risorse dello stack hanno un AWSEB nome, significa che sono state create tramite AWS Elastic Beanstalk. In questo caso, è necessario aggiornare l'ambiente Beanstalk in modo che Elastic Beanstalk rimuova la configurazione di avvio e la sostituisca con un modello di avvio.

Aggiornare uno stack per utilizzare un modello di avvio

Segui i passaggi indicati in questa sezione per effettuare le seguenti operazioni:

- Riscrivi la configurazione di lancio come modello di avvio utilizzando le proprietà equivalenti del modello di avvio.
- Associa il nuovo modello di avvio con il gruppo con dimensionamento automatico.
- Distribuisci questi aggiornamenti.

Per modificare il modello dello stack e aggiornare lo stack

1. Segui le stesse procedure generali per modificare il modello di stack descritte in [Modificare un modello di stack](#) nella AWS CloudFormation Guida per l'utente.
2. Riscrivi la configurazione di avvio come modello di avvio. Fai riferimento al file di esempio seguente:

Esempio: una semplice configurazione di avvio

```
---
Resources:
  myLaunchConfig:
    Type: AWS::AutoScaling::LaunchConfiguration
    Properties:
      ImageId: ami-02354e95b3example
      InstanceType: t3.micro
      SecurityGroups:
        - !Ref EC2SecurityGroup
      KeyName: MyKeyPair
      BlockDeviceMappings:
        - DeviceName: /dev/xvda
          Ebs:
            VolumeSize: 150
            DeleteOnTermination: true
      UserData:
        Fn::Base64: !Sub |
          #!/bin/bash -xe
          yum install -y aws-cfn-bootstrap
          /opt/aws/bin/cfn-signal -e $? --stack ${AWS::StackName} --resource myASG
          --region ${AWS::Region}
```

Esempio: l'equivalente del modello di avvio

```
---
Resources:
```

```
myLaunchTemplate:
  Type: AWS::EC2::LaunchTemplate
  Properties:
    LaunchTemplateName: !Sub ${AWS::StackName}-launch-template
    LaunchTemplateData:
      ImageId: ami-02354e95b3example
      InstanceType: t3.micro
      SecurityGroupIds:
        - Ref! EC2SecurityGroup
      KeyName: MyKeyPair
      BlockDeviceMappings:
        - DeviceName: /dev/xvda
          Ebs:
            VolumeSize: 150
            DeleteOnTermination: true
      UserData:
        Fn::Base64: !Sub |
          #!/bin/bash -x
          yum install -y aws-cfn-bootstrap
          /opt/aws/bin/cfn-signal -e $? --stack ${AWS::StackName} --resource
myASG --region ${AWS::Region}
```

Per informazioni di riferimento su tutte le proprietà supportate da Amazon EC2, consulta [AWS::EC2::LaunchTemplate](#) la Guida per l'AWS CloudFormation utente.

Nota come il modello di avvio include la proprietà `LaunchTemplateName` con un valore di `!Sub ${AWS::StackName}-launch-template`. Questo è necessario se si desidera che il nome del modello di avvio includa il nome dello stack.

3. Se la proprietà **IamInstanceProfile** è presente nella configurazione di avvio, dovrai convertirla in una struttura e specificare il nome o l'ARN del profilo dell'istanza. Per vedere un esempio, consulta [AWS::EC2::LaunchTemplate](#).
4. Se le proprietà **AssociatePublicIpAddress**, **InstanceMonitoring** o **PlacementTenancy** sono presenti nella configurazione di avvio, è necessario convertirle in una struttura. Per alcuni esempi, consulta [AWS::EC2::LaunchTemplate](#).

Un'eccezione si verifica quando il valore della proprietà `MapPublicIpOnLaunch` nelle sottoreti utilizzate per il gruppo con dimensionamento automatico corrisponde al valore della proprietà `AssociatePublicIpAddress` nella configurazione di avvio. In questo caso, puoi ignorare la proprietà `AssociatePublicIpAddress`. La proprietà `AssociatePublicIpAddress` viene

utilizzata solo per sovrascrivere la proprietà `MapPublicIpOnLaunch` e modificare se le istanze ricevono un indirizzo IPv4 pubblico all'avvio.

- È possibile copiare i gruppi di sicurezza dalla proprietà **SecurityGroups** in una delle due posizioni del modello di avvio. Normalmente, si copiano i gruppi di sicurezza nella proprietà `SecurityGroupIds`. Tuttavia, se si crea una struttura `NetworkInterfaces` all'interno del modello di avvio per specificare la proprietà `AssociatePublicIpAddress`, è necessario copiare invece i gruppi di sicurezza nella proprietà `Groups` dell'interfaccia di rete.
- Se nella configurazione di avvio sono presenti strutture `BlockDeviceMapping` impostate su **NoDevice** o `true` è necessario specificare una stringa vuota per `NoDevice` nel modello di avvio per far sì che Amazon EC2 ometta il dispositivo.
- Se la proprietà **SpotPrice** è presente nella configurazione di avvio, ti consigliamo di ometterla dal modello di avvio. Le tue istanze spot verranno avviate al prezzo Spot corrente. Questo prezzo non supererà mai il prezzo on demand.

Per richiedere le istanze spot, hai due opzioni che si escludono a vicenda:

- La prima consiste nell'utilizzare la struttura `InstanceMarketOptions` del modello di avvio (scelta non consigliata). Per ulteriori informazioni, consulta [AWS::EC2::LaunchTemplate InstanceMarketOptions](#) la Guida per l'AWS CloudFormation utente.
 - L'altro è aggiungere una struttura `MixedInstancesPolicy` al gruppo con dimensionamento automatico. In questo modo avrai a disposizione più opzioni su come effettuare la richiesta. Una richiesta di istanza Spot nel modello di avvio non supporta la selezione di più di un tipo di istanza per gruppo con dimensionamento automatico. Tuttavia, una policy di istanze miste supporta più di una selezione del tipo di istanza per gruppo con dimensionamento automatico. Le richieste di istanze spot traggono vantaggio dalla possibilità di scegliere tra più di un tipo di istanza. Per ulteriori informazioni, consulta [AWS::AutoScaling::AutoScalingGroup MixedInstancesPolicy](#) [AWS::AutoScaling::AutoScaling](#) nella Guida AWS CloudFormation per l'utente.
- Rimuovi la **LaunchConfigurationName** proprietà dalla risorsa [AWS::AutoScaling::AutoScalingGroup](#) . Aggiungi il modello di avvio al suo posto.

Negli esempi seguenti, la funzione intrinseca [Ref](#) ottiene l'ID della [AWS::EC2::LaunchTemplate](#) risorsa con l'ID logico. `myLaunchTemplate` La [GetAtt](#) funzione ottiene il numero di versione più recente (ad esempio 1) del modello di avvio per la `Version` proprietà.

Esempio: senza una policy a istanze miste

```
---
Resources:
  myASG:
    Type: AWS::AutoScaling::AutoScalingGroup
    Properties:
      LaunchTemplate:
        LaunchTemplateId: !Ref myLaunchTemplate
        Version: !GetAtt myLaunchTemplate.LatestVersionNumber
    ...
```

Esempio: con una policy a istanze miste

```
---
Resources:
  myASG:
    Type: AWS::AutoScaling::AutoScalingGroup
    Properties:
      MixedInstancesPolicy:
        LaunchTemplate:
          LaunchTemplateSpecification:
            LaunchTemplateId: !Ref myLaunchTemplate
            Version: !GetAtt myLaunchTemplate.LatestVersionNumber
    ...
```

Per informazioni di riferimento per tutte le proprietà supportate da Amazon EC2 Auto Scaling, [AWS::AutoScaling::AutoScaling](#) consulta [AWS::AutoScaling::AutoScaling](#) nella AWS CloudFormation Guida per l'utente.

9. Quando sei pronto per distribuire questi aggiornamenti, segui le CloudFormation procedure per aggiornare lo stack con il modello di stack modificato. Per ulteriori informazioni, consultare [Modifica di un ruolo](#) nella AWS CloudFormation Guida per l'utente.

Comprendi l'aggiornamento dei comportamenti delle risorse stack

CloudFormation aggiorna le risorse dello stack confrontando le modifiche tra il modello aggiornato fornito e le configurazioni delle risorse descritte nella versione precedente del modello di stack. Le

configurazioni delle risorse non interessate da modifiche rimangono inalterate durante il processo di aggiornamento.

CloudFormation supporta l'[UpdatePolicy](#) attributo per i gruppi Auto Scaling. Durante un aggiornamento, se UpdatePolicy è impostato su AutoScalingRollingUpdate, CloudFormation sostituisce InService le istanze dopo aver eseguito i passaggi di questa procedura. Se UpdatePolicy è impostato su AutoScalingReplacingUpdate, CloudFormation sostituisce il gruppo Auto Scaling e il relativo pool caldo (se esistente).

Se non hai specificato un UpdatePolicy attributo per il tuo gruppo Auto Scaling, viene verificata la correttezza del modello di avvio, ma CloudFormation non implementa alcuna modifica tra le istanze del gruppo Auto Scaling. Tutte le nuove istanze riceveranno il tuo modello di avvio, ma le istanze esistenti continuano a essere eseguite con la configurazione di avvio con cui sono state originariamente avviate (nonostante la configurazione di avvio non esista). L'eccezione si verifica quando si modificano le opzioni di acquisto, ad esempio aggiungendo una policy per le istanze miste. In questo caso, il gruppo con dimensionamento automatico sostituisce gradualmente le istanze esistenti con nuove istanze in base alle nuove opzioni di acquisto.

Monitora la migrazione

per Monitorare la migrazione

1. Nella [console AWS CloudFormation](#), selezionare lo stack aggiornato, quindi selezionare la scheda Eventi per visualizzare gli eventi dello stack.
2. Per aggiornare l'elenco degli eventi con gli eventi più recenti, scegli il pulsante di aggiornamento nella console. CloudFormation
3. Durante l'aggiornamento dello stack, noterai più eventi per ogni aggiornamento delle risorse. Se nella colonna Motivo dello stato vedi un'eccezione che indica un problema durante il tentativo di creare il modello di avvio, consulta [Risoluzione dei problemi di Amazon EC2 Auto Scaling: modelli di avvio](#) per le possibili cause.
4. (Facoltativo) A seconda dell'utilizzo dell'attributo UpdatePolicy, puoi monitorare l'avanzamento del tuo gruppo con dimensionamento automatico dalla pagina dei [gruppi con dimensionamento automatico](#) della console Amazon EC2. Seleziona il gruppo Auto Scaling. Nella scheda Activity (Attività), in Activity history (Cronologia attività), la colonna Status (Stato) indica se il gruppo con dimensionamento automatico ha avviato o terminato correttamente le istanze o se l'attività di dimensionamento è ancora in corso.

5. Quando l'aggiornamento dello stack è completo, CloudFormation genera un evento UPDATE_COMPLETE stack. Per ulteriori informazioni, consulta [Monitoraggio dell'avanzamento di un aggiornamento dello stack](#) nella AWS CloudFormation Guida per l'utente.
6. Una volta completato l'aggiornamento dello stack, apri [la pagina dei modelli di avvio](#) e [la pagina delle configurazioni di avvio](#) della console Amazon EC2. Noterai che viene creato un nuovo modello di avvio e la configurazione di avvio viene eliminata.

Riferimento della mappatura di avvio

A scopo di riferimento, la tabella seguente elenca tutte le proprietà di primo livello della [AWS::AutoScaling::LaunchConfiguration](#) risorsa con le proprietà corrispondenti nella risorsa [AWS::EC2::LaunchTemplate](#)

Avvia la proprietà di origine della configurazione	Proprietà target del modello di avvio
AssociatePublicIpAddress	NetworkInterfaces.AssociatePublicIpAddress
BlockDeviceMappings	BlockDeviceMappings
ClassicLinkVPCId	Non disponibile ¹
ClassicLinkVPCSecurityGroups	Non disponibile ¹
EbsOptimized	EbsOptimized
IamInstanceProfile	Specifica <code>IamInstanceProfile.Arn</code> o <code>IamInstanceProfile.Name</code> , ma non entrambi
ImageId	ImageId
InstanceId	InstanceId
InstanceMonitoring	Monitoring.Enabled
InstanceType	InstanceType

Avvia la proprietà di origine della configurazione	Proprietà target del modello di avvio
KernelId	KernelId
KeyName	KeyName
LaunchConfigurationName	LaunchTemplateName
MetadataOptions	MetadataOptions
PlacementTenancy	Placement.Tenancy
RamDiskId	RamDiskId
SecurityGroups	Specifica SecurityGroupIds o NetworkInterfaces.Groups , ma non entrambi
SpotPrice	InstanceMarketOptions.SpotOptions.MaxPrice
UserData	UserData

¹ Le ClassicLinkVPCSecurityGroups proprietà ClassicLinkVPCId and non possono essere utilizzate in un modello di avvio perché EC2-Classic non è più disponibile.

Esempi di creazione e gestione di modelli di avvio con () AWS Command Line Interface AWS CLI

Puoi creare e gestire modelli di lancio tramite AWS Management Console AWS CLI, o SDK. Questa sezione mostra esempi di creazione e gestione di modelli di lancio per Amazon EC2 Auto Scaling da AWS CLI

Indice

- [Esempio di utilizzo](#)
- [Creare un modello di avvio di base](#)
- [Specificare dei tag che etichettano le istanze all'avvio](#)

- [Specificare un ruolo IAM da passare alle istanze](#)
- [Assegnare un indirizzo IP pubblico](#)
- [Specificare uno script di dati utente che configura le istanze all'avvio](#)
- [Specificare di una mappatura dei dispositivi a blocchi di un'AMI](#)
- [Specificare gli host dedicati per portare licenze software da fornitori esterni](#)
- [Specificare un'interfaccia di rete esistente](#)
- [Creare più interfacce di rete](#)
- [Gestire i modelli di avvio](#)
- [Aggiornare un gruppo con dimensionamento automatico per utilizzare un modello di avvio](#)

Esempio di utilizzo

```
{
  "LaunchTemplateName": "my-template-for-auto-scaling",
  "VersionDescription": "test description",
  "LaunchTemplateData": {
    "ImageId": "ami-04d5cc9b88example",
    "InstanceType": "t2.micro",
    "SecurityGroupIds": [
      "sg-903004f88example"
    ],
    "KeyName": "MyKeyPair",
    "Monitoring": {
      "Enabled": true
    },
    "Placement": {
      "Tenancy": "dedicated"
    },
    "CreditSpecification": {
      "CpuCredits": "unlimited"
    },
    "MetadataOptions": {
      "HttpTokens": "required",
      "HttpPutResponseHopLimit": 1,
      "HttpEndpoint": "enabled"
    }
  }
}
```

Creare un modello di avvio di base

Per creare un modello di lancio di base, usa il [create-launch-template](#) comando come segue, con queste modifiche:

- Replace (Sostituisci) `ami-04d5cc9b88example` con l'ID dell'AMI da cui avviare le istanze.
- Sostituisci `t2.micro` con un tipo di istanza compatibile con l'AMI specificata.

Questo esempio crea un modello di lancio con il nome *my-template-for-auto-scaling*. Se le istanze create da questo modello di avvio vengono avviate in un VPC predefinito, ricevono un indirizzo IP pubblico per impostazione predefinita. Se le istanze vengono avviate in un VPC non predefinito, non ricevono un indirizzo IP pubblico per impostazione predefinita.

```
aws ec2 create-launch-template --launch-template-name my-template-for-auto-scaling --  
version-description version1 \  
  --launch-template-data  
'{"ImageId": "ami-04d5cc9b88example", "InstanceType": "t2.micro"}'
```

Per ulteriori informazioni sulla citazione di parametri in formato JSON, consulta [Utilizzo di virgolette con stringhe nella AWS CLI](#) nella Guida per l'utente della AWS Command Line Interface).

In alternativa, puoi specificare i parametri formattati JSON in un file di configurazione.

Nell'esempio seguente viene creato un modello di avvio di base, facendo riferimento a un file di configurazione per i valori dei parametri del modello di avvio.

```
aws ec2 create-launch-template --launch-template-name my-template-for-auto-scaling --  
version-description version1 \  
  --launch-template-data file://config.json
```

Contenuto di `config.json`.

```
{  
  "ImageId": "ami-04d5cc9b88example",  
  "InstanceType": "t2.micro"  
}
```

Specificare dei tag che etichettano le istanze all'avvio

Nell'esempio seguente viene aggiunto un tag (ad esempio `purpose=webserver`) alle istanze al momento dell'avvio.

```
aws ec2 create-launch-template --launch-template-name my-template-for-auto-scaling --
version-description version1 \
  --launch-template-data '{"TagSpecifications":[{"ResourceType":"instance","Tags":
[{"Key": "purpose", "Value": "webserver"}]}], "ImageId": "ami-04d5cc9b88example", "InstanceType": "t2.
```

Note

Se si specificano i tag delle istanze nel modello di avvio e si sceglie di propagare i tag del gruppo con dimensionamento automatico alle relative istanze, tutti i tag vengono uniti. Se viene specificata la stessa chiave tag per un tag nel modello di avvio e un tag nel gruppo con dimensionamento automatico, il valore del tag del gruppo ha la precedenza.

Specificare un ruolo IAM da passare alle istanze

L'esempio seguente specifica il nome del profilo di istanza associato al ruolo IAM da passare alle istanze all'avvio. Per ulteriori informazioni, consulta [Ruoli IAM per le applicazioni in esecuzione sulle istanze Amazon EC2](#).

```
aws ec2 create-launch-template --launch-template-name my-template-for-auto-scaling --
version-description version1 \
  --launch-template-data '{"IamInstanceProfile":{"Name": "my-instance-
profile"}, "ImageId": "ami-04d5cc9b88example", "InstanceType": "t2.micro"}'
```

Assegnare un indirizzo IP pubblico

L'[create-launch-template](#) esempio seguente configura il modello di avvio per assegnare indirizzi pubblici alle istanze avviate in un VPC non predefinito.

Note

Quando specifichi un'interfaccia di rete, specifica un valore per `Groups` che corrisponda ai gruppi di sicurezza per il VPC nel quale il gruppo con dimensionamento automatico avvierà

le istanze. Specificare il VPC e le sottoreti come proprietà del gruppo con dimensionamento automatico.

```
aws ec2 create-launch-template --launch-template-name my-template-for-auto-scaling --
version-description version1 \
  --launch-template-data '{"NetworkInterfaces":
[{"DeviceIndex":0,"AssociatePublicIpAddress":true,"Groups":
["sg-903004f88example"],"DeleteOnTermination":true]}],"ImageId":"ami-04d5cc9b88example","InstanceType":"t2.micro"}'
```

Specificare uno script di dati utente che configura le istanze all'avvio

Nell'esempio seguente viene specificato uno script di dati utente come stringa con codifica base64 che configura le istanze all'avvio. Il [create-launch-template](#) comando richiede dati utente con codifica base64.

```
aws ec2 create-launch-template --launch-template-name my-template-for-auto-scaling --
version-description version1 \
--launch-template-data
'{"UserData":"IyEvYmLuL2Jhc...", "ImageId":"ami-04d5cc9b88example", "InstanceType":"t2.micro"}'
```

Specificare di una mappatura dei dispositivi a blocchi di un'AMI

L'[create-launch-template](#) esempio seguente crea un modello di avvio con una mappatura del dispositivo a blocchi: un volume EBS da 22 gigabyte mappato su. /dev/xvdcz Il volume /dev/xvdcz utilizza il tipo di volume a scopo generico SSD (gp2) e viene eliminato quando termina l'istanza a cui è allegato.

```
aws ec2 create-launch-template --launch-template-name my-template-for-auto-scaling --
version-description version1 \
  --launch-template-data '{"BlockDeviceMappings":[{"DeviceName":"/dev/xvdcz","Ebs":
{"VolumeSize":22,"VolumeType":"gp2","DeleteOnTermination":true}}],"ImageId":"ami-04d5cc9b88example","InstanceType":"t2.micro"}'
```

Specificare gli host dedicati per portare licenze software da fornitori esterni

Se specifichi un tenancy host, è possibile specificare un gruppo di risorse host e una configurazione delle licenze di License Manager per ottenere licenze software idonee da fornitori esterni. Quindi, puoi utilizzare le licenze sulle istanze EC2 utilizzando il seguente comando. [create-launch-template](#)

```
aws ec2 create-launch-template --launch-template-name my-template-for-auto-scaling --
version-description version1 \
  --launch-template-data '{"Placement":
{"Tenancy":"host","HostResourceGroupArn":"arn"}, "LicenseSpecifications":
[{"LicenseConfigurationArn":"arn"}], "ImageId":"ami-04d5cc9b88example", "InstanceType":"t2.micro"
```

Specificare un'interfaccia di rete esistente

L'[create-launch-template](#) esempio seguente configura l'interfaccia di rete principale per utilizzare un'interfaccia di rete esistente.

```
aws ec2 create-launch-template --launch-template-name my-template-for-auto-scaling --
version-description version1 \
  --launch-template-data '{"NetworkInterfaces":
[{"DeviceIndex":0, "NetworkInterfaceId":"eni-
b9a5ac93", "DeleteOnTermination":false}], "ImageId":"ami-04d5cc9b88example", "InstanceType":"t2.mi
```

Creare più interfacce di rete

L'[create-launch-template](#) esempio seguente aggiunge un'interfaccia di rete secondaria. L'interfaccia di rete primaria ha un indice dispositivo pari a 0 e l'interfaccia di rete secondaria ha un indice dispositivo pari a 1.

```
aws ec2 create-launch-template --launch-template-name my-template-for-auto-scaling --
version-description version1 \
  --launch-template-data '{"NetworkInterfaces":[{"DeviceIndex":0, "Groups":
["sg-903004f88example"], "DeleteOnTermination":true}, {"DeviceIndex":1, "Groups":
["sg-903004f88example"], "DeleteOnTermination":true}], "ImageId":"ami-04d5cc9b88example", "Instance
```

Se utilizzi un tipo di istanza che supporta più schede di rete ed Elastic Fabric Adapters (EFA), puoi aggiungere un'interfaccia secondaria a una scheda di rete secondaria e abilitare EFA utilizzando il comando seguente. [create-launch-template](#) Per ulteriori informazioni, consulta [Aggiunta di un EFA a un modello di avvio](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux).

```
aws ec2 create-launch-template --launch-template-name my-template-for-auto-scaling --
version-description version1 \
  --launch-template-data '{"NetworkInterfaces":
[{"NetworkCardIndex":0, "DeviceIndex":0, "Groups":
["sg-7c2270198example"], "InterfaceType":"efa", "DeleteOnTermination":true},
```



```
{"NetworkCardIndex":1,"DeviceIndex":1,"Groups":  
["sg-7c2270198example"],"InterfaceType":"efa","DeleteOnTermination":true}], "ImageId":"ami-09d95
```

Warning

Il tipo di istanza p4d.24xlarge comporta costi superiori, rispetto agli altri esempi in questa sezione. Per ulteriori informazioni sui prezzi delle istanze P4d, consulta [Prezzi delle istanze P4d di Amazon EC2](#).

Note

Allegare più interfacce di rete della stessa sottorete a un'istanza può introdurre il routing asimmetrico, specialmente nelle istanze che utilizzano una variante di Linux non Amazon. Se hai bisogno di questo tipo di configurazione, dovrai anche configurare l'interfaccia di rete secondaria all'interno del sistema operativo. Per un esempio, vedi [Come posso far funzionare la mia interfaccia di rete secondaria nella mia istanza Ubuntu EC2?](#) nel AWS Knowledge Center.

Gestire i modelli di avvio

AWS CLI Include diversi altri comandi che consentono di gestire i modelli di lancio.

Indice

- [Pubblicare e descrivere i modelli di avvio](#)
- [Creazione di una versione del modello di avvio](#)
- [Eliminare una versione del modello di avvio](#)
- [Eliminare un modello di avvio](#)

Pubblicare e descrivere i modelli di avvio

Puoi usare due AWS CLI comandi per ottenere informazioni sui tuoi modelli di lancio: [describe-launch-templates](#) e [describe-launch-template-versions](#).

Il [describe-launch-templates](#) comando consente di ottenere un elenco di tutti i modelli di avvio creati. È possibile utilizzare un'opzione per filtrare i risultati in base al nome di un modello di avvio, creare

ora, chiave tag o combinazione chiave-valore tag. Questo comando restituisce informazioni di riepilogo su uno qualsiasi dei modelli di avvio, inclusi l'identificativo, la versione più recente e la versione di default del modello di avvio.

L'esempio seguente fornisce un riepilogo del modello di avvio specificato.

```
aws ec2 describe-launch-templates --launch-template-names my-template-for-auto-scaling
```

Di seguito è riportata una risposta di esempio.

```
{
  "LaunchTemplates": [
    {
      "LaunchTemplateId": "lt-068f72b729example",
      "LaunchTemplateName": "my-template-for-auto-scaling",
      "CreateTime": "2020-02-28T19:52:27.000Z",
      "CreatedBy": "arn:aws:iam::123456789012:user/Bob",
      "DefaultVersionNumber": 1,
      "LatestVersionNumber": 1
    }
  ]
}
```

Se non utilizzi l'opzione `--launch-template-names` per limitare l'output a un modello di avvio, vengono restituite le informazioni su tutti i modelli di avvio.

Il [describe-launch-template-versions](#) comando seguente fornisce informazioni che descrivono le versioni del modello di avvio specificato.

```
aws ec2 describe-launch-template-versions --launch-template-id lt-068f72b729example
```

Di seguito è riportata una risposta di esempio.

```
{
  "LaunchTemplateVersions": [
    {
      "VersionDescription": "version1",
      "LaunchTemplateId": "lt-068f72b729example",
      "LaunchTemplateName": "my-template-for-auto-scaling",
      "VersionNumber": 1,
      "CreatedBy": "arn:aws:iam::123456789012:user/Bob",
    }
  ]
}
```

```

    "LaunchTemplateData": {
      "TagSpecifications": [
        {
          "ResourceType": "instance",
          "Tags": [
            {
              "Key": "purpose",
              "Value": "webserver"
            }
          ]
        }
      ],
      "ImageId": "ami-04d5cc9b88example",
      "InstanceType": "t2.micro",
      "NetworkInterfaces": [
        {
          "DeviceIndex": 0,
          "DeleteOnTermination": true,
          "Groups": [
            "sg-903004f88example"
          ],
          "AssociatePublicIpAddress": true
        }
      ],
      "DefaultVersion": true,
      "CreateTime": "2020-02-28T19:52:27.000Z"
    }
  ]
}

```

Creazione di una versione del modello di avvio

Il [create-launch-template-version](#) comando seguente crea una nuova versione del modello di avvio basata sulla versione 1 del modello di avvio e specifica un ID AMI diverso.

```

aws ec2 create-launch-template-version --launch-template-id lt-068f72b729example --
version-description version2 \
--source-version 1 --launch-template-data "ImageId=ami-c998b6b2example"

```

Per impostare la versione predefinita del modello di lancio, utilizzate il [modify-launch-template](#) comando.

Eliminare una versione del modello di avvio

Il [delete-launch-template-versions](#) comando seguente elimina la versione del modello di lancio specificata.

```
aws ec2 delete-launch-template-versions --launch-template-id lt-068f72b729example --versions 1
```

Eliminare un modello di avvio

Se non è più necessario un modello di avvio, è possibile eliminarlo utilizzando il [delete-launch-template](#) comando seguente. L'eliminazione di un modello di avvio ne elimina tutte le versioni.

```
aws ec2 delete-launch-template --launch-template-id lt-068f72b729example
```

Aggiornare un gruppo con dimensionamento automatico per utilizzare un modello di avvio

È possibile utilizzare il [update-auto-scaling-group](#) comando per aggiungere un modello di avvio a un gruppo Auto Scaling esistente.

Aggiornare un gruppo con dimensionamento automatico affinché utilizzi la versione più recente di un modello di avvio

Il [update-auto-scaling-group](#) comando seguente aggiorna il gruppo Auto Scaling specificato per utilizzare la versione più recente del modello di avvio specificato.

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg \
--launch-template LaunchTemplateId=lt-068f72b729example,Version='$Latest'
```

Aggiornare un gruppo con dimensionamento automatico affinché utilizzi la versione specifica di un modello di avvio

Il [update-auto-scaling-group](#) comando seguente aggiorna il gruppo Auto Scaling specificato per utilizzare una versione specifica del modello di avvio specificato.

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg \
```

```
--launch-template LaunchTemplateName=my-template-for-auto-scaling,Version='2'
```

Usa AWS Systems Manager i parametri anziché gli ID AMI nei modelli di avvio

Questa sezione mostra come creare un modello di avvio che specifichi un AWS Systems Manager parametro che fa riferimento a un ID Amazon Machine Image (AMI). È possibile utilizzare un parametro memorizzato nello stesso Account AWS, un parametro condiviso da un altro Account AWS o un parametro pubblico per un'AMI pubblica gestita da AWS.

Con i parametri Systems Manager, puoi aggiornare i gruppi con scalabilità automatica per utilizzare nuovi ID AMI senza dover creare nuovi modelli di avvio o nuove versioni dei modelli di avvio ogni volta che viene modificato un ID AMI. Tali ID possono variare regolarmente, ad esempio quando un'AMI viene aggiornata con gli ultimi aggiornamenti del sistema operativo o del software.

È possibile creare, aggiornare o eliminare i propri parametri di Systems Manager utilizzando [Parameter Store, una funzionalità di AWS Systems Manager](#). A tal fine, devi creare un parametro di Systems Manager prima di poterlo utilizzare in un modello di avvio. Per iniziare, crea un parametro con il tipo di dati `aws:ec2:image` e, per il relativo valore, immetti l'ID di un'AMI. Il formato dell'ID AMI è `ami-<identifier>`, ad esempio `ami-123example456`. L'ID AMI corretto dipende dal tipo di istanza e dalla Regione AWS da cui stai avviando il gruppo con dimensionamento automatico.

Per ulteriori informazioni sulla creazione di un parametro valido per un ID AMI, vedere [Creazione dei parametri di Systems Manager](#).

Crea un modello di avvio che specifichi un parametro per l'AMI

Per creare un modello di avvio che specifichi un parametro per l'AMI, utilizzate uno dei seguenti metodi:

Console

Per creare un modello di lancio utilizzando un parametro AWS Systems Manager

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nel riquadro di navigazione, seleziona Modelli di avvio quindi Crea modello di avvio.
3. In nome modello di avvio, inserire un nome descrittivo per il modello di avvio.

4. In Applicazioni e immagini SO (Amazon Machine Image), scegli Sfoglia altre AMI.
5. Scegli il pulsante con la freccia a destra della barra di ricerca, quindi scegli Specifica valore personalizzato/parametro Systems Manager.
6. Nella finestra di dialogo Specifica valore personalizzato o parametro Systems Manager, segui questi passaggi:
 - a. Per la stringa ID AMI o parametro Manager Systems, inserisci il nome del parametro Systems Manager utilizzando uno dei seguenti formati:

Per fare riferimento a un parametro pubblico:

- **resolve:ssm:*public-parameter***

Per fare riferimento a un parametro memorizzato nello stesso account:

- **resolve:ssm:*parameter-name***
- **resolve:ssm:*parameter-name:version-number***
- **resolve:ssm:*parameter-name:label***

Per fare riferimento a un parametro condiviso da un altro Account AWS:

- **resolve:ssm:*parameter-ARN***
- **resolve:ssm:*parameter-ARN:version-number***
- **resolve:ssm:*parameter-ARN:label***

- b. Selezionare Salva.

7. Configura qualsiasi altro parametro del modello di avvio, se necessario, quindi scegli Crea modello di avvio. Per ulteriori informazioni, consulta [Creare un modello di avvio per un gruppo con dimensionamento automatico](#).

AWS CLI

Per creare un modello di avvio che specifichi un parametro Systems Manager, è possibile utilizzare uno dei seguenti comandi di esempio. Sostituisci ciascun *placeholder input dell'utente* con le tue informazioni.

Esempio: creare un modello di avvio che specifichi un parametro pubblico di AWS proprietà

Utilizza la sintassi seguente: `resolve:ssm:public-parameter`, dove `resolve:ssm` è il prefisso standard e `public-parameter` è il percorso e il nome del parametro pubblico.

In questo esempio, il modello di avvio utilizza un parametro pubblico AWS fornito per avviare istanze Regione AWS che utilizzano l'AMI Amazon Linux 2 più recente configurata per il tuo profilo.

```
aws ec2 create-launch-template --launch-template-name my-template-for-auto-scaling
--version-description version1 \
--launch-template-data file://config.json
```

Contenuto di `config.json`.

```
{
  "ImageId": "resolve:ssm:/aws/service/ami-amazon-linux-latest/amzn2-ami-hvm-
x86_64-gp2",
  "InstanceType": "t2.micro"
}
```

Di seguito è riportata una risposta di esempio.

```
{
  "LaunchTemplate": {
    "LaunchTemplateId": "lt-089c023a30example",
    "LaunchTemplateName": "my-template-for-auto-scaling",
    "CreateTime": "2022-12-28T19:52:27.000Z",
    "CreatedBy": "arn:aws:iam::123456789012:user/Bob",
    "DefaultVersionNumber": 1,
    "LatestVersionNumber": 1
  }
}
```

Esempio: crea un modello di lancio che specifichi un parametro memorizzato nello stesso account

Utilizza la sintassi seguente: `resolve:ssm:parameter-name`, dove `resolve:ssm` è il prefisso standard e `parameter-name` è il nome del parametro Systems Manager.

Nell'esempio seguente viene creato un modello di avvio che ottiene l'ID AMI da un parametro Systems Manager esistente, denominato `golden-ami`.

```
aws ec2 create-launch-template --launch-template-name my-template-for-auto-scaling \
```

```
--launch-template-data file://config.json
```

Contenuto di config.json.

```
{
  "ImageId": "resolve:ssm:golden-ami",
  "InstanceType": "t2.micro"
}
```

Quando non è specificata alcuna versione, viene utilizzata la versione più recente del parametro.

L'esempio seguente fa riferimento a una versione specifica del parametro *golden-ami*.

L'esempio seguente utilizza la versione *3* del parametro *golden-ami*, ma puoi utilizzare qualsiasi versione valida.

```
{
  "ImageId": "resolve:ssm:golden-ami:3",
  "InstanceType": "t2.micro"
}
```

L'esempio simile riportato di seguito fa riferimento all'etichetta *prod* del parametro che corrisponde a una versione specifica del parametro *golden-ami*.

```
{
  "ImageId": "resolve:ssm:golden-ami:prod",
  "InstanceType": "t2.micro"
}
```

Di seguito è riportato un output di esempio.

```
{
  "LaunchTemplate": {
    "LaunchTemplateId": "lt-068f72b724example",
    "LaunchTemplateName": "my-template-for-auto-scaling",
    "CreateTime": "2022-12-27T17:11:21.000Z",
    "CreatedBy": "arn:aws:iam::123456789012:user/Bob",
    "DefaultVersionNumber": 1,
    "LatestVersionNumber": 1
  }
}
```


Esempio: crea un modello di lancio che specifica un parametro condiviso da un altro Account AWS

Utilizzare la sintassi seguente: `resolve:ssm:parameter-ARN`, where `resolve:ssm` è il prefisso standard ed `parameter-ARN` è l'ARN del parametro Systems Manager.

L'esempio seguente crea un modello di avvio che ottiene l'ID AMI da un parametro Systems Manager esistente con l'ARN di. `arn:aws:ssm:us-east-2:123456789012:parameter/MyParameter`

```
aws ec2 create-launch-template --launch-template-name my-template-for-auto-scaling
--version-description version1 \
--launch-template-data file://config.json
```

Contenuto di `config.json`.

```
{
  "ImageId": "resolve:ssm:arn:aws:ssm:us-east-2:123456789012:parameter/MyParameter",
  "InstanceType": "t2.micro"
}
```

Quando non è specificata alcuna versione, viene utilizzata la versione più recente del parametro.

L'esempio seguente fa riferimento a una versione specifica del parametro `MyParameter`. L'esempio seguente utilizza la versione `3` del parametro `MyParameter`, ma puoi utilizzare qualsiasi versione valida.

```
{
  "ImageId": "resolve:ssm:arn:aws:ssm:us-east-2:123456789012:parameter/MyParameter:3",
  "InstanceType": "t2.micro"
}
```

L'esempio simile riportato di seguito fa riferimento all'etichetta `prod` del parametro che corrisponde a una versione specifica del parametro `MyParameter`.

```
{
  "ImageId": "resolve:ssm:arn:aws:ssm:us-east-2:123456789012:parameter/MyParameter:prod",
  "InstanceType": "t2.micro"
}
```

```
}
```

Di seguito è riportata una risposta di esempio.

```
{
  "LaunchTemplate": {
    "LaunchTemplateId": "lt-00f93d4588example",
    "LaunchTemplateName": "my-template-for-auto-scaling",
    "CreateTime": "2024-01-08T12:43:21.000Z",
    "CreatedBy": "arn:aws:iam::123456789012:user/Bob",
    "DefaultVersionNumber": 1,
    "LatestVersionNumber": 1
  }
}
```

Per specificare un parametro dal Parameter Store in un modello di avvio, è necessario disporre dell'`ssm:GetParameters` autorizzazione per il parametro specificato. Chiunque utilizzi il modello di lancio necessita inoltre dell'`ssm:GetParameters` autorizzazione per convalidare il valore del parametro. Per ulteriori informazioni, consulta [Limitazione dell'accesso ai parametri di Systems Manager utilizzando le policy IAM](#) nella Guida per l'AWS Systems Manager utente.

Verifica che un modello di avvio ottenga l'ID AMI corretto

Usa il [describe-launch-template-versions](#) comando e includi l'`--resolve-alias` opzione per risolvere il parametro nell'ID AMI effettivo.

```
aws ec2 describe-launch-template-versions --launch-template-name my-template-for-auto-scaling \
  --versions $Default --resolve-alias
```

L'esempio restituisce l'ID AMI per `ImageId`. Quando un'istanza viene avviata utilizzando questo modello di avvio, l'ID AMI si risolve nell'`ami-0ac394d6a3example`.

```
{
  "LaunchTemplateVersions": [
    {
      "LaunchTemplateId": "lt-089c023a30example",
      "LaunchTemplateName": "my-template-for-auto-scaling",
      "VersionNumber": 1,
      "CreateTime": "2022-12-28T19:52:27.000Z",
```

```
    "CreatedBy": "arn:aws:iam::123456789012:user/Bob",
    "DefaultVersion": true,
    "LaunchTemplateData": {
      "ImageId": "ami-0ac394d6a3example",
      "InstanceType": "t2.micro",
    }
  }
]
```

Risorse correlate

Per maggiori dettagli sulla specificazione di un parametro Systems Manager nel modello di avvio, consulta [Utilizzare un parametro Systems Manager anziché un ID AMI](#) nella Amazon EC2 User Guide for Linux Instances.

Per ulteriori informazioni sull'utilizzo dei parametri di Systems Manager, vedere i seguenti materiali di riferimento nella documentazione di Systems Manager.

- Per creare versioni ed etichette dei parametri, vedere [Utilizzo delle versioni dei parametri](#) e [Utilizzo delle etichette dei parametri](#).
- Per informazioni su come cercare i parametri pubblici dell'AMI supportati da Amazon EC2, consulta Chiamata dei parametri [pubblici dell'AMI](#).
- Per informazioni sulla condivisione dei parametri con altri AWS account o tramite AWS Organizations, consulta [Utilizzo dei parametri condivisi](#).
- Per informazioni sul monitoraggio della corretta creazione dei parametri, consulta [Supporto nativo dei parametri per Amazon Machine Image ID](#).

Limitazioni

Quando lavorate con i parametri di Systems Manager, tenete presente le seguenti limitazioni:

- Dimensionamento automatico Amazon EC2 supporta solo la specifica di ID AMI come parametri.
- La creazione o l'aggiornamento di [gruppi di istanze miste](#) utilizzando un modello di avvio che specifica un parametro Systems Manager non è attualmente supportato.
- Se il gruppo Auto Scaling utilizza un modello di avvio che specifica un parametro Systems Manager, non sarà possibile avviare un aggiornamento dell'istanza con la configurazione desiderata o utilizzando skip matching.

- Quando si effettua una chiamata per creare o aggiornare il gruppo con dimensionamento automatico, Dimensionamento automatico Amazon EC2 risolve il parametro Systems Manager nel modello di avvio. Se utilizzi parametri avanzati o limiti di velocità di trasmissione effettiva più elevati, le chiamate frequenti all'archivio dei parametri (ovvero, l'operazione `GetParameters`) possono comportare un aumento dei costi per Systems Manager, poiché vengono addebitati costi per ogni interazione con l'API dell'archivio dei parametri. Per ulteriori informazioni, consultare [Prezzi di AWS Systems Manager](#).

Configurazioni di avvio

Important

Non puoi effettuare chiamate `CreateLaunchConfiguration` con nuovi tipi di istanze Amazon EC2 rilasciati dopo il 31 dicembre 2022. Inoltre, qualsiasi nuovo account creato dopo il 1° giugno 2023 non avrà la possibilità di creare nuove configurazioni di avvio tramite la console. In futuro, i nuovi account non saranno in grado di creare nuove configurazioni di avvio utilizzando la console, l'API, la CLI e CloudFormation. Esegui la migrazione ai modelli di lancio per assicurarti di non dover creare nuove configurazioni di lancio ora o in futuro. Per ulteriori informazioni sulla migrazione di gruppi con dimensionamento automatico ai modelli di avvio, consulta [Migra i tuoi gruppi di Auto Scaling ai modelli di avvio](#).

Una configurazione di avvio è un modello di configurazione di istanze che un gruppo Auto Scaling utilizza per avviare le istanze EC2. Quando crei una configurazione di avvio, specifichi le informazioni sulle istanze. Includi l'ID dell'Amazon Machine Image (AMI), il tipo di istanza, una coppia di chiavi, uno o più gruppi di sicurezza e una mappatura dei dispositivi a blocchi. Se hai avviato un'istanza EC2 in precedenza, hai specificato le stesse informazioni per avviare l'istanza.

È possibile specificare la configurazione di avvio con più gruppi Auto Scaling. Tuttavia, è possibile specificare una sola configurazione di avvio alla volta, per un gruppo con scalabilità automatica, e non è possibile modificarla dopo averla creata. Per modificare la configurazione di avvio per un gruppo Auto Scaling, è necessario crearne una e quindi aggiornare il gruppo con scalabilità automatica con la configurazione di avvio creata.

Indice

- [Creazione di una configurazione di avvio](#)
- [Modifica la configurazione di avvio per un gruppo con scalabilità automatica](#)

Creazione di una configurazione di avvio

Important

Non puoi effettuare chiamate `CreateLaunchConfiguration` con nuovi tipi di istanze Amazon EC2 rilasciati dopo il 31 dicembre 2022. Inoltre, qualsiasi nuovo account creato

dopo il 1° giugno 2023 non avrà la possibilità di creare nuove configurazioni di avvio tramite la console. In futuro, i nuovi account non saranno in grado di creare nuove configurazioni di avvio utilizzando la console, l'API, la CLI e CloudFormation. Esegui la migrazione ai modelli di lancio per assicurarti di non dover creare nuove configurazioni di lancio ora o in futuro. Per ulteriori informazioni sulla migrazione di gruppi con dimensionamento automatico ai modelli di avvio, consulta [Migra i tuoi gruppi di Auto Scaling ai modelli di avvio](#).

Questo argomento descrive come creare una configurazione di avvio.

Dopo aver creato una configurazione di avvio, non è possibile modificarla. È invece necessario creare una nuova configurazione di avvio.

Per associare una nuova configurazione di avvio a un gruppo Auto Scaling esistente, vedere. [Modifica la configurazione di avvio per un gruppo con scalabilità automatica](#) Per creare un nuovo gruppo Auto Scaling, vedere. [Creare un gruppo con dimensionamento automatico utilizzando una configurazione di avvio](#)

Indice

- [Creazione di una configurazione di avvio](#)
- [Configura le opzioni dei metadati dell'istanza](#)
- [Creazione di una configurazione di avvio utilizzando un'istanza EC2](#)

Creazione di una configurazione di avvio


Come creare una configurazione di avvio (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nella barra di navigazione in alto, seleziona la tua AWS regione.
3. Nel pannello sinistro di navigazione in Dimensionamento automatico, scegliere Gruppi con dimensionamento automatico.
4. Nella parte superiore della pagina, scegli Configurazioni di avvio. Quando viene richiesta la conferma, scegli Visualizza configurazioni di avvio per confermare che desideri visualizzare la pagina delle configurazioni di avvio.
5. Scegli Create launch configuration (Crea una configurazione di avvio) e inserisci un nome per la configurazione di avvio.

6. In Amazon machine image (AMI) scegli un'AMI. Per trovare un'AMI specifica, puoi usare [find a suitable AMI \(trova un'AMI adatta\)](#), annota il relativo ID e inseriscilo come criterio di ricerca.

Come ottenere l'ID dell'AMI Amazon Linux 2:

- a. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
 - b. Nel pannello di navigazione a sinistra, in Istanze, scegli Istanze e poi Avvia istanze.
 - c. Nella scheda Quick Start alla pagina Choose an Amazon Machine Image (Scegli un'Amazon Machine Image), annota l'ID dell'AMI accanto ad Amazon Linux 2 AMI (HVM) (AMI Amazon Linux 2 (HVM)).
7. In Instance type (Tipo di istanza), scegli una configurazione hardware per le istanze.
 8. In Additional configuration (Configurazione aggiuntiva), presta particolare attenzione ai campi seguenti:
 - a. (Facoltativo) In Purchasing option (Opzione di acquisto), è possibile scegliere tra: Request Spot Instances (Richiesta Istanze Spot) per richiedere Istanze Spot al prezzo Spot, limitato al prezzo on demand. Eventualmente, è possibile specificare un prezzo massimo per ora di istanza per le istanze Spot.

 Note

Le istanze Spot sono una scelta conveniente rispetto alle istanze on demand, se puoi scegliere con flessibilità quando eseguire le applicazioni e se è possibile interromperle. Per ulteriori informazioni, consulta [Richiedi istanze spot per applicazioni flessibili e con tolleranza ai guasti](#).

- b. (Facoltativo) Per IAM instance profile (Profilo istanza IAM) seleziona un ruolo da associare alle istanze. Per ulteriori informazioni, consulta [Ruoli IAM per le applicazioni in esecuzione sulle istanze Amazon EC2](#).
- c. (Facoltativo) Per il monitoraggio, scegli se consentire alle istanze di pubblicare dati metrici a intervalli di 1 minuto su CloudWatch Amazon abilitando il monitoraggio dettagliato. Vengono applicati costi aggiuntivi. Per ulteriori informazioni, consulta [Configurare il monitoraggio per le istanze Auto Scaling](#).
- d. (Facoltativo) In Advanced details (Dettagli avanzati), User data (Dati utente), puoi specificare i dati utente per configurare un'istanza durante l'avvio o per eseguire uno script di configurazione dopo l'avvio dell'istanza.

- e. (Facoltativo) In **Advanced details** (Dettagli avanzati), **IP address type** (Tipo di indirizzo IP), scegli se assegnare un [indirizzo IP pubblico](#) alle istanze del gruppo. Se non imposti un valore, l'impostazione di default consiste nell'assegnare automaticamente le impostazioni IP pubbliche delle sottoreti in cui vengono avviate le istanze.
9. (Facoltativo) Se non hai bisogno di spazio di archiviazione aggiuntivo in **Storage** (volumes) (Archiviazione (volumi)), puoi ignorare questa sezione. In caso contrario, per specificare i volumi da allegare alle istanze oltre ai volumi specificati dall'AMI, seleziona **Add new volume** (Aggiungi nuovo volume). Quindi scegli le opzioni desiderate e i valori associati in **Devices** (Dispositivi), **Snapshot**, **Size** (Dimensioni), **Volume type** (Tipo di volume), **IOPS**, **Throughput** (Prestazioni), **Delete on termination** (Elimina alla terminazione), e **Encrypted** (Crittografato).
10. In **Security groups** (Gruppi di sicurezza), crea o seleziona il gruppo di sicurezza da associare alle istanze del gruppo. Se lasci **Create new security group** (Crea nuovo gruppo di sicurezza) selezionato, viene configurata una regola SSH predefinita per le istanze Amazon EC2 che eseguono Linux. Viene configurato un ruolo RDP di default per le istanze Amazon EC2 che eseguono Windows.
11. In **Key pair (login)** (Coppia di chiavi (login)), scegli un'opzione in **Key pair options** (Opzioni di coppia di chiavi).

Se hai già configurato una coppia di chiavi dell'istanza Amazon EC2 instance, è possibile sceglierla in questa pagina.

Se invece non hai ancora una coppia di chiavi dell'istanza Amazon EC2, scegli **Create a new key pair** (Crea una nuova coppia di chiavi) e assegnale un nome riconoscibile. Scegli **Download key pair** (Scarica coppia di chiavi) per scaricare la coppia di chiavi sul computer.

 **Important**

Se hai bisogno di connetterti alle istanze, non scegliere **Proceed without a key pair** (Procedi senza una coppia di chiavi).

12. Selezionare la casella di controllo di conferma, quindi scegliere **Create launch configuration** (Crea configurazione di avvio).

Per creare una configurazione di avvio da una configurazione di avvio esistente (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nella barra di navigazione in alto, seleziona la tua AWS regione.

3. Nel pannello sinistro di navigazione in Dimensionamento automatico, scegliere Gruppi con dimensionamento automatico.
4. Nella parte superiore della pagina, scegli Configurazioni di avvio. Quando viene richiesta la conferma, scegli Visualizza configurazioni di avvio per confermare che desideri visualizzare la pagina delle configurazioni di avvio.
5. Seleziona la configurazione di avvio e scegli Actions (Operazioni), Copy launch configuration (Copia configurazione di avvio). Ciò imposta una nuova configurazione di avvio con le stesse opzioni dell'originale, ma con "Copy" aggiunto al nome.
6. Nella pagina Copy Launch Configuration (Copia configurazione di avvio), modifica le opzioni di configurazione in base alle esigenze e scegli Create launch configuration (Crea configurazione di avvio).

Come creare una configurazione di avvio tramite la riga di comando

È possibile utilizzare uno dei seguenti comandi:

- [create-launch-configuration](#) (AWS CLI)
- [Nuovo-AS \(LaunchConfiguration\)](#) AWS Tools for Windows PowerShell

Configura le opzioni dei metadati dell'istanza

Amazon EC2 Auto Scaling supporta la configurazione del servizio IMDS (Instance Metadata Service) nelle configurazioni di avvio. In questo modo è possibile utilizzare le configurazioni di avvio per configurare le istanze Amazon EC2 nei gruppi Auto Scaling affinché richiedano l'Instance Metadata Service versione 2 (IMDSv2), che è un metodo orientato alla sessione per richiedere i metadati dell'istanza. Per ulteriori informazioni sui vantaggi di IMDSv2, consulta questo articolo nel blog AWS sui [miglioramenti per aggiungere protezione in profondità al servizio metadati dell'istanza EC2](#).

È possibile configurare IMDS per supportare sia IMDSv2 che IMDSv1 (impostazione di default) che per richiedere l'utilizzo di IMDSv2. Se si utilizza lo AWS CLI o uno degli SDK per configurare IMDS, è necessario utilizzare la versione più recente di AWS CLI o l'SDK per richiedere l'uso di IMDSv2.

È possibile configurare la configurazione di avvio per i seguenti scenari:

- Richiesta dell'uso di IMDSv2 quando richiedi i metadati dell'istanza
- Specifica del limite di hop di risposta PUT
- Disattivazione dell'accesso ai metadati dell'istanza

Per ulteriori dettagli sulla configurazione del servizio metadati dell'istanza, consulta l'argomento seguente: [Configurazione del servizio di metadati dell'istanza](#) nella (Guida per l'utente di Amazon EC2 User Guide per le istanze Linux).

Utilizza la procedura seguente per configurare le opzioni IMDS in una configurazione di avvio. Dopo aver creato la configurazione di avvio, è possibile associarla al gruppo Auto Scaling. Se associ la configurazione di avvio a un gruppo Auto Scaling esistente, questa viene dissociata dal gruppo Auto Scaling e le istanze esistenti richiedono la sostituzione per utilizzare le opzioni IMDS specificate nella nuova configurazione di avvio. Per ulteriori informazioni, consulta [Modifica la configurazione di avvio per un gruppo con scalabilità automatica](#).

Come configurare IMDS in una configurazione di avvio (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nella barra di navigazione in alto, seleziona la tua regione. AWS
3. Nel pannello sinistro di navigazione in Dimensionamento automatico, scegliere Gruppi con dimensionamento automatico.
4. Nella parte superiore della pagina, scegli Configurazioni di avvio. Quando viene richiesta la conferma, scegli Visualizza configurazioni di avvio per confermare che desideri visualizzare la pagina delle configurazioni di avvio.
5. Scegli Create launch configuration (Crea configurazione di avvio) e crea la configurazione di avvio nel modo usuale. Includi l'ID dell'Amazon Machine Image (AMI), il tipo di istanza e, facoltativamente, una coppia di chiavi, uno o più gruppi di sicurezza e altri volumi EBS o volumi di archiviazione delle istanze aggiuntivi.
6. Per configurare le opzioni dei metadati dell'istanza per tutte le istanze associate a questa configurazione di avvio, in Additional configuration (Configurazione aggiuntiva), Advanced details (Dettagli avanzati), effettua le seguenti operazioni:
 - a. In Metadata accessible (Accesso ai metadati), scegli se abilitare o disabilitare l'accesso all'endpoint HTTP del servizio metadati dell'istanza. Per impostazione predefinita, l'endpoint HTTP è abilitato. Se scegli di disabilitare l'endpoint, l'accesso ai metadati dell'istanza viene disattivato. È possibile specificare la condizione per richiedere IMDSv2 solo quando l'endpoint HTTP è abilitato.
 - b. In Metadata version (Versione dei metadati), puoi scegliere se richiedere l'utilizzo di Instance Metadata Service Version 2 (IMDSv2) quando richiedi i metadati dell'istanza. Se non specifichi un valore, l'impostazione di default è quella di supportare sia IMDSv1 che IMDSv2.

- c. In Metadata token response hop limit (Limite di hop di risposta token metadati), puoi impostare il numero consentito di hop di rete per il token dei metadati. Se non specifichi un valore, il valore di default è 1.
7. Al termine, scegli Create launch configuration (Crea configurazione di avvio).

Come richiedere l'utilizzo di IMDSv2 in una configurazione di avvio utilizzando il parametro AWS CLI

Usa il seguente [create-launch-configuration](#) comando con `--metadata-options set toHttpTokens=required`. Quando specifichi un valore per `HttpTokens`, devi anche impostare `HttpEndpoint` anche su `enabled` (abilitato). Poiché l'installazione `secure token` è impostata su `required` (richiesto) per il recupero dei metadati, questo acconsente a che l'istanza richieda l'uso di IMDSv2, quando richiede i metadati dell'istanza.

```
aws autoscaling create-launch-configuration \  
  --launch-configuration-name my-lc-with-imdsv2 \  
  --image-id ami-01e24be29428c15b2 \  
  --instance-type t2.micro \  
  ...  
  --metadata-options "HttpEndpoint=enabled,HttpTokens=required"
```

Come disattivare l'accesso ai metadati dell'istanza

Utilizzate il [create-launch-configuration](#) comando seguente per disattivare l'accesso ai metadati dell'istanza. È possibile riattivare l'accesso in un secondo momento utilizzando il [modify-instance-metadata-options](#) comando.

```
aws autoscaling create-launch-configuration \  
  --launch-configuration-name my-lc-with-ims-disabled \  
  --image-id ami-01e24be29428c15b2 \  
  --instance-type t2.micro \  
  ...  
  --metadata-options "HttpEndpoint=disabled"
```

Creazione di una configurazione di avvio utilizzando un'istanza EC2

Hai anche la possibilità di creare una configurazione di avvio utilizzando gli attributi di un'istanza EC2 in esecuzione.

Ci sono delle differenze tra la creazione di una configurazione di avvio da zero e la creazione di una configurazione di avvio da un'istanza EC2 esistente. Quando si crea una configurazione di avvio da zero, è necessario specificare l'ID dell'immagine, il tipo di istanza, le risorse facoltative (ad esempio i dispositivi di archiviazione) e le impostazioni facoltative (ad esempio il monitoraggio). Invece, quando crei una configurazione di avvio da un'istanza in esecuzione, Amazon EC2 Auto Scaling ricava gli attributi per la configurazione di avvio dall'istanza specificata. Gli attributi si ricavano anche dalla mappatura dei dispositivi a blocchi per l'AMI da cui l'istanza è stata avviata, ignorando altri eventuali dispositivi a blocchi aggiunti dopo l'avvio.

Quando crei una configurazione di avvio utilizzando un'istanza in esecuzione, puoi sostituire i seguenti attributi specificandoli nell'ambito della stessa richiesta: AMI, dispositivi a blocchi, coppia di chiavi, profilo dell'istanza, tipo di istanza, kernel, monitoraggio delle istanze, tenancy di posizionamento, ramdisk, gruppi di sicurezza, prezzo Spot (max), dati utente, se l'istanza dispone di un indirizzo IP pubblico associato e se l'istanza è ottimizzata per EBS.

Note

Se l'istanza specificata dispone di proprietà che non sono attualmente supportate da configurazioni di avvio, le istanze avviate dal gruppo Auto Scaling potrebbero non essere identiche all'istanza EC2 originale.

Important

L'AMI utilizzata per avviare l'istanza specificata deve essere tuttora esistente.

Argomenti

- [Creazione di una configurazione di avvio da un'istanza EC2 \(AWS CLI\)](#)
- [Creazione una configurazione di avvio da un'istanza e sovrascrittura dei dispositivi a blocchi \(AWS CLI\)](#)
- [Creazione di una configurazione di avvio e sovrascrittura del tipo di istanza \(AWS CLI\)](#)

Creazione di una configurazione di avvio da un'istanza EC2 (AWS CLI)

Utilizza il seguente [create-launch-configuration](#) comando per creare una configurazione di avvio da un'istanza utilizzando gli stessi attributi dell'istanza. I dispositivi a blocchi aggiunti dopo l'avvio vengono ignorati.

```
aws autoscaling create-launch-configuration --launch-configuration-name my-lc-from-instance --instance-id i-a8e09d9c
```

È possibile utilizzare il [describe-launch-configurations](#) comando seguente per descrivere la configurazione di avvio e verificare che i relativi attributi corrispondano a quelli dell'istanza.

```
aws autoscaling describe-launch-configurations --launch-configuration-names my-lc-from-instance
```

Di seguito è riportata una risposta di esempio.

```
{
  "LaunchConfigurations": [
    {
      "UserData": null,
      "EbsOptimized": false,
      "LaunchConfigurationARN": "arn",
      "InstanceMonitoring": {
        "Enabled": false
      },
      "ImageId": "ami-05355a6c",
      "CreatedTime": "2014-12-29T16:14:50.382Z",
      "BlockDeviceMappings": [],
      "KeyName": "my-key-pair",
      "SecurityGroups": [
        "sg-8422d1eb"
      ],
      "LaunchConfigurationName": "my-lc-from-instance",
      "KernelId": "null",
      "RamdiskId": null,
      "InstanceType": "t1.micro",
      "AssociatePublicIpAddress": true
    }
  ]
}
```

Creazione una configurazione di avvio da un'istanza e sovrascrittura dei dispositivi a blocchi (AWS CLI)

Per impostazione predefinita, Amazon EC2 Auto Scaling utilizza gli attributi dell'istanza EC2 specificati per creare la configurazione di avvio. Tuttavia, i dispositivi a blocchi provengono dall'AMI utilizzata per avviare l'istanza, non dall'istanza. Per aggiungere dei dispositivi a blocchi alla configurazione di avvio, sovrascrivi la mappatura dei dispositivi a blocchi per la configurazione di avvio.

Usa il seguente [create-launch-configuration](#) comando per creare una configurazione di avvio utilizzando un'istanza EC2 ma con una mappatura personalizzata dei dispositivi a blocchi.

```
aws autoscaling create-launch-configuration --launch-configuration-name my-lc-from-instance-bdm --instance-id i-a8e09d9c \
  --block-device-mappings "[{\\"DeviceName\\":\\"/dev/sda1\\",\\"Ebs\\":{\\"SnapshotId\\":\\"snap-3decf207\\"}},{\\"DeviceName\\":\\"/dev/sdf\\",\\"Ebs\\":{\\"SnapshotId\\":\\"snap-eed6ac86\\"}}]"
```

Usa il [describe-launch-configurations](#) comando seguente per descrivere la configurazione di avvio e verificare che utilizzi la mappatura personalizzata dei dispositivi a blocchi.

```
aws autoscaling describe-launch-configurations --launch-configuration-names my-lc-from-instance-bdm
```

Il seguente esempio di risposta descrive la configurazione di avvio.

```
{
  "LaunchConfigurations": [
    {
      "UserData": null,
      "EbsOptimized": false,
      "LaunchConfigurationARN": "arn",
      "InstanceMonitoring": {
        "Enabled": false
      },
      "ImageId": "ami-c49c0dac",
      "CreatedTime": "2015-01-07T14:51:26.065Z",
      "BlockDeviceMappings": [
        {
          "DeviceName": "/dev/sda1",
```

```

        "Ebs": {
            "SnapshotId": "snap-3decf207"
        }
    },
    {
        "DeviceName": "/dev/sdf",
        "Ebs": {
            "SnapshotId": "snap-eed6ac86"
        }
    }
],
"KeyName": "my-key-pair",
"SecurityGroups": [
    "sg-8637d3e3"
],
"LaunchConfigurationName": "my-lc-from-instance-bdm",
"KernelId": null,
"RamdiskId": null,
"InstanceType": "t1.micro",
"AssociatePublicIpAddress": true
}
]
}

```

Creazione di una configurazione di avvio e sovrascrittura del tipo di istanza (AWS CLI)

Per impostazione predefinita, Amazon EC2 Auto Scaling utilizza gli attributi dell'istanza EC2 specificati per creare la configurazione di avvio. A seconda delle tue esigenze, puoi scegliere di sovrascrivere gli attributi dall'istanza e utilizzare i valori che ti servono. Ad esempio, puoi sovrascrivere il tipo di istanza.

Utilizza il [create-launch-configuration](#) comando seguente per creare una configurazione di avvio utilizzando un'istanza EC2 ma con un tipo di istanza diverso (ad esempio `t2.medium`) rispetto all'istanza (ad esempio `t1.micro`).

```

aws autoscaling create-launch-configuration --launch-configuration-name my-lc-from-
instance-changetype \
  --instance-id i-a8e09d9c --instance-type t2.medium

```

Usa il [describe-launch-configurations](#) comando seguente per descrivere la configurazione di avvio e verificare che il tipo di istanza sia stato sovrascritto.

```
aws autoscaling describe-launch-configurations --launch-configuration-names my-lc-from-instance-changetype
```

Il seguente esempio di risposta descrive la configurazione di avvio.

```
{
  "LaunchConfigurations": [
    {
      "UserData": null,
      "EbsOptimized": false,
      "LaunchConfigurationARN": "arn",
      "InstanceMonitoring": {
        "Enabled": false
      },
      "ImageId": "ami-05355a6c",
      "CreatedTime": "2014-12-29T16:14:50.382Z",
      "BlockDeviceMappings": [],
      "KeyName": "my-key-pair",
      "SecurityGroups": [
        "sg-8422d1eb"
      ],
      "LaunchConfigurationName": "my-lc-from-instance-changetype",
      "KernelId": "null",
      "RamdiskId": null,
      "InstanceType": "t2.medium",
      "AssociatePublicIpAddress": true
    }
  ]
}
```

Modifica la configurazione di avvio per un gruppo con scalabilità automatica

Important

Forniamo le informazioni sulle configurazioni di avvio per i clienti che non hanno ancora eseguito la migrazione dalle configurazioni di avvio ai modelli di avvio. Per ulteriori informazioni sulla migrazione di gruppi con dimensionamento automatico ai modelli di avvio, consulta [Migra i tuoi gruppi di Auto Scaling ai modelli di avvio](#).

Questo argomento descrive come associare una configurazione di avvio diversa al gruppo Auto Scaling.

Dopo aver modificato la configurazione di avvio, tutte le nuove istanze vengono lanciate utilizzando le nuove opzioni di configurazione, ma le istanze esistenti non ne risentono. Per ulteriori informazioni, consulta [Aggiornamento di istanze di dimensionamento automatico](#).

Modifica della configurazione di avvio per un gruppo con scalabilità automatica (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nel pannello sinistro di navigazione in Dimensionamento automatico, scegliere Gruppi con dimensionamento automatico.
3. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

4. Nella scheda Details (Dettagli) scegli Launch configurations (Configurazioni di avvio), Edit (Modifica).
5. Per la configurazione di avvio, scegli la configurazione di avvio.
6. Al termine, scegli Update (Aggiorna).

Per modificare la configurazione di avvio per un gruppo Auto Scaling utilizzando la riga di comando

È possibile utilizzare uno dei seguenti comandi:

- [update-auto-scaling-group](#) (AWS CLI)
- [Aggiorna come \(AutoScalingGroup\)](#) AWS Tools for Windows PowerShell

Gruppi Auto Scaling

Note

Se non conosci i gruppi di Auto Scaling, segui i passaggi del tutorial [Crea il tuo primo gruppo di Auto Scaling](#) per iniziare e scopri come reagisce un gruppo di Auto Scaling quando un'istanza del gruppo termina.

Un gruppo con scalabilità automatica contiene una raccolta di istanze EC2, che vengono trattate come raggruppamento logico ai fini della scalabilità automatica e della gestione. I gruppi con scalabilità automatica consentono inoltre di utilizzare caratteristiche di Dimensionamento automatico Amazon EC2, quali le sostituzioni dei controlli di integrità e le policy di dimensionamento. Il mantenimento del numero di istanze in un gruppo con scalabilità automatica e la scalabilità automatica sono entrambe funzionalità di base del servizio Dimensionamento automatico Amazon EC2.

La dimensione di un gruppo con scalabilità automatica dipende dal numero di istanze impostato come capacità desiderata. È possibile regolare la dimensione in base alla domanda, manualmente o con la scalabilità automatica.

Un gruppo con scalabilità automatica inizia avviando un numero di istanze sufficiente a soddisfare la capacità desiderata. Mantiene questo numero di istanze eseguendo controlli d'integrità periodici sulle istanze del gruppo. Il gruppo con scalabilità automatica continua a mantenere un numero fisso di istanze anche nel caso di un'istanza non integra. Se un'istanza perde di integrità, il gruppo la termina e ne avvia un'altra per sostituirla. Per ulteriori informazioni, consulta [Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico](#).

Per aumentare o diminuire dinamicamente il numero di istanze nel gruppo in base al cambiamento delle condizioni è possibile utilizzare le policy di dimensionamento. Quando la policy di dimensionamento è attiva, il gruppo con scalabilità automatica adegua la capacità desiderata tra i valori di capacità minima e massima specificati e avvia o termina le istanze in base alle esigenze. È inoltre possibile dimensionare le risorse in base a una pianificazione. Per ulteriori informazioni, consulta [Scegli il tuo metodo di dimensionamento](#).

Quando crei un gruppo con dimensionamento automatico puoi scegliere di avviare istanze on demand, istanze Spot o entrambe. È possibile specificare più opzioni di acquisto per il gruppo con

scalabilità automatica solo quando utilizzi un modello di avvio. Per ulteriori informazioni, consulta [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#).

Le istanze Spot consentono di impiegare la capacità EC2 inutilizzata a prezzi significativamente inferiori rispetto alla tariffe on demand. Per ulteriori informazioni, consulta [Istanze spot Amazon EC2](#). Tra le istanze Spot e le istanze On demand ci sono differenze fondamentali:

- Il prezzo per le istanze Spot varia in base alla richiesta
- Amazon EC2 può terminare una singola istanza Spot se la disponibilità o il prezzo delle istanze Spot cambia

Quando un'istanza Spot viene terminata, il gruppo con scalabilità automatica tenta di avviare un'istanza sostitutiva per mantenere la capacità desiderata per il gruppo.

Quando le istanze vengono avviate, se avrai specificato più zone di disponibilità, la capacità desiderata viene distribuita su queste zone. Se si verifica un'operazione di dimensionamento, Dimensionamento automatico Amazon EC2 mantiene automaticamente il bilanciamento tra tutte le zone di disponibilità specificate.

Indice

- [Creare gruppi con dimensionamento automatico utilizzando modelli di avvio](#)
- [Creare gruppi con dimensionamento automatico utilizzando configurazioni di avvio](#)
- [Aggiornamento di un gruppo con dimensionamento automatico](#)
- [Tag di gruppi e istanze Auto Scaling](#)
- [Policy di manutenzione delle istanze](#)
- [Hook del ciclo di vita di Amazon EC2 Auto Scaling](#)
- [Warm pool per Dimensionamento automatico Amazon EC2](#)
- [Scollegare o collegare istanze](#)
- [Rimuovere temporaneamente le istanze dal gruppo con scalabilità automatica](#)
- [Eliminazione dell'infrastruttura Auto Scaling](#)
- [Esempi di creazione e gestione di gruppi Auto Scaling con gli SDK AWS](#)

Creare gruppi con dimensionamento automatico utilizzando modelli di avvio

Se hai creato un modello di avvio, puoi creare un gruppo con dimensionamento automatico che utilizza un modello di avvio come modello di configurazione per le sue istanze EC2. Il modello di avvio specifica informazioni quali l'ID AMI, il tipo di istanza, la coppia di chiavi, i gruppi di sicurezza e la mappatura dei dispositivi a blocchi per le istanze. Per informazioni sulla creazione dei modelli di lancio, consulta [Creare un modello di avvio per un gruppo con dimensionamento automatico](#).

Per creare un gruppo con dimensionamento automatico devi disporre di autorizzazioni sufficienti. Se non esiste ancora, per creare il ruolo collegato al servizio utilizzato da Dimensionamento automatico Amazon EC2 per eseguire azioni per conto tuo, devi disporre di autorizzazioni sufficienti. Per esempi di policy IAM che un amministratore può utilizzare come riferimento per concederti le autorizzazioni, consulta [Esempi di policy basate su identità](#) e [Supporto modello di avvio](#).

Indice

- [Creare un gruppo con dimensionamento automatico utilizzando un modello di avvio](#)
- [Creare un gruppo con dimensionamento automatico utilizzando la procedura guidata di Amazon EC2](#)
- [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#)

Creare un gruppo con dimensionamento automatico utilizzando un modello di avvio

Quando crei un gruppo con dimensionamento automatico, devi specificare le informazioni necessarie a configurare le istanze Amazon EC2, le zone di disponibilità e le sottoreti VPC per le istanze, la capacità desiderata e i limiti di capacità minima e massima.

Per configurare le istanze Amazon EC2 avviate dal gruppo con dimensionamento automatico, puoi specificare un modello di avvio o una configurazione di avvio. La procedura seguente illustra come creare un gruppo con dimensionamento automatico utilizzando un modello di avvio.

Prerequisiti

- È necessario aver creato un modello di avvio. Per ulteriori informazioni, consulta [Creare un modello di avvio per un gruppo con dimensionamento automatico](#).

Come creare un gruppo con dimensionamento automatico utilizzando un modello d' avvio (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Nella barra di navigazione nella parte superiore dello schermo, scegli lo stesso Regione AWS che hai usato per creare il modello di lancio.
3. Selezionare Crea un gruppo con dimensionamento automatico.
4. Nella pagina Choose launch template or configuration (Scegli configurazione o modello di avvio) eseguire le operazioni seguenti:
 - a. In Nome gruppo con dimensionamento automatico, inserisci un nome per il gruppo con dimensionamento automatico.
 - b. Come Launch template (Modello di avvio) scegli un modello di avvio esistente.
 - c. Per la Versione del modello di avvio, scegli se durante il dimensionamento orizzontale il gruppo con dimensionamento automatico debba usare la versione di default, quella più recente o una versione specifica del modello di avvio.
 - d. Verificare che il modello di avvio supporti tutte le opzioni che si intende utilizzare e quindi scegliere Next (Successivo).
5. Nella pagina Scegli le opzioni di avvio dell'istanza, se non utilizzi più tipi di istanza, puoi saltare la sezione Requisiti del tipo di istanza per utilizzare il tipo di istanza EC2 specificato nel modello di lancio.

Per utilizzare più tipi di istanze, consulta [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#).

6. In Rete, per VPC, scegli un VPC. Il gruppo con dimensionamento automatico deve essere creato nello stesso VPC del gruppo di sicurezza specificato nel modello di avvio.
7. Per (Sottoreti) Availability Zones and subnets (Zone di disponibilità e sottoreti), scegli una o più sottoreti nel VPC specificato. Per un'elevata disponibilità, utilizza sottoreti in più zone di disponibilità. Per ulteriori informazioni, consulta [Considerazioni sulla scelta di sottoreti VPC](#).
8. Se è stato creato un modello di avvio con un tipo di istanza specificato, è possibile continuare con il passaggio successivo per creare un gruppo con dimensionamento automatico che utilizzi il tipo di istanza nel modello di avvio.

In alternativa, è possibile scegliere l'opzione Override launch template (Sostituisci modello di avvio) se non è specificato alcun tipo di istanza nel modello di avvio o se si desidera utilizzare

più tipi di istanza per la scalabilità automatica. Per ulteriori informazioni, consulta [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#).

9. Scegli Next (Avanti) per andare al passaggio successivo.

In alternativa, è possibile accettare il resto dei valori predefiniti e scegliere Skip to review (Passa a revisione).

10. (Facoltativo) Nella pagina Configure group size and scaling policies (Configura policy di dimensionamento e dimensione del gruppo configura le seguenti opzioni e quindi scegli Next (Successivo):
- Per registrare le istanze Amazon EC2 con un load balancer, scegli Enable load balancing (Abilita bilanciamento del carico), quindi scegli un load balancer esistente o creane uno nuovo. Per ulteriori informazioni, consulta [Utilizzare sistema di bilanciamento del carico elastico per distribuire il traffico tra le istanze nel gruppo con dimensionamento automatico](#). Per creare un nuovo load balancer, segui la procedura descritta in [Configurare un Application Load Balancer o Network Load Balancer dalla console di Dimensionamento automatico Amazon EC2](#).
 - (Facoltativo) Per i controlli dell'integrità, tipi di controlli dell'integrità aggiuntivi, seleziona Attiva i controlli dell'integrità di Elastic Load Balancing.
 - (Facoltativo) In Periodo di tolleranza del controllo dell'integrità, inserisci il tempo in secondi. Questo è il periodo di tempo che Dimensionamento automatico Amazon EC2 deve attendere prima di controllare lo stato di integrità di un'istanza dopo che questa è entrata nello stato InService. Per ulteriori informazioni, consulta [Impostazione del periodo di tolleranza del controllo dell'integrità per un gruppo con scalabilità automatica](#).
 - In Impostazioni aggiuntive, Monitoraggio, scegli se abilitare la raccolta di metriche di CloudWatch gruppo. Questi parametri forniscono misurazioni che possono essere indicatori di un potenziale problema, come ad esempio il numero di istanze di terminazione o il numero di istanze in sospeso. Per ulteriori informazioni, consulta [Monitorare i parametri di CloudWatch per gruppi e istanze Auto Scaling](#).
 - Per Abilita il riscaldamento predefinito dell'istanza, seleziona questa opzione e scegli l'orario di riscaldamento per la tua applicazione. Se stai creando un gruppo Auto Scaling con una politica di scalabilità, la funzionalità di riscaldamento delle istanze predefinita migliora i CloudWatch parametri Amazon utilizzati per la scalabilità dinamica. Per ulteriori informazioni, consulta [Impostazione della preparazione di default dell'istanza per un gruppo con scalabilità automatica](#).

11. (Facoltativo) Nella pagina Configure group size and scaling policies (Configura policy di dimensionamento e dimensione del gruppo configurare le seguenti opzioni e quindi scegliere Next (Successivo):
 - a. In Dimensioni gruppo, per Capacità desiderata, immettere il numero iniziale di istanze da avviare.
 - b. Nella sezione Dimensionamento, in Limiti di dimensionamento, se il nuovo valore per la capacità desiderata è maggiore della capacità minima desiderata e della capacità massima desiderata, la capacità massima desiderata viene automaticamente aumentata fino al nuovo valore di capacità desiderato. È possibile modificare questi limiti se necessario. Per ulteriori informazioni, consulta [Impostare i limiti di dimensionamento per il gruppo con dimensionamento automatico](#).
 - c. Per Dimensionamento automatico, scegli se desideri creare una policy di dimensionamento di monitoraggio degli obiettivi. È possibile creare questa policy anche dopo aver creato il gruppo con dimensionamento automatico.

Se scegli la policy di dimensionamento del monitoraggio target, segui le istruzioni riportate in [Creazione di una policy di dimensionamento con monitoraggio degli obiettivi](#) per creare la policy.
 - d. Per la policy di manutenzione dell'istanza, scegli se desideri creare una policy di manutenzione dell'istanza. È possibile creare questa policy anche dopo aver creato il gruppo con dimensionamento automatico. Per creare una policy, segui le indicazioni fornite in [Imposta una policy di manutenzione dell'istanza](#).
 - e. In Instance scale-in protection (Protezione per la riduzione delle istanze) scegliere se abilitare la protezione per la riduzione delle istanze. Per ulteriori informazioni, consulta [Utilizzare la protezione di riduzione orizzontale dell'istanza](#).
12. (Facoltativo) Per ricevere le notifiche, scegliere Add notification (Aggiungi notifica), configurare la notifica, quindi scegliere Next (Successivo). Per ulteriori informazioni, consulta [Opzioni di notifica Amazon SNS per Amazon EC2 Auto Scaling](#).
13. (Facoltativo) Per aggiungere tag, scegliere Add tag (Aggiungi tag), fornire una chiave e un valore di tag, quindi scegliere Next (Successivo). Per ulteriori informazioni, consulta [Tag di gruppi e istanze Auto Scaling](#).
14. Nella pagina Verifica, scegliere Crea gruppo con dimensionamento automatico.

Per creare un gruppo con dimensionamento automatico tramite la riga di comando

È possibile utilizzare uno dei seguenti comandi:

- [create-auto-scaling-group](#) (AWS CLI)
- [Nuovo come \(\) AutoScalingGroup](#) AWS Tools for Windows PowerShell

Creare un gruppo con dimensionamento automatico utilizzando la procedura guidata di Amazon EC2

La procedura seguente mostra come creare un gruppo con dimensionamento automatico tramite la procedura guidata Launch instance (Avvia istanza) nella console Amazon EC2. Questa opzione popola automaticamente un modello di avvio con specifici dettagli di configurazione derivati dalla procedura guidata Launch instance (Avvia istanza).

Note

La procedura guidata non popola il gruppo con dimensionamento automatico con il numero di istanze specificate, ma compila solo il modello di avvio con l'ID Amazon Machine Image (AMI) e il tipo di istanza. Utilizzo della procedura guidata Crea gruppo con dimensionamento automatico per specificare il numero di istanze da avviare.

Un'AMI fornisce le informazioni necessarie per configurare una nuova istanza. Puoi avviare più istanze da un'unica AMI quando devi disporre di più istanze con la stessa configurazione. Si consiglia di utilizzare un'AMI personalizzata che ha già installato l'applicazione per evitare di terminare le istanze se si riavvia un'istanza appartenente a un gruppo con dimensionamento automatico. Per utilizzare un'AMI personalizzata con Amazon EC2 Auto Scaling, devi prima creare l'AMI da un'istanza personalizzata e quindi utilizzare l'AMI per creare un modello di avvio per il tuo gruppo AutoconScaling.dimensionamento automatico.

Prerequisiti

- Devi aver creato un'AMI personalizzata nella stessa area in Regione AWS cui intendi creare il gruppo Auto Scaling. Per ulteriori informazioni, consulta [Creare un'AMI](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Utilizza un'AMI personalizzata come modello

In questa sezione, utilizzi la procedura guidata di avvio di Amazon EC2 per compilare automaticamente un modello di avvio con la tua AMI personalizzata. In alternativa, per configurare il modello di lancio da zero o per ulteriori informazioni sui parametri che puoi configurare per il modello di lancio, consulta [Crea il tuo modello di avvio \(console\)](#).

Per utilizzare un'AMI personalizzata come modello

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Sulla barra di navigazione nella parte superiore dello schermo, Regione AWS viene visualizzata la corrente. Seleziona una regione in cui avviare il gruppo con dimensionamento automatico.
3. Nel riquadro di navigazione, seleziona Istanze.
4. Scegli Avvia istanza e procedi come segue:
 - a. In Nome e tag, lascia il campo Nome vuoto. Il nome non fa parte dei dati utilizzati per creare un modello di avvio.
 - b. In Immagini di applicazioni e sistema operativo (Amazon Machine Image), scegli Sfoglia altre AMI per sfogliare il catalogo completo di AMI.
 - c. Scegli Le mie AMI, trova l'AMI creata in precedenza e quindi scegli Seleziona.
 - d. In Tipo di istanza, seleziona un tipo di istanza.

Note

Seleziona lo stesso tipo di istanza utilizzato quando hai creato l'AMI o uno più potente

- e. Sul lato destro dello schermo, in Riepilogo, inserisci un numero qualsiasi nel campo Numero di istanze. Il numero inserito qui non è rilevante. Dovrai specificare il numero di istanze che desideri avviare quando crei il gruppo con dimensionamento automatico.

Nel campo Numero di istanze, viene visualizzato un messaggio che dice Quando si lanciano più di un'istanza, prendere in considerazione EC2 Auto Scaling.

- f. Seleziona il collegamento ipertestuale considerare EC2 Auto Scaling.
- g. Nella finestra di dialogo di conferma Avvio in gruppo con dimensionamento automatico, scegli Continua per passare alla pagina Crea modello di avvio con l'AMI e il tipo di istanza selezionati nella procedura guidata dell'istanza di avvio già popolata.

Dopo aver scelto Continua, si apre la pagina Crea modello di avvio. Seguire questa procedura per terminare la creazione di un modello di avvio.

Per creare un modello di avvio

1. In Nome e descrizione del modello di avvio immettere un nome e una descrizione per il nuovo modello di avvio.
2. (Facoltativo) In Coppia di chiavi (login), Nome della coppia di chiavi, scegli il nome della coppia di chiavi creata in precedenza da utilizzare durante il collegamento alle istanze, ad esempio tramite SSH.
3. (Facoltativo) Per Impostazioni di rete, Gruppi di sicurezza, scegli uno o più [gruppi di sicurezza](#) creati in precedenza.
4. (Facoltativo) In Configurare l'archiviazione, aggiorna la configurazione dell'archiviazione. La configurazione di archiviazione di default è determinata dall'AMI e dal tipo di istanza.
5. Al termine della configurazione del modello di avvio, scegli Create launch template (Crea modello di avvio).
6. Nella pagina di conferma, scegli Crea gruppo con dimensionamento automatico).

Creazione di un gruppo con dimensionamento automatico

Note

Il resto di questo argomento descrive la procedura di base per la creazione di un gruppo con dimensionamento automatico. Per ulteriori informazioni sui parametri che puoi configurare per il gruppo con dimensionamento automatico, consulta [Creare un gruppo con dimensionamento automatico utilizzando un modello di avvio](#).

Dopo aver scelto Crea gruppo con dimensionamento automatico, si apre la procedura guidata Crea gruppo con dimensionamento automatico. Segui questa procedura per creare un gruppo con dimensionamento automatico.

Creazione di un gruppo con dimensionamento automatico


1. Nella pagina Scegli il modello o la configurazione di avvio immetti un nome per il gruppo con dimensionamento automatico.
2. Il modello di avvio creato è già selezionato.

Per la Versione del modello di avvio, scegli se durante il dimensionamento orizzontale il gruppo con dimensionamento automatico debba usare la versione di default, quella più recente o una versione specifica del modello di avvio.

3. Scegli Avanti per andare al passaggio successivo.
4. Nella pagina Scegli le opzioni di avvio dell'istanza, se non utilizzi più tipi di istanza, puoi saltare la sezione Requisiti del tipo di istanza per utilizzare il tipo di istanza EC2 specificato nel modello di lancio.

Per utilizzare più tipi di istanze, consulta [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#).

5. In Rete, per VPC, scegli un VPC. Il gruppo con dimensionamento automatico deve essere creato nello stesso VPC del gruppo di sicurezza specificato nel modello di avvio.

 Tip

Se non hai specificato un gruppo di sicurezza nel modello di avvio, le istanze vengono avviate con un gruppo di sicurezza predefinito del VPC specificato. Per impostazione predefinita, questo gruppo di sicurezza non permette il traffico in ingresso da reti esterne.

6. Per Zone di disponibilità e sottoreti, scegli una o più sottoreti nel VPC specificato.
7. Scegli Avanti due volte per passare alla pagina Configurazione delle dimensioni del gruppo e delle policy di dimensionamento.
8. In Dimensione del gruppo, definisci la capacità desiderata (numero iniziale di istanze da avviare immediatamente dopo la creazione del gruppo con dimensionamento automatico).
9. Nella sezione Dimensionamento, in Limiti di dimensionamento, se il nuovo valore per la capacità desiderata è maggiore della capacità minima desiderata e della capacità massima desiderata, la capacità massima desiderata viene automaticamente aumentata fino al nuovo valore di capacità desiderato. È possibile modificare questi limiti se necessario. Per ulteriori informazioni, consulta [Impostare i limiti di dimensionamento per il gruppo con dimensionamento automatico](#).
10. Selezionare Passa a revisione.
11. Nella pagina Verifica, scegliere Crea gruppo con dimensionamento automatico.

Passaggi successivi

Puoi verificare che il gruppo con dimensionamento automatico sia stato creato correttamente visualizzando la cronologia delle attività. Nella scheda Attività, in Cronologia attività, la colonna Stato indica se il gruppo con dimensionamento automatico ha avviato correttamente le istanze. Se le istanze non vengono avviate o si avviano ma terminano immediatamente, consulta i seguenti argomenti per individuare le possibili cause e le soluzioni:

- [Risoluzione dei problemi di Amazon EC2 Auto Scaling: errori di avvio dell'istanza EC2](#)
- [Risoluzione dei problemi Amazon EC2 Auto Scaling: problemi AMI](#)
- [Risoluzione dei problemi relativi ai controlli dell'integrità in Amazon EC2 Auto Scaling.](#)

Se lo si desidera, è possibile collegare un load balancer nella stessa regione del gruppo con dimensionamento automatico. Per ulteriori informazioni, consulta [Utilizzare sistema di bilanciamento del carico elastico per distribuire il traffico tra le istanze nel gruppo con dimensionamento automatico.](#)

Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto

È possibile avviare e dimensionare automaticamente un parco di istanze on demand e istanze Spot all'interno di un singolo gruppo con dimensionamento automatico. Oltre a ricevere sconti per l'utilizzo di Istanze Spot, è possibile utilizzare Istanze riservate o Savings Plan per ricevere sconti sul normale prezzo delle istanze on demand. Questi fattori ti aiutano a risparmiare sui costi per le istanze EC2 e ti assicurano il dimensionamento e le prestazioni desiderate per l'applicazione.

Le istanze Spot sono capacità di riserva disponibile a forti sconti rispetto al prezzo On-Demand di EC2. Le istanze spot sono una scelta economicamente vantaggiosa se si può essere flessibili riguardo all'orario di esecuzione delle applicazioni e se queste possono essere interrotte. Possono essere utilizzate per varie applicazioni flessibili e tolleranti ai guasti. Gli esempi includono server Web stateless, endpoint API, applicazioni di big data e analisi, carichi di lavoro containerizzati, pipeline CI/CD, elaborazione ad alte prestazioni e throughput elevato (HPC/HTC), carichi di lavoro di rendering e altri carichi di lavoro flessibili.

Per ulteriori informazioni, consulta le [opzioni di acquisto delle istanze](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Argomenti

- [Panoramica della configurazione](#)
- [Strategie di allocazione](#)

- [Crea un gruppo di istanze miste utilizzando la selezione del tipo di istanza basata su attributi](#)
- [Crea un gruppo di istanze miste scegliendo manualmente i tipi di istanze](#)
- [Configurare un gruppo Auto Scaling per utilizzare i pesi delle istanze](#)
- [Utilizza un modello di avvio diverso per un tipo di istanza](#)

Panoramica della configurazione

Questo capitolo fornisce una panoramica e le best practice per la creazione di un gruppo di istanze miste.

Indice

- [Panoramica](#)
- [Flessibilità tipologia istanza](#)
- [Flessibilità zona di disponibilità](#)
- [Prezzo istanza Spot massimo](#)
- [Ribilanciamento proattivo della capacità](#)
- [Comportamento del dimensionamento](#)
- [Disponibilità regionale dei tipi di istanze](#)
- [Risorse correlate](#)
- [Limitazioni](#)

Panoramica

Per creare un gruppo di istanze miste, sono disponibili due opzioni:

- [Selezione del tipo di istanza basata sugli attributi](#): definisci i requisiti di calcolo per scegliere automaticamente i tipi di istanza in base agli attributi specifici delle istanze.
- [Selezione manuale del tipo di istanza](#): scegli manualmente i tipi di istanza adatti al tuo carico di lavoro.

Manual selection

La procedura seguente descrive come creare un gruppo di istanze miste scegliendo manualmente i tipi di istanza:

1. Scegliere un modello di avvio che contenga i parametri per lanciare un'istanza EC2. I parametri nei modelli di lancio sono facoltativi, ma Dimensionamento automatico Amazon EC2 non può avviare un'istanza se l'ID Amazon Machine Image (AMI) non è presente nel modello di avvio.
2. Scegli l'opzione per sovrascrivere il modello di avvio.
3. Scegli manualmente i tipi di istanza adatti al tuo carico di lavoro.
4. Specifica le percentuali di istanze on demand e istanze spot da avviare.
5. Scegli le strategie di allocazione determinano il modo in cui Dimensionamento automatico Amazon EC2 soddisfa la capacità on demand e spot desiderate dai possibili tipi di istanza.
6. Scegli le zone di disponibilità e le sottoreti VPC in cui avviare le istanze.
7. Specificate la dimensione iniziale del gruppo (la capacità desiderata) e la dimensione minima e massima del gruppo.

Le sostituzioni sono necessarie per sovrascrivere il tipo di istanza dichiarato nel modello di avvio e utilizzare più tipi di istanze incorporati nella definizione delle risorse del gruppo con dimensionamento automatico. Per ulteriori informazioni sulle famiglie e i tipi di istanza disponibili, consulta [Instance types \(Tipi di istanza\)](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

È inoltre possibile configurare i seguenti parametri opzionali per ogni tipo di istanza:

- **LaunchTemplateSpecification**— È possibile assegnare un modello di lancio diverso a un tipo di istanza in base alle esigenze. Questa opzione non è al momento disponibile dalla console. Per ulteriori informazioni, consulta [Utilizza un modello di avvio diverso per un tipo di istanza](#).
- **WeightedCapacity**— Siete voi a decidere in che misura l'istanza conta ai fini della capacità desiderata rispetto al resto delle istanze del gruppo. Se specifichi un valore **WeightedCapacity** per un tipo di istanza, devi specificare un valore **WeightedCapacity** anche per tutti gli altri tipi. Per impostazione predefinita, ogni istanza conta come un'unica istanza ai fini della capacità desiderata. Per ulteriori informazioni, consulta [Configurare un gruppo Auto Scaling per utilizzare i pesi delle istanze](#).

Attribute-based selection

Per consentire a Dimensionamento automatico Amazon EC2 di scegliere automaticamente i tipi di istanza in base agli attributi specifici dell'istanza, utilizza i seguenti passaggi per creare un gruppo di istanze miste specificando i requisiti di elaborazione:

1. Scegliere un modello di avvio che contenga i parametri per lanciare un'istanza EC2. I parametri nei modelli di lancio sono facoltativi, ma Dimensionamento automatico Amazon EC2 non può avviare un'istanza se l'ID Amazon Machine Image (AMI) non è presente nel modello di avvio.
2. Scegli l'opzione per sovrascrivere il modello di avvio.
3. Specifica gli attributi di istanza che corrispondono ai requisiti di calcolo, come ad esempio vCPU e requisiti di memoria.
4. Specifica le percentuali di istanze on demand e istanze spot da avviare.
5. Scegli le strategie di allocazione determinano il modo in cui Dimensionamento automatico Amazon EC2 soddisfa la capacità on demand e spot desiderate dai possibili tipi di istanza.
6. Scegli le zone di disponibilità e le sottoreti VPC in cui avviare le istanze.
7. Specificate la dimensione iniziale del gruppo (la capacità desiderata) e la dimensione minima e massima del gruppo.

Le sostituzioni sono necessarie per sovrascrivere il tipo di istanza dichiarato nel modello di avvio e utilizzare un set di attributi dell'istanza che descrivono i requisiti di calcolo. Per gli attributi supportati, consulta il riferimento [InstanceRequirements](#) all'API Amazon EC2 Auto Scaling. In alternativa, puoi utilizzare un modello di avvio che contiene già la definizione degli attributi dell'istanza.

Puoi anche configurare il parametro `LaunchTemplateSpecification` all'interno della struttura delle sostituzioni per assegnare un modello di avvio diverso a una serie di requisiti dell'istanza, in base alle esigenze. Questa opzione non è al momento disponibile dalla console. Per ulteriori informazioni, consulta il riferimento [LaunchTemplateOverrides](#) all'API Amazon EC2 Auto Scaling.

Per impostazione predefinita, viene impostato il valore per la capacità desiderata del gruppo con dimensionamento automatico come numero di istanze.

Alternativamente, è possibile impostare il valore della capacità desiderata come numero di vCPU o quantità di memoria. A tale scopo, utilizza le proprietà `DesiredCapacityType` nell'operazione API `CreateAutoScalingGroup` o il campo a discesa del tipo di capacità desiderato in AWS Management Console. Si tratta di un'alternativa utile ai [pesi delle istanze](#).

Flessibilità tipologia istanza

Per aumentare la disponibilità, implementa l'applicazione su più tipi di istanza. È consigliabile utilizzare più tipi di istanza per soddisfare i requisiti di capacità. In questo modo, Dimensionamento

automatico Amazon EC2 può avviare un altro tipo di istanza se la capacità dell'istanza non è sufficiente nelle zone di disponibilità scelte.

Con le istanze spot, se la capacità dell'istanza è insufficiente, Dimensionamento automatico Amazon EC2 continua a tentare di avviarle da altri pool di istanza Spot. (I pool che utilizza sono determinati dalla scelta dei tipi di istanza e della strategia di allocazione.) Dimensionamento automatico Amazon EC2 ti aiuta a sfruttare i risparmi sui costi delle istanze spot lanciandole al posto delle istanze on demand.

Noi suggeriamo di essere flessibili su almeno 10 tipi di istanza per ogni carico di lavoro. Quando scegli i tipi di istanza, non limitarti solo ai nuovi tipi di istanza più comuni. La scelta di tipi di istanza di vecchia generazione consente tendenzialmente di ottenere un minor numero di interruzioni, in quanto sono meno richiesti dai clienti on demand.

Flessibilità zona di disponibilità

Distribuisci il gruppo con dimensionamento automatico su più zone di disponibilità. Con più zone di disponibilità, è possibile progettare applicazioni che eseguono il failover automatico tra zone di disponibilità per una maggiore resilienza.

Come ulteriore vantaggio, puoi accedere a un pool di capacità Amazon EC2 più profondo rispetto ai gruppi in una singola zona di disponibilità. Poiché in ciascuna zona di disponibilità la capacità varia indipendentemente per ogni tipo di istanza, spesso è possibile ottenere più capacità di calcolo per lo stesso prezzo se si ha una certa flessibilità per entrambi i tipi di istanza e per la Zona di disponibilità.

Per ulteriori informazioni sull'utilizzo delle zone di disponibilità multiple, consulta [Esempio: distribuzione di istanze tra le zone di disponibilità](#).

Prezzo istanza Spot massimo

Quando create il gruppo Auto Scaling utilizzando AWS CLI o un SDK, potete specificare il parametro `SpotMaxPrice`. Il parametro `SpotMaxPrice` determina il prezzo massimo che si è disposti a pagare per un'ora di istanza spot.

Quando specifichi il parametro `WeightedCapacity` nelle sostituzioni ("`DesiredCapacityType`": "`vcpu`" o "`DesiredCapacityType`": "`memory-mib`" a livello di gruppo), il prezzo massimo rappresenta il prezzo unitario massimo, non il prezzo massimo per un'intera istanza.

Noi suggeriamo di non specificare un prezzo massimo. L'applicazione potrebbe non essere eseguita se non si riceve alcuna istanza spot, ad esempio nel caso in cui il prezzo massimo sia troppo basso.

Se non si specifica un prezzo massimo, il prezzo massimo predefinito è il prezzo on demand. Si paga solo il prezzo Spot per le istanze Spot che vengono avviate. Ricevi comunque i forti sconti offerti dalle istanze spot. Questi sconti sono possibili a causa dei prezzi spot stabili resi disponibili con il [modello di prezzi spot](#). Per ulteriori informazioni, consulta [Prezzi e risparmio](#) nella Guida per l'utente di Amazon EC2 per istanze Linux.

Ribilanciamento proattivo della capacità

Se il tuo caso d'uso lo consente, ti consigliamo il ribilanciamento della capacità. Il ribilanciamento della capacità consente di mantenere la disponibilità del carico di lavoro aumentando in modo proattivo il parco istanze con una nuova istanza spot prima che un'istanza spot in esecuzione riceva l'avviso di interruzione dell'istanza spot di due minuti.

Quando il ribilanciamento della capacità viene abilitato, il dimensionamento automatico Amazon EC2 cerca di sostituire proattivamente le istanze Spot per cui è suggerito il ribilanciamento. È possibile decidere di ribilanciare il carico di lavoro su Istanze Spot nuove o esistenti che non presentano un rischio elevato di interruzione.

Per ulteriori informazioni, consulta [Utilizza il ribilanciamento della capacità per gestire le interruzioni spot di Amazon EC2](#).

Comportamento del dimensionamento

Quando crei un gruppo di istanze miste, per impostazione predefinita vengono utilizzate le istanze on demand. Per utilizzare le istanze spot, è necessario modificare la percentuale del gruppo da avviare come istanze on demand. Per la percentuale on demand, puoi specificare qualsiasi numero tra 0 e 100.

Facoltativamente, puoi stabilire un numero di base di istanze on demand da cui iniziare. Se procedi in tal senso, Dimensionamento automatico Amazon EC2 avvia le istanze Spot solo dopo aver avviato questa capacità di base sulle istanze on demand nella fase di aumento orizzontale del gruppo. Qualsiasi cosa oltre la capacità di base utilizza la percentuale on demand per stabilire quante istanze on demand e istanze Spot avviare.

Amazon EC2 Auto Scaling converte la percentuale nel numero equivalente di istanze. Se il risultato crea un numero frazionario, arrotonda al numero intero successivo a favore delle istanze on demand.

La seguente tabella mostra il comportamento del gruppo con dimensionamento automatico quando aumentano le dimensioni.

Esempio: comportamento di dimensionamento

Opzioni di acquisto Dimensione del gruppo e numero totale di istanze in esecuzione tra le opzioni d'acquisto

10 20 30 40

Esempio 1: base di 10, 50/50% on demand/spot

On-Demand Instances (base amount) 10 10 10 10

On-Demand Instances 0 5 10 15

Spot Instances 0 5 10 15

Esempio 2: base di 0, 0/100% on demand/spot

On-Demand Instances (base amount) 0 0 0 0

On-Demand Instances 0 0 0 0

Spot Instances 10 20 30 40

Esempio 3: base di 0, 60/40% on demand/spot

On-Demand Instances (base amount) 0 0 0 0

Opzioni di acquisto	Dimensione del gruppo e numero totale di istanze in esecuzione tra le opzioni d'acquisto			
On-Demand Instances	6	12	18	24
Spot Instances	4	8	12	16
Esempio 4: base di 0, 0/100% on demand/spot				
On-Demand Instances (base amount)	0	0	0	0
On-Demand Instances	10	20	30	40
Spot Instances	0	0	0	0
Esempio 5: base di 12, 0/100% on demand/spot				
On-Demand Instances (base amount)	10	12	12	12
On-Demand Instances	0	0	0	0
Spot Instances	0	8	18	28

Quando la dimensione del gruppo aumenta, Dimensionamento automatico Amazon EC2 cerca di assicurare che la capacità sia equilibrata in modo uniforme tra le zone di disponibilità specificate. Quindi, avvia i tipi di istanza in base alla strategia di allocazione specificata.

Quando la dimensione del gruppo diminuisce, Dimensionamento automatico Amazon EC2 identifica innanzitutto quale dei due tipi (spot oppure on demand) debba essere terminato. Quindi, tenta di terminare le istanze in modo equilibrato nelle zone di disponibilità specificate. Inoltre, favorisce la terminazione delle istanze in modo da allinearsi maggiormente alle strategie di allocazione. Per ulteriori informazioni sulla terminazione automatica, consulta [Configurazione delle politiche di terminazione per Amazon EC2 Auto Scaling](#).

Disponibilità regionale dei tipi di istanze

La disponibilità dei tipi di istanze EC2 varia a seconda del tipo di istanza. Regione AWS Ad esempio, in una determinata Regione i tipi di istanza di ultima generazione potrebbero non essere ancora disponibili. A causa delle variazioni nella disponibilità delle istanze tra le regioni, potresti riscontrare problemi durante le richieste programmatiche se nella tua regione non sono disponibili più tipi di istanze nelle sostituzioni. L'utilizzo di più tipi di istanze non disponibili nella propria regione potrebbe causare il completo fallimento della richiesta. Per risolvere il problema, riprova la richiesta con diversi tipi di istanza, assicurandoti che ogni tipo di istanza sia disponibile nella regione. Per cercare i tipi di istanza offerti per posizione, usa il [describe-instance-type-offerings](#) comando. Per informazioni sulle famiglie e i tipi di istanze Amazon EC2, consulta [Trova tipi di istanza Amazon EC2](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Risorse correlate

Per ulteriori best practice di istanze Spot, consulta [Best practice per Spot EC2](#) nella Guida per l'utente Amazon EC2 per le istanze Linux.

Limitazioni

Dopo aver aggiunto le sostituzioni a un gruppo di Auto Scaling utilizzando [una politica di istanze miste](#), puoi aggiornare le sostituzioni con UpdateAutoScalingGroup la chiamata API ma non eliminarle. Per rimuovere completamente le sostituzioni, devi prima cambiare il gruppo Auto Scaling in modo che utilizzi un modello di avvio o una configurazione di avvio anziché una politica di istanze miste. Quindi, puoi aggiungere nuovamente una politica per le istanze miste senza alcuna sostituzione.

Strategie di allocazione

Quando utilizzi più tipi di istanza, gestisci il modo in cui Dimensionamento automatico Amazon EC2 soddisfa le capacità on demand e spot dai possibili tipi di istanza. A tale scopo, si specificano strategie di allocazione.

Per esaminare le best practice per un gruppo di istanze miste, consulta [Panoramica della configurazione](#)

Indice

- [Spot Instances](#)
- [Istanze on demand](#)
- [Come funzionano le strategie di allocazione con i pesi](#)

Spot Instances

Dimensionamento automatico Amazon EC2 fornisce le seguenti strategie di allocazione per le istanze Spot:

price-capacity-optimized (consigliato)

La strategia di allocazione ottimizzata per prezzo e capacità esamina sia il prezzo che la capacità per selezionare i pool di istanze spot che hanno il prezzo più basso possibile e meno probabilità di interruzioni.

Quando inizi, ti consigliamo questa strategia. Per ulteriori informazioni, consulta la sezione [Introduzione alla strategia di price-capacity-optimized allocazione per le istanze Spot EC2](#) nel blog. AWS

capacity-optimized

Amazon EC2 Auto Scaling richiede le istanze spot dal pool con capacità ottimale per il numero di istanze in fase di avvio.

Con le istanze Spot, i prezzi cambiano lentamente nel tempo in base ai trend a lungo termine dell'offerta e della domanda. La capacità varia, tuttavia, in tempo reale. La strategia capacity-optimized avvia automaticamente Istanze spot nei pool più disponibili esaminando i dati di capacità in tempo reale e prevedendo quali sono le più disponibili. Ciò consente di ridurre al minimo le possibili interruzioni per i carichi di lavoro che potrebbero avere un costo di interruzione più elevato associato al riavvio del lavoro e al checkpoint. Per dare a determinati tipi di istanza una maggiore possibilità di essere avviate per prime, utilizzare capacity-optimized-prioritized.

capacity-optimized-prioritized

È possibile impostare l'ordine dei tipi di istanza delle sostituzioni del modello di avvio dalla priorità più alta alla più bassa (dalla prima all'ultima nell'elenco). Amazon EC2 Auto Scaling rispetta le

priorità del tipo di istanza sulla base del miglior tentativo, ma prima ottimizza la capacità. Questa è una buona opzione per i carichi di lavoro in cui è necessario ridurre al minimo la possibilità di interruzioni, ma la preferenza per determinati tipi di istanza è anch'essa importante. Se la strategia di allocazione on demand è impostata su `prioritized`, viene applicata la stessa priorità quando si soddisfa la capacità on demand.

lowest-price

Dimensionamento automatico Amazon EC2 richiede che le istanze spot utilizzino i pool con il prezzo più basso all'interno di una zona di disponibilità, attraverso il numero N di pool Spot specificato nell'impostazione Pool con prezzi più bassi. Ad esempio, se specifichi quattro tipi di istanze e quattro zone di disponibilità, il gruppo con dimensionamento automatico ha accesso a 16 pool Spot. (quattro in ogni zona di disponibilità). Se specifichi due pool Spot (N=2) per la strategia di allocazione, il gruppo con dimensionamento automatico può utilizzare i due pool più economici per zona di disponibilità per soddisfare la capacità Spot.

Poiché questa strategia considera solo il prezzo dell'istanza e non la capacità disponibile, potrebbe comportare tassi di interruzione elevati.

Tieni presente che Dimensionamento automatico Amazon EC2 si impegna per prelevare istanze Spot dal numero N di pool specificati. Tuttavia, se un pool esaurisce la capacità Spot prima di soddisfare la capacità desiderata, Dimensionamento automatico Amazon EC2 continua a soddisfare la tua richiesta attingendo al prossimo pool più economico. Per soddisfare la capacità desiderata, potresti ricevere istanze spot da più pool rispetto al numero N specificato. Allo stesso modo, se la maggior parte dei pool non dispone di capacità Spot, è possibile ricevere la capacità desiderata completa da un numero N di pool inferiore a quello specificato.

Note

Se si configurano le istanze Spot per l'avvio con [AMD SEV-SNP](#) attivato, viene addebitata una tariffa di utilizzo oraria aggiuntiva equivalente al 10% della [tariffa oraria on demand](#) del tipo di istanza selezionato. Se la strategia di allocazione utilizza il prezzo come input, il dimensionamento automatico Amazon EC2 non include questa tariffa aggiuntiva; viene utilizzato solo il prezzo spot.

Istanze on demand

Amazon EC2 Auto Scaling fornisce le seguenti strategie di allocazione utilizzabili per le istanze on demand:

lowest-price

Dimensionamento automatico Amazon EC2 implementa automaticamente il tipo di istanza più economico in ciascuna zona di disponibilità in base al prezzo on demand corrente.

Per garantire che la capacità desiderata sia soddisfatta, puoi ricevere più tipi di istanze on demand in ciascuna zona di disponibilità, a seconda della quantità di capacità richiesta.

prioritized

Al fine di soddisfare la capacità on demand, Dimensionamento automatico Amazon EC2 utilizza l'ordine dei tipi di istanza nell'elenco delle sostituzioni dei modelli di avvio per determinare quale tipo di istanza utilizzare per prima. Ad esempio, supponiamo che siano state specificate tre sostituzioni del modello di avvio nell'ordine seguente: `c5.large`, `c4.large` e `c3.large`. Quando le istanze on demand vengono avviate, il gruppo con dimensionamento automatico soddisfa la capacità on demand a partire da `c5.large`, `c4.large` e infine `c3.large`.

Quando gestisci l'ordine di priorità delle istanze on demand, tieni presente quanto segue:

- È possibile pagare per l'utilizzo in anticipo per ottenere sconti significativi sulle istanze on demand utilizzando i Savings Plans o le istanze riservate. Per ulteriori informazioni, consulta la pagina [Prezzi di Amazon EC2](#).
- Con le istanze riservate, si applica la tariffa scontata del normale prezzo delle istanze on demand, se Amazon EC2 Auto Scaling avvia tipi di istanza corrispondenti. Quindi, se hai istanze riservate inutilizzate per `c4.large`, puoi impostare la priorità del tipo di istanza in modo da assegnare la priorità più alta per le istanze riservate a un tipo di istanza `c4.large`. Quando viene avviata un'istanza `c4.large`, vengono addebitati i prezzi delle istanze riservate.
- Con i Savings Plan, quando utilizzi Instance Savings Plans Amazon EC2 o Savings Plans per il calcolo, si applica la tariffa scontata del normale prezzo dell'istanza on demand. Con Savings Plans, hai una maggiore flessibilità nell'assegnare priorità ai tipi di istanza. Se utilizzi i tipi di istanza coperti dal tuo Savings Plan, puoi impostarli in qualsiasi ordine di priorità. Puoi anche modificare occasionalmente l'intero ordine dei tipi di istanze, pur continuando a ricevere la tariffa scontata Savings Plan. Per ulteriori informazioni sui Savings Plans, consulta la [Guida per l'utente di Savings Plans](#).

Come funzionano le strategie di allocazione con i pesi

Quando specificate il `WeightedCapacity` parametro nelle sostituzioni (`"DesiredCapacityType": "vcpu"` o `"DesiredCapacityType": "memory-mib"` a livello di gruppo), le strategie di allocazione funzionano esattamente come per gli altri gruppi di Auto Scaling.

L'unica differenza è che quando scegli la `price-capacity-optimized` strategia `lowest-price` o `or`, le istanze provengono dai pool di istanze con il prezzo unitario più basso in ogni zona di disponibilità. Per ulteriori informazioni, consulta [Configurare un gruppo Auto Scaling per utilizzare i pesi delle istanze](#).

Ad esempio, immagina di disporre di un gruppo con dimensionamento automatico con diversi tipi di istanza con diverse quantità di vCPU. Usi `lowest-price` per le tue strategie di allocazione Spot e on demand. Se scegli di assegnare pesi in base al numero di vCPU di ciascun tipo di istanza, Amazon EC2 Auto Scaling avvia i tipi di istanza con il prezzo più basso per i valori di peso assegnati (ad esempio, per vCPU) al momento dell'adempimento. Se si tratta di un'istanza Spot, significa il prezzo per l'istanza Spot più basso per vCPU. Se si tratta di un'istanza on demand, significa il prezzo per l'istanza on demand più basso per vCPU.

Crea un gruppo di istanze miste utilizzando la selezione del tipo di istanza basata su attributi

In alternativa alla scelta manuale dei tipi di istanza di un gruppo di istanze miste, puoi specificare un set di attributi dell'istanza che descrivono i requisiti di calcolo. Quando Dimensionamento automatico Amazon EC2 avvia istanze, qualsiasi tipo di istanza utilizzato dal gruppo con scalabilità automatica deve rispettare agli attributi dell'istanza richiesti. Questa opzione è nota come selezione del tipo di istanza basata su attributi.

Questo approccio è ideale per carichi di lavoro e framework che possono essere flessibili su quali tipi di istanza utilizzano, come ad esempio container, big data e CI/CD.

Di seguito sono elencati i vantaggi della selezione del tipo di istanza basata su attributi:

- Flessibilità ottimale per le istanze Spot: Amazon EC2 Auto Scaling può scegliere tra un'ampia gamma di tipi di istanze per il lancio di istanze Spot. Ciò soddisfa la best practice delle istanze Spot che indica di essere flessibili sui tipi di istanza, il che offre al servizio Amazon EC2 Spot maggiori possibilità di trovare e allocare la quantità necessaria di capacità di calcolo.

- Usa facilmente i tipi di istanza giusti: con così tanti tipi di istanze disponibili, trovare i tipi di istanza giusti per il tuo carico di lavoro può richiedere molto tempo. Quando si specificano gli attributi dell'istanza, i tipi di istanza avranno automaticamente gli attributi richiesti per il carico di lavoro.
- Uso automatico di nuovi tipi di istanze: i gruppi di Auto Scaling possono utilizzare tipi di istanze di nuova generazione non appena vengono rilasciati. I tipi di istanza di nuova generazione vengono utilizzati automaticamente quando soddisfano i requisiti e si allineano alle strategie di allocazione scelte per il gruppo con scalabilità automatica.

Argomenti

- [Come funziona la selezione del tipo di istanza basata su attributi](#)
- [Protezione del prezzo](#)
- [Prerequisiti](#)
- [Crea un gruppo di istanze miste con la selezione del tipo di istanza basata sugli attributi \(console\)](#)
- [Crea un gruppo di istanze miste con la selezione del tipo di istanza basata sugli attributi \(AWS CLI\)](#)
- [Configurazione di esempio](#)
- [Visualizza in anteprima i tipi di istanza](#)
- [Risorse correlate](#)

Come funziona la selezione del tipo di istanza basata su attributi

Con la selezione del tipo di istanza basata sugli attributi, anziché fornire un elenco di tipi di istanza specifici, è possibile fornire un elenco di attributi di istanza richiesti dalle istanze, ad esempio:

- Numero di vCPU: il numero minimo e massimo di vCPU per istanza.
- Memoria: il numero minimo e massimo GiBs di memoria per istanza.
- Archiviazione locale: se utilizzare EBS o i volumi di Instance Store per l'archiviazione locale.
- Prestazioni affidabili: se utilizzare la famiglia di istanze T, inclusi i tipi T4g, T3a, T3 e T2.

Sono disponibili molte opzioni per definire i requisiti dell'istanza. Per una descrizione di ciascuna opzione e dei valori predefiniti, consulta il riferimento [InstanceRequirements](#) all'API Amazon EC2 Auto Scaling.

Quando il gruppo Auto Scaling deve avviare un'istanza, cercherà i tipi di istanza che corrispondono agli attributi specificati e sono disponibili in quella zona di disponibilità. La strategia di allocazione

determina quindi quali tipi di istanza corrispondenti avviare. Per impostazione predefinita, la selezione del tipo di istanza basata sugli attributi ha una funzione di protezione del prezzo abilitata per impedire al gruppo Auto Scaling di lanciare tipi di istanze che superano le soglie di budget.

Per impostazione predefinita, si utilizza il numero di istanze come unità di misura per impostare la capacità desiderata del gruppo Auto Scaling, il che significa che ogni istanza conta come un'unità.

Alternativamente, è possibile impostare il valore della capacità desiderata come numero di vCPU o quantità di memoria. A tale scopo, utilizzate il campo a discesa Tipo di capacità desiderato AWS Management Console o la `DesiredCapacityType` proprietà nell'operazione `CreateAutoScalingGroup` o `UpdateAutoScalingGroup` API. Amazon EC2 Auto Scaling avvia quindi il numero di istanze necessarie per soddisfare la vCPU o la capacità di memoria desiderata. Ad esempio, se si utilizzano vCPU come tipo di capacità desiderato e si utilizzano istanze con 2 vCPU ciascuna, una capacità desiderata di 10 vCPU lancerebbe 5 istanze. Si tratta di un'alternativa utile ai [pesi delle istanze](#).

Protezione del prezzo

Con la protezione del prezzo, puoi specificare il prezzo massimo che sei disposto a pagare per le istanze EC2 lanciate dal tuo gruppo Auto Scaling. La protezione del prezzo è una funzionalità che impedisce al gruppo Auto Scaling di utilizzare tipi di istanze che considereresti troppo costosi anche se corrispondono agli attributi specificati.

La protezione del prezzo è abilitata per impostazione predefinita e prevede soglie di prezzo separate per le istanze On-Demand e le istanze Spot. Quando Amazon EC2 Auto Scaling deve lanciare nuove istanze, qualsiasi tipo di istanza con un prezzo superiore alla soglia pertinente non viene lanciato.

Argomenti

- [Protezione dei prezzi su richiesta](#)
- [Protezione del prezzo spot](#)
- [Personalizza la protezione del prezzo](#)

Protezione dei prezzi su richiesta

Per le istanze On-Demand, definisci il prezzo On-Demand massimo che sei disposto a pagare come percentuale superiore a un prezzo On-Demand identificato. Il prezzo On-Demand identificato è il prezzo del tipo di istanza C, M o R dell'attuale generazione con il prezzo più basso con gli attributi specificati.

Se un valore di protezione del prezzo On-Demand non è definito in modo esplicito, verrà utilizzato un prezzo On-Demand massimo predefinito del 20% superiore al prezzo On-Demand identificato.

Protezione del prezzo spot

Per impostazione predefinita, Amazon EC2 Auto Scaling applicherà automaticamente una protezione ottimale del prezzo delle istanze Spot per scegliere in modo coerente tra un'ampia gamma di tipi di istanze. Puoi anche impostare manualmente la protezione del prezzo. Tuttavia, lasciare che Amazon EC2 Auto Scaling lo faccia per te può aumentare la probabilità che la tua capacità Spot venga soddisfatta.

Puoi specificare manualmente la protezione del prezzo utilizzando una delle seguenti opzioni. Se imposti manualmente la protezione del prezzo, ti consigliamo di utilizzare la prima opzione.

- Una percentuale di un prezzo On-Demand identificato: il prezzo On-Demand identificato è il prezzo del tipo di istanza C, M o R dell'attuale generazione con il prezzo più basso con gli attributi specificati.
- Una percentuale superiore a un prezzo Spot identificato: il prezzo Spot identificato è il prezzo del tipo di istanza C, M o R dell'attuale generazione con il prezzo più basso con gli attributi specificati. Non consigliamo di utilizzare questa opzione perché i prezzi Spot possono variare e pertanto anche la soglia di protezione del prezzo potrebbe variare.

Personalizza la protezione del prezzo

Puoi personalizzare le soglie di protezione del prezzo nella console Amazon EC2 Auto Scaling o utilizzando gli SDK o. AWS CLI

- Nella console, utilizza le impostazioni di protezione del prezzo On-Demand e Protezione del prezzo Spot negli attributi aggiuntivi dell'istanza.
- Nella [InstanceRequirements](#) struttura, per specificare la soglia di protezione del prezzo delle istanze On-Demand, utilizza la `OnDemandMaxPricePercentageOverLowestPrice` proprietà. Per specificare la soglia di protezione del prezzo dell'istanza Spot, utilizza la proprietà `MaxSpotPriceAsPercentageOfOptimalOnDemandPrice` o la `SpotMaxPricePercentageOverLowestPrice` proprietà.

Se si imposta `DesiredCapacityType` su `vCPU` o `Memory GiB`, la protezione del prezzo si applica in base al prezzo per vCPU o per memoria anziché al prezzo per istanza.

Puoi anche disattivare la protezione del prezzo. Per indicare l'assenza di una soglia di protezione del prezzo, specifica un valore percentuale elevato, ad esempio 999999.

Note

Se nessun tipo di istanza C, M o R dell'attuale generazione corrisponde agli attributi specificati, la protezione del prezzo è comunque applicabile. Se non viene trovata alcuna corrispondenza, il prezzo identificato proviene dai tipi di istanza della generazione corrente con il prezzo più basso o, in mancanza, dai tipi di istanza della generazione precedente con il prezzo più basso, che corrispondono ai tuoi attributi.

Prerequisiti

- Creazione di un modello di avvio. Per ulteriori informazioni, consulta [Creare un modello di avvio per un gruppo con dimensionamento automatico](#).
- Assicurati che il modello di avvio non richieda già istanze Spot.

Crea un gruppo di istanze miste con la selezione del tipo di istanza basata sugli attributi (console)

Usa la procedura seguente per creare un gruppo di istanze miste utilizzando la selezione del tipo di istanza basata su attributi. Per aiutarti a eseguire i passaggi in modo efficiente, alcune sezioni facoltative vengono saltate.


Per la maggior parte dei carichi per scopi generici, è sufficiente specificare il numero di vCPU e di memoria necessari. Per casi d'uso avanzati, è possibile specificare attributi come tipo di archiviazione, interfacce di rete, produttore della CPU e tipo di acceleratore.

Per esaminare le best practice per un gruppo di istanze miste, consulta [Panoramica della configurazione](#)

Come creare gruppi di istanze

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Sulla barra di navigazione nella parte superiore della schermata, scegli la stessa Regione AWS che utilizzata durante la creazione del modello di avvio.
3. Selezionare Crea un gruppo con dimensionamento automatico.

4. Nella pagina Scegli il modello o la configurazione di avvio, in Nome gruppo con dimensionamento automatico, inserisci un nome per il gruppo con dimensionamento automatico.
5. Per scegliere il modello di avvio, completa le seguenti operazioni:
 - a. Come Launch template (Modello di avvio) scegli un modello di avvio esistente.
 - b. Per la Versione del modello di avvio, scegli se durante il dimensionamento orizzontale il gruppo con dimensionamento automatico debba usare la versione di default, quella più recente o una versione specifica del modello di avvio.
 - c. Verificare che il modello di avvio supporti tutte le opzioni che si intende utilizzare e quindi scegliere Next (Successivo).
6. Nella pagina Scegli le opzioni di avvio istanza, procedi come segue:
 - a. Per Instance type requirements (Requisiti del tipo di istanza), scegli Override launch template (Sostituisci modello di avvio).

 Note

Se scegli un modello di avvio che contiene già un set di attributi di istanza, come vCPU e memoria, vengono visualizzati gli attributi dell'istanza. Questi attributi vengono aggiunti alle proprietà del gruppo con dimensionamento automatico, dove puoi aggiornarli dalla console Amazon EC2 in qualsiasi momento.

- b. In Specify instance attributes (Specifica attributi dell'istanza), inizia inserendo le vCPU e i requisiti di memoria.
 - Per vCPUs (vCPU) inserire il numero minimo e massimo desiderato di vCPU. Per non specificare alcun limite, selezionare No minimum (Nessun minimo), No maximum (Nessun massimo) o entrambe le opzioni.
 - Per Memory (GiB) (Memoria [GiB]) inserire la quantità minima e massima di memoria desiderata. Per non specificare alcun limite, selezionare No minimum (Nessun minimo), No maximum (Nessun massimo) o entrambe le opzioni.
- c. (Facoltativo) Per Additional instance attributes (Attributi istanza aggiuntivi), facoltativamente, è possibile specificare uno o più attributi per esprimere i requisiti di calcolo in modo più dettagliato. Ogni attributo aggiuntivo aggiunge ulteriori vincoli alla propria richiesta.
- d. Espandete Anteprema dei tipi di istanze corrispondenti per visualizzare i tipi di istanza con gli attributi specificati.

- e. In Opzioni di acquisto delle istanze, per Distribuzione delle istanze, specifica le percentuali del gruppo da avviare rispettivamente come istanze on demand e istanze spot. Se l'applicazione è senza stato, tollerante ai guasti e può gestire un'istanza interrotta, è possibile specificare una percentuale più alta di istanze Spot.
 - f. (Facoltativo) Quando specifichi una percentuale per le istanze spot, seleziona Includi capacità di base on demand e specifica quindi la quantità minima di capacità iniziale del gruppo con dimensionamento automatico che deve essere raggiunta dalle istanze on demand. Qualsiasi cosa oltre la capacità di base utilizza le impostazioni Instances distribution (Distribuzione delle istanze) per stabilire quante istanze on demand e istanze Spot avviare.
 - g. In Allocation strategies (Strategie di allocazione), Lowest price (Prezzo più basso) viene selezionato automaticamente per On-Demand allocation strategy (Strategia di allocazione on demand) e non può essere cambiato.
 - h. Per Spot allocation strategy (Strategia di allocazione Spot), scegli una strategia di allocazione. L'opzione Price capacity optimized (Ottimizzazione per prezzo e capacità) è selezionata per impostazione predefinita. Per impostazione predefinita, l'opzione Lowest price (Prezzo più basso) è nascosta e viene visualizzata solo quando scegli Show all strategies (Mostra tutte le strategie). Se hai scelto Prezzo più basso inserisci il numero di pool con il prezzo più basso da diversificare tra i Pool con prezzi più bassi.
 - i. Per il ribilanciamento della capacità, scegli se abilitare o disabilitare il ribilanciamento della capacità. Utilizza il ribilanciamento della capacità per rispondere automaticamente quando le tue istanze spot stanno per terminare a causa di un'interruzione spot. Per ulteriori informazioni, consulta [Utilizza il ribilanciamento della capacità per gestire le interruzioni spot di Amazon EC2](#).
 - j. In Rete, per VPC, scegli un VPC. Il gruppo con dimensionamento automatico deve essere creato nello stesso VPC del gruppo di sicurezza specificato nel modello di avvio.
 - k. Per Availability Zones and subnets (Zone di disponibilità e sottoreti), scegli una o più sottoreti nel VPC specificato. Per un'elevata disponibilità, utilizza sottoreti in più zone di disponibilità. Per ulteriori informazioni, consulta [Considerazioni sulla scelta di sottoreti VPC](#).
 - l. Scegli Avanti, Avanti.
7. Per la fase Configure group size and scaling policies (Configura le dimensioni dei gruppi e le policy di scalabilità), procedi come riportato di seguito:
- a. Per misurare la capacità desiderata in unità diverse dalle istanze, scegli l'opzione appropriata per Dimensioni gruppo, Tipo di capacità desiderata. Sono supportati Units

- (Unità), vCPUs (vCPU) e Memory GiB (GiB di memoria). Per impostazione predefinita, Dimensionamento automatico Amazon EC2 specifica Units (Unità), che si traduce in un numero di istanze.
- b. Per capacità desiderata, le dimensioni iniziali del tuo gruppo con dimensionamento automatico.
 - c. Nella sezione Dimensionamento, in Limiti di dimensionamento, se il nuovo valore per la capacità desiderata è maggiore della capacità minima desiderata e della capacità massima desiderata, la capacità massima desiderata viene automaticamente aumentata fino al nuovo valore di capacità desiderato. È possibile modificare questi limiti se necessario. Per ulteriori informazioni, consulta [Impostare i limiti di dimensionamento per il gruppo con dimensionamento automatico](#).
8. Selezionare Passa a revisione.
 9. Nella pagina Verifica, scegliere Crea gruppo con dimensionamento automatico.

Crea un gruppo di istanze miste con la selezione del tipo di istanza basata sugli attributi (AWS CLI)

Per creare un gruppo di istanze miste tramite la riga di comando

Utilizzare uno dei seguenti comandi:

- [create-auto-scaling-group](#) (AWS CLI)
- [AWS Tools for Windows PowerShell Nuovo-come \(\) AutoScalingGroup](#)

Configurazione di esempio

Per creare un gruppo Auto Scaling con selezione del tipo di istanza basata sugli attributi utilizzando il AWS CLI, utilizzare il comando seguente. [create-auto-scaling-group](#)

Vengono specificati i seguenti attributi di istanza:

- VCpuCount: i tipi di istanza devono avere un minimo di quattro e un massimo di otto vCPU.
- MemoryMiB: i tipi di istanza devono avere un minimo di 16.384 MiB di memoria.
- CpuManufacturers: i tipi di istanza devono avere una CPU prodotta da Intel.

JSON

```
aws autoscaling create-auto-scaling-group --cli-input-json file://~/config.json
```

Di seguito è riportato un esempio del file config.json.

```
{
  "AutoScalingGroupName": "my-asg",
  "DesiredCapacityType": "units",
  "MixedInstancesPolicy": {
    "LaunchTemplate": {
      "LaunchTemplateSpecification": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "$Default"
      },
      "Overrides": [{
        "InstanceRequirements": {
          "VCpuCount": {"Min": 4, "Max": 8},
          "MemoryMiB": {"Min": 16384},
          "CpuManufacturers": ["intel"]
        }
      }]
    },
    "InstancesDistribution": {
      "OnDemandPercentageAboveBaseCapacity": 50,
      "SpotAllocationStrategy": "price-capacity-optimized"
    }
  },
  "MinSize": 0,
  "MaxSize": 100,
  "DesiredCapacity": 4,
  "DesiredCapacityType": "units",
  "VPCZoneIdentifier": "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782"
}
```

Per impostare il valore della capacità desiderata come numero di vCPU o quantità di memoria, specifica "DesiredCapacityType": "vcpu" o "DesiredCapacityType": "memory-mib" nel file. Il tipo di capacità di default desiderato è units, che imposta il valore della capacità desiderata come numero di istanze.

YAML

In alternativa, è possibile utilizzare il seguente [create-auto-scaling-group](#) comando per creare il gruppo Auto Scaling. Questo fa riferimento a un file YAML, non a un file JSON, come unico parametro per il gruppo con dimensionamento automatico.

```
aws autoscaling create-auto-scaling-group --cli-input-yaml file://~/config.yaml
```

Di seguito è riportato un esempio del file `config.yaml`.

```
---
AutoScalingGroupName: my-asg
DesiredCapacityType: units
MixedInstancesPolicy:
  LaunchTemplate:
    LaunchTemplateSpecification:
      LaunchTemplateName: my-launch-template
      Version: $Default
    Overrides:
      - InstanceRequirements:
          VCpuCount:
            Min: 2
            Max: 4
          MemoryMiB:
            Min: 2048
          CpuManufacturers:
            - intel
      InstancesDistribution:
        OnDemandPercentageAboveBaseCapacity: 50
        SpotAllocationStrategy: price-capacity-optimized
  MinSize: 0
  MaxSize: 100
  DesiredCapacity: 4
DesiredCapacityType: units
VPCZoneIdentifier: subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782
```

Per impostare il valore della capacità desiderata come numero di vCPU o quantità di memoria, specifica `DesiredCapacityType: vcpu` o `DesiredCapacityType: memory-mib` nel file. Il tipo di capacità di default desiderato è `units`, che imposta il valore della capacità desiderata come numero di istanze.

Visualizza in anteprima i tipi di istanza

È possibile visualizzare in anteprima i tipi di istanza che corrispondono ai requisiti di calcolo senza avviarli e modificare i requisiti, se necessario. Quando crei il gruppo con scalabilità automatica nella console Dimensionamento automatico Amazon EC2, viene visualizzata un'anteprima dei tipi di istanza nella sezione Preview matching instance types (Anteprima tipi di istanza corrispondenti) alla pagina Choose instance launch options (Scelta delle opzioni di avvio dell'istanza).

In alternativa, puoi visualizzare in anteprima i tipi di istanza effettuando una chiamata [GetInstanceTypesFromInstanceRequirements](#) API Amazon EC2 utilizzando AWS CLI o un SDK. Trasferisci i parametri InstanceRequirements nella richiesta nel formato esatto da utilizzare per creare o aggiornare un gruppo con scalabilità automatica. Per ulteriori informazioni, consulta [Preview instance types with specified attributes](#) (Anteprima di tipi di istanza con attributi specificati) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Risorse correlate

Per ulteriori informazioni sulla selezione del tipo di istanza basata sugli attributi, consulta Selezione del tipo di [istanza basata sugli attributi per EC2 Auto Scaling and EC2 Fleet](#) sul blog. AWS

Puoi dichiarare la selezione del tipo di istanza basata su attributi quando crei un gruppo con dimensionamento automatico usando AWS CloudFormation. Per ulteriori informazioni, consulta lo snippet di esempio nella sezione [Snippet del modello di dimensionamento automatico](#) della Guida per l'utente AWS CloudFormation .

Crea un gruppo di istanze miste scegliendo manualmente i tipi di istanze

Questo argomento mostra come avviare più tipi di istanza in un singolo gruppo con dimensionamento automatico scegliendo manualmente i tipi di istanza.

Se preferisci utilizzare gli attributi dell'istanza come criteri per la selezione dei tipi di istanza, consulta [Crea un gruppo di istanze miste utilizzando la selezione del tipo di istanza basata su attributi](#).

Indice

- [Prerequisiti](#)
- [Crea un gruppo di istanze miste \(console\)](#)
- [Crea un gruppo di istanze miste \(AWS CLI\)](#)
- [Configurazioni di esempio](#)

Prerequisiti

- Creazione di un modello di avvio. Per ulteriori informazioni, consulta [Creare un modello di avvio per un gruppo con dimensionamento automatico](#).
- Assicurati che il modello di avvio non richieda già istanze Spot.

Crea un gruppo di istanze miste (console)

Utilizza la procedura seguente per creare un gruppo di istanze miste scegliendo manualmente i tipi di istanza che il gruppo può avviare. Per aiutarti a eseguire i passaggi in modo efficiente, alcune sezioni facoltative vengono saltate.

Per esaminare le best practice per un gruppo di istanze miste, consulta [Panoramica della configurazione](#)

Come creare gruppi di istanze

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Sulla barra di navigazione nella parte superiore della schermata, scegli la stessa Regione AWS che utilizzata durante la creazione del modello di avvio.
3. Selezionare Crea un gruppo con dimensionamento automatico.
4. Nella pagina Scegli il modello o la configurazione di avvio, in Nome gruppo con dimensionamento automatico, inserisci un nome per il gruppo con dimensionamento automatico.
5. Per scegliere il modello di avvio, completa le seguenti operazioni:
 - a. Come Launch template (Modello di avvio) scegli un modello di avvio esistente.
 - b. Per la Versione del modello di avvio, scegli se durante il dimensionamento orizzontale il gruppo con dimensionamento automatico debba usare la versione di default, quella più recente o una versione specifica del modello di avvio.
 - c. Verificare che il modello di avvio supporti tutte le opzioni che si intende utilizzare e quindi scegliere Next (Successivo).
6. Nella pagina Scegli le opzioni di avvio istanza, procedi come segue:
 - a. Per Instance type requirements (Requisiti del tipo di istanza), scegli Override launch template (Sostituisci modello di avvio) e quindi Manually add instance types (Aggiungi manualmente tipi di istanza).

- b. Scegli i tipi di istanza. Puoi usare i nostri suggerimenti come punto di partenza. L'opzione Family and generation flexible (Flessibile per famiglia e generazione) è selezionata per impostazione predefinita.
- Per modificare l'ordine dei tipi di istanza, utilizzare le frecce. Se scegli una strategia di allocazione basata sulla priorità, l'ordine del tipo di istanza imposta la priorità di avvio.
 - Per rimuovere un tipo di istanza, scegli X.
 - (Facoltativo), puoi assegnare a ciascun tipo di istanza un peso relativo utilizzando le caselle della colonna Peso. A tale scopo, inserisci il numero di unità che un'istanza di quel tipo deve contare rispetto alla capacità desiderata del gruppo. Ciò potrebbe essere utile se istanze tipi di istanza offrono diverse funzionalità di vCPU, memoria, archiviazione o larghezza di banda della rete. Per ulteriori informazioni, consulta [Configurare un gruppo Auto Scaling per utilizzare i pesi delle istanze](#).

Considera che se hai scelto di utilizzare i consigli Flessibile per dimensioni, tutti i tipi di istanze che fanno parte di questa sezione hanno automaticamente un valore di ponderazione. Se non desideri specificare alcun valore, deseleziona le caselle nella colonna Weight (Peso) per tutti i tipi di istanze.

- c. In Instance purchase options (Opzioni di acquisto delle istanze), per Instances distribution (Distribuzione delle istanze), specifica le percentuali del gruppo da avviare rispettivamente come istanze on demand e istanze spot. Se l'applicazione è senza stato, tollerante ai guasti e può gestire un'istanza interrotta, è possibile specificare una percentuale più alta di istanze Spot.
- d. (Facoltativo) Quando specifichi una percentuale per le istanze spot, seleziona Includi capacità di base on demand e specifica quindi la quantità minima di capacità iniziale del gruppo con dimensionamento automatico che deve essere raggiunta dalle istanze on demand. Qualsiasi cosa oltre la capacità di base utilizza le impostazioni Instances distribution (Distribuzione delle istanze) per stabilire quante istanze on demand e istanze Spot avviare.
- e. In Allocation strategies (Strategie di allocazione), per On-Demand allocation strategy (Strategia di allocazione on demand), scegli una strategia di allocazione. Quando scegli manualmente i tipi di istanza, l'opzione Prioritized (Priorità) è selezionata per impostazione predefinita.
- f. Per Spot allocation strategy (Strategia di allocazione Spot), scegli una strategia di allocazione. L'opzione Price capacity optimized (Ottimizzazione per prezzo e capacità) è selezionata per impostazione predefinita. Per impostazione predefinita, l'opzione Lowest

price (Prezzo più basso) è nascosta e viene visualizzata solo quando scegli Show all strategies (Mostra tutte le strategie).

- Se hai scelto Prezzo più basso inserisci il numero di pool con il prezzo più basso da diversificare tra i Pool con prezzi più bassi.
 - Puoi scegliere Capacità ottimizzata, puoi facoltativamente selezionare la casella Dai priorità ai tipi di istanza per far sì che Dimensionamento automatico Amazon EC2 scelga il tipo di istanza da avviare per primo in base all'ordine in cui sono elencati i tipi di istanza.
- g. Per il ribilanciamento della capacità, scegli se abilitare o disabilitare il ribilanciamento della capacità. Utilizza il ribilanciamento della capacità per rispondere automaticamente quando le tue istanze spot stanno per terminare a causa di un'interruzione spot. Per ulteriori informazioni, consulta [Utilizza il ribilanciamento della capacità per gestire le interruzioni spot di Amazon EC2](#).
 - h. In Rete, per VPC, scegli un VPC. Il gruppo con dimensionamento automatico deve essere creato nello stesso VPC del gruppo di sicurezza specificato nel modello di avvio.
 - i. Per Availability Zones and subnets (Zone di disponibilità e sottoreti), scegli una o più sottoreti nel VPC specificato. Per un'elevata disponibilità, utilizza sottoreti in più zone di disponibilità. Per ulteriori informazioni, consulta [Considerazioni sulla scelta di sottoreti VPC](#).
 - j. Scegli Avanti, Avanti.
7. Per la fase Configure group size and scaling policies (Configura le dimensioni dei gruppi e le policy di scalabilità), procedi come riportato di seguito:
 - a. In Dimensioni gruppo, per Capacità desiderata, immettere il numero iniziale di istanze da avviare.

Per impostazione predefinita, la capacità desiderata è espressa come numero di istanze. Se hai assegnato dei pesi ai tuoi tipi di istanze, devi convertire questo valore nella stessa unità di misura utilizzata per assegnare i pesi, ad esempio il numero di vCPU.
 - b. Nella sezione Dimensionamento, in Limiti di dimensionamento, se il nuovo valore per la capacità desiderata è maggiore della capacità minima desiderata e della capacità massima desiderata, la capacità massima desiderata viene automaticamente aumentata fino al nuovo valore di capacità desiderato. È possibile modificare questi limiti se necessario. Per ulteriori informazioni, consulta [Impostare i limiti di dimensionamento per il gruppo con dimensionamento automatico](#).
 8. Selezionare Passa a revisione.
 9. Nella pagina Verifica, scegliere Crea gruppo con dimensionamento automatico.

Crea un gruppo di istanze miste (AWS CLI)

Per creare un gruppo di istanze miste tramite la riga di comando

Utilizzare uno dei seguenti comandi:

- [create-auto-scaling-group](#) (AWS CLI)
- [Nuovi come \(AutoScalingGroup\)](#) AWS Tools for Windows PowerShell

Configurazioni di esempio

Le seguenti configurazioni di esempio mostrano come creare gruppi di istanze miste utilizzando le diverse strategie di allocazione Spot.

Note

Questi esempi mostrano come utilizzare un file di configurazione formattato in JSON o YAML. Se si utilizza la AWS CLI versione 1, è necessario specificare un file di configurazione in formato JSON. Se si utilizza la AWS CLI versione 2, è possibile specificare un file di configurazione formattato in YAML o JSON.

Esempi

- [Esempio 1: avvio di Istanze Spot con la strategia di allocazione capacity-optimized](#)
- [Esempio 2: avvio di Istanze Spot con la strategia di allocazione capacity-optimized-prioritized](#)
- [Esempio 3: avvio di istanze Spot utilizzando la strategia di allocazione lowest-price diversificata su due pool](#)
- [Esempio 4: Lancio di Istanze spot con la strategia di allocazione price-capacity-optimized](#)

Esempio 1: avvio di Istanze Spot con la strategia di allocazione **capacity-optimized**

Il [create-auto-scaling-group](#) comando seguente crea un gruppo Auto Scaling che specifica quanto segue:

- La percentuale del gruppo da avviare come istanze on demand (0) e il numero di base di istanze on demand con cui iniziare (1).
- I tipi di istanza da avviare in ordine di priorità (c5.large, c5a.large, m5.large, m5a.large, c4.large, m4.large, c3.large, m3.large).

- Le sottoreti in cui avviare le istanze (subnet-5ea0c127, subnet-6194ea3b, subnet-c934b782). Ognuna di queste corrisponde a una zona di disponibilità diversa.
- Il modello di avvio (my-launch-template) e la versione del modello di avvio (\$Default)

Quando Amazon EC2 Auto Scaling tenta di soddisfare la capacità on demand, avvia prima il tipo di istanza c5.large. Le istanze Spot provengono dal pool Spot ottimale in ciascuna zona di disponibilità in base alla capacità delle istanze Spot.

JSON

```
aws autoscaling create-auto-scaling-group --cli-input-json file://~/config.json
```

Il file config.json contiene il seguente contenuto.

```
{
  "AutoScalingGroupName": "my-asg",
  "MixedInstancesPolicy": {
    "LaunchTemplate": {
      "LaunchTemplateSpecification": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "$Default"
      },
      "Overrides": [
        {
          "InstanceType": "c5.large"
        },
        {
          "InstanceType": "c5a.large"
        },
        {
          "InstanceType": "m5.large"
        },
        {
          "InstanceType": "m5a.large"
        },
        {
          "InstanceType": "c4.large"
        },
        {
          "InstanceType": "m4.large"
        }
      ]
    }
  }
}
```

```

        {
            "InstanceType": "c3.large"
        },
        {
            "InstanceType": "m3.large"
        }
    ]
},
"InstancesDistribution": {
    "OnDemandBaseCapacity": 1,
    "OnDemandPercentageAboveBaseCapacity": 0,
    "SpotAllocationStrategy": "capacity-optimized"
}
},
"MinSize": 1,
"MaxSize": 5,
"DesiredCapacity": 3,
"VPCZoneIdentifier": "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782"
}

```

YAML

In alternativa, è possibile utilizzare il seguente [create-auto-scaling-group](#) comando per creare il gruppo Auto Scaling. Questo fa riferimento a un file YAML, non a un file JSON, come unico parametro per il gruppo con dimensionamento automatico.

```
aws autoscaling create-auto-scaling-group --cli-input-yaml file://~/config.yaml
```

Il file `config.yaml` contiene il seguente contenuto.

```

---
AutoScalingGroupName: my-asg
MixedInstancesPolicy:
  LaunchTemplate:
    LaunchTemplateSpecification:
      LaunchTemplateName: my-launch-template
      Version: $Default
    Overrides:
      - InstanceType: c5.large
      - InstanceType: c5a.large
      - InstanceType: m5.large
      - InstanceType: m5a.large
      - InstanceType: c4.large

```



```
- InstanceType: m4.large
- InstanceType: c3.large
- InstanceType: m3.large
InstancesDistribution:
  OnDemandBaseCapacity: 1
  OnDemandPercentageAboveBaseCapacity: 0
  SpotAllocationStrategy: capacity-optimized
MinSize: 1
MaxSize: 5
DesiredCapacity: 3
VPCZoneIdentifier: subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782
```

Esempio 2: avvio di Istanze Spot con la strategia di allocazione **capacity-optimized-prioritized**

Il [create-auto-scaling-group](#) comando seguente crea un gruppo Auto Scaling che specifica quanto segue:

- La percentuale del gruppo da avviare come istanze on demand (0) e il numero di base di istanze on demand con cui iniziare (1).
- I tipi di istanza da avviare in ordine di priorità (c5.large, c5a.large, m5.large, m5a.large, c4.large, m4.large, c3.large, m3.large).
- Le sottoreti in cui avviare le istanze (subnet-5ea0c127, subnet-6194ea3b, subnet-c934b782). Ognuna di queste corrisponde a una zona di disponibilità diversa.
- Il modello di avvio (my-launch-template) e la versione del modello di avvio (\$Latest)

Quando Amazon EC2 Auto Scaling tenta di soddisfare la capacità on demand, avvia prima il tipo di istanza c5.large. Quando Dimensionamento automatico Amazon EC2 cerca di soddisfare la tua capacità Spot, rispetta le priorità del tipo di istanza sulla base del miglior tentativo. Tuttavia, prima ottimizza la capacità.

JSON

```
aws autoscaling create-auto-scaling-group --cli-input-json file://~/config.json
```

Il file config.json contiene il seguente contenuto.

```
{
  "AutoScalingGroupName": "my-asg",
```

```
"MixedInstancesPolicy": {
  "LaunchTemplate": {
    "LaunchTemplateSpecification": {
      "LaunchTemplateName": "my-launch-template",
      "Version": "$Latest"
    },
    "Overrides": [
      {
        "InstanceType": "c5.large"
      },
      {
        "InstanceType": "c5a.large"
      },
      {
        "InstanceType": "m5.large"
      },
      {
        "InstanceType": "m5a.large"
      },
      {
        "InstanceType": "c4.large"
      },
      {
        "InstanceType": "m4.large"
      },
      {
        "InstanceType": "c3.large"
      },
      {
        "InstanceType": "m3.large"
      }
    ]
  },
  "InstancesDistribution": {
    "OnDemandBaseCapacity": 1,
    "OnDemandPercentageAboveBaseCapacity": 0,
    "SpotAllocationStrategy": "capacity-optimized-prioritized"
  }
},
"MinSize": 1,
"MaxSize": 5,
"DesiredCapacity": 3,
"VPCZoneIdentifier": "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782"
```

```
}
```

YAML

In alternativa, è possibile utilizzare il seguente [create-auto-scaling-group](#) comando per creare il gruppo Auto Scaling. Questo fa riferimento a un file YAML, non a un file JSON, come unico parametro per il gruppo con dimensionamento automatico.

```
aws autoscaling create-auto-scaling-group --cli-input-yaml file://~/config.yaml
```

Il file `config.yaml` contiene il seguente contenuto.

```
---
AutoScalingGroupName: my-asg
MixedInstancesPolicy:
  LaunchTemplate:
    LaunchTemplateSpecification:
      LaunchTemplateName: my-launch-template
      Version: $Default
    Overrides:
      - InstanceType: c5.large
      - InstanceType: c5a.large
      - InstanceType: m5.large
      - InstanceType: m5a.large
      - InstanceType: c4.large
      - InstanceType: m4.large
      - InstanceType: c3.large
      - InstanceType: m3.large
    InstancesDistribution:
      OnDemandBaseCapacity: 1
      OnDemandPercentageAboveBaseCapacity: 0
      SpotAllocationStrategy: capacity-optimized-prioritized
  MinSize: 1
  MaxSize: 5
  DesiredCapacity: 3
  VPCZoneIdentifier: subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782
```

Esempio 3: avvio di istanze Spot utilizzando la strategia di allocazione **lowest-price** diversificata su due pool

Il [create-auto-scaling-group](#) comando seguente crea un gruppo Auto Scaling che specifica quanto segue:

- La percentuale del gruppo da avviare come istanze on demand (50). (Questa impostazione non specifica anche il numero di base di istanze on demand con cui iniziare).
- I tipi di istanza da avviare in ordine di priorità (c5.large, c5a.large, m5.large, m5a.large, c4.large, m4.large, c3.large, m3.large).
- Le sottoreti in cui avviare le istanze (subnet-5ea0c127, subnet-6194ea3b, subnet-c934b782). Ognuna di queste corrisponde a una zona di disponibilità diversa.
- Il modello di avvio (my-launch-template) e la versione del modello di avvio (\$Latest)

Quando Amazon EC2 Auto Scaling tenta di soddisfare la capacità on demand, avvia prima il tipo di istanza c5.large. Per la capacità Spot, Dimensionamento automatico Amazon EC2 tenta di avviare le istanze Spot in modo uniforme nei due pool con il prezzo più basso in ciascuna zona di disponibilità.

JSON

```
aws autoscaling create-auto-scaling-group --cli-input-json file://~/config.json
```

Il file config.json contiene il seguente contenuto.

```
{
  "AutoScalingGroupName": "my-asg",
  "MixedInstancesPolicy": {
    "LaunchTemplate": {
      "LaunchTemplateSpecification": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "$Latest"
      },
      "Overrides": [
        {
          "InstanceType": "c5.large"
        },
        {
          "InstanceType": "c5a.large"
        },
        {
          "InstanceType": "m5.large"
        },
        {
          "InstanceType": "m5a.large"
        }
      ]
    }
  }
}
```

```

    {
      "InstanceType": "c4.large"
    },
    {
      "InstanceType": "m4.large"
    },
    {
      "InstanceType": "c3.large"
    },
    {
      "InstanceType": "m3.large"
    }
  ]
},
"InstancesDistribution": {
  "OnDemandPercentageAboveBaseCapacity": 50,
  "SpotAllocationStrategy": "lowest-price",
  "SpotInstancePools": 2
}
},
"MinSize": 1,
"MaxSize": 5,
"DesiredCapacity": 3,
"VPCZoneIdentifier": "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782"
}

```

YAML

In alternativa, è possibile utilizzare il seguente [create-auto-scaling-group](#) comando per creare il gruppo Auto Scaling. Questo fa riferimento a un file YAML, non a un file JSON, come unico parametro per il gruppo con dimensionamento automatico.

```
aws autoscaling create-auto-scaling-group --cli-input-yaml file://~/config.yaml
```

Il file `config.yaml` contiene il seguente contenuto.

```

---
AutoScalingGroupName: my-asg
MixedInstancesPolicy:
  LaunchTemplate:
    LaunchTemplateSpecification:
      LaunchTemplateName: my-launch-template
      Version: $Default

```

Overrides:

- InstanceType: *c5.large*
- InstanceType: *c5a.large*
- InstanceType: *m5.large*
- InstanceType: *m5a.large*
- InstanceType: *c4.large*
- InstanceType: *m4.large*
- InstanceType: *c3.large*
- InstanceType: *m3.large*

InstancesDistribution:

OnDemandPercentageAboveBaseCapacity: *50*

SpotAllocationStrategy: *lowest-price*

SpotInstancePools: *2*

MinSize: *1*

MaxSize: *5*

DesiredCapacity: *3*

VPCZoneIdentifier: *subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782*

Esempio 4: Lancio di Istanze spot con la strategia di allocazione `price-capacity-optimized`

Il [create-auto-scaling-group](#) comando seguente crea un gruppo Auto Scaling che specifica quanto segue:

- La percentuale del gruppo da avviare come istanze on demand (30). (Questa impostazione non specifica anche il numero di base di istanze on demand con cui iniziare).
- I tipi di istanza da avviare in ordine di priorità (*c5.large*, *c5a.large*, *m5.large*, *m5a.large*, *c4.large*, *m4.large*, *c3.large*, *m3.large*).
- Le sottoreti in cui avviare le istanze (*subnet-5ea0c127*, *subnet-6194ea3b*, *subnet-c934b782*). Ognuna di queste corrisponde a una zona di disponibilità diversa.
- Il modello di avvio (*my-launch-template*) e la versione del modello di avvio (*\$Latest*)

Quando Amazon EC2 Auto Scaling tenta di soddisfare la capacità on demand, avvia prima il tipo di istanza *c5.large*. Per quanto riguarda la capacità spot, Dimensionamento automatico Amazon EC2 tenta di avviare le istanze spot da pool di istanze spot con il prezzo più basso possibile, ma anche con capacità ottimale per il numero di istanze che vengono avviate.

JSON

```
aws autoscaling create-auto-scaling-group --cli-input-json file://~/config.json
```

Il file `config.json` contiene il seguente contenuto.

```
{
  "AutoScalingGroupName": "my-asg",
  "MixedInstancesPolicy": {
    "LaunchTemplate": {
      "LaunchTemplateSpecification": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "$Latest"
      },
      "Overrides": [
        {
          "InstanceType": "c5.large"
        },
        {
          "InstanceType": "c5a.large"
        },
        {
          "InstanceType": "m5.large"
        },
        {
          "InstanceType": "m5a.large"
        },
        {
          "InstanceType": "c4.large"
        },
        {
          "InstanceType": "m4.large"
        },
        {
          "InstanceType": "c3.large"
        },
        {
          "InstanceType": "m3.large"
        }
      ]
    },
    "InstancesDistribution": {
      "OnDemandPercentageAboveBaseCapacity": 30,
      "SpotAllocationStrategy": "price-capacity-optimized"
    }
  },
  "MinSize": 1,
  "MaxSize": 5,
}
```

```
"DesiredCapacity": 3,  
"VPCZoneIdentifier": "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782"  
}
```

YAML

In alternativa, è possibile utilizzare il seguente [create-auto-scaling-group](#) comando per creare il gruppo Auto Scaling. Questo fa riferimento a un file YAML, non a un file JSON, come unico parametro per il gruppo con dimensionamento automatico.

```
aws autoscaling create-auto-scaling-group --cli-input-yaml file://~/config.yaml
```

Il file `config.yaml` contiene il seguente contenuto.

```
---  
AutoScalingGroupName: my-asg  
MixedInstancesPolicy:  
  LaunchTemplate:  
    LaunchTemplateSpecification:  
      LaunchTemplateName: my-launch-template  
      Version: $Default  
    Overrides:  
      - InstanceType: c5.large  
      - InstanceType: c5a.large  
      - InstanceType: m5.large  
      - InstanceType: m5a.large  
      - InstanceType: c4.large  
      - InstanceType: m4.large  
      - InstanceType: c3.large  
      - InstanceType: m3.large  
    InstancesDistribution:  
      OnDemandPercentageAboveBaseCapacity: 30  
      SpotAllocationStrategy: price-capacity-optimized  
  MinSize: 1  
  MaxSize: 5  
  DesiredCapacity: 3  
  VPCZoneIdentifier: subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782
```


Configurare un gruppo Auto Scaling per utilizzare i pesi delle istanze

Quando si utilizzano più tipi di istanza, è possibile specificare quante unità associare a ciascun tipo di istanza e quindi specificare la capacità del gruppo con la stessa unità di misura. Questa opzione di specificazione della capacità è nota come pesi.

Ad esempio, supponiamo che di eseguire un'applicazione ad alta intensità di calcolo che funziona meglio con almeno 8 VCPU e 15 GiB di RAM. Se utilizzi `c5.2xlarge` come unità base, uno dei seguenti tipi di istanza EC2 soddisfa le esigenze dell'applicazione.

Esempio di tipi di istanza

Tipo di istanza	VCPU	Memoria (GiB)
<code>c5.2xlarge</code>	8	16
<code>c5.4xlarge</code>	16	32
<code>c5.12xlarge</code>	48	96
<code>c5.18xlarge</code>	72	144
<code>c5.24xlarge</code>	96	192

Per impostazione predefinita, tutti i tipi di istanze hanno lo stesso peso indipendentemente dalla dimensione. In altre parole, sia che il dimensionamento automatico Amazon EC2 avvii un tipo di istanza grande o piccola, ciascuna conta per la capacità desiderata del gruppo con dimensionamento automatico.

Con i pesi, tuttavia, si assegna un valore numerico che specifica quante unità associare a ciascun tipo di istanza. Ad esempio, se le istanze sono di dimensioni diverse, un'istanza `c5.2xlarge` potrebbe avere il peso di 2, una `c5.4xlarge` (che è due volte più grande) potrebbe avere il peso di 4 e così via. Quindi, quando Dimensionamento automatico Amazon EC2 riduce orizzontalmente il gruppo, questi pesi si traducono nel numero di unità che ciascuna istanza conta per la capacità desiderata.

Il peso non cambia il tipo di istanza che Dimensionamento automatico Amazon EC2 sceglie di avviare, ma le strategie di allocazione lo fanno. Per ulteriori informazioni, consulta [Strategie di allocazione](#).

Important

Per configurare un gruppo con dimensionamento automatico affinché soddisfi la capacità desiderata utilizzando il numero di vCPU o la quantità di memoria di ogni tipo di istanza, consigliamo di utilizzare la selezione del tipo di istanza basata su attributi. L'impostazione del `DesiredCapacityType` parametro specifica automaticamente il numero di unità da associare a ciascun tipo di istanza in base al valore impostato per questo parametro. Per ulteriori informazioni, consulta [Crea un gruppo di istanze miste utilizzando la selezione del tipo di istanza basata su attributi](#).

Indice

- [Considerazioni](#)
- [Comportamenti relativi al peso delle istanze](#)
- [Configurazione di un gruppo con dimensionamento automatico per utilizzare i pesi](#)
- [Esempio di prezzo Spot per unità all'ora](#)

Considerazioni

Questa sezione illustra le considerazioni chiave per un'implementazione efficace dei pesi.

- Scegliete alcuni tipi di istanze che soddisfano le esigenze prestazionali della vostra applicazione. Decidi il peso che ogni tipo di istanza deve contare per la capacità desiderata del gruppo Auto Scaling in base alle sue capacità. Questi pesi si applicano alle istanze attuali e future.
- Evita ampi intervalli tra i pesi. Ad esempio, non specificate il peso di 1 per un tipo di istanza quando il tipo di istanza successivo più grande ha un peso di 200. Anche la differenza tra i pesi più piccoli e più grandi non dovrebbe essere eccessiva. Differenze di peso estreme possono influire negativamente sull'ottimizzazione dei costi e delle prestazioni.
- Specificate la capacità desiderata del gruppo in unità, non in istanze. Ad esempio, se utilizzi pesi basati su vCPU, imposta il numero di core desiderato e anche il minimo e il massimo.
- Imposta i pesi e la capacità desiderata in modo che quest'ultima sia almeno due o tre volte più grande del peso maggiore.

Quando si aggiornano i gruppi esistenti, tenere presente quanto segue:

- Quando aggiungi pesi a un gruppo esistente, includi i pesi per tutti i tipi di istanza attualmente in uso.
- Quando aggiungi o modifichi pesi, Amazon EC2 Auto Scaling avvia o interrompe le istanze per raggiungere la capacità desiderata in base ai nuovi valori di peso.
- Se rimuovi un tipo di istanza, le istanze in esecuzione di quel tipo mantengono il loro peso finale, anche se non sono più definite.

Comportamenti relativi al peso delle istanze

Quando utilizzi i pesi delle istanze, Dimensionamento automatico Amazon EC2 si comporta nel modo seguente:

- la capacità attuale sarà la capacità desiderata o al di sopra di essa. La capacità attuale può superare la capacità desiderata se vengono avviate istanze che superano le unità di capacità rimanenti desiderate. Ad esempio, supponiamo di specificare due tipi di istanza `c5.2xlarge` e `c5.12xlarge`, di assegnare pesi di istanza di 2 per `c5.2xlarge` e 12 per `c5.12xlarge`. Se ci sono cinque unità rimanenti per soddisfare la capacità desiderata e dimensionamento automatico Amazon EC2 esegue il provisioning di `c5.12xlarge`, la capacità desiderata viene superata di sette unità.
- All'avvio delle istanze, Amazon EC2 Auto Scaling dà priorità alla distribuzione della capacità tra le zone di disponibilità e al rispetto delle strategie di allocazione rispetto al superamento della capacità desiderata.
- Amazon EC2 Auto Scaling può superare il limite di capacità massima per mantenere l'equilibrio tra le zone di disponibilità, utilizzando le tue strategie di allocazione preferite. Il limite rigido imposto da Amazon EC2 Auto Scaling è la capacità desiderata più il peso massimo.

Configurazione di un gruppo con dimensionamento automatico per utilizzare i pesi

È possibile configurare un gruppo con dimensionamento automatico per utilizzare i pesi, come illustrato negli esempi seguenti. AWS CLI Per istruzioni su come usare la console, consulta [Crea un gruppo di istanze miste scegliendo manualmente i tipi di istanze](#).

Per configurare un nuovo gruppo con dimensionamento automatico per utilizzare i pesi (AWS CLI)

Utilizza il comando [create-auto-scaling-group](#). Ad esempio, il comando seguente crea un nuovo gruppo con dimensionamento automatico e assegna la ponderazione specificando quanto segue:

- La percentuale del gruppo da avviare come istanze on demand (0)

- La strategia di allocazione per le istanze Spot in ciascuna zona di disponibilità (*capacity-optimized*)
- I tipi di istanza da avviare in ordine di priorità (*m4.16xlarge*, *m5.24xlarge*)
- I pesi dell'istanza che corrispondono alla differenza di dimensione relativa (vCPU) tra i tipi di istanza (16, 24)
- Le sottoreti in cui avviare le istanze (*subnet-5ea0c127*, *subnet-6194ea3b*, *subnet-c934b782*), ciascuna corrispondente a una zona di disponibilità diversa
- Il modello di avvio (*my-launch-template*) e la versione del modello di avvio (*\$Latest*)

```
aws autoscaling create-auto-scaling-group --cli-input-json file://~/config.json
```

Il file `config.json` contiene il seguente contenuto.

```
{
  "AutoScalingGroupName": "my-asg",
  "MixedInstancesPolicy": {
    "LaunchTemplate": {
      "LaunchTemplateSpecification": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "$Latest"
      },
      "Overrides": [
        {
          "InstanceType": "m4.16xlarge",
          "WeightedCapacity": "16"
        },
        {
          "InstanceType": "m5.24xlarge",
          "WeightedCapacity": "24"
        }
      ]
    },
    "InstancesDistribution": {
      "OnDemandPercentageAboveBaseCapacity": 0,
      "SpotAllocationStrategy": "capacity-optimized"
    }
  },
  "MinSize": 160,
  "MaxSize": 720,
```

```
"DesiredCapacity": 480,  
"VPCZoneIdentifier": "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782",  
"Tags": []  
}
```

Per configurare un gruppo con dimensionamento automatico esistente per utilizzare i pesi (AWS CLI)

Utilizza il comando [update-auto-scaling-group](#). Ad esempio, il comando seguente assegna la ponderazione ai tipi di istanza in un gruppo con dimensionamento automatico esistente specificando quanto segue:

- I tipi di istanza da avviare in ordine di priorità (c5.18xlarge, c5.24xlarge, c5.2xlarge, c5.4xlarge)
- I pesi dell'istanza che corrispondono alla differenza di dimensione relativa (vCPU) tra i tipi di istanza (18, 24, 2, 4)
- La nuova maggiore capacità desiderata, che è più grande del peso più grande

```
aws autoscaling update-auto-scaling-group --cli-input-json file://~/config.json
```

Il file `config.json` contiene il seguente contenuto.

```
{  
  "AutoScalingGroupName": "my-existing-asg",  
  "MixedInstancesPolicy": {  
    "LaunchTemplate": {  
      "Overrides": [  
        {  
          "InstanceType": "c5.18xlarge",  
          "WeightedCapacity": "18"  
        },  
        {  
          "InstanceType": "c5.24xlarge",  
          "WeightedCapacity": "24"  
        },  
        {  
          "InstanceType": "c5.2xlarge",  
          "WeightedCapacity": "2"  
        },  
        {  
          "InstanceType": "c5.4xlarge",
```

```

        "WeightedCapacity": "4"
      }
    ]
  }
},
"MinSize": 0,
"MaxSize": 100,
"DesiredCapacity": 100
}

```

Come eseguire la verifica utilizzando la riga di comando

Utilizzare uno dei seguenti comandi:

- [describe-auto-scaling-groups](#) (AWS CLI)
- [Get-as \(\) AutoScalingGroup](#) AWS Tools for Windows PowerShell

Esempio di prezzo Spot per unità all'ora

Nella tabella seguente viene confrontato il prezzo orario per le istanze Spot in diverse zone di disponibilità negli Stati Uniti orientali (Virginia settentrionale) con il prezzo per le istanze on demand nella stessa Regione. I prezzi indicati sono prezzi di esempio e non prezzi correnti. Questi sono i costi per ora di istanza.

Esempio: prezzo Spot per ora di istanza

Tipo di istanza	us-east-1a	us-east-1b	us-east-1c	Prezzi on demand
c5.2xlarge	0,180 USD	0,191 USD	0,170 USD	0,34 USD
c5.4xlarge	0,341 USD	0,361 USD	0,318 USD	0,68 USD
c5.12xlarge	0,779 USD	0,777 USD	0,777 USD	2,04 USD
c5.18xlarge	1,207 USD	1,475 USD	1,357 USD	3,06 USD

Tipo di istanza	us-east-1a	us-east-1b	us-east-1c	Prezzi on demand
c5.24xlarge	1,555 USD	1,555 USD	1,555 USD	4,08 USD

Con la ponderazione delle istanze, è possibile valutare i costi in base a ciò che si utilizza per unità all'ora. È possibile calcolare il prezzo all'ora per unità dividendo il prezzo di un tipo di istanza per il numero di unità che essa rappresenta. Per le istanze on demand, il prezzo per unità all'ora è lo stesso per l'implementazione di un tipo di istanza e per quella di una dimensione diversa dello stesso tipo di istanza. Tuttavia, il prezzo di istanza Spot per unità all'ora varia in base al pool Spot.

L'esempio seguente mostra come funziona il calcolo del prezzo spot per unità oraria con i pesi delle istanze. Quindi, per facilità di calcolo, supponiamo di voler avviare istanze Spot solo in us-east-1a. Il prezzo per unità oraria è riportato nella tabella seguente.

Esempio: esempio di prezzo istanza Spot per unità all'ora

Tipo di istanza	us-east-1a	Peso dell'istanza	Prezzo all'ora per unità
c5.2xlarge	0,180 USD	2	0,090 USD
c5.4xlarge	0,341 USD	4	0,085 USD
c5.12xlarge	0,779 USD	12	0,065 USD
c5.18xlarge	1,207 USD	18	0,067 USD
c5.24xlarge	1,555 USD	24	0,065 USD

Utilizza un modello di avvio diverso per un tipo di istanza

Oltre a utilizzare più tipi di istanza, puoi anche utilizzare più modelli di avvio.

Supponiamo ad esempio di configurare un gruppo con dimensionamento automatico per applicazioni a calcolo intensivo e di voler includere una combinazione di tipi di istanza C5, C5a e C6g. Tuttavia, le istanze C6g sono dotate di un processore AWS Graviton basato sull'architettura Arm a 64 bit, mentre le istanze C5 e C5a funzionano su processori Intel x86 a 64 bit. L'AMI per le istanze C5 e C5a funziona su entrambe queste istanze, ma non su istanze C6g. Per risolvere questo problema, utilizza

un modello di avvio diverso per le istanze C6g. È comunque possibile utilizzare lo stesso modello di avvio per le istanze C5 e C5a.

Questa sezione contiene le procedure per l'utilizzo di per eseguire attività relative all'utilizzo di più modelli AWS CLI di avvio. Al momento, questa funzionalità è disponibile solo se utilizzi la AWS CLI o un SDK e non è disponibile dalla console.

Indice

- [Configurazione di un gruppo con dimensionamento automatico per utilizzare più modelli di avvio](#)
- [Risorse correlate](#)

Configurazione di un gruppo con dimensionamento automatico per utilizzare più modelli di avvio

È possibile configurare un gruppo con dimensionamento automatico per utilizzare molteplici modelli di avvio, come illustrato negli esempi seguenti.

Configurazione di un gruppo con dimensionamento automatico per utilizzare molteplici modelli di avvio (AWS CLI)

Utilizza il comando [create-auto-scaling-group](#). Ad esempio, il seguente comando crea un gruppo con dimensionamento automatico. Specifica il valore dei tipi di istanza `c5.large`, `c5a.large`, e `c6g.large` e definisce un nuovo modello di avvio per il tipo di istanza `c6g.large` per garantire che venga utilizzato un AMI appropriato per avviare le istanze Arm. Amazon EC2 Auto Scaling utilizza l'ordine dei tipi di istanza per determinare quale tipo di istanza utilizzare per prima al fine di soddisfare la capacità on demand.

```
aws autoscaling create-auto-scaling-group --cli-input-json file://~/config.json
```

Il file `config.json` contiene il seguente contenuto.

```
{
  "AutoScalingGroupName": "my-asg",
  "MixedInstancesPolicy": {
    "LaunchTemplate": {
      "LaunchTemplateSpecification": {
        "LaunchTemplateName": "my-launch-template-for-x86",
        "Version": "$Latest"
      },
    },
    "Overrides": [
      {
```



```

    "InstanceType": "c6g.large",
    "LaunchTemplateSpecification": {
      "LaunchTemplateName": "my-launch-template-for-arm",
      "Version": "$Latest"
    }
  },
  {
    "InstanceType": "c5.large"
  },
  {
    "InstanceType": "c5a.large"
  }
]
},
"InstancesDistribution": {
  "OnDemandBaseCapacity": 1,
  "OnDemandPercentageAboveBaseCapacity": 50,
  "SpotAllocationStrategy": "capacity-optimized"
}
},
"MinSize": 1,
"MaxSize": 5,
"DesiredCapacity": 3,
"VPCZoneIdentifier": "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782",
"Tags": [ ]
}

```

Configurazione di un gruppo con dimensionamento automatico esistente per utilizzare molteplici modelli di avvio (AWS CLI)

Utilizza il comando [update-auto-scaling-group](#). Ad esempio, il comando seguente assegna il modello di avvio denominato *my-launch-template-for-arm* al tipo di *c6g.large* istanza per il gruppo con dimensionamento automatico denominato *my-asg*

```
aws autoscaling update-auto-scaling-group --cli-input-json file://~/config.json
```

Il file `config.json` contiene il seguente contenuto.

```

{
  "AutoScalingGroupName": "my-asg",
  "MixedInstancesPolicy": {
    "LaunchTemplate": {

```

```
"Overrides":[
  {
    "InstanceType":"c6g.large",
    "LaunchTemplateSpecification": {
      "LaunchTemplateName": "my-launch-template-for-arm",
      "Version": "$Latest"
    }
  },
  {
    "InstanceType":"c5.large"
  },
  {
    "InstanceType":"c5a.large"
  }
]
```

Come verificare i modelli di avvio per un gruppo con dimensionamento automatico

Utilizzare uno dei seguenti comandi:

- [describe-auto-scaling-groups](#) (AWS CLI)
- [Get-as \(AutoScalingGroup\)](#) AWS Tools for Windows PowerShell

Risorse correlate

[Puoi trovare un esempio di specificazione di più modelli di lancio utilizzando la selezione del tipo di istanza basata sugli attributi in un modello su re:POST. AWS CloudFormationAWS](#)

Creare gruppi con dimensionamento automatico utilizzando configurazioni di avvio

Important

Non puoi effettuare chiamate `CreateLaunchConfiguration` con nuovi tipi di istanze Amazon EC2 rilasciati dopo il 31 dicembre 2022. Inoltre, qualsiasi nuovo account creato dopo il 1° giugno 2023 non avrà la possibilità di creare nuove configurazioni di avvio tramite la console. In futuro, i nuovi account non saranno in grado di creare nuove configurazioni di

avvio utilizzando la console, l'API, la CLI e. CloudFormation Esegui la migrazione ai modelli di lancio per assicurarti di non dover creare nuove configurazioni di lancio ora o in futuro. Per ulteriori informazioni sulla migrazione di gruppi con dimensionamento automatico ai modelli di avvio, consulta [Migra i tuoi gruppi di Auto Scaling ai modelli di avvio](#).

Se hai creato una configurazione di avvio o un'istanza EC2, puoi creare un gruppo con dimensionamento automatico che utilizza una configurazione di avvio come modello di configurazione per le sue istanze EC2. La configurazione di avvio specifica informazioni quali l'ID AMI, il tipo di istanza, la coppia di chiavi, i gruppi di sicurezza e la mappatura dei dispositivi a blocchi per le istanze. Per informazioni sulla creazione di configurazioni di avvio, consulta [Creazione di una configurazione di avvio](#).

Per creare un gruppo con dimensionamento automatico devi disporre di autorizzazioni sufficienti. Se non esiste ancora, per creare il ruolo collegato al servizio utilizzato da Dimensionamento automatico Amazon EC2 per eseguire azioni per conto tuo, devi disporre di autorizzazioni sufficienti. Per esempi di policy IAM che un amministratore può utilizzare come riferimento per concederti le autorizzazioni, consulta [Esempi di policy basate su identità](#).

Indice

- [Creare un gruppo con dimensionamento automatico utilizzando una configurazione di avvio](#)
- [Creare un gruppo con dimensionamento automatico utilizzando parametri di un'istanza esistente](#)

Creare un gruppo con dimensionamento automatico utilizzando una configurazione di avvio

Important

Forniamo le informazioni sulle configurazioni di avvio per i clienti che non hanno ancora eseguito la migrazione dalle configurazioni di avvio ai modelli di avvio. Per ulteriori informazioni sulla migrazione di gruppi con dimensionamento automatico ai modelli di avvio, consulta [Migra i tuoi gruppi di Auto Scaling ai modelli di avvio](#).

Quando crei un gruppo con dimensionamento automatico, devi specificare le informazioni necessarie a configurare le istanze Amazon EC2, le zone di disponibilità e le sottoreti VPC per le istanze, la capacità desiderata e i limiti di capacità minima e massima.

La procedura seguente illustra come creare un gruppo con dimensionamento automatico utilizzando una configurazione di avvio. Non è possibile modificare una configurazione di avvio dopo la creazione, ma puoi sostituire la configurazione di avvio di un gruppo con dimensionamento automatico. Per ulteriori informazioni, consulta [Modifica la configurazione di avvio per un gruppo con scalabilità automatica](#).

Prerequisiti

- È necessario aver creato una configurazione di avvio. Per ulteriori informazioni, consulta [Creazione di una configurazione di avvio](#).

Come creare un gruppo con dimensionamento automatico utilizzando un modello di avvio (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Nella barra di navigazione nella parte superiore dello schermo, scegli la stessa Regione AWS che hai usato quando hai creato la configurazione di avvio.
3. Selezionare Crea un gruppo con dimensionamento automatico.
4. Nella pagina Scegli il modello o la configurazione di avvio, in Nome gruppo con dimensionamento automatico, inserisci un nome per il gruppo con dimensionamento automatico.
5. Per scegliere una configurazione di avvio, procedi come segue:
 - a. Per Launch Template (Modello di avvio) scegli Switch to launch configuration (Passa alla configurazione di avvio).
 - b. Per Launch configuration (Configurazione di avvio) scegli una configurazione di avvio esistente.
 - c. Verifica che la configurazione di avvio supporti tutte le opzioni che intendi utilizzare, quindi scegli Next (Successivo).
6. Alla pagina (Configura impostazioni)Configure instance launch options (Configura opzioni di avvio dell'istanza), in Network (Rete), per VPC, scegli un VPC. Il gruppo con scalabilità automatica deve essere creato nello stesso VPC del gruppo di sicurezza specificato nella configurazione di avvio.
7. Per (Sottoreti) Availability Zones and subnets (Zone di disponibilità e sottoreti), scegli una o più sottoreti nel VPC specificato. Per un'elevata disponibilità, utilizza sottoreti in più zone di disponibilità. Per ulteriori informazioni, consulta la pagina [Considerazioni sulla scelta di sottoreti VPC](#).

8. Seleziona Next (Successivo).

In alternativa, è possibile accettare il resto dei valori predefiniti e scegliere Skip to review (Passa a revisione).

9. (Facoltativo) Nella pagina Configure group size and scaling policies (Configura policy di dimensionamento e dimensione del gruppo configura le seguenti opzioni e quindi scegli Next (Successivo):
 - a. Per registrare le istanze Amazon EC2 con un load balancer, scegli Enable load balancing (Abilita bilanciamento del carico), quindi scegli un load balancer esistente o creane uno nuovo. Per ulteriori informazioni, consulta [Utilizzare sistema di bilanciamento del carico elastico per distribuire il traffico tra le istanze nel gruppo con dimensionamento automatico](#). Per creare un nuovo load balancer, segui la procedura descritta in [Configurare un Application Load Balancer o Network Load Balancer dalla console di Dimensionamento automatico Amazon EC2](#).
 - b. (Facoltativo) Per i controlli dell'integrità, tipi di controlli dell'integrità aggiuntivi, seleziona Attiva i controlli dell'integrità di Elastic Load Balancing.
 - c. (Facoltativo) In Periodo di tolleranza del controllo dell'integrità, inserisci il tempo in secondi. Questo è il periodo di tempo che Dimensionamento automatico Amazon EC2 deve attendere prima di controllare lo stato di integrità di un'istanza dopo che questa è entrata nello stato InService. Per ulteriori informazioni, consulta [Impostazione del periodo di tolleranza del controllo dell'integrità per un gruppo con scalabilità automatica](#).
 - d. In Impostazioni aggiuntive, Monitoraggio, scegli se abilitare la raccolta di metriche di CloudWatch gruppo. Questi parametri forniscono misurazioni che possono essere indicatori di un potenziale problema, come ad esempio il numero di istanze di terminazione o il numero di istanze in sospeso. Per ulteriori informazioni, consulta [Monitorare i parametri di CloudWatch per gruppi e istanze Auto Scaling](#).
 - e. Per Abilita il riscaldamento predefinito dell'istanza, seleziona questa opzione e scegli l'orario di riscaldamento per la tua applicazione. Se stai creando un gruppo Auto Scaling con una politica di scalabilità, la funzionalità di riscaldamento delle istanze predefinita migliora i CloudWatch parametri Amazon utilizzati per la scalabilità dinamica. Per ulteriori informazioni, consulta [Impostazione della preparazione di default dell'istanza per un gruppo con scalabilità automatica](#).
10. (Facoltativo) Nella pagina Configure group size and scaling policies (Configura policy di dimensionamento e dimensione del gruppo configurare le seguenti opzioni e quindi scegliere Next (Successivo):

- a. In Dimensioni gruppo, per Capacità desiderata, immettere il numero iniziale di istanze da avviare.
- b. Nella sezione Dimensionamento, in Limiti di dimensionamento, se il nuovo valore per la capacità desiderata è maggiore della capacità minima desiderata e della capacità massima desiderata, la capacità massima desiderata viene automaticamente aumentata fino al nuovo valore di capacità desiderato. È possibile modificare questi limiti se necessario. Per ulteriori informazioni, consulta [Impostare i limiti di dimensionamento per il gruppo con dimensionamento automatico](#).
- c. Per Dimensionamento automatico, scegli se desideri creare una policy di dimensionamento di monitoraggio degli obiettivi. È possibile creare questa policy anche dopo aver creato il gruppo con dimensionamento automatico.

Se scegli la policy di dimensionamento del monitoraggio target, segui le istruzioni riportate in [Creazione di una policy di dimensionamento con monitoraggio degli obiettivi](#) per creare la policy.

- d. Per la policy di manutenzione dell'istanza, scegli se desideri creare una policy di manutenzione dell'istanza. È possibile creare questa policy anche dopo aver creato il gruppo con dimensionamento automatico. Per creare una policy, segui le indicazioni fornite in [Imposta una policy di manutenzione dell'istanza](#).
 - e. In Instance scale-in protection (Protezione per la riduzione delle istanze) scegliere se abilitare la protezione per la riduzione delle istanze. Per ulteriori informazioni, consulta [Utilizzare la protezione di riduzione orizzontale dell'istanza](#).
11. (Facoltativo) Per ricevere le notifiche, scegliere Add notification (Aggiungi notifica), configurare la notifica, quindi scegliere Next (Successivo). Per ulteriori informazioni, consulta [Opzioni di notifica Amazon SNS per Amazon EC2 Auto Scaling](#).
 12. (Facoltativo) Per aggiungere tag, scegliere Add tag (Aggiungi tag), fornire una chiave e un valore di tag, quindi scegliere Next (Successivo). Per ulteriori informazioni, consulta [Tag di gruppi e istanze Auto Scaling](#).
 13. Nella pagina Verifica, scegliere Crea gruppo con dimensionamento automatico.

Per creare un gruppo con dimensionamento automatico tramite la riga di comando

È possibile utilizzare uno dei seguenti comandi:

- [create-auto-scaling-group](#) (AWS CLI)

- [Nuovo-AS \(\) AutoScalingGroup](#) AWS Tools for Windows PowerShell

Creare un gruppo con dimensionamento automatico utilizzando parametri di un'istanza esistente

Important

Forniamo le informazioni sulle configurazioni di avvio per i clienti che non hanno ancora eseguito la migrazione dalle configurazioni di avvio ai modelli di avvio. Per ulteriori informazioni sulla migrazione di gruppi con dimensionamento automatico ai modelli di avvio, consulta [Migra i tuoi gruppi di Auto Scaling ai modelli di avvio](#).

Se è la prima volta che crei un gruppo con scalabilità automatica, ti consigliamo di utilizzare la console per creare un modello di avvio da un'istanza EC2 esistente. Quindi usa il modello di avvio per creare un nuovo gruppo con scalabilità automatica. Per questa procedura, consultare [Creare un gruppo con dimensionamento automatico utilizzando la procedura guidata di Amazon EC2](#).

La procedura seguente illustra come creare un gruppo con scalabilità automatica specificando un'istanza esistente da utilizzare come base per l'avvio di altre istanze. Per creare un'istanza EC2 sono necessari più parametri, ad esempio l'ID dell'Amazon Machine Image (AMI), il tipo di istanza, la coppia di chiavi e il gruppo di sicurezza. Tutte queste informazioni vengono usate anche da Amazon EC2 Auto Scaling per avviare istanze per tuo conto quando c'è la necessità di dimensionare. Queste informazioni sono archiviate in un modello o in una configurazione di avvio.

Quando utilizzi un'istanza esistente, Dimensionamento automatico Amazon EC2 crea un gruppo con scalabilità automatica che avvia le istanze in base a una configurazione di avvio creata nello stesso momento. La nuova configurazione di avvio ha lo stesso nome del gruppo con scalabilità automatica e include alcuni dettagli di configurazione dell'istanza identificata.

I seguenti dettagli di configurazione vengono copiati dall'istanza selezionata alla configurazione di avvio:

- ID AMI
- Tipo di istanza
- Coppia di chiavi
- Gruppi di sicurezza

- Tipo di indirizzo IP (pubblico o privato)
- Profilo dell'istanza IAM, se applicabile
- Monitoraggio (true o false)
- EBS ottimizzato (true o false)
- Impostazione relativa alla tenancy, in caso di avvio in un VPC (condiviso o dedicato)
- ID kernel e ID disco RAM, se applicabili
- Dati utente, se specificati
- Prezzo istanza Spot (massimo)

La sottorete VPC e la zona di disponibilità vengono copiate dall'istanza identificata nella definizione di risorse del gruppo con dimensionamento automatico.

Se l'istanza identificata si trova in un gruppo di collocazione, il nuovo gruppo con scalabilità automatica avvia istanze nello stesso gruppo di posizionamento dell'istanza identificata. Poiché le impostazioni di configurazione di avvio non consentono di specificare un gruppo di posizionamento, quest'ultimo viene copiato nell'attributo `PlacementGroup` del nuovo gruppo con scalabilità automatica.

I seguenti dettagli di configurazione non vengono copiati dall'istanza identificata:

- Archiviazione: i dispositivi a blocchi (volumi EBS e volumi dell'archivio istanze) non vengono copiati dall'istanza identificata. Invece, la mappatura dei dispositivi a blocchi creata nell'ambito della creazione dell'AMI determina quali dispositivi vengono utilizzati.
- Numero di interfacce di rete: le interfacce di rete non vengono copiate dall'istanza identificata. Al contrario, Amazon EC2 Auto Scaling utilizza le impostazioni di default per creare un'interfaccia di rete, ovvero l'interfaccia di rete primaria (`eth0`).
- Opzioni dei metadati dell'istanza: le impostazioni accessibili ai metadati, la versione dei metadati e il limite di hop di risposta token non vengono copiate dall'istanza identificata. Al contrario, Amazon EC2 Auto Scaling utilizza le impostazioni di default. Per ulteriori informazioni, consulta [Configura le opzioni dei metadati dell'istanza](#).
- Load balancer: se l'istanza identificata è registrata con uno o più load balancer, le informazioni questi ultimi non vengono copiate nel load balancer o nell'attributo del gruppo di destinazione del nuovo gruppo con scalabilità automatica.
- Tag: se l'istanza identificata ha tag, questi non vengono copiati nell'attributo `Tags` del nuovo gruppo con scalabilità automatica.

Prerequisiti

L'istanza EC2 deve soddisfare i seguenti criteri:

- L'istanza non è un membro di un altro gruppo con scalabilità automatica.
- L'istanza è nello stato `running`.
- L'AMI utilizzata per avviare l'istanza deve essere ancora esistente.

Creazione di un gruppo con dimensionamento automatico da un'istanza EC2 (console)

Creazione di un gruppo con scalabilità automatica da un'istanza EC2

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nel pannello di navigazione, in Instances (Istanze) selezionare Instances (Istanze) e poi selezionarne una.
3. Scegli Actions (Operazioni), Instance settings (Impostazioni istanza), Attach to Auto Scaling Group (Allega al gruppo con scalabilità automatica).
4. Nella pagina Attach to Auto Scaling Group (Allega al gruppo con scalabilità automatica), per Auto Scaling Group (gruppo con scalabilità automatica), inserisci un nome per il gruppo e scegli Attach (Allega).

Una volta collegata l'istanza, questa viene considerata parte del gruppo con scalabilità automatica. Il nuovo gruppo con scalabilità automatica viene creato utilizzando una nuova configurazione di avvio con lo stesso nome specificato per il gruppo con scalabilità automatica. Il gruppo con scalabilità automatica ha una capacità desiderata e una dimensione massima di 1.

5. (Facoltativo) Per modificare le impostazioni per il gruppo con scalabilità automatica, nel pannello di navigazione, vai in Auto Scaling e scegli Auto Scaling Groups (Gruppi Auto Scaling). Seleziona la casella di controllo accanto al nuovo gruppo con scalabilità automatica, scegli il pulsante Edit (Modifica) sopra l'elenco dei gruppi, modifica le impostazioni in base alle esigenze e quindi scegli Update (Aggiorna).

Creazione di un gruppo con dimensionamento automatico con due istanze EC2 (AWS CLI)

Nelle procedure seguenti viene illustrato come utilizzare il comando CLI per creare un gruppo con dimensionamento automatico da un'istanza EC2.

Questa procedura non aggiunge l'istanza al gruppo con scalabilità automatica. Affinché l'istanza possa essere collegata, devi eseguire il comando [attach-instances](#) dopo la creazione del gruppo con scalabilità automatica.

Prima di iniziare, individua l'ID dell'istanza EC2 utilizzando la console Amazon EC2 o il comando [describe-instances](#).

Per utilizzare l'istanza corrente come modello

- Usa il seguente [create-auto-scaling-group](#) comando per creare un gruppo Auto Scaling `my-asg-from-instance`, dall'istanza EC2. `i-0e69cc3f05f825f4f`

```
aws autoscaling create-auto-scaling-group --auto-scaling-group-name my-asg-from-instance \  
  --instance-id i-0e69cc3f05f825f4f --min-size 1 --max-size 2 --desired-capacity 2
```

Come verificare se il gruppo con scalabilità automatica abbia avviato delle istanze

- Utilizzate il seguente [describe-auto-scaling-groups](#) comando per verificare che il gruppo Auto Scaling sia stato creato correttamente.

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg-from-instance
```

La risposta di esempio seguente mostra che la capacità desiderata del gruppo è 2, che il gruppo ha due istanze in esecuzione e che la configurazione di avvio è chiamata `my-asg-from-instance`.

```
{  
  "AutoScalingGroups": [  
    {  
      "AutoScalingGroupName": "my-asg-from-instance",  
      "AutoScalingGroupARN": "arn",  
      "LaunchConfigurationName": "my-asg-from-instance",  
      "MinSize": 1,  
      "MaxSize": 2,  
      "DesiredCapacity": 2,  
      "DefaultCooldown": 300,  
      "AvailabilityZones": [  
        "us-west-2a"      ]  
    }  
  ]  
}
```

```
],
"LoadBalancerNames": [],
"TargetGroupARNs": [],
"HealthCheckType": "EC2",
"HealthCheckGracePeriod": 0,
"Instances": [
  {
    "InstanceId": "i-06905f55584de02da",
    "InstanceType": "t2.micro",
    "AvailabilityZone": "us-west-2a",
    "LifecycleState": "InService",
    "HealthStatus": "Healthy",
    "LaunchConfigurationName": "my-asg-from-instance",
    "ProtectedFromScaleIn": false
  },
  {
    "InstanceId": "i-087b42219468eacde",
    "InstanceType": "t2.micro",
    "AvailabilityZone": "us-west-2a",
    "LifecycleState": "InService",
    "HealthStatus": "Healthy",
    "LaunchConfigurationName": "my-asg-from-instance",
    "ProtectedFromScaleIn": false
  }
],
"CreatedTime": "2020-10-28T02:39:22.152Z",
"SuspendedProcesses": [ ],
"VPCZoneIdentifier": "subnet-6bea5f06",
"EnabledMetrics": [ ],
"Tags": [ ],
"TerminationPolicies": [
  "Default"
],
"NewInstancesProtectedFromScaleIn": false,
"ServiceLinkedRoleARN": "arn",
"TrafficSources": []
}
]
```

Come visualizzare la configurazione di avvio

- Utilizzate il [describe-launch-configurations](#) comando seguente per visualizzare i dettagli della configurazione di avvio.

```
aws autoscaling describe-launch-configurations --launch-configuration-names my-asg-from-instance
```

Di seguito è riportato un output di esempio:

```
{
  "LaunchConfigurations": [
    {
      "LaunchConfigurationName": "my-asg-from-instance",
      "LaunchConfigurationARN": "arn",
      "ImageId": "ami-0528a5175983e7f28",
      "KeyName": "my-key-pair-uswest2",
      "SecurityGroups": [
        "sg-05eaec502fcdadc2e"
      ],
      "ClassicLinkVPCSecurityGroups": [ ],
      "UserData": "",
      "InstanceType": "t2.micro",
      "KernelId": "",
      "RamdiskId": "",
      "BlockDeviceMappings": [ ],
      "InstanceMonitoring": {
        "Enabled": true
      },
      "CreatedTime": "2020-10-28T02:39:22.321Z",
      "EbsOptimized": false,
      "AssociatePublicIpAddress": true
    }
  ]
}
```

Come terminare l'istanza

- Se non hai più bisogno dell'istanza, puoi decidere di terminarla. Il seguente comando, [terminate-instance](#), termina l'istanza `i-0e69cc3f05f825f4f`.

```
aws ec2 terminate-instances --instance-ids i-0e69cc3f05f825f4f
```

Dopo aver terminato un'istanza Amazon EC2, non è possibile riavviarla. Dopo la terminazione, i dati non esistono più e il volume non può essere allegato a nessuna istanza. Per ulteriori informazioni sulla terminazione delle istanze, consulta [Terminazione delle istanze](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Aggiornamento di un gruppo con dimensionamento automatico

Puoi aggiornare la maggior parte dei dati del tuo gruppo con dimensionamento automatico. Non è possibile aggiornare il nome di un gruppo di Auto Scaling o modificarlo. Regione AWS

Come aggiornare un gruppo con dimensionamento automatico

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Scegli il tuo gruppo con dimensionamento automatico per visualizzare le informazioni sul gruppo, con le schede per Dettagli, Attività, dimensionamento automatico, Gestione delle istanze, Monitoraggio e Aggiornamento dell'istanza.
3. Scegli le schede per le aree di configurazione che ti interessano e aggiorna le impostazioni secondo necessità. Per ogni impostazione che modifichi, scegli **Aggiorna** per salvare le modifiche alla configurazione del gruppo con dimensionamento automatico.

- Scheda Dettagli

Queste sono le impostazioni generali per il tuo gruppo con dimensionamento automatico. Puoi modificarli e gestirli nello stesso modo in cui crei un gruppo con dimensionamento automatico.

La sezione Configurazioni avanzate contiene alcune opzioni che non sono disponibili durante la creazione del gruppo, come le [policy di terminazione](#), l'[attesa](#), i [processi sospesi](#) e la [durata massima dell'istanza](#). Puoi inoltre visualizzare, ma non modificare, il gruppo di posizionamento e il [ruolo collegato al servizio](#) del gruppo con dimensionamento automatico.

Se il gruppo è associato a risorse Elastic Load Balancing, consulta [Aggiungere o rimuovere zone di disponibilità](#) prima di modificare le zone di disponibilità. Alcune restrizioni sul sistema di bilanciamento del carico potrebbero impedirti di applicare le modifiche alle zone di disponibilità del tuo gruppo alle zone di disponibilità del sistema di bilanciamento del carico.

- Scheda Attività
 - Notifiche di attività: notifiche [Amazon SNS](#)
- Scheda Dimensionamento automatico
 - Policy di scalabilità dinamica: politiche di scalabilità [dinamica](#)
 - [Politiche di scalabilità predittiva: politiche di scalabilità predittiva](#)
 - [Azioni pianificate: azioni pianificate](#)
- Scheda Gestione delle istanze
 - Ganci del ciclo di vita: ganci del ciclo [di vita](#)
 - Piscina calda — [Piscine calde](#)
- Scheda Monitoraggio
 - C'è solo un'opzione in questa scheda, che consente di abilitare o disabilitare la [raccolta di metriche di CloudWatch gruppo](#).

Per aggiornare un gruppo con dimensionamento automatico tramite la riga di comando

È possibile utilizzare uno dei seguenti comandi:

- [update-auto-scaling-group](#) (AWS CLI)
- [Aggiorna come \(AutoScalingGroup\)](#) AWS Tools for Windows PowerShell

Aggiornamento di istanze di dimensionamento automatico

Associando un nuovo modello di avvio o configurazione di avvio a un gruppo con dimensionamento automatico, si verificherà l'aggiornamento della configurazione di tutte le nuove istanze. Le istanze esistenti continuano a funzionare con la configurazione con cui sono state originariamente avviate. Per applicare le tue modifiche alle istanze esistenti hai le seguenti opzioni:

- Avvia un aggiornamento dell'istanza per sostituire le istanze precedenti. Per ulteriori informazioni, consulta [Usa un aggiornamento dell'istanza per aggiornare le istanze in un gruppo di Auto Scaling](#).
- Consenti alle operazioni di dimensionamento di sostituire gradualmente istanze meno recenti con istanze più recenti in base alle [policy di terminazione](#).
- Terminale manualmente in modo che vengano sostituite dal tuo gruppo con dimensionamento automatico.

Note

Puoi modificare i seguenti attributi dell'istanza specificandoli come parte del modello di avvio o della configurazione di avvio:

- Amazon Machine Image (AMI)
- dispositivi a blocchi
- coppia di chiavi
- tipo di istanza
- gruppi di sicurezza
- dati utente
- monitoraggio
- Profilo dell'istanza IAM
- Tenancy di posizionamento
- kernel
- ramdisk
- se l'istanza dispone di un indirizzo IP pubblico

Tag di gruppi e istanze Auto Scaling

Un tag è un'etichetta di attributo personalizzata che si assegna o che si assegna a AWS una risorsa. AWS Ogni tag è costituito da due parti:

- Una chiave del tag (ad esempio, `costcenter`, `environment`, o `project`).
- Un campo facoltativo, noto come valore del tag (ad esempio, `111122223333` o `production`)

I tag consentono di:

- Tieni traccia dei costi. AWS Attivi questi tag sulla AWS Billing and Cost Management dashboard. AWS utilizza i tag per classificare i costi e fornirti un rapporto mensile sull'allocazione dei costi. Per ulteriori informazioni, consulta [Utilizzo dei tag](#) per l'allocazione dei costi nella Guida per l'utente di AWS Billing .

- Controllo dell'accesso ai gruppi Auto Scaling in base ai tag. È possibile utilizzare le condizioni nelle policy IAM per controllare l'accesso ai gruppi Auto Scaling in base ai tag su quel determinato gruppo. Per ulteriori informazioni, consulta [Tag di sicurezza](#).
- Filtra e cerca i gruppi con scalabilità automatica in base ai tag aggiunti. Per ulteriori informazioni, consulta [Usa i tag per filtrare i gruppi Auto Scaling](#).
- Identifica e organizza le tue risorse. AWS Molti Servizi AWS supportano l'etichettatura, quindi puoi assegnare lo stesso tag a risorse di servizi diversi per indicare che le risorse sono correlate.

È possibile aggiungere tag a gruppi Auto Scaling nuovi o già esistenti. Puoi anche propagare tag da un gruppo con scalabilità automatica alle istanze EC2 avviate dal gruppo.

I tag non vengono propagati ai volumi Amazon EBS. Per aggiungere tag ai volumi Amazon EBS, specifica i tag in un modello di avvio. Per ulteriori informazioni, consulta [Creare un modello di avvio per un gruppo con dimensionamento automatico](#).

Puoi creare e gestire i tag tramite AWS Management Console AWS CLI, o SDK.

Indice

- [Limitazioni all'uso e alla denominazione dei tag](#)
- [Ciclo di vita dell'assegnazione di tag all'istanza EC2](#)
- [Assegnazione dei tag ai gruppi Auto Scaling](#)
- [Eliminazione dei tag](#)
- [Tag di sicurezza](#)
- [Controllo dell'accesso ai tag](#)
- [Usa i tag per filtrare i gruppi Auto Scaling](#)

Limitazioni all'uso e alla denominazione dei tag

Ai tag si applicano le seguenti limitazioni di base:

- Il numero massimo di tag per risorsa è 50.
- Il numero massimo di tag che è possibile aggiungere o rimuovere con una singola chiamata è 25.
- La lunghezza massima della chiave è di 128 caratteri Unicode.
- La lunghezza massima del valore è di 256 caratteri Unicode.

- Per le chiavi e i valori dei tag viene fatta la distinzione tra maiuscole e minuscole. Come best practice, è consigliabile definire una strategia per l'uso delle lettere maiuscole e minuscole nei tag e implementarla sistematicamente in tutti i tipi di risorse.
- Non utilizzare il `aws:` prefisso nei nomi o nei valori dei tag, poiché è AWS riservato all'uso. Non puoi modificare o cancellare i nomi o i valori dei tag con questo prefisso, e questi non sono considerati nella quota di tag per risorsa.

Ciclo di vita dell'assegnazione di tag all'istanza EC2

Se hai scelto di propagare i tag nelle istanze EC2, i tag sono gestiti come segue:

- Nella maggior parte dei casi, quando un gruppo con scalabilità automatica avvia le istanze, vi aggiunge i tag durante la creazione delle risorse, non successivamente.
- Il gruppo con scalabilità automatica aggiunge automaticamente un tag alle istanze con una chiave di `aws:autoscaling:groupName` e un valore del nome del gruppo con scalabilità automatica.
- Se si specificano i tag delle istanze nel modello di avvio e si sceglie di propagare i tag del gruppo alle relative istanze, tutti i tag vengono uniti. Se viene specificata la stessa chiave tag per un tag nel modello di avvio e un tag nel gruppo con dimensionamento automatico, il valore del tag del gruppo ha la precedenza.
- Quando alleggi le istanze esistenti, il gruppo con scalabilità automatica aggiunge i tag alle istanze, sovrascrivendo quelli esistenti con la stessa chiave tag. Inoltre, aggiunge un tag con una chiave di `aws:autoscaling:groupName` e un valore del nome del gruppo con scalabilità automatica.
- Quando distacchi un'istanza da un gruppo con scalabilità automatica, viene rimosso solo il tag `aws:autoscaling:groupName`.

Assegnazione dei tag ai gruppi Auto Scaling

Quando aggiungi un tag al gruppo con scalabilità automatica, puoi specificare se questo debba essere aggiunto alle istanze avviate nel gruppo con scalabilità automatica. Se modifichi un tag, la sua versione aggiornata viene aggiunta alle istanze avviate nel gruppo con scalabilità automatica in seguito alla modifica. Se crei o modifichi un tag per un gruppo con scalabilità automatica, le modifiche non vengono apportate alle istanze già in esecuzione nel gruppo con scalabilità automatica.

Indice

- [Aggiunta o modifica dei tag \(console\)](#)

- [Come aggiungere o modificare tag \(AWS CLI\)](#)

Aggiunta o modifica dei tag (console)

Assegnazione di tag a un gruppo con scalabilità automatica al momento della creazione

Quando si utilizza la console Amazon EC2 per creare un gruppo con dimensionamento automatico, è possibile specificare chiavi e valori di tag nella pagina Add tags (Aggiungi tag) della procedura guidata per la creazione di un gruppo con dimensionamento automatico. Per propagare un tag alle istanze avviate nel gruppo con scalabilità automatica, assicurati di mantenere l'opzione Tag new instances (Assegna tag alle nuove istanze) per il tag selezionato. In alternativa, è possibile deselegzionarla.

Come aggiungere o modificare i tag a un gruppo con scalabilità automatica esistente

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina Auto Scaling groups (Gruppi Auto Scaling).

3. Nella scheda Details (Dettagli) scegliere Tags (Tag), Edit (Modifica).
4. Per modificare i tag esistenti, modifica Key (Chiave) e Value (Valore).
5. Per aggiungere un nuovo tag, scegli Add tag (Aggiungi tag) e modifica Key (Chiave) e Value (Valore). È possibile lasciare selezionato Tag new instances (Aggiungi tag a nuove istanze) per assegnare il tag alle istanze avviate automaticamente nel gruppo con scalabilità automatica, o altrimenti deselegzionarlo.
6. Una volta aggiunti i tag, scegli Update (Aggiorna).

Come aggiungere o modificare tag (AWS CLI)

Gli esempi seguenti mostrano come utilizzare per AWS CLI aggiungere tag quando si creano gruppi di Auto Scaling e come aggiungere o modificare tag per gruppi Auto Scaling esistenti.

Assegnazione di tag a un gruppo con scalabilità automatica al momento della creazione

Utilizzate il [create-auto-scaling-group](#) comando per creare un nuovo gruppo Auto Scaling e aggiungere un tag, ad esempio **environment=production**, al gruppo Auto Scaling. Il tag viene assegnato anche a tutte le istanze avviate nel gruppo con scalabilità automatica.

```
aws autoscaling create-auto-scaling-group --auto-scaling-group-name my-asg \  
  --launch-configuration-name my-launch-config --min-size 1 --max-size 3 \  
  --vpc-zone-identifier "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782" \  
  --tags Key=environment,Value=production,PropagateAtLaunch=true
```

Come creare o modificare i tag per un gruppo con scalabilità automatica esistente

Utilizzate il [create-or-update-tags](#) comando per creare o modificare un tag. Ad esempio, il comando seguente aggiunge i tag **Name=my-asg** e **costcenter=cc123**. I tag vengono aggiunti anche a qualsiasi istanza avviata nel gruppo con scalabilità automatica in seguito a questa modifica. Se un tag con questa chiave esiste già, il tag esistente viene sostituito. La console Amazon EC2 associa il nome visualizzato per ogni istanza con il nome specificato per la chiave Name (distinzione tra lettere maiuscole e minuscole).

```
aws autoscaling create-or-update-tags \  
  --tags ResourceId=my-asg,ResourceType=auto-scaling-group,Key=Name,Value=my-  
asg,PropagateAtLaunch=true \  
  ResourceId=my-asg,ResourceType=auto-scaling-  
group,Key=costcenter,Value=cc123,PropagateAtLaunch=true
```

Descrizione dei tag per un gruppo con scalabilità automatica (AWS CLI)

Se desideri visualizzare i tag applicati a un gruppo con scalabilità automatica specifico, puoi utilizzare uno dei comandi seguenti:

- [describe-tags](#) — Fornisci il nome del gruppo Auto Scaling per visualizzare un elenco dei tag per il gruppo specificato.

```
aws autoscaling describe-tags --filters Name=auto-scaling-group,Values=my-asg
```

Di seguito è riportata una risposta di esempio.

```
{  
  "Tags": [  
    {  
      "ResourceType": "auto-scaling-group",
```

```

        "ResourceId": "my-asg",
        "PropagateAtLaunch": true,
        "Value": "production",
        "Key": "environment"
    }
]
}

```

- [describe-auto-scaling-groups](#)— Fornite il nome del gruppo Auto Scaling per visualizzare gli attributi del gruppo specificato, inclusi eventuali tag.

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

Di seguito è riportata una risposta di esempio.

```

{
  "AutoScalingGroups": [
    {
      "AutoScalingGroupName": "my-asg",
      "AutoScalingGroupARN": "arn",
      "LaunchTemplate": {
        "LaunchTemplateId": "lt-0b97f1e282EXAMPLE",
        "LaunchTemplateName": "my-launch-template",
        "Version": "$Latest"
      },
      "MinSize": 1,
      "MaxSize": 5,
      "DesiredCapacity": 1,
      ...
      "Tags": [
        {
          "ResourceType": "auto-scaling-group",
          "ResourceId": "my-asg",
          "PropagateAtLaunch": true,
          "Value": "production",
          "Key": "environment"
        }
      ],
      ...
    }
  ]
}

```

}

Eliminazione dei tag

Puoi eliminare un tag associato al gruppo con scalabilità automatica in qualsiasi momento.

Indice

- [Eliminazione dei tag \(console\)](#)
- [Eliminazione dei tag \(AWS CLI\)](#)

Eliminazione dei tag (console)

Come Per eliminare un tag

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto a un gruppo esistente.

Si aprirà un riquadro diviso nella parte inferiore della pagina Auto Scaling groups (Gruppi Auto Scaling).

3. Nella scheda Details (Dettagli) scegliere Tags (Tag), Edit (Modifica).
4. Scegli Remove (Rimuovi) accanto al tag.
5. Scegli Aggiorna.

Eliminazione dei tag (AWS CLI)

Per eliminare un tag, utilizza il comando [delete-tags](#). Ad esempio, il comando seguente elimina un tag con la chiave di **environment**.

```
aws autoscaling delete-tags --tags "ResourceId=my-asg,ResourceType=auto-scaling-group,Key=environment"
```

È necessario specificare la chiave del tag, ma non è necessario specificarne il valore. Se specifichi un valore non corretto, il tag non verrà eliminato.

Tag di sicurezza

Usa i tag per verificare che il richiedente (ad esempio un utente o un ruolo IAM) disponga delle autorizzazioni necessarie per creare, modificare o eliminare gruppi con scalabilità automatica specifici. Fornire informazioni sui tag nell'elemento condizione di una policy IAM utilizzando una o più delle seguenti chiavi di condizione:

- Utilizza `autoscaling:ResourceTag/tag-key: tag-value` per concedere (o negare) agli utenti operazioni su gruppi Auto Scaling con tag specifici.
- Utilizza `aws:RequestTag/tag-key: tag-value` per richiedere che un tag specifico sia presente (o non presente) in una richiesta.
- Utilizza `aws:TagKeys [tag-key, ...]` per richiedere che chiavi tag specifiche siano presenti (o non presenti) in una richiesta.

Ad esempio, è possibile negare l'accesso a tutti i gruppi Auto Scaling che includono un tag con la chiave **environment** e il valore **production**, come mostrato nell'esempio seguente.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Deny",
      "Action": [
        "autoscaling:CreateAutoScalingGroup",
        "autoscaling:UpdateAutoScalingGroup",
        "autoscaling>DeleteAutoScalingGroup"
      ],
      "Resource": "*",
      "Condition": {
        "StringEquals": {"autoscaling:ResourceTag/environment": "production"}
      }
    }
  ]
}
```

Per ulteriori informazioni sull'uso delle chiavi di condizione per controllare l'accesso ai gruppi con dimensionamento automatico, consulta [Funzionamento di Dimensionamento automatico Amazon EC2 con IAM](#).

Controllo dell'accesso ai tag

Utilizza i tag per verificare che il richiedente (ad esempio un utente o un ruolo IAM) disponga delle autorizzazioni necessarie per aggiungere, modificare o eliminare i tag per i gruppi con scalabilità automatica.

La seguente policy IAM di esempio ti offre l'autorizzazione principale per rimuovere solo il tag con la chiave **temporary** dai gruppi con dimensionamento automatico.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "autoscaling:DeleteTags",
      "Resource": "*",
      "Condition": {
        "ForAllValues:StringEquals": { "aws:TagKeys": ["temporary"] }
      }
    }
  ]
}
```

Per altri esempi di policy IAM che applicano vincoli sui tag specificati per i gruppi con dimensionamento automatico, consulta [Controllo di quali tag chiavi e valori tag possono essere utilizzati](#).

Note

Anche se hai una policy che limita l'assegnazione (o la rimozione) di tag da parte dei tuoi utenti su un gruppo con scalabilità automatica, ciò non ne impedisce la modifica manuale sulle istanze in seguito al loro avvio. Per visualizzare esempi che controllano l'accesso ai tag nelle istanze EC2, consulta [Esempio: assegnazione di tag alle risorse](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Usa i tag per filtrare i gruppi Auto Scaling

Gli esempi seguenti mostrano come utilizzare i filtri con il [describe-auto-scaling-groups](#) comando per descrivere i gruppi di Auto Scaling con tag specifici. Il filtraggio per tag è limitato a AWS CLI o a un SDK e non è disponibile dalla console.

Considerazioni sui filtri

- È possibile specificare più filtri e più valori filtro in una singola richiesta.
- Non puoi utilizzare caratteri jolly con i valori di filtro.
- I valori di filtro fanno distinzione tra maiuscole e minuscole.

Esempio: descrivi i gruppi Auto Scaling con una chiave tag e una coppia di valori specifici

Il comando seguente mostra come filtrare i risultati per mostrare solo i gruppi Auto Scaling con la chiave tag e la coppia di valori di **environment=production**.

```
aws autoscaling describe-auto-scaling-groups \
  --filters Name=tag-key,Values=environment Name=tag-value,Values=production
```

Di seguito è riportata una risposta di esempio.

```
{
  "AutoScalingGroups": [
    {
      "AutoScalingGroupName": "my-asg",
      "AutoScalingGroupARN": "arn",
      "LaunchTemplate": {
        "LaunchTemplateId": "lt-0b97f1e282EXAMPLE",
        "LaunchTemplateName": "my-launch-template",
        "Version": "$Latest"
      },
      "MinSize": 1,
      "MaxSize": 5,
      "DesiredCapacity": 1,
      "...",
      "Tags": [
        {
          "ResourceType": "auto-scaling-group",
          "ResourceId": "my-asg",
          "PropagateAtLaunch": true,
```



```

        "Value": "production",
        "Key": "environment"
    }
  ],
  ...
},
... additional groups ...
]
}

```

In alternativa, puoi specificare i tag usando un filtro tag: **<key>**. Ad esempio, il comando seguente mostra come filtrare i risultati per mostrare solo i gruppi Auto Scaling con una chiave tag e la coppia di valori di **environment=production**. Questo filtro è formattato come segue: Name=tag: **<key>**, Values=**<value>**, con **<key>** e **<value>** che rappresentano una coppia chiave tag e valore.

```

aws autoscaling describe-auto-scaling-groups \
  --filters Name=tag:environment,Values=production

```

Puoi anche filtrare l' AWS CLI output utilizzando l' --query opzione. L'esempio seguente mostra come limitare l' AWS CLI output del comando precedente solo al nome del gruppo, alla dimensione minima, alla dimensione massima e agli attributi di capacità desiderati.

```

aws autoscaling describe-auto-scaling-groups \
  --filters Name=tag:environment,Values=production \
  --query "AutoScalingGroups[].{AutoScalingGroupName: AutoScalingGroupName, MinSize: MinSize, MaxSize: MaxSize, DesiredCapacity: DesiredCapacity}"

```

Di seguito è riportata una risposta di esempio.

```

[
  {
    "AutoScalingGroupName": "my-asg",
    "MinSize": 0,
    "MaxSize": 10,
    "DesiredCapacity": 1
  },
  ... additional groups ...
]

```

]

Per ulteriori informazioni sul filtraggio, vedere [Filtraggio dell' AWS CLI output](#) nella Guida dell'AWS Command Line Interface utente.

Esempio: descrivi i gruppi Auto Scaling con tag corrispondenti alla chiave tag specificata

Il comando seguente mostra come filtrare i risultati per mostrare solo i gruppi Auto Scaling con il tag **environment**, indipendentemente dal valore del tag.

```
aws autoscaling describe-auto-scaling-groups \  
  --filters Name=tag-key,Values=environment
```

Esempio: descrivi i gruppi Auto Scaling con tag corrispondenti al set di chiavi tag specificato

Il comando seguente mostra come filtrare i risultati per mostrare solo i gruppi Auto Scaling con i tag **environment** e **project**, indipendentemente dai valori del tag.

```
aws autoscaling describe-auto-scaling-groups \  
  --filters Name=tag-key,Values=environment Name=tag-key,Values=project
```

Esempio: descrivi i gruppi Auto Scaling con tag corrispondenti ad almeno una delle chiavi tag specificate

Il comando seguente mostra come filtrare i risultati per mostrare solo i gruppi Auto Scaling con i tag **environment** o **project**, indipendentemente dai valori del tag.

```
aws autoscaling describe-auto-scaling-groups \  
  --filters Name=tag-key,Values=environment,project
```

Esempio: descrivi i gruppi Auto Scaling con il valore del tag specificato

Il comando seguente mostra come filtrare i risultati per mostrare solo i gruppi Auto Scaling con un valore del tag di **production**, indipendentemente dalla chiave tag.

```
aws autoscaling describe-auto-scaling-groups \  
  --filters Name=tag-value,Values=production
```

Esempio: descrivi i gruppi Auto Scaling con il set di valori del tag specificato

Il comando seguente mostra come filtrare i risultati per mostrare solo i gruppi Auto Scaling con i valori del tag **production** e **development**, indipendentemente dalla chiave tag.

```
aws autoscaling describe-auto-scaling-groups \  
  --filters Name=tag-value,Values=production Name=tag-value,Values=development
```

Esempio: descrivi i gruppi Auto Scaling con tag corrispondenti ad almeno una dei valori del tag specificati

Il comando seguente mostra come filtrare i risultati per mostrare solo i gruppi Auto Scaling con un valore del tag di **production** o **development**, indipendentemente dalla chiave tag.

```
aws autoscaling describe-auto-scaling-groups \  
  --filters Name=tag-value,Values=production,development
```

Esempio: descrivi i gruppi Auto Scaling con tag corrispondenti a più chiavi tag e valori

È inoltre possibile combinare filtri per creare logiche AND e OR personalizzate per eseguire filtri più complessi.

Il comando seguente mostra come filtrare i risultati per mostrare solo i gruppi Auto Scaling con un set di tag specifico. Una chiave di tag è **environment** AND il valore del tag è (**production** OR **development**) AND l'altra chiave di tag è **costcenter** AND il valore del tag è **cc123**.

```
aws autoscaling describe-auto-scaling-groups \  
  --filters Name=tag:environment,Values=production,development \  
  Name=tag:costcenter,Values=cc123
```

Policy di manutenzione delle istanze

È possibile configurare una policy di manutenzione delle istanze per il gruppo con dimensionamento automatico per soddisfare requisiti di capacità specifici durante gli eventi che causano la sostituzione delle istanze, come l'aggiornamento dell'istanza o il processo di controllo dell'integrità.

Ad esempio, supponiamo di disporre di un gruppo con dimensionamento automatico con un numero limitato di istanze. Desideri evitare le potenziali interruzioni dovute alla chiusura e alla successiva sostituzione di un'istanza quando i controlli dell'integrità indicano che si tratta di un'istanza danneggiata. Con una policy di manutenzione delle istanze, puoi assicurarti che Dimensionamento

automatico Amazon EC2 avvii prima una nuova istanza e poi attenda che sia completamente pronta prima di terminare l'istanza non integra.

Una policy di manutenzione delle istanze consente inoltre di ridurre al minimo le potenziali interruzioni nei casi in cui più istanze vengono sostituite contemporaneamente. Si impostano i parametri minimi e massimi della percentuale di integrità per la policy e il gruppo con dimensionamento automatico può aumentare e diminuire la capacità solo all'interno di questo intervallo minimo-massimo quando sostituisce le istanze. Un intervallo più ampio aumenta il numero di istanze che possono essere sostituite contemporaneamente.

Indice

- [Panoramica della policy di manutenzione delle istanze](#)
- [Impostare una policy di manutenzione delle istanze per il gruppo con dimensionamento automatico](#)

Panoramica della policy di manutenzione delle istanze

Questo capitolo offre una panoramica delle opzioni disponibili e descrive gli aspetti da considerare quando si crea una policy di manutenzione delle istanze.

Indice

- [Panoramica](#)
- [Concetti principali](#)
- [Preparazione dell'istanza](#)
- [Periodo di tolleranza dei controlli dell'integrità](#)
- [Aumentare il gruppo con dimensionamento automatico](#)
- [Scenari di esempio](#)

Panoramica

Quando crei una policy di manutenzione delle istanze per il tuo gruppo con dimensionamento automatico, la policy influisce sugli eventi di Dimensionamento automatico Amazon EC2 che causano la sostituzione delle istanze. Ciò si traduce in comportamenti di sostituzione più coerenti all'interno dello stesso gruppo con dimensionamento automatico. Inoltre, consente di ottimizzare il gruppo in termini di disponibilità o costi a seconda delle esigenze.

Nella console sono disponibili le seguenti opzioni di configurazione:

- **Avvio prima della chiusura:** è necessario eseguire il provisioning di una nuova istanza prima di poter terminare un'istanza esistente. Questo approccio è una buona scelta per le applicazioni che privilegiano la disponibilità rispetto al risparmio sui costi.
- **Termina e avvia:** le nuove istanze vengono fornite contemporaneamente alla chiusura delle istanze esistenti. Questo approccio è una buona scelta per le applicazioni che privilegiano il risparmio sui costi rispetto alla disponibilità. È anche una buona scelta per le applicazioni che non dovrebbero avviare una capacità superiore a quella attualmente disponibile, anche quando si sostituiscono le istanze.
- **Policy personalizzata:** questa opzione consente di impostare la propria policy con un intervallo minimo e massimo personalizzato per la quantità di capacità che si desidera rendere disponibile quando si sostituiscono le istanze. Questo approccio può aiutarti a raggiungere il giusto equilibrio tra costi e disponibilità.

L'impostazione predefinita per un gruppo con dimensionamento automatico è l'assenza di una policy di manutenzione dell'istanza, il che fa sì che risponda agli eventi di manutenzione dell'istanza con i comportamenti predefiniti. I comportamenti predefiniti sono descritti nella tabella seguente.

Comportamenti predefiniti degli eventi di manutenzione dell'istanza

Evento	Descrizione	Comportamento predefinito
Controlli dell'integrità non andati a buon fine	Si verifica automaticamente quando le istanze non superano i controlli dell'integrità. Dimensionamento automatico Amazon EC2 sostituisce le istanze che non superano i controlli dell'integrità. Per comprendere le cause degli errori dei controlli 'integrità, consulta Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico .	Termina e avvia.
Aggiornamento istanza	Succede quando si avvia un aggiornamento dell'istanza. A	Termina e avvia.

Evento	Descrizione	Comportamento predefinito
	seconda della configurazione, un aggiornamento di istanza sostituisce le istanze una alla volta, più istanze alla volta o tutte insieme. Per ulteriori informazioni, consulta Usa un aggiornamento dell'istanza per aggiornare le istanze in un gruppo di Auto Scaling.	
Durata massima dell'istanza	Si verifica automaticamente quando le istanze raggiungono la durata massima dell'istanza specificata per il gruppo con dimensionamento automatico. Dimensionamento automatico o Amazon EC2 sostituisce le istanze che hanno raggiunto la durata massima di vita delle istanze. Per ulteriori informazioni, consulta Sostituire le istanze Auto Scaling basate sulla durata massima dell'istanza.	Termina e avvia.

Evento	Descrizione	Comportamento predefinito
Ribilanciamento	<p>Si verifica automaticamente se ci sono modifiche sottostanti che causano lo squilibrio del gruppo. Dimensionamento automatico Amazon EC2 ribilancia il gruppo nelle seguenti situazioni:</p> <ul style="list-style-type: none">• Una zona di disponibilità che aveva in precedenza una capacità insufficiente viene ripristinata oppure si aggiunge o si rimuove una zona di disponibilità dal gruppo. Quando ciò accade, il gruppo con dimensionamento automatico cerca di bilanciarsi in modo uniforme tra le zone di disponibilità. Per ulteriori informazioni, consulta Attività di ribilanciamento.• Attiva il ribilanciamento della capacità nel tuo gruppo con dimensionamento automatico, che tenta di avviare nuove istanze spot prima che quelle esistenti vengano interrotte quando la disponibilità delle istanze spot cambia. Per ulteriori informazioni, consulta Utilizza il ribilanciamento della capacità per gestire le	<p>Avvia prima di terminare.</p> <p>Dimensionamento automatico Amazon EC2 è in grado di superare i limiti di dimensione del gruppo fino al 10% della capacità massima. Tuttavia, se si utilizza il ribilanciamento della capacità, è possibile superare questi limiti solo fino al 10% della capacità desiderata.</p>

Evento	Descrizione	Comportamento predefinito
	<p>interruzioni spot di Amazon EC2.</p> <ul style="list-style-type: none"> • Aggiorna il tuo gruppo con dimensionamento automatico e questo sostituisce gradualmente le istanze in base alle nuove opzioni di acquisto scelte durante l'aggiornamento di una policy di istanze miste. Per ulteriori informazioni, consulta Aggiornamento di un gruppo con dimensionamento automatico. 	

Dimensionamento automatico Amazon EC2 continuerà a funzionare come impostazione predefinita per terminare e avviare nelle seguenti situazioni. Pertanto, quando si verifica una di queste situazioni, la capacità del gruppo potrebbe essere inferiore alla soglia inferiore della policy di manutenzione delle istanze.

- Quando un'istanza si interrompe in modo imprevisto, ad esempio a causa dell'azione umana. Dimensionamento automatico Amazon EC2 sostituisce immediatamente una qualsiasi istanza se questa non è più in esecuzione. Per ulteriori informazioni, consulta [Controlli dell'integrità Amazon EC2](#).
- Quando Amazon EC2 riavvia, arresta o ritira un'istanza come parte di un evento pianificato prima che Dimensionamento automatico Amazon EC2 possa avviare l'istanza sostitutiva. Per ulteriori informazioni su tali eventi, consulta [Eventi pianificati per le istanze](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.
- Quando il servizio Spot Amazon EC2 avvia un'interruzione di un'istanza spot e un'istanza spot viene quindi interrotta forzatamente.

Con le istanze spot, se hai abilitato il ribilanciamento della capacità nel tuo gruppo con dimensionamento automatico, l'istanza potrebbe avere già un'istanza in sospenso proveniente da un

altro pool spot che abbiamo lanciato prima che iniziassimo l'interruzione spot. Per dettagli su come funziona il ribilanciamento della capacità, consulta [Utilizza il ribilanciamento della capacità per gestire le interruzioni spot di Amazon EC2](#).

Tuttavia, poiché non è garantito che le istanze spot rimangano disponibili e possono essere terminate con un avviso di interruzione dell'istanza spot di due minuti, la soglia inferiore prevista dalla policy di manutenzione delle istanze può essere superata se le istanze vengono interrotte prima del lancio delle nuove istanze.

Concetti principali

Prima di iniziare, acquisisci familiarità con i concetti principali e i termini di base riportati di seguito:

Capacità desiderata

La Capacità desiderata rappresenta la capacità iniziale del gruppo con dimensionamento automatico nel momento in cui viene creato. È anche la capacità che il gruppo cerca di mantenere quando non ci sono condizioni di scalabilità associate al gruppo.

Policy di manutenzione delle istanze

Una policy di manutenzione dell'istanza controlla se un'istanza viene fornita prima che un'istanza esistente venga terminata per gli eventi di manutenzione dell'istanza. Determina inoltre quanto al di sotto e al di sopra della capacità desiderata il gruppo con dimensionamento automatico può arrivare a sostituire più istanze contemporaneamente.

Percentuale massima di integrità

La percentuale massima integra è la percentuale della capacità desiderata che il gruppo con dimensionamento automatico può raggiungere quando si sostituiscono le istanze. Rappresenta la percentuale massima del gruppo che può essere in servizio e integro, o in sospenso, per supportare il carico di lavoro. Nella console, è possibile impostare la percentuale massima di integrità quando si utilizza l'opzione Avvia prima del termine o Policy personalizzata. I valori validi sono 100-200 percento.

Percentuale minima di integrità

La percentuale minima di integrità è la percentuale della capacità desiderata da mantenere in servizio, integra e pronta all'uso per supportare il carico di lavoro in caso di sostituzione delle istanze. Un'istanza è considerata integra e pronta per l'uso dopo aver completato con successo il primo controllo di integrità e dopo che è trascorso il tempo di riscaldamento specificato. Nella

console, è possibile impostare la percentuale minima di integrità quando si utilizza l'opzione Termina e avvia o Policy personalizzata. I valori validi sono 0-100 percento.

Note

Per sostituire le istanze più velocemente, puoi specificare una percentuale minima di integrità bassa. Tuttavia, se non ci sono abbastanza istanze integre in esecuzione, ciò può ridurre la disponibilità. Ti consigliamo di selezionare un valore ragionevole per mantenere la disponibilità in situazioni in cui verranno sostituite più istanze.

Preparazione dell'istanza

Se le istanze hanno bisogno di tempo per inicializzarsi dopo l'ingresso nello stato InService, abilita il riscaldamento predefinito dell'istanza per il gruppo con dimensionamento automatico. Con il riscaldamento predefinito delle istanze, puoi impedire che le istanze vengano conteggiate ai fini della percentuale minima di integrità prima che siano pronte. Ciò garantisce che Dimensionamento automatico Amazon EC2 consideri il tempo necessario per disporre di una capacità sufficiente per supportare il carico di lavoro prima di terminare le istanze esistenti.

Come ulteriore vantaggio, puoi migliorare le CloudWatch metriche di Amazon utilizzate per la scalabilità dinamica abilitando il riscaldamento predefinito dell'istanza. Se il gruppo Auto Scaling dispone di politiche di ridimensionamento, quando il gruppo si ridimensiona, utilizza lo stesso periodo di riscaldamento predefinito per evitare che le istanze vengano conteggiate ai fini delle CloudWatch metriche prima del completamento dell'inizializzazione.

Per ulteriori informazioni, consulta [Impostazione della preparazione di default dell'istanza per un gruppo con scalabilità automatica](#).

Periodo di tolleranza dei controlli dell'integrità

Dimensionamento automatico Amazon EC2 determina se un'istanza è integra in base allo stato dei controlli dell'integrità utilizzati dal gruppo con dimensionamento automatico. Per ulteriori informazioni, consulta [Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico](#).

Per assicurarsi che questi controlli dell'integrità vengano avviati il prima possibile, non impostare un periodo di tolleranza del controllo dell'integrità del gruppo troppo alto, ma abbastanza alto da consentire ai controlli dell'integrità di Elastic Load Balancing di determinare se una destinazione è

disponibile per gestire le richieste. Per ulteriori informazioni, consulta [Impostazione del periodo di tolleranza del controllo dell'integrità per un gruppo con scalabilità automatica](#).

Aumentare il gruppo con dimensionamento automatico

Una policy di manutenzione delle istanze si applica solo agli eventi di manutenzione delle istanze e non impedisce il ridimensionamento manuale o automatico del gruppo.

Se al gruppo con dimensionamento automatico sono associate policy di dimensionamento o azioni pianificate, queste possono essere eseguite in parallelo mentre si verificano gli eventi di manutenzione dell'istanza. In tal caso, potrebbero aumentare o diminuire la capacità desiderata del gruppo, ma solo entro i limiti di scalabilità definiti. Per ulteriori informazioni su questi limiti, consulta [Impostare i limiti di dimensionamento per il gruppo con dimensionamento automatico](#).

Scenari di esempio

In uno scenario tipico, la policy di manutenzione dell'istanza e la capacità desiderata potrebbero essere simili alle seguenti:

- Percentuale minima di integrità = 90 per cento
- Percentuale massima di integrità = 120 per cento
- Capacità desiderata = 100

Durante qualsiasi evento di manutenzione dell'istanza, il gruppo con dimensionamento automatico può avere da un minimo di 90 istanze a un massimo di 120. Dopo l'evento, il gruppo torna ad avere 100 istanze.

Quando si utilizza una policy di manutenzione dell'istanza con un gruppo con dimensionamento automatico che dispone di un warm pool, le percentuali di integrità minima e massima vengono applicate separatamente al gruppo dimensionamento automatico e al pool caldo.

Ad esempio, supponiamo che questa sia la tua configurazione:

- Percentuale minima di integrità = 90 per cento
- Percentuale massima di integrità = 120 per cento
- Capacità desiderata = 100
- Dimensione pool caldo = 10

Se si avvia un aggiornamento dell'istanza per riciclare le istanze del gruppo, Dimensionamento automatico Amazon EC2 sostituisce prima le istanze nel gruppo con dimensionamento automatico e poi le istanze nel pool caldo. Sebbene Dimensionamento automatico Amazon EC2 stia ancora lavorando alla sostituzione delle istanze nel gruppo con dimensionamento automatico, il gruppo potrebbe avere solo 90 istanze e fino a 120. Una volta terminato il lavoro con il gruppo, Dimensionamento automatico Amazon EC2 può iniziare a sostituire le istanze nel pool caldo. Mentre ciò accade, il pool caldo potrebbe avere da un minimo di 9 istanze a un massimo di 12.

Impostare una policy di manutenzione delle istanze per il gruppo con dimensionamento automatico

È possibile creare una policy di manutenzione dell'istanza al momento della creazione di un gruppo con dimensionamento automatico. È possibile crearla anche per i gruppi esistenti.

Impostando una policy di manutenzione dell'istanza nel gruppo con dimensionamento automatico, non è più necessario specificare i valori dei parametri di integrità minima e massima delle istanze, a meno che non si desideri ignorare la policy di manutenzione dell'istanza.

Nella console, Dimensionamento automatico Amazon EC2 offre opzioni per aiutarti a iniziare.

Indice

- [Imposta una policy di manutenzione dell'istanza](#)
- [Rimuovere una policy di manutenzione delle istanze](#)

Imposta una policy di manutenzione dell'istanza

Per impostare una policy di manutenzione dell'istanza su un gruppo con dimensionamento automatico, utilizza uno dei seguenti metodi:

Console

Impostare una policy di manutenzione dell'istanza su un nuovo gruppo (console)

1. Segui le istruzioni in [Creare un gruppo con dimensionamento automatico utilizzando un modello di avvio](#) e completa ogni passaggio della procedura, fino al passaggio 11.
2. Nella sezione Configura le policy di dimensione e dimensionamento del gruppo, per Capacità desiderata, inserisci il numero iniziale di istanze da avviare.

3. Nella sezione Dimensionamento, in Limiti di dimensionamento, se il nuovo valore per la capacità desiderata è maggiore della capacità minima desiderata e della capacità massima desiderata, la capacità massima desiderata viene automaticamente aumentata fino al nuovo valore di capacità desiderato. È possibile modificare questi limiti se necessario.
4. Per Dimensionamento automatico, scegli se desideri creare una policy di dimensionamento di monitoraggio degli obiettivi. È possibile creare questa policy anche dopo aver creato il gruppo con dimensionamento automatico.

Se scegli la policy di dimensionamento del monitoraggio target, segui le istruzioni riportate in [Creazione di una policy di dimensionamento con monitoraggio degli obiettivi](#) per creare la policy.

5. Nella sezione Policy di manutenzione dell'istanza, scegli una delle opzioni disponibili:
 - **Avvio prima del termine:** è necessario eseguire il provisioning di una nuova istanza prima di poter chiudere un'istanza esistente. Si tratta di una buona scelta per le applicazioni che privilegiano la disponibilità rispetto al risparmio.
 - **Termina e avvia:** le nuove istanze vengono fornite contemporaneamente alla chiusura delle istanze esistenti. Si tratta di una buona scelta per le applicazioni che privilegiano il risparmio rispetto alla disponibilità. È anche una buona scelta per le applicazioni che non dovrebbero avviare una capacità superiore a quella attualmente disponibile.
 - **Policy personalizzata:** questa opzione consente di impostare una policy con un intervallo minimo e massimo personalizzato per la quantità di capacità che si desidera rendere disponibile quando si sostituiscono le istanze. Questo approccio può aiutarti a raggiungere il giusto equilibrio tra costi e disponibilità.
6. Per Imposta una percentuale di integrità adeguata, inserisci i valori per uno o entrambi i campi seguenti. I campi abilitati variano a seconda dell'opzione scelta nel passaggio precedente.
 - **Min:** imposta la percentuale minima di integrità necessaria per procedere con la sostituzione delle istanze.
 - **Max:** imposta la percentuale massima di integrità possibile quando si sostituiscono le istanze.
7. Espandi la sezione Visualizza la capacità durante le sostituzioni in base alla capacità desiderata per confermare in che modo i valori di Min e Max si applicano al gruppo. I valori esatti utilizzati dipendono dal valore di capacità desiderato, che cambierà se il gruppo cresce.

8. Continuare con le fasi in [Creare un gruppo con dimensionamento automatico utilizzando un modello di avvio](#).

AWS CLI

Come impostare una policy di manutenzione dell'istanza su un nuovo gruppo (AWS CLI)

Aggiungere l'opzione al comando. `--instance-maintenance-policy` [create-auto-scaling-group](#) L'esempio seguente imposta una policy di manutenzione dell'istanza su un nuovo gruppo con dimensionamento automatico denominato *my-asg*.

```
aws autoscaling create-auto-scaling-group \  
  --launch-template LaunchTemplateName=my-launch-template,Version='1' \  
  --auto-scaling-group-name my-asg \  
  --min-size 1 \  
  --max-size 10 \  
  --desired-capacity 5 \  
  --default-instance-warmup 20 \  
  --instance-maintenance-policy '{  
    "MinHealthyPercentage": 90,  
    "MaxHealthyPercentage": 120  
  }' \  
  --vpc-zone-identifier "subnet-5e6example,subnet-613example,subnet-c93example"
```

Console

Come impostare una policy di manutenzione dell'istanza su un nuovo gruppo (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Sulla barra di navigazione nella parte superiore della schermata, seleziona la Regione AWS in cui è stato creato il gruppo con dimensionamento automatico.
3. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

4. Nella scheda Dettagli, scegli Policy di manutenzione dell'istanza, Modifica.
5. Per impostare una policy di manutenzione dell'istanza sul gruppo, scegli una delle opzioni disponibili:

- **Avvio prima del termine:** è necessario eseguire il provisioning di una nuova istanza prima di poter chiudere un'istanza esistente. Si tratta di una buona scelta per le applicazioni che privilegiano la disponibilità rispetto al risparmio.
 - **Termina e avvia:** le nuove istanze vengono fornite contemporaneamente alla chiusura delle istanze esistenti. Si tratta di una buona scelta per le applicazioni che privilegiano il risparmio rispetto alla disponibilità. È anche una buona scelta per le applicazioni che non dovrebbero avviare una capacità superiore a quella attualmente disponibile.
 - **Policy personalizzata:** questa opzione consente di impostare una policy con un intervallo minimo e massimo personalizzato per la quantità di capacità che si desidera rendere disponibile quando si sostituiscono le istanze. Questo approccio può aiutarti a raggiungere il giusto equilibrio tra costi e disponibilità.
6. Per Imposta una percentuale di integrità adeguata, inserisci i valori per uno o entrambi i campi seguenti. I campi abilitati variano a seconda dell'opzione scelta nel passaggio precedente.
 - **Min:** imposta la percentuale minima di integrità necessaria per procedere con la sostituzione delle istanze.
 - **Max:** imposta la percentuale massima di integrità possibile quando si sostituiscono le istanze.
 7. Espandi la sezione Visualizza la capacità durante le sostituzioni in base alla capacità desiderata per confermare in che modo i valori di Min e Max si applicano al gruppo. I valori esatti utilizzati dipendono dal valore di capacità desiderato, che cambierà se il gruppo cresce.
 8. Scegli Aggiorna.

AWS CLI

Come impostare una policy di manutenzione dell'istanza su un gruppo esistente (AWS CLI)

Aggiungete l'`--instance-maintenance-policy` opzione al [update-auto-scaling-group](#) comando. L'esempio seguente imposta una policy di manutenzione dell'istanza per il gruppo con dimensionamento automatico specificato.

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg \  
--instance-maintenance-policy '{  
  "MinHealthyPercentage": 90,  
  "MaxHealthyPercentage": 120
```

```
}'
```

Rimuovere una policy di manutenzione delle istanze

Se si desidera interrompere l'uso di un criterio di manutenzione dell'istanza con il gruppo con dimensionamento automatico, è possibile rimuoverlo.

Console

Per rimuovere una policy di manutenzione delle istanze (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Sulla barra di navigazione nella parte superiore della schermata, seleziona la Regione AWS in cui è stato creato il gruppo con dimensionamento automatico.
3. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

4. Nella scheda Dettagli, scegli Policy di manutenzione dell'istanza, Modifica.
5. Scegli Nessuna policy di manutenzione dell'istanza.
6. Scegli Aggiorna.

AWS CLI

Come rimuovere una policy di manutenzione dell'istanza (AWS CLI)

Aggiungete l'`--instance-maintenance-policy` opzione al [update-auto-scaling-group](#) comando. L'esempio seguente rimuove la policy di manutenzione dell'istanza dal gruppo con dimensionamento automatico specificato.

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg \  
  --instance-maintenance-policy '{  
    "MinHealthyPercentage": -1,  
    "MaxHealthyPercentage": -1  
  }'
```


Hook del ciclo di vita di Amazon EC2 Auto Scaling

Amazon EC2 Auto Scaling offre la possibilità di aggiungere hook del ciclo di vita ai gruppi Auto Scaling. Questi hook consentono di creare soluzioni che conoscono gli eventi nel ciclo di vita dell'istanza Auto Scaling e quindi eseguire un'operazione personalizzata sulle istanze quando si verifica l'evento del ciclo di vita corrispondente. Un hook del ciclo di vita fornisce un periodo di tempo specificato (un'ora, per impostazione predefinita) per attendere il completamento dell'operazione prima che l'istanza passi allo stato successivo.

Esempio di utilizzo degli hook del ciclo di vita con istanze Auto Scaling:

- Quando si verifica un evento di aumento orizzontale, l'istanza appena avviata completa la sequenza di avvio e passa allo stato di attesa. Quando l'istanza è in stato di attesa, esegue uno script per scaricare e installare il software necessario per l'applicazione, in modo che l'istanza sia del tutto pronta per iniziare a ricevere traffico. Dopo aver terminato l'installazione del software, lo script, invia il comando `complete-lifecycle-action` per continuare.
- Quando si verifica un evento `scale-in`, un lifecycle hook mette in pausa l'istanza prima che venga terminata e ti invia una notifica tramite Amazon EventBridge. Mentre l'istanza è in stato di attesa, puoi richiamare una AWS Lambda funzione o connetterti all'istanza per scaricare log o altri dati prima che l'istanza venga completamente terminata.

Un uso diffuso degli hook del ciclo di vita consiste nel controllare quando le istanze sono registrate con Elastic Load Balancing. Aggiungendo un hook del ciclo di vita di avvio al gruppo con scalabilità automatica, è possibile assicurarsi che gli script di bootstrap siano stati completati correttamente e che le applicazioni nelle istanze siano pronte per accettare il traffico prima che vengano registrate nel load balancer al termine dell'hook del ciclo di vita.

Indice

- [Disponibilità dell'hook del ciclo di vita](#)
- [Considerazioni e limitazioni per hook del ciclo di vita](#)
- [Risorse correlate](#)
- [Come funzionano gli hook del ciclo di vita](#)
- [Preparazione all'aggiunta di un hook del ciclo di vita al gruppo con scalabilità automatica](#)
- [Recupero dello stato del ciclo di vita di destinazione tramite i metadati dell'istanza](#)
- [Aggiungere un hook del ciclo di vita](#)

- [Completare un'operazione del ciclo di vita](#)
- [Tutorial: configurazione dei dati utente per recuperare lo stato del ciclo di vita di destinazione tramite i metadati dell'istanza](#)
- [Tutorial : Configurazione di un hook del ciclo di vita che richiama una funzione Lambda](#)

Disponibilità dell'hook del ciclo di vita

Nella tabella seguente sono elencati gli hook del ciclo di vita disponibili per vari scenari.

Evento	Avvio o terminazione ¹ delle istanze	Durata massima dell'istanza : istanze sostitutive	Aggiornamento delle istanze : istanze sostitutive	Ribilanciamento della capacità : istanze sostitutive	Warm pool : istanze che entrano e escono dal warm pool
Avvio delle istanze	✓	✓	✓	✓	✓
Interruzione delle istanze	✓	✓	✓	✓	✓

¹ Si applica a tutti gli avvii e le terminazioni, indipendentemente dal fatto che vengano avviati automaticamente o manualmente, ad esempio quando si chiamano le operazioni `SetDesiredCapacity` o `TerminateInstanceInAutoScalingGroup`. Non si applica quando si allegano o si distaccano istanze, si spostano istanze in modalità standby o si elimina il gruppo con l'opzione Forza eliminazione.

Considerazioni e limitazioni per hook del ciclo di vita

Quando lavori con gli hook del ciclo di vita, tieni presente le seguenti note e limitazioni:

- Amazon EC2 Auto Scaling fornisce il proprio ciclo di vita per facilitare la gestione dei gruppi Auto Scaling. Questo ciclo di vita differisce da quello delle altre istanze EC2. Per ulteriori informazioni, consulta [Ciclo di vita delle istanze di Amazon EC2 Auto Scaling](#). Anche le istanze in un warm pool hanno il proprio ciclo di vita, come descritto in [Transizioni dello stato del ciclo di vita delle istanze in un warm pool](#).

- Puoi utilizzare hook del ciclo di vita con istanze Spot, ma questi non impediscono a un'istanza di terminare nel caso in cui quella capacità non sia più disponibile, situazione che può verificarsi in qualsiasi momento, ricevendo un avviso di interruzione due minuti prima. Per ulteriori informazioni, consulta [Interruzioni istanze Spot](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux. Tuttavia, puoi abilitare il ribilanciamento della capacità per sostituire in modo proattivo le istanze Spot che hanno ricevuto una raccomandazione di ribilanciamento dal servizio Spot di Amazon EC2, un segnale che viene inviato quando un'istanza Spot presenta un elevato rischio di interruzione. Per ulteriori informazioni, consulta [Utilizza il ribilanciamento della capacità per gestire le interruzioni spot di Amazon EC2](#).
- Le istanze possono rimanere in stato di attesa per un determinato periodo di tempo. Il timeout di default per un hook del ciclo di vita è di un'ora (timeout dell'heartbeat). È disponibile anche un timeout globale che specifica il periodo di tempo massimo durante il quale è possibile mantenere un'istanza in uno stato di attesa. Il timeout globale è pari a 48 ore o 100 volte il timeout dell'heartbeat, a seconda di quale sia il valore minore.
- Il risultato dell'hook del ciclo di vita può essere `Abbandona` o `Continua`. Se un'istanza viene avviata, `"continue"` (continua) indica che le operazioni hanno avuto esito positivo e che Dimensionamento automatico Amazon EC2 può mettere l'istanza in servizio. In caso contrario, `ABBANDONA` indica che le operazioni personalizzate hanno avuto esito negativo e che l'istanza può essere terminata e sostituita. Se un'istanza sta terminando, sia `abandon` (abbandona) che `continue` (continua) permettono la terminazione dell'istanza. Tuttavia, `abandon` (abbandona) interrompe qualsiasi operazione restante, ad esempio altri hook del ciclo di vita, mentre `continue` (continua) permette il completamento di qualsiasi altro hook del ciclo di vita.
- Se gli hook del ciclo di vita hanno continuamente esito negativo, Amazon EC2 Auto Scaling limita la velocità con cui consente l'avvio delle istanze, pertanto assicurati di verificare e risolvere eventuali problemi permanenti nelle operazioni dei tuoi cicli di vita.
- La creazione e l'aggiornamento degli hook del ciclo di vita utilizzando o un SDK forniscono opzioni non disponibili quando si crea un lifecycle hook da AWS CLI AWS CloudFormation AWS Management Console Ad esempio, il campo per specificare l'ARN di un argomento SNS o di una coda SQS non viene visualizzato nella console, perché Amazon EC2 Auto Scaling invia già eventi ad Amazon EventBridge. Questi eventi possono essere filtrati e reindirizzati a AWS servizi come Lambda, Amazon SNS e Amazon SQS in base alle esigenze.
- Puoi aggiungere più hook del ciclo di vita a un gruppo di Auto Scaling durante la creazione, chiamando l'[CreateAutoScalingGroup](#) API utilizzando, o un SDK. AWS CLI AWS CloudFormation Tuttavia, ogni hook deve avere la stessa destinazione di notifica e ruolo IAM, se specificati. Per

creare lifecycle hook con obiettivi di notifica diversi e ruoli diversi, crea i lifecycle hook uno alla volta in chiamate separate all'API. [PutLifecycleHook](#)

- Se aggiungi un hook del ciclo di vita per l'avvio, il periodo di tolleranza del controllo dell'integrità inizia non appena l'istanza raggiunge lo stato `InService`. Per ulteriori informazioni, consulta [Impostazione del periodo di tolleranza del controllo dell'integrità per un gruppo con scalabilità automatica](#).

Considerazioni sul dimensionamento

- Le politiche di scalabilità dinamica si espandono e riducono in risposta ai dati CloudWatch metrici, come CPU e I/O di rete, aggregati su più istanze. Durante il dimensionamento orizzontale, Dimensionamento automatico Amazon EC2 non conta immediatamente una nuova istanza per i parametri aggregati delle istanze del gruppo con dimensionamento automatico. Attende che l'istanza raggiunga lo stato `InService` e che la preparazione dell'istanza sia terminata. Per ulteriori informazioni, consulta l'argomento della preparazione dell'istanza di default, consulta [Considerazioni sulle prestazioni di dimensionamento](#).
- Su scala ridotta, i parametri aggregati delle istanze potrebbero non riflettere immediatamente la rimozione di un'istanza in fase di terminazione. L'istanza in fase di terminazione smette di contare per i parametri aggregati delle istanze del gruppo poco dopo l'inizio del flusso di lavoro di terminazione di Dimensionamento automatico Amazon EC2.
- Nella maggior parte dei casi, se vengono richiamati hook del ciclo di vita, le attività di dimensionamento basate su policy di dimensionamento semplice vengono sospese fino a quando le operazioni del ciclo di vita non sono completate e il tempo di raffreddamento è trascorso. Se imposti un intervallo lungo per il tempo di raffreddamento, la ripresa del dimensionamento richiede più tempo. Per ulteriori informazioni, consulta [Ulteriori ritardi causati dagli hook del ciclo di vita](#) nell'argomento `Attesa`. In generale, ti consigliamo di non utilizzare policy di dimensionamento semplici se puoi utilizzare invece policy di dimensionamento a fasi o di tracciamento mirate.

Risorse correlate

Per un video introduttivo, vedi [AWS re:Invent 2018: la gestione della capacità semplificata con Amazon EC2 Auto Scaling on YouTube](#)

Forniamo alcuni frammenti di template JSON e YAML che puoi usare per capire come dichiarare gli hook del ciclo di vita nei tuoi modelli di stack. AWS CloudFormation Per ulteriori

informazioni, consulta il riferimento nella Guida per l'utente. [AWS::AutoScaling::LifecycleHook](#) AWS CloudFormation

Puoi anche visitare il nostro [GitHub repository](#) per scaricare modelli di esempio e script di dati utente per i lifecycle hook.

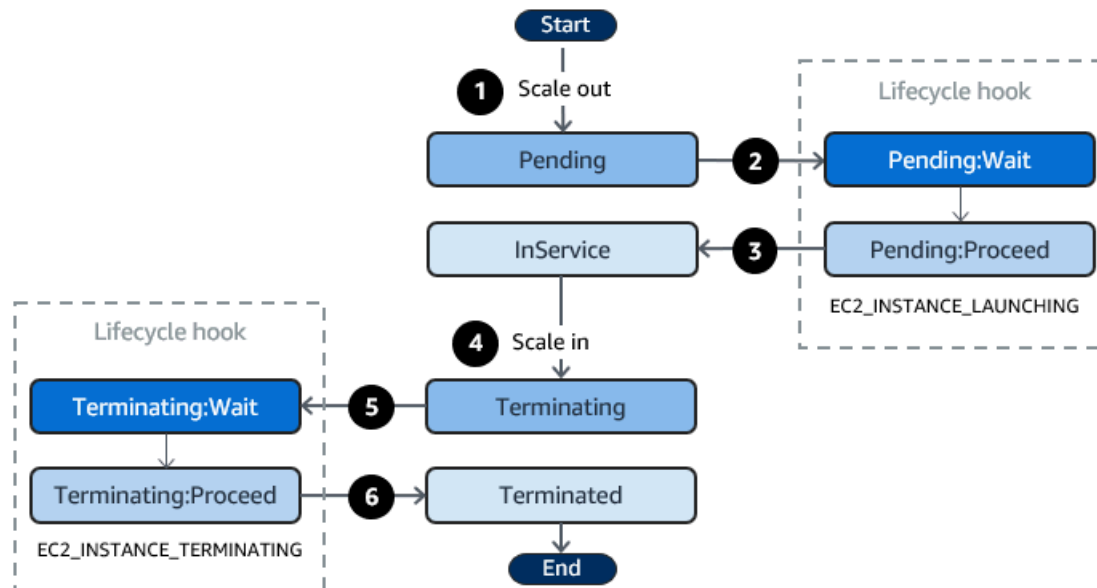
Per esempi sull'uso degli hook del ciclo di vita, consulta il seguente post sul blog.

- [Creazione di un sistema di backup per istanze dimensionate utilizzando il comando di esecuzione Lambda e Amazon EC2](#)
- [Esegui il codice prima di terminare un'istanza di dimensionamento automatico EC2.](#)

Come funzionano gli hook del ciclo di vita

Dal momento in cui la si avvia fino alla sua terminazione, un'istanza Amazon EC2 passa attraverso stati diversi. Puoi creare operazioni personalizzate affinché il tuo gruppo con dimensionamento automatico esegua operazioni quando un'istanza passa in uno stato di attesa a causa di un hook del ciclo di vita.

L'illustrazione seguente mostra le transizioni tra gli stati delle istanze di Auto Scaling quando si utilizzano i lifecycle hook per la scalabilità orizzontale e la scalabilità in entrata.



Come mostrato nel diagramma precedente:

1. Il gruppo con scalabilità automatica risponde a un evento di aumento orizzontale e inizia l'avvio di un'istanza.

2. L'hook del ciclo di vita mette l'istanza in uno stato di attesa (`Pending:Wait`) e poi esegue un'operazione personalizzata.

L'istanza rimane in stato di attesa fino a quando non hai completato l'operazione del ciclo di vita oppure finché non termina il periodo di timeout. Per impostazione predefinita, l'istanza rimane in stato di attesa per un'ora, dopo di che il gruppo con scalabilità automatica continua il processo di avvio o di terminazione (`Pending:Proceed`). Se ti serve più tempo, puoi riavviare il periodo di timeout mediante la registrazione di un heartbeat. Se completi l'operazione del ciclo di vita quando l'operazione personalizzata è stata completata e prima della scadenza del periodo di timeout, il periodo termina e il gruppo con scalabilità automatica continua il processo di avvio.

3. L'istanza entra nello stato `InService` e inizia il periodo di tolleranza del controllo dell'integrità. Tuttavia, prima che l'istanza raggiunga lo stato `InService`, se il gruppo con scalabilità automatica è associato a un load balancer Elastic Load Balancer, l'istanza viene registrata con il load balancer, che inizia a controllarne l'integrità. Al termine del periodo di tolleranza del controllo dell'integrità, Amazon EC2 Auto Scaling inizia a controllare lo stato d'integrità dell'istanza.
4. Il gruppo con scalabilità automatica risponde a un evento di riduzione orizzontale e inizia l'avvio di un'istanza. Se il gruppo con scalabilità automatica viene usato con Elastic Load Balancing, l'istanza di terminazione viene prima deregistrata dal load balancer. Se per il load balancer è abilitato lo svuotamento connessione, prima di completare il processo di deregistrazione l'istanza interrompe l'accettazione di nuove connessioni e attende che quelle esistenti vengano scaricate.
5. L'hook del ciclo di vita mette l'istanza in uno stato di attesa (`Terminating:Wait`) e poi esegue un'operazione personalizzata.

L'istanza rimarrà in stato di attesa finché non avrai completato l'operazione del ciclo di vita oppure finché il periodo di timeout (un'ora per impostazione predefinita) non sarà terminato. Dopo aver completato l'hook del ciclo di vita o il periodo di timeout sarà scaduto, l'istanza passa allo stato successivo (`Terminating:Proceed`).

6. L'istanza è terminata.

Important

Anche le istanze in un warm pool hanno il proprio ciclo di vita con gli stati di attesa corrispondenti, come descritto in [Transizioni dello stato del ciclo di vita delle istanze in un warm pool](#).

Preparazione all'aggiunta di un hook del ciclo di vita al gruppo con scalabilità automatica

Prima di aggiungere un hook del ciclo di vita al gruppo con scalabilità automatica, assicurati che lo script di dati utente o la destinazione di notifica siano configurati correttamente.

- Per utilizzare uno script di dati utente che consenta l'esecuzione di operazioni personalizzate sulle istanze durante l'avvio, non è necessario configurare una destinazione di notifica. Tuttavia, devi aver già creato il modello di avvio o la configurazione di avvio che specifichino lo script di dati utente e averli associati al gruppo con scalabilità automatica. Per maggiori informazioni sugli script di dati utente, consulta [Run commands on your Linux instance at launch](#) (Esecuzione di comandi sull'istanza Linux all'avvio) nella Guida dell'utente di Amazon EC2 per le istanze Linux.
- Per segnalare ad Amazon EC2 Auto Scaling quando l'azione del ciclo di vita è completa, devi aggiungere [CompleteLifecycleAction](#) la chiamata API allo script e devi creare manualmente un ruolo IAM con una policy che consenta alle istanze di Auto Scaling di chiamare questa API. Il modello di avvio o la configurazione di avvio devono specificare questo ruolo utilizzando un profilo dell'istanza IAM associato alle istanze Amazon EC2 al momento dell'avvio. Per ulteriori informazioni, consulta [Completare un'operazione del ciclo di vita](#) e [Ruoli IAM per le applicazioni in esecuzione sulle istanze Amazon EC2](#).
- Per utilizzare un servizio come Lambda per eseguire un'azione personalizzata, è necessario aver già creato una EventBridge regola e specificato una funzione Lambda come destinazione. Per ulteriori informazioni, consulta [Configurazione di una destinazione di notifica per le notifiche del ciclo di vita](#).
- Per consentire a Lambda di segnalare Amazon EC2 Auto Scaling quando l'azione del ciclo di vita è completa, devi [CompleteLifecycleAction](#) aggiungere la chiamata API al codice della funzione. Inoltre, devi avere associato al ruolo di esecuzione della funzione una policy IAM che conceda a Lambda l'autorizzazione per completare le operazioni del ciclo di vita. Per ulteriori informazioni, consulta [Tutorial : Configurazione di un hook del ciclo di vita che richiama una funzione Lambda](#).
- Per utilizzare un servizio come Amazon SNS o Amazon SQS per eseguire un'operazione personalizzata, devi aver già creato l'argomento SNS o la coda SQS e avere pronto il relativo Amazon Resource Name (ARN). Inoltre, devi aver già creato il ruolo IAM che offra ad Amazon EC2 Auto Scaling l'accesso all'argomento SNS o alla destinazione SQS e avere pronto il relativo ARN. Per ulteriori informazioni, consulta [Configurazione di una destinazione di notifica per le notifiche del ciclo di vita](#).

Note

Per impostazione predefinita, quando aggiungi un hook del ciclo di vita nella console, Amazon EC2 Auto Scaling invia notifiche sugli eventi del ciclo di vita ad Amazon EventBridge. L'utilizzo EventBridge o di uno script di dati utente è una best practice consigliata. Per creare un lifecycle hook che invii notifiche direttamente ad Amazon SNS o Amazon SQS, usa AWS CloudFormation, o un SDK per aggiungere AWS CLI l'hook del ciclo di vita.

Configurazione di una destinazione di notifica per le notifiche del ciclo di vita

È possibile aggiungere hook del ciclo di vita a un gruppo con scalabilità automatica per eseguire operazioni personalizzate quando un'istanza entra in stato di attesa. Puoi scegliere un servizio di destinazione in modo che queste operazioni vengano eseguite secondo il tuo approccio di sviluppo preferito.

Il primo approccio utilizza Amazon EventBridge per richiamare una funzione Lambda che esegue l'azione desiderata. Il secondo approccio prevede la creazione di un argomento Amazon Simple Notification Service (Amazon SNS) su cui vengono pubblicate le notifiche. I client possono sottoscrivere all'argomento SNS e ricevere messaggi pubblicati utilizzando un protocollo supportato. L'ultimo approccio prevede l'utilizzo di Amazon Simple Queue Service (Amazon SQS), un sistema di messaggistica utilizzato dalle applicazioni distribuite per scambiare messaggi tramite un modello di polling.

Come best practice, ti consigliamo di utilizzare EventBridge. Le notifiche inviate ad Amazon SNS e Amazon SQS contengono le stesse informazioni delle notifiche a cui invia Amazon EC2 Auto Scaling. In precedenza EventBridge, la pratica standard consisteva nell'invia una notifica a SNS o SQS e integrare un altro servizio con SNS o SQS per eseguire azioni programmatiche. Oggi, EventBridge offre più opzioni per i servizi a cui rivolgersi e semplifica la gestione degli eventi utilizzando un'architettura serverless.

Le procedure seguenti illustrano come impostare la destinazione di notifica.

Ricorda che se disponi di uno script di dati utente nel modello di avvio o nella configurazione di avvio che configura le istanze al momento dell'avvio, non è necessario ricevere una notifica per eseguire operazioni personalizzate sulle istanze.

Indice

- [Indirizza le notifiche a Lambda utilizzando EventBridge](#)
- [Ricezione di notifiche tramite Amazon SNS](#)
- [Ricezione di notifiche tramite Amazon SQS](#)
- [Esempio di messaggio di notifica per Amazon SNS e Amazon SQS](#)

Important

La EventBridge regola, la funzione Lambda, l'argomento Amazon SNS e la coda Amazon SQS che usi con i lifecycle hook devono sempre trovarsi nella stessa regione in cui hai creato il gruppo Auto Scaling.

Indirizza le notifiche a Lambda utilizzando EventBridge

È possibile configurare una EventBridge regola per richiamare una funzione Lambda quando un'istanza entra in uno stato di attesa. Amazon EC2 Auto Scaling invia una notifica degli eventi del ciclo di vita sull'istanza in fase di avvio o chiusura e un token che puoi usare EventBridge per controllare l'azione del ciclo di vita. Per esempi di questi eventi, consulta [Riferimento eventi Dimensionamento automatico Amazon EC2](#).

Note

Quando usi AWS Management Console per creare una regola di evento, la console aggiunge automaticamente le autorizzazioni IAM necessarie per concedere l' EventBridge autorizzazione a chiamare la tua funzione Lambda. Se crei una regola di evento utilizzando la AWS CLI, dovrai concedere tali autorizzazioni in modo esplicito.

Per informazioni su come creare regole di evento nella EventBridge console, consulta [Creazione di EventBridge regole Amazon che reagiscono agli eventi](#) nella Amazon EventBridge User Guide.

- oppure -

Per un tutorial introduttivo rivolto agli utenti della console, consulta [Tutorial : Configurazione di un hook del ciclo di vita che richiama una funzione Lambda](#). Questo tutorial mostra come creare una semplice funzione Lambda che ascolta gli eventi di lancio e li scrive in un CloudWatch registro Logs.

Per creare una EventBridge regola che richiami una funzione Lambda

1. Crea una funzione Lambda utilizzando la [console Lambda](#) e prendi nota del relativo Amazon Resource Name (ARN). Ad esempio, `arn:aws:lambda:region:123456789012:function:my-function`. È necessario l'ARN per creare un EventBridge obiettivo. Per ulteriori informazioni, consulta [Nozioni di base su Lambda](#) nella Guida per gli sviluppatori AWS Lambda .
2. Per creare una regola che corrisponda agli eventi per l'avvio delle istanze, utilizza il seguente comando: [put-rule](#).

```
aws events put-rule --name my-rule --event-pattern file://pattern.json --state ENABLED
```

Gli esempi seguenti mostrano `pattern.json` per un'operazione del ciclo di vita per l'avvio di un'istanza. Sostituisci il testo in *corsivo* con il nome del gruppo con scalabilità automatica.

```
{
  "source": [ "aws.autoscaling" ],
  "detail-type": [ "EC2 Instance-launch Lifecycle Action" ],
  "detail": {
    "AutoScalingGroupName": [ "my-asg" ]
  }
}
```

Se il comando viene eseguito correttamente, EventBridge risponde con l'ARN della regola. Prendi nota di questo ARN. Dovrai inserirlo una volta arrivato alla fase 4.

Per creare una regola che corrisponda ad altri eventi, modifica il modello di evento. Per ulteriori informazioni, consulta [Utilizzato EventBridge per gestire gli eventi di Auto Scaling](#).

3. Per specificare la funzione Lambda da utilizzare come destinazione per la regola, usa il seguente comando: [put-target](#).

```
aws events put-targets --rule my-rule --targets
  Id=1,Arn=arn:aws:lambda:region:123456789012:function:my-function
```

Nel comando precedente, *my-rule* è il nome specificato per la regola alla fase 2 e il valore per il parametro `Arn` è l'ARN della funzione che hai creato alla fase 1.

- Per aggiungere autorizzazioni che permettono alla regola di richiamare la funzione Lambda, utilizza il seguente comando Lambda: [add-permission](#). Questo comando affida il EventBridge service principal (`events.amazonaws.com`) e assegna le autorizzazioni alla regola specificata.

```
aws lambda add-permission --function-name my-function --statement-id my-unique-id \
  --action 'lambda:InvokeFunction' --principal events.amazonaws.com --source-arn
  arn:aws:events:region:123456789012:rule/my-rule
```

Nel precedente comando:

- *my-function*, è il nome della funzione Lambda che desideri utilizzi la regola come destinazione.
- *my-unique-id* è un identificatore univoco definito per descrivere l'istruzione nella policy della funzione Lambda.
- `source-arn` è l'ARN della EventBridge regola.

Se eseguirai il comando correttamente, riceverai un output simile al seguente.

```
{
  "Statement": "{\"Sid\":\"my-unique-id\",
    \"Effect\":\"Allow\",
    \"Principal\":{\"Service\":\"events.amazonaws.com\"},
    \"Action\":\"lambda:InvokeFunction\",
    \"Resource\":\"arn:aws:lambda:us-west-2:123456789012:function:my-function\",
    \"Condition\":
      {\"ArnLike\":
        {\"AWS:SourceArn\":
          \"arn:aws:events:us-west-2:123456789012:rule/my-rule\"}}}"
}
```

Il valore di `Statement` è una versione in formato stringa JSON dell'istruzione aggiunta alla policy della funzione Lambda.

- Dopo aver seguito queste istruzioni, passa a [Aggiungere un hook del ciclo di vita](#) come fase successiva.

Ricezione di notifiche tramite Amazon SNS

Per impostare una destinazione di notifica (un argomento SNS) in modo da riceverla quando avviene un'operazione del ciclo di vita, puoi utilizzare Amazon SNS. Amazon SNS invia quindi le notifiche ai destinatari sottoscritti. Fino alla conferma della sottoscrizione, nessuna delle notifiche pubblicate nell'argomento viene inviata alla coda.

Come impostare le notifiche tramite Amazon SNS

1. Crea un argomento Amazon SNS utilizzando la [console Amazon SNS](#) o il seguente comando: [create-topic](#). Assicurati che l'argomento sia nella stessa Regione del gruppo con scalabilità automatica che stai utilizzando. Per ulteriori informazioni, consulta [Nozioni di base su Amazon SNS](#) nella Guida per gli sviluppatori di Amazon Simple Notification Service.

```
aws sns create-topic --name my-sns-topic
```

2. Prendi nota dell'Amazon Resource Name (ARN) dell'argomento, per esempio, `arn:aws:sns:region:123456789012:my-sns-topic`. Ne avrai bisogno per creare l'hook del ciclo di vita.
3. Crea un ruolo di servizio IAM per consentire ad Amazon EC2 Auto Scaling l'accesso al tuo destinatario della notifica Amazon SNS.

Come concedere ad Amazon EC2 Auto Scaling l'accesso al tuo argomento SNS

- a. Apertura della console IAM <https://console.aws.amazon.com/iam/>
- b. Nel riquadro di navigazione sulla sinistra, scegliere Roles (Ruoli).
- c. Scegli Crea ruolo.
- d. In Seleziona tipo di entità attendibile, scegli Servizio AWS .
- e. Per il caso d'uso, in Use cases for other AWS services (Casi d'uso per altri servizi AWS), scegli EC2 Auto Scaling e poi EC2 Auto Scaling Notification Access (Accesso alle notifiche di EC2 Auto Scaling).
- f. Scegli Next (Avanti) due volte per andare alla pagina Name, review, and create (Nomina, verifica e crea).
- g. Per Nome ruolo, immettere un nome per il ruolo, (ad esempio **my-notification-role**), quindi scegliere Crea ruolo.
- h. Nella pagina Roles (Ruoli) scegli il ruolo appena creato per aprire la pagina Summary (Riepilogo). Prendere nota del ruolo ARN. Ad esempio,

`arn:aws:iam::123456789012:role/my-notification-role`. Ne avrai bisogno per creare l'hook del ciclo di vita.

4. Dopo aver seguito queste istruzioni, passa a [Aggiunta di hook del ciclo di vita \(AWS CLI\)](#) come fase successiva.

Ricezione di notifiche tramite Amazon SQS

Per impostare una destinazione di notifica in modo da ricevere le notifiche quando avviene un'operazione del ciclo di vita, puoi utilizzare Amazon SQS. Una coda di tipo consumer deve quindi eseguire il polling di una coda SQS per eseguire operazioni su queste notifiche.

Important

Le code FIFO non sono compatibili con gli hook del ciclo di vita.

Come impostare le notifiche tramite Amazon SQS

1. Crea una coda SQS usando la [console Amazon SQS](#). Assicurati che la coda sia nella stessa Regione del gruppo con scalabilità automatica che stai utilizzando. Per ulteriori informazioni, consulta [Nozioni di base su Amazon SQS](#) nella Guida per gli sviluppatori di Amazon Simple Queue Service.
2. Prendi nota dell'ARN della coda, ad esempio `arn:aws:sqs:us-west-2:123456789012:my-sqs-queue`. Ne avrai bisogno per creare l'hook del ciclo di vita.
3. Crea un ruolo di servizio IAM per consentire ad Amazon EC2 Auto Scaling l'accesso alla tua destinazione di notifica Amazon SQS.

Come concedere ad Amazon EC2 Auto Scaling l'accesso alla tua coda SQS

- a. Apertura della console IAM <https://console.aws.amazon.com/iam/>
- b. Nel riquadro di navigazione sulla sinistra, scegliere Roles (Ruoli).
- c. Scegli Crea ruolo.
- d. In Seleziona tipo di entità attendibile, scegli Servizio AWS .
- e. Per il caso d'uso, in Use cases for other AWS services (Casi d'uso per altri servizi AWS), scegli EC2 Auto Scaling e poi EC2 Auto Scaling Notification Access (Accesso alle notifiche di EC2 Auto Scaling).

- f. Scegli Next (Avanti) due volte per andare alla pagina Name, review, and create (Nomina, verifica e crea).
 - g. Per Nome ruolo, immettere un nome per il ruolo, (ad esempio **my-notification-role**), quindi scegliere Crea ruolo.
 - h. Nella pagina Roles (Ruoli) scegli il ruolo appena creato per aprire la pagina Summary (Riepilogo). Prendere nota del ruolo ARN. Ad esempio, `arn:aws:iam::123456789012:role/my-notification-role`. Ne avrai bisogno per creare l'hook del ciclo di vita.
4. Dopo aver seguito queste istruzioni, passa a [Aggiunta di hook del ciclo di vita \(AWS CLI\)](#) come fase successiva.

Esempio di messaggio di notifica per Amazon SNS e Amazon SQS

Mentre l'istanza è in uno stato di attesa, un messaggio viene pubblicato alla destinazione di notifica Amazon SNS o Amazon SQS. Il messaggio include le informazioni seguenti:

- `LifecycleActionToken`: il token dell'azione del ciclo di vita.
- `AccountId`— L' Account AWS ID.
- `AutoScalingGroupName`: il nome del gruppo Auto Scaling.
- `LifecycleHookName`: il nome dell'hook del ciclo di vita.
- `EC2InstanceId`: l'ID dell'istanza EC2.
- `LifecycleTransition`: il tipo di hook del ciclo di vita.
- `NotificationMetadata`: i metadati di notifica.

Di seguito è riportato un esempio di messaggio di notifica.

```
Service: AWS Auto Scaling
Time: 2021-01-19T00:36:26.533Z
RequestId: 18b2ec17-3e9b-4c15-8024-ff2e8ce8786a
LifecycleActionToken: 71514b9d-6a40-4b26-8523-05e7ee35fa40
AccountId: 123456789012
AutoScalingGroupName: my-asg
LifecycleHookName: my-hook
EC2InstanceId: i-0598c7d356eba48d7
LifecycleTransition: autoscaling:EC2_INSTANCE_LAUNCHING
NotificationMetadata: hook message metadata
```

Esempio di messaggio di notifica di test

Quando si aggiunge per la prima volta un hook del ciclo di vita, viene pubblicato un messaggio di notifica di test nella destinazione di notifica. Di seguito è riportato un esempio di messaggio di notifica di test.

```
Service: AWS Auto Scaling
Time: 2021-01-19T00:35:52.359Z
RequestId: 18b2ec17-3e9b-4c15-8024-ff2e8ce8786a
Event: autoscaling:TEST_NOTIFICATION
AccountId: 123456789012
AutoScalingGroupName: my-asg
AutoScalingGroupARN: arn:aws:autoscaling:us-west-2:123456789012:autoScalingGroup:042cba90-ad2f-431c-9b4d-6d9055bcc9fb:autoScalingGroupName/my-asg
```

Note

Per esempi di eventi forniti da Amazon EC2 Auto Scaling EventBridge a, consulta.

[Riferimento eventi Dimensionamento automatico Amazon EC2](#)

Recupero dello stato del ciclo di vita di destinazione tramite i metadati dell'istanza

Ogni istanza Auto Scaling avviata passa attraverso diversi stati del ciclo di vita. Per richiamare operazioni personalizzate dall'interno di un'istanza che esegue operazioni su specifici passaggi allo stato del ciclo di vita specifico, devi recuperare lo stato del ciclo di vita di destinazione tramite i metadati dell'istanza.

Ad esempio, potrebbe essere necessario un meccanismo per rilevare la terminazione dell'istanza dall'interno dell'istanza per eseguire del codice sull'istanza prima che venga terminata. Puoi farlo scrivendo il codice che esamina lo stato del ciclo di vita di un'istanza direttamente dall'istanza. Se desideri continuare, puoi quindi aggiungere un hook del ciclo di vita al gruppo con dimensionamento automatico per mantenere l'istanza in esecuzione fino all'invio del comando `complete-lifecycle-action` da parte del codice.

Il ciclo di vita dell'istanza Auto Scaling ha due stati stabili primari, `InService` e `Terminated`, e due stati stabili secondari, `Detached` e `Standby`. Se utilizzi un warm pool, il ciclo di vita ha

quattro stati stabili aggiuntivi: `Warmed:Hibernated`, `Warmed:Running`, `Warmed:Stopped` e `Warmed:Terminated`.

Quando un'istanza si prepara a passare a uno degli stati stabili precedenti, Amazon EC2 Auto Scaling aggiorna il valore dell'elemento di metadati dell'istanza `autoscaling/target-lifecycle-state`. Per ottenere lo stato del ciclo di vita di destinazione dall'interno dell'istanza, è necessario utilizzare il servizio di metadati di istanza per recuperarlo dai metadati dell'istanza.

Note

I metadati dell'istanza sono dati relativi a un'istanza Amazon EC2 che le applicazioni possono utilizzare per interrogare l'istanza. Il servizio di metadati di istanza è un componente su istanza utilizzato dal codice locale per accedere ai metadati dell'istanza. Il codice locale può includere script di dati utente o applicazioni in esecuzione sull'istanza.

Il codice locale può accedere ai metadati di istanza da un'istanza in esecuzione utilizzando due metodi alternativi: Instance Metadata Service versione 1 (IMDSv1) o Instance Metadata Service versione 2 (IMDSv2). IMDSv2 utilizza le richieste orientate alla sessione e attenua diversi tipi di vulnerabilità che potrebbero essere utilizzate per tentare di accedere ai metadati dell'istanza. Per informazioni dettagliate su questi due metodi, consulta la sezione [Utilizzo di IMDSv2](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

IMDSv2

```
TOKEN=`curl -X PUT "http://169.254.169.254/latest/api/token" -H "X-aws-ec2-metadata-token-ttl-seconds: 21600" ` \  
&& curl -H "X-aws-ec2-metadata-token: $TOKEN" -v http://169.254.169.254/latest/meta-data/autoscaling/target-lifecycle-state
```

IMDSv1

```
curl http://169.254.169.254/latest/meta-data/autoscaling/target-lifecycle-state
```

Di seguito è riportato un output di esempio.

```
InService
```


Lo stato del ciclo di vita di destinazione è lo stato a cui l'istanza sta passando. Lo stato del ciclo di vita corrente è lo stato in cui si trova l'istanza. I due stati possono coincidere quando l'operazione del ciclo di vita viene completata e l'istanza termina la transizione allo stato del ciclo di vita di destinazione. Non è possibile recuperare lo stato del ciclo di vita corrente dell'istanza dai metadati dell'istanza.

Amazon EC2 Auto Scaling ha iniziato a generare lo stato del ciclo di vita di destinazione il 10 marzo 2022. Se l'istanza passa a uno degli stati del ciclo di vita di destinazione dopo tale data, l'elemento dello stato del ciclo di vita di destinazione è presente nei metadati dell'istanza. In caso contrario, non è presente e ricevi un errore HTTP 404.

Per informazioni sul recupero dei metadati dell'istanza, consulta la sezione [Recupero dei metadati dell'istanza](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Consulta la sezione [Tutorial: configurazione dei dati utente per recuperare lo stato del ciclo di vita di destinazione tramite i metadati dell'istanza](#) per un tutorial che mostra come creare un hook del ciclo di vita con un'operazione personalizzata in uno script di dati utente che utilizza lo stato del ciclo di vita di destinazione.

Important

Per assicurarti di poter richiamare un'azione personalizzata il prima possibile, il codice locale deve interrogare spesso IMDS e riprovare a correggere gli errori.

Aggiungere un hook del ciclo di vita

Per mettere le istanze Auto Scaling in uno stato di attesa ed eseguire operazioni personalizzate su di esse, puoi aggiungere hook del ciclo di vita al gruppo con scalabilità automatica. Le operazioni personalizzate vengono eseguite all'avvio o prima della fine delle istanze. Le istanze rimarranno in stato di attesa finché non avrai completato l'operazione del ciclo di vita oppure finché il periodo di timeout non sarà terminato.

Dopo aver creato un gruppo Auto Scaling da AWS Management Console, è possibile aggiungervi uno o più hook del ciclo di vita, fino a un totale di 50 hook del ciclo di vita. Puoi anche utilizzare AWS CLI, AWS CloudFormation, o un SDK per aggiungere gli hook del ciclo di vita a un gruppo Auto Scaling durante la creazione.

Per impostazione predefinita, quando aggiungi un hook del ciclo di vita nella console, Amazon EC2 Auto Scaling invia notifiche sugli eventi del ciclo di vita ad Amazon. EventBridge L'utilizzo

EventBridge o di uno script di dati utente è una best practice consigliata. Per creare un hook del ciclo di vita che invii notifiche direttamente ad Amazon SNS o Amazon SQS, puoi utilizzare [put-lifecycle-hook](#) il comando, come mostrato negli esempi in questo argomento.

Indice

- [Come aggiungere hook del ciclo di vita \(console\)](#)
- [Aggiunta di hook del ciclo di vita \(AWS CLI\)](#)

Come aggiungere hook del ciclo di vita (console)

Per aggiungere un hook del ciclo di vita a un gruppo con dimensionamento automatico, segui questa procedura. Per aggiungere hook del ciclo di vita per il dimensionamento orizzontale (avvio delle istanze) o il dimensionamento verticale (terminazione delle istanze o restituzione a un pool caldo), devi creare due hook separati.

Prima di iniziare, assicurati di avere impostato un'operazione personalizzata, a seconda delle necessità, come descritto in [Preparazione all'aggiunta di un hook del ciclo di vita al gruppo con scalabilità automatica](#).

Per aggiungere un hook del ciclo di vita per il dimensionamento orizzontale

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico. Si aprirà un riquadro diviso nella parte inferiore della pagina.
3. Nella scheda Instance management (Gestione istanze) in Lifecycle hooks (Hook del ciclo di vita), scegliere Create lifecycle hook (Crea hook del ciclo di vita).
4. Per definire un hook del ciclo di vita per il dimensionamento orizzontale (avvio delle istanze), procedi come descritto di seguito:
 - a. Per Lifecycle hook name (Nome hook del ciclo di vita) specificare un nome per l'hook del ciclo di vita.
 - b. In Lifecycle transition (Transizione ciclo di vita), scegli Instance launch (Avvia istanza).
 - c. Per Timeout heartbeat, specifica quanto a lungo, in secondi, le istanze debbano rimanere in uno stato di attesa durante il dimensionamento orizzontale prima che l'hook scada. L'intervallo è tra 30 e 7200 secondi. L'impostazione di un periodo di timeout esteso fornisce più tempo per il completamento dell'operazione personalizzata. Quindi, se finisci prima

della fine del periodo di timeout, invia il [complete-lifecycle-action](#) comando per consentire all'istanza di passare allo stato successivo.

- d. Per Default result (Risultato di default), specifica l'operazione da intraprendere quando gli hook del ciclo di vita scadono o si verifica un errore imprevisto. Puoi scegliere di CONTINUARE o ABBANDONARE.
 - Se scegli CONTINUA, il gruppo con dimensionamento automatico può procedere con qualsiasi altro hook del ciclo di vita e quindi mettere l'istanza in servizio.
 - Se scegli ABBANDONA, il gruppo con dimensionamento automatico arresta tutte le operazioni rimanenti e termina l'istanza immediatamente.
 - e. (Facoltativo) In Metadati delle notifiche, specifica le altre informazioni che desideri includere ogni volta che Dimensionamento automatico Amazon EC2 invia un messaggio alla destinazione di notifica.
5. Scegli Crea.

Per aggiungere un hook del ciclo di vita per il dimensionamento verticale

1. Scegli Crea hook del ciclo di vita per continuare da dove avevi interrotto dopo aver creato un hook del ciclo di vita per il dimensionamento orizzontale.
2. Per definire un hook del ciclo di vita per il dimensionamento (istanze che terminano o ritornano a un pool caldo), procedi come segue:
 - a. Per Lifecycle hook name (Nome hook del ciclo di vita) specificare un nome per l'hook del ciclo di vita.
 - b. In Lifecycle transition (Transizione ciclo di vita) scegli Instance terminated (Terminazione istanza).
 - c. Per Timeout heartbeat, specifica quanto a lungo, in secondi, le istanze debbano rimanere in uno stato di attesa durante il dimensionamento orizzontale prima che l'hook scada. Ti consigliamo un breve periodo di timeout, 30 da 1 a 120 2 secondi, a seconda del tempo necessario per eseguire le attività finali, come ad esempio estrarre i log EC2. CloudWatch
 - d. In Default result (Risultato di default), specifica l'operazione che il gruppo con scalabilità automatica deve intraprendere una volta scaduto il timeout o se si verifica un errore imprevisto. Sia ABANDON (ABBANDONA) sia CONTINUE (CONTINUA) permettono all'istanza di terminare.

- Se scegli CONTINUE (CONTINUA), il gruppo con scalabilità automatica può procedere con qualsiasi operazione restante, ad esempio altri hook del ciclo di vita, prima della terminazione.
 - Se scegli ABBANDONA, il gruppo con dimensionamento automatico termina l'istanza immediatamente.
- e. (Facoltativo) In Metadati delle notifiche, specifica le altre informazioni che desideri includere ogni volta che Dimensionamento automatico Amazon EC2 invia un messaggio alla destinazione di notifica.
3. Scegli Crea.

Aggiunta di hook del ciclo di vita (AWS CLI)

Crea e aggiorna gli hook del ciclo di vita utilizzando il comando. [put-lifecycle-hook](#)

Per eseguire un'operazione di aumento orizzontale, utilizza il comando seguente.

```
aws autoscaling put-lifecycle-hook --lifecycle-hook-name my-launch-hook \  
  --auto-scaling-group-name my-asg \  
  --lifecycle-transition autoscaling:EC2_INSTANCE_LAUNCHING
```

Per eseguire un'operazione di riduzione orizzontale, utilizza invece il comando seguente.

```
aws autoscaling put-lifecycle-hook --lifecycle-hook-name my-termination-hook \  
  --auto-scaling-group-name my-asg \  
  --lifecycle-transition autoscaling:EC2_INSTANCE_TERMINATING
```

Per ricevere notifiche utilizzando Amazon SNS o Amazon SQS, aggiungi le opzioni `--notification-target-arn` e `--role-arn`.

Nell'esempio seguente viene creato un hook del ciclo di vita che specifica un argomento SNS denominato *my-sns-topic* come destinazione di notifica.

```
aws autoscaling put-lifecycle-hook --lifecycle-hook-name my-termination-hook \  
  --auto-scaling-group-name my-asg \  
  --lifecycle-transition autoscaling:EC2_INSTANCE_TERMINATING \  
  --notification-target-arn arn:aws:sns:region:123456789012:my-sns-topic \  
  --role-arn arn:aws:iam::123456789012:role/my-notification-role
```

L'argomento riceve una notifica di test con la seguente coppia chiave-valore:

```
"Event": "autoscaling:TEST_NOTIFICATION"
```

Per impostazione predefinita, il [put-lifecycle-hook](#) comando crea un hook del ciclo di vita con un timeout del battito cardiaco di secondi (un'ora). 3600

Per modificare il timeout dell'heartbeat per un hook del ciclo di vita esistente, aggiungere l'opzione `--heartbeat-timeout`, come mostrato nell'esempio seguente.

```
aws autoscaling put-lifecycle-hook --lifecycle-hook-name my-termination-hook \  
  --auto-scaling-group-name my-asg --heartbeat-timeout 120
```

Se un'istanza è già in stato di attesa, puoi impedire il timeout del lifecycle hook registrando un heartbeat, utilizzando il comando CLI. [record-lifecycle-action-heartbeat](#) Ciò estende il periodo di timeout del valore di timeout specificato al momento della creazione dell'hook del ciclo di vita. Se finisci prima della fine del periodo di timeout, puoi inviare il comando [complete-lifecycle-action](#) CLI per consentire all'istanza di passare allo stato successivo. Per maggiori informazioni ed esempi, consulta [Completare un'operazione del ciclo di vita](#).

Completare un'operazione del ciclo di vita

Quando un gruppo con scalabilità automatica risponde a un evento del ciclo di vita, mette l'istanza in stato di attesa e invia una notifica di evento. Mentre l'istanza è in stato di attesa, è possibile eseguire un'operazione personalizzata.

Completare l'azione del ciclo di vita con il risultato di CONTINUE è utile se termini prima della scadenza del periodo di timeout. Se non completi l'azione del ciclo di vita, l'hook del ciclo di vita passa allo stato specificato per Risultato predefinito al termine del periodo di timeout.

Indice

- [Completare un'operazione del ciclo di vita \(manuale\)](#)
- [Completare un'operazione del ciclo di vita \(automatico\)](#)

Completare un'operazione del ciclo di vita (manuale)

La procedura seguente è relativa all'interfaccia della riga di comando e non è supportata nella console. Le informazioni che devono essere sostituite, ad esempio l'ID dell'istanza o il nome di un gruppo con scalabilità automatica, vengono visualizzate in corsivo.

Completamento di un'operazione del ciclo di vita (AWS CLI)

1. Se hai bisogno di più tempo per completare l'azione personalizzata, usa il [record-lifecycle-action-heartbeat](#) comando per riavviare il periodo di timeout e mantenere l'istanza in uno stato di attesa. Ad esempio, se il periodo di timeout è di un'ora e si chiama questo comando dopo 30 minuti, l'istanza rimane in uno stato di attesa per un'ulteriore ora, ossia per un totale di 90 minuti.

Puoi specificare il token di operazione del ciclo di vita ricevuto con la [notifica](#), come illustrato nel comando seguente.

```
aws autoscaling record-lifecycle-action-heartbeat --lifecycle-hook-name my-launch-hook \  
  --auto-scaling-group-name my-asg --lifecycle-action-token bcd2f1b8-9a78-44d3-8a7a-4dd07d7cf635
```

In alternativa, puoi specificare l'ID dell'istanza ricevuta con la [notifica](#) come mostrato nel comando seguente.

```
aws autoscaling record-lifecycle-action-heartbeat --lifecycle-hook-name my-launch-hook \  
  --auto-scaling-group-name my-asg --instance-id i-1a2b3c4d
```

2. Se completate l'azione personalizzata prima della fine del periodo di timeout, utilizzate il [complete-lifecycle-action](#) comando in modo che il gruppo Auto Scaling possa continuare ad avviare o terminare l'istanza. Puoi specificare il token dell'operazione del ciclo di vita, come mostrato nel comando seguente.

```
aws autoscaling complete-lifecycle-action --lifecycle-action-result CONTINUE \  
  --lifecycle-hook-name my-launch-hook --auto-scaling-group-name my-asg \  
  --lifecycle-action-token bcd2f1b8-9a78-44d3-8a7a-4dd07d7cf635
```

In alternativa, puoi specificare l'ID dell'istanza, come mostrato nel comando seguente.

```
aws autoscaling complete-lifecycle-action --lifecycle-action-result CONTINUE \  
  --instance-id i-1a2b3c4d --lifecycle-hook-name my-launch-hook \  
  --auto-scaling-group-name my-asg
```

Completare un'operazione del ciclo di vita (automatico)

Se disponi di uno script di dati utente che configura le istanze dopo l'avvio, non è necessario completare manualmente le operazioni del ciclo di vita. È possibile aggiungere il [complete-lifecycle-action](#) comando allo script. Lo script può recuperare l'ID dell'istanza dai metadati dell'istanza e segnalare ad Amazon EC2 Auto Scaling quando gli script di bootstrap sono stati completati correttamente.

Se non lo stai già facendo, aggiorna lo script per recuperare l'ID istanza dai metadati dell'istanza. Per ulteriori informazioni, consulta la sezione [Recupero dei metadati dell'istanza](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

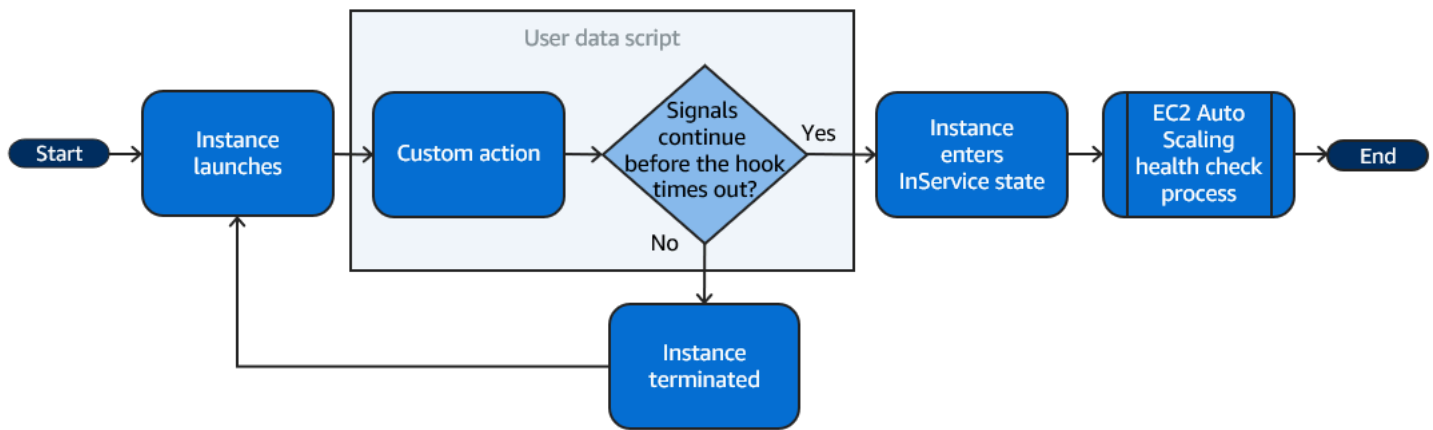
Se utilizzi Lambda, puoi anche impostare un callback nel codice della funzione per consentire al ciclo di vita dell'istanza di procedere se l'operazione personalizzata viene eseguita correttamente. Per ulteriori informazioni, consulta [Tutorial : Configurazione di un hook del ciclo di vita che richiama una funzione Lambda](#).

Tutorial: configurazione dei dati utente per recuperare lo stato del ciclo di vita di destinazione tramite i metadati dell'istanza

Un modo comune per creare azioni personalizzate per i lifecycle hook consiste nell'utilizzare le notifiche che Amazon EC2 Auto Scaling invia ad altri servizi, come Amazon EventBridge. Tuttavia, puoi evitare di dover creare un'infrastruttura aggiuntiva utilizzando invece uno script di dati utente per spostare il codice che configura le istanze e completa l'operazione del ciclo di vita nelle istanze stesse.

Il seguente tutorial illustra le nozioni di base su come utilizzare uno script di dati utente e i metadati dell'istanza. È possibile creare una configurazione di base del gruppo con scalabilità automatica con uno script di dati utente che legge lo [stato del ciclo di vita di destinazione](#) delle istanze del gruppo ed esegue un'operazione di callback in una fase specifica del ciclo di vita di un'istanza per continuare il processo di avvio.

La seguente illustrazione riassume il flusso di un evento di scalabilità orizzontale quando si utilizza uno script di dati utente per eseguire un'azione personalizzata. Dopo l'avvio di un'istanza, il ciclo di vita dell'istanza viene sospeso fino al completamento dell'hook del ciclo di vita, tramite timeout o facendo sì che Amazon EC2 Auto Scaling riceva un segnale di continuazione.



Indice

- [Passaggio 1: creazione di un ruolo IAM con autorizzazioni per completare le operazioni del ciclo di vita](#)
- [Passaggio 2: creazione di un modello di avvio e inclusione del ruolo IAM e di uno script di dati utente](#)
- [Passaggio 3: creazione di un gruppo con dimensionamento automatico](#)
- [Passaggio 4: aggiunta di un hook del ciclo di vita](#)
- [Passaggio 5: test e verifica della funzionalità](#)
- [Fase 6: pulizia](#)
- [Risorse correlate](#)

Passaggio 1: creazione di un ruolo IAM con autorizzazioni per completare le operazioni del ciclo di vita

Quando utilizzi l'SDK AWS CLI o un AWS SDK per inviare un callback per completare le azioni del ciclo di vita, devi utilizzare un ruolo IAM con autorizzazioni per completare le azioni del ciclo di vita.

Come creare la policy

1. Nella console IAM, apri [Policies \(Policy\)](#), quindi seleziona Create policy (Crea policy).
2. Scegli la scheda JSON.
3. Nella casella Policy Document (Documento della policy), copia e incolla il documento della policy seguente nella casella. Sostituisci il *testo di esempio* con il numero di account e il nome del gruppo con scalabilità automatica che desideri creare (**TestAutoScalingEvent-group**).


```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "autoscaling:CompleteLifecycleAction"
      ],
      "Resource":
        "arn:aws:autoscaling:*:123456789012:autoScalingGroup:*:autoScalingGroupName/TestAutoScalingEvent-group"
    }
  ]
}
```

4. Seleziona Successivo.
5. In Policy name (Nome policy), inserisci **TestAutoScalingEvent-policy**. Scegli Create Policy (Crea policy).

Dopo aver creato la policy, puoi creare un ruolo che la utilizzi.

Come creare il ruolo

1. Nel riquadro di navigazione sulla sinistra, scegliere Roles (Ruoli).
2. Scegli Crea ruolo.
3. In Seleziona tipo di entità attendibile, scegli Servizio AWS .
4. Per il tuo caso d'uso, scegli EC2 quindi scegli Successivo.
5. In Aggiungi autorizzazioni, scegli la policy che hai creato (-policy). TestAutoScalingEvent Quindi, seleziona Next (Successivo).
6. Nella pagina Nome, revisione e creazione, per Nome ruolo, inserisci **TestAutoScalingEvent-role** e scegli Crea ruolo.

Passaggio 2: creazione di un modello di avvio e inclusione del ruolo IAM e di uno script di dati utente

Crea un modello di avvio da utilizzare con un gruppo con scalabilità automatica. Includi il ruolo IAM creato e lo script di dati utente di esempio fornito.

Per creare un modello di avvio

1. Apri la pagina [Configurazioni di avvio](#) della console Amazon EC2.
2. Scegli Crea modello di avvio.
3. In Launch template name (Nome del modello di) avvio, inserisci **TestAutoScalingEvent-template**.
4. In Auto Scaling guidance (Guida Auto Scaling), seleziona la casella di controllo.
5. Per Application and OS Images (Amazon Machine Image) (Immagini di applicazioni e SO (Amazon Machine Image)), scegli Amazon Linux 2 (HVM), tipo di volume SSD, 64 bit (x86) dall'elenco Avvio rapido.
6. Per Instance type (Tipo di istanza), scegli un tipo di istanza Amazon EC2 (ad esempio, "t2.micro").
7. In Advanced details (Dettagli avanzati), espandi la sezione per visualizzare i campi.
8. Per il profilo dell'istanza IAM, scegli il nome del profilo dell'istanza IAM del tuo ruolo IAM (TestAutoScalingEvent-role). Un profilo dell'istanza è un container per un ruolo IAM che permette ad Amazon EC2 di trasferire il ruolo IAM a un'istanza quando quest'ultima viene avviata.

Se hai creato un ruolo IAM utilizzando la console IAM, la console ha creato automaticamente un profilo dell'istanza assegnandogli lo stesso nome del ruolo a cui corrisponde.

9. Nel campo User data (Dati utente), incolla lo script riportato di seguito. Sostituisci il testo di esempio `group_name` con il nome del gruppo Auto Scaling che desiderate creare e `region` con quello che Regione AWS desiderate venga utilizzato dal gruppo Auto Scaling.

```
#!/bin/bash

function get_target_state {
    echo $(curl -s http://169.254.169.254/latest/meta-data/autoscaling/target-lifecycle-state)
}

function get_instance_id {
    echo $(curl -s http://169.254.169.254/latest/meta-data/instance-id)
}

function complete_lifecycle_action {
    instance_id=$(get_instance_id)
    group_name='TestAutoScalingEvent-group'
    region='us-west-2'
```

```
echo $instance_id
echo $region
echo $(aws autoscaling complete-lifecycle-action \
  --lifecycle-hook-name TestAutoScalingEvent-hook \
  --auto-scaling-group-name $group_name \
  --lifecycle-action-result CONTINUE \
  --instance-id $instance_id \
  --region $region)
}

function main {
  while true
  do
    target_state=$(get_target_state)
    if [ \"$target_state\" = \"InService\" ]; then
      # Change hostname
      export new_hostname=\"${group_name}-${instance_id}\"
      hostname $new_hostname
      # Send callback
      complete_lifecycle_action
      break
    fi
    echo $target_state
    sleep 5
  done
}

main
```

Questo script di dati utente d'esempio svolge le seguenti operazioni:

- Chiama i metadati dell'istanza per recuperare lo stato del ciclo di vita di destinazione e l'ID dell'istanza dai metadati dell'istanza
- Recupera ripetutamente lo stato del ciclo di vita di destinazione fino a quando non cambia in InService
- Cambia il nome host dell'istanza antepoendo l'ID istanza al nome del gruppo con scalabilità automatica, se lo stato del ciclo di vita di destinazione è InService
- Invia un callback chiamando il comando complete-lifecycle-action della CLI per segnalare ad Amazon EC2 Auto Scaling di CONTINUE (continuare) il processo di avvio di EC2

10. Scegli Crea modello di avvio.

11. Nella pagina di conferma, scegli Crea gruppo con dimensionamento automatico).

Note

Per altri esempi che puoi usare come riferimento per lo sviluppo dello script dei dati utente, consulta il [GitHub repository](#) per Amazon EC2 Auto Scaling.

Passaggio 3: creazione di un gruppo con dimensionamento automatico

Dopo avere creato il modello di avvio, crea un gruppo con scalabilità automatica.

Creazione di un gruppo con scalabilità automatica

1. Nella pagina Choose launch template or configuration (Scegli il modello o la configurazione di avvio), per Auto Scaling group name (Nome gruppo con scalabilità automatica) inserisci un nome per il gruppo con scalabilità automatica (**TestAutoScalingEvent-group**).
2. Scegli Next (Successivo) per andare alla pagina Choose instance launch options (Scegli le opzioni di avvio dell'istanza).
3. In Network (Rete), scegli un VPC.
4. Per Availability Zones and subnets (Zone di disponibilità e sottoreti), scegli una o più sottoreti da una o più zone di disponibilità.
5. Nella sezione Instance type requirements (Requisiti del tipo di istanza), utilizza l'impostazione di default per semplificare questo passaggio. (Non sovrascrivere il modello di avvio.) Per questo tutorial, avvierai solo un'istanza on demand utilizzando il tipo di istanza specificato nel modello di avvio.
6. Scegli Skip to review (Passa alla revisione) nella parte inferiore della schermata.
7. Nella pagina Review (Revisione), rivedi le impostazioni del gruppo con scalabilità automatica, quindi scegli Create Auto Scaling group (Crea gruppo con scalabilità automatica).

Passaggio 4: aggiunta di un hook del ciclo di vita

Aggiungi un hook del ciclo di vita per mantenere l'istanza in uno stato di attesa fino al completamento dell'operazione del ciclo di vita.

Per aggiungere un hook del ciclo di vita

1. Apri la [pagina Gruppi con dimensionamento automatico](#) nella console Amazon EC2.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico. Si aprirà un riquadro diviso nella parte inferiore della pagina.
3. Nel riquadro inferiore, nella scheda Instance management (Gestione istanze) in Lifecycle hooks (Hook del ciclo di vita), scegli Create lifecycle hook (Crea hook del ciclo di vita).
4. Per definire un hook del ciclo di vita per il dimensionamento orizzontale (avvio delle istanze), procedi come descritto di seguito:
 - a. In Lifecycle hook name (Nome hook del ciclo di vita), inserisci **TestAutoScalingEvent-hook**.
 - b. In Lifecycle transition (Transizione ciclo di vita), scegli Instance launch (Avvia istanza).
 - c. In Heartbeat timeout (Timeout dell'heartbeat), inserisci **300** per il numero di secondi di attesa del callback dallo script di dati utente.
 - d. In Default result (Risultato di default), scegli ABANDON (ABBANDONA). Se l'hook va in timeout senza ricevere un callback dallo script dei dati utente, il gruppo con scalabilità automatica termina la nuova istanza.
 - e. (Facoltativo) Lascia vuoto il campo Notification metadata (Metadati della notifica).
5. Scegli Crea.

Passaggio 5: test e verifica della funzionalità

Per effettuare il test della funzionalità, aggiorna il gruppo con scalabilità automatica aumentando di 1 la capacità desiderata del gruppo con scalabilità automatica. Lo script dei dati utente viene eseguito e inizia a controllare lo stato del ciclo di vita di destinazione dell'istanza subito dopo l'avvio dell'istanza. Lo script modifica il nome host e invia un'operazione di callback quando lo stato del ciclo di vita di destinazione è InService. Solitamente, il completamento di questo processo richiede solo pochi secondi.

Per aumentare le dimensioni del gruppo con dimensionamento automatico

1. Apri la [pagina Gruppi con dimensionamento automatico](#) nella console Amazon EC2.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico. Puoi visualizzare i dettagli in un pannello inferiore mentre continui a visualizzare le righe superiori del pannello superiore.

3. Nel riquadro inferiore, nella scheda Details (Dettagli) scegli Group details (Dettagli gruppo), Edit (Modifica).
4. Per Desired capacity (Capacità desiderata), aumentare il valore corrente di 1.
5. Scegli Aggiorna. Durante l'avvio o la terminazione di un'istanza, la colonna Status (Stato) visualizza lo stato Updating capacity (Aggiornamento della capacità).

Dopo avere aumentato la capacità desiderata, puoi verificare che l'istanza sia stata avviata correttamente e che non sia terminata dalla descrizione delle attività di dimensionamento.

Come visualizzare le attività di dimensionamento

1. Dalla pagina Auto Scaling groups (Gruppi Auto Scaling), seleziona il gruppo.
2. Nella scheda Activity (Attività) in Activity history (Cronologia attività), la colonna Status (Stato) indica se il gruppo con scalabilità automatica ha avviato correttamente le istanze.
3. Se lo script di dati utente non riesce, al termine del periodo di timeout viene visualizzata un'attività di dimensionamento con uno stato di Canceled e un messaggio di stato di Instance failed to complete user's Lifecycle Action: Lifecycle Action with token e85eb647-4fe0-4909-b341-a6c42EXAMPLE was abandoned: Lifecycle Action Completed with ABANDON Result.

Fase 6: pulizia

Se hai finito di utilizzare le risorse create appositamente per il tutorial, segui questa procedura per eliminarle.

Come eliminare l'hook del ciclo di vita

1. Apri la [pagina Gruppi con dimensionamento automatico](#) nella console Amazon EC2.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.
3. Nella scheda Instance management (Gestione istanze) in Lifecycle hooks (Hook del ciclo di vita), scegli Create lifecycle hook (Crea hook del ciclo di vita)(TestAutoScalingEvent-hook).
4. Scegli Actions (Operazioni), Delete (Elimina).
5. Per confermare, scegli di nuovo Delete (Elimina).

Per eliminare il modello di avvio

1. Apri la pagina [Configurazioni di avvio](#) della console Amazon EC2.
2. Seleziona il modello di avvio (TestAutoScalingEvent-template), quindi scegli Actions (Operazioni), Delete template (Elimina modello).
3. Quando viene richiesta la conferma, digitare **Delete** per confermare l'eliminazione del modello di avvio specificato e scegliere Delete (Elimina).

Se hai finito di lavorare con il gruppo con scalabilità automatica di esempio, eliminalo. Puoi anche eliminare il ruolo IAM e la policy di autorizzazione creati.

Per eliminare il gruppo con scalabilità automatica

1. Apri la [pagina Gruppi con dimensionamento automatico](#) nella console Amazon EC2.
2. Seleziona la casella di controllo accanto al gruppo con scalabilità automatica (TestAutoScalingEvent-group), quindi scegli Delete (Elimina).
3. Quando viene richiesta la conferma, digita **delete** per confermare l'eliminazione del gruppo con dimensionamento automatico specificato e scegli Elimina.

Un'icona di caricamento nella colonna Nome indica che il gruppo con dimensionamento automatico è in fase di eliminazione. Sono necessari alcuni minuti per terminare le istanze ed eliminare il gruppo.

Per eliminare il ruolo IAM

1. Apri la [pagina Roles \(Ruoli\)](#) della console IAM.
2. Seleziona il ruolo della funzione (TestAutoScalingEvent-role).
3. Scegli Elimina.
4. Quando viene richiesta la conferma, digita il nome del ruolo e scegli Delete (Elimina).

Come eliminare la policy IAM

1. Apri la pagina [Policies \(Policy\)](#) nella console IAM.
2. Seleziona la policy creata (TestAutoScalingEvent-policy).
3. Scegli Operazioni > Elimina.

4. Quando viene richiesta la conferma, digita il nome della policy e scegli Delete (Elimina).

Risorse correlate

I seguenti argomenti correlati possono essere utili per lo sviluppo di codice che richiama azioni sulle istanze in base ai dati disponibili nei metadati dell'istanza.

- [Recupero dello stato del ciclo di vita di destinazione tramite i metadati dell'istanza](#). Questa sezione descrive lo stato del ciclo di vita per altri casi d'uso, come la chiusura dell'istanza.
- [Come aggiungere hook del ciclo di vita \(console\)](#). Questa procedura mostra come aggiungere lifecycle hook sia per il dimensionamento orizzontale (avvio delle istanze) che per il dimensionamento verticale (istanze che terminano o ritornano a un pool caldo).
- [Categorie di metadati delle istanze](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux. Questo argomento elenca tutte le categorie di metadati delle istanze che puoi utilizzare per richiamare azioni sulle istanze EC2.

Per un tutorial che mostra come usare Amazon per EventBridge creare regole che richiamano le funzioni Lambda in base agli eventi che si verificano alle istanze del tuo gruppo Auto Scaling, consulta [Tutorial : Configurazione di un hook del ciclo di vita che richiama una funzione Lambda](#)

Tutorial : Configurazione di un hook del ciclo di vita che richiama una funzione Lambda

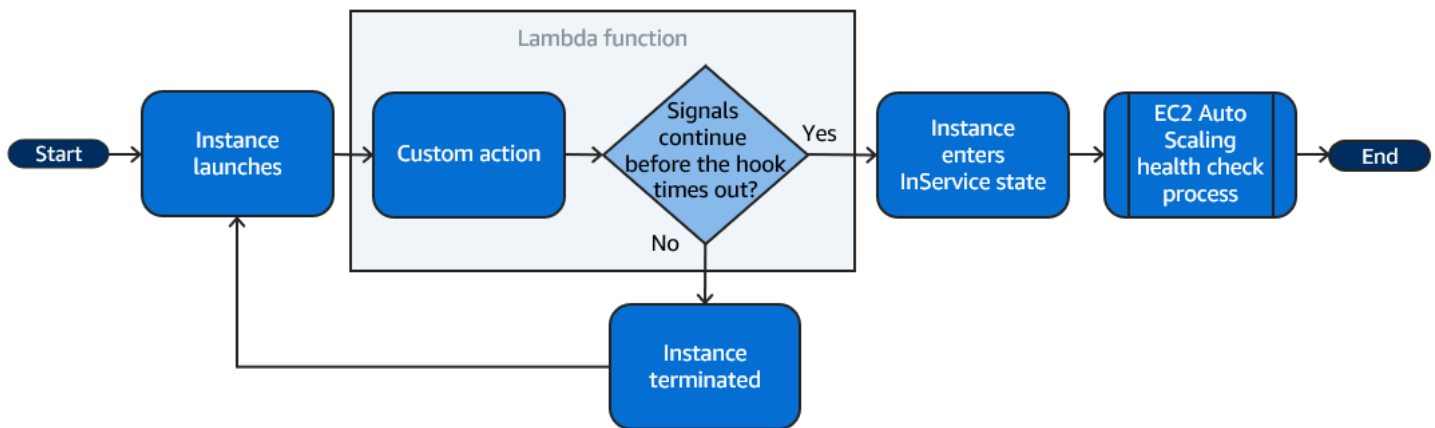
In questo esercizio, creerai una EventBridge regola Amazon che include uno schema di filtro che, se abbinato, richiama una AWS Lambda funzione come obiettivo della regola. Forniamo il modello di filtro e il codice di funzione di esempio da utilizzare.

Se tutto è configurato correttamente, al termine di questo tutorial, la funzione Lambda eseguirà un'operazione personalizzata all'avvio delle istanze. L'azione personalizzata registra semplicemente l'evento nel flusso di log CloudWatch Logs associato alla funzione Lambda.

Se questa operazione ha esito positivo, la funzione Lambda esegue anche un callback per consentire al ciclo di vita dell'istanza di procedere, ma consente all'istanza di abbandonare l'avvio e terminare se l'operazione fallisce.

La seguente illustrazione riassume il flusso di un evento di scalabilità orizzontale quando si utilizza una funzione Lambda per eseguire un'azione personalizzata. Dopo l'avvio di un'istanza, il ciclo di vita

dell'istanza viene sospeso fino al completamento dell'hook del ciclo di vita, tramite timeout o facendo sì che Amazon EC2 Auto Scaling riceva un segnale di continuazione.



Indice

- [Prerequisiti](#)
- [Passaggio 1: creazione di un ruolo IAM con autorizzazioni per completare le operazioni del ciclo di vita](#)
- [Passaggio 2: creazione di una funzione Lambda](#)
- [Fase 3: Creare una regola EventBridge](#)
- [Passaggio 4: aggiunta di un hook del ciclo di vita](#)
- [Passaggio 5: test e verifica dell'evento](#)
- [Fase 6: pulizia](#)
- [Risorse correlate](#)

Prerequisiti

Prima di iniziare questa esercitazione, crea un gruppo con scalabilità automatica, se non hai già uno. Per creare un gruppo con dimensionamento automatico, apri la [pagina dei gruppi Auto Scaling](#) nella console Amazon EC2 e scegli Crea gruppo con dimensionamento automatico.

Passaggio 1: creazione di un ruolo IAM con autorizzazioni per completare le operazioni del ciclo di vita

Prima di creare una funzione Lambda, è necessario innanzitutto creare un ruolo di esecuzione e una policy di autorizzazioni per permettere a Lambda di completare gli hook del ciclo di vita.

Come creare la policy

1. Nella console IAM, apri [Policies \(Policy\)](#), quindi seleziona Create policy (Crea policy).
2. Scegli la scheda JSON.
3. Nella casella Policy Document (Documento policy), incolla il documento della policy nella casella, sostituendo il testo in *corsivo* con il numero di account e il nome del gruppo con scalabilità automatica.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "autoscaling:CompleteLifecycleAction"
      ],
      "Resource":
        "arn:aws:autoscaling:*:123456789012:autoScalingGroup:*:autoScalingGroupName/my-  
asg"
    }
  ]
}
```

4. Seleziona Successivo.
5. In Policy name (Nome policy), inserisci **LogAutoScalingEvent-policy**. Scegli Create Policy (Crea policy).

Dopo aver creato la policy, puoi creare un ruolo che la utilizzi.

Come creare il ruolo

1. Nel riquadro di navigazione sulla sinistra, scegliere Roles (Ruoli).
2. Scegli Crea ruolo.
3. In Seleziona tipo di entità attendibile, scegli Servizio AWS .
4. Per il tuo caso d'uso, scegli Lambda quindi scegli Successivo.
5. In Aggiungi autorizzazioni, scegli la politica che hai creato (LogAutoScalingEvent-policy) e la politica denominata. AWSLambdaBasicExecutionRole Quindi, seleziona Next (Successivo).

Note

La `AWSLambdaBasicExecutionRolepolicy` dispone delle autorizzazioni necessarie alla funzione per scrivere i log in Logs. CloudWatch

6. Nella pagina Nome, revisione e creazione, per Nome ruolo, inserisci **LogAutoScalingEvent-role** e scegli Crea ruolo.

Passaggio 2: creazione di una funzione Lambda

Crea una funzione Lambda affinché serva da destinazione per gli eventi. La funzione Lambda di esempio, scritta in Node.js, viene richiamata da EventBridge quando viene emesso un evento corrispondente da Amazon EC2 Auto Scaling.

Per creare una funzione Lambda

1. Aprire la [pagina Functions \(Funzioni\)](#) nella console Lambda.
2. Scegli Create function (Crea funzione) e Author from scratch (Crea da zero).
3. In Basic information (Informazioni di base) , per Function name (Nome funzione) , inserisci **LogAutoScalingEvent**.
4. Per Runtime, scegli Node.js 18.x.
5. Scorri verso il basso e scegli Cambia ruolo di esecuzione predefinito e in Ruolo di esecuzione scegli Utilizza un ruolo esistente.
6. Per Ruolo esistente, scegli -role. LogAutoScalingEvent
7. Non modificare gli altri valori di default.
8. Scegli Crea funzione. Verrai riportato alla configurazione e al codice della funzione.
9. Con la funzione LogAutoScalingEvent ancora aperta nella console, in Codice della funzione, nell'editor, incolla il seguente codice di esempio nel file denominato index.mjs.

```
import { AutoScalingClient, CompleteLifecycleActionCommand } from "@aws-sdk/client-auto-scaling";
export const handler = async(event) => {
  console.log('LogAutoScalingEvent');
  console.log('Received event:', JSON.stringify(event, null, 2));
  var autoscaling = new AutoScalingClient({ region: event.region });
  var eventDetail = event.detail;
```

```
var params = {
  AutoScalingGroupName: eventDetail['AutoScalingGroupName'], /* required */
  LifecycleActionResult: 'CONTINUE', /* required */
  LifecycleHookName: eventDetail['LifecycleHookName'], /* required */
  InstanceId: eventDetail['EC2InstanceId'],
  LifecycleActionToken: eventDetail['LifecycleActionToken']
};
var response;
const command = new CompleteLifecycleActionCommand(params);
try {
  var data = await autoscaling.send(command);
  console.log(data); // successful response
  response = {
    statusCode: 200,
    body: JSON.stringify('SUCCESS'),
  };
} catch (err) {
  console.log(err, err.stack); // an error occurred
  response = {
    statusCode: 500,
    body: JSON.stringify('ERROR'),
  };
}
return response;
};
```

Questo codice registra semplicemente l'evento in modo che, alla fine di questo tutorial, sia possibile vedere apparire un evento nel flusso di log CloudWatch Logs associato a questa funzione Lambda.

10. Seleziona Deploy (Implementa).

Fase 3: Creare una regola EventBridge

Crea una EventBridge regola per eseguire la funzione Lambda. Per ulteriori informazioni sull'utilizzo EventBridge, consulta [Utilizzato EventBridge per gestire gli eventi di Auto Scaling](#).

Creazione di una regola tramite la console

1. Apri la [EventBridge console](#).
2. Nel pannello di navigazione, scegli Regole.
3. Scegli Crea regola.

4. Per **Define rule detail** (Definisci dettagli della regola), effettua le seguenti operazioni:
 - a. Per **Nome**, immetti **LogAutoScalingEvent-rule**.
 - b. Per **Event bus** (Bus di eventi), scegli **default**. Quando un Servizio AWS utente genera un evento, questo passa sempre al bus eventi predefinito del tuo account.
 - c. Per **Rule type** (Tipo di regola), scegli **Rule with an event pattern** (Regola con un modello di eventi).
 - d. Seleziona **Successivo**.
5. Per **Build event pattern** (Crea modello di eventi), procedi come segue:
 - a. Per **Event source**, scegli **AWS eventi** o **eventi EventBridge partner**.
 - b. Scorri verso il basso fino a **Modello di evento** ed esegui le seguenti operazioni:
 - i. In **Event source** (Origine eventi), selezionare **Servizi AWS**.
 - ii. Per **Servizio AWS**, scegli **Auto Scaling** (Scalabilità automatica).
 - iii. In **Event Type** (Tipo di evento), seleziona **Instance Launch and Terminate** (Avvio e arresto dell'istanza).
 - iv. Per impostazione predefinita, la regola corrisponde a qualsiasi evento di aumento o riduzione orizzontale. Per creare una regola che ti avvisa quando esiste un evento di dimensionamento e un'istanza viene messa in stato di attesa a causa di un hook del ciclo di vita, scegli **Specific instance event(s)** (Evento/i di istanza specifici) e seleziona **EC2 Instance-launch Lifecycle Action** (Operazione del ciclo di vita per l'avvio dell'istanza EC2).
 - v. Per impostazione predefinita, la regola corrisponde a qualsiasi gruppo con scalabilità automatica nella Regione. Affinché la regola corrisponda a un gruppo con dimensionamento automatico specifico, seleziona **Nomi gruppi specifici**, quindi seleziona il gruppo.
 - vi. Seleziona **Successivo**.
6. Per **Select target(s)** (Seleziona destinazione/i), esegui queste operazioni:
 - a. Per **Target types** (Tipi di target), scegli **Servizio AWS**.
 - b. Per **Select a target** (Seleziona destinazione), scegli **Lambda function** (Funzione Lambda).
 - c. Per **Funzione**, scegli **LogAutoScalingEvent**.
 - d. Scegliere **Next** (Successivo) due volte.

7. Nella pagina **Rivedi e crea**, scegli **Crea regola**

Passaggio 4: aggiunta di un hook del ciclo di vita

In questa sezione, si aggiunge un hook del ciclo di vita in modo che Lambda esegua la funzione sulle istanze al momento dell'avvio.

Per aggiungere un hook del ciclo di vita

1. Apri la [pagina Gruppi con dimensionamento automatico](#) nella console Amazon EC2.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico. Si aprirà un riquadro diviso nella parte inferiore della pagina.
3. Nel riquadro inferiore, nella scheda Instance management (Gestione istanze) in Lifecycle hooks (Hook del ciclo di vita), scegli Create lifecycle hook (Crea hook del ciclo di vita).
4. Per definire un hook del ciclo di vita per il dimensionamento orizzontale (avvio delle istanze), procedi come descritto di seguito:
 - a. In Lifecycle hook name (Nome hook del ciclo di vita), inserisci **LogAutoScalingEvent-hook**.
 - b. In Lifecycle transition (Transizione ciclo di vita), scegli Instance launch (Avvia istanza).
 - c. In Heartbeat timeout (Timeout heartbeat), inserisci **300** per il numero di secondi di attesa della richiamata dalla funzione Lambda.
 - d. In Default result (Risultato di default), scegli ABANDON (ABBANDONA). Ciò significa che se l'hook va in timeout senza ricevere un callback dalla funzione Lambda, il gruppo con scalabilità automatica terminerà una nuova istanza.
 - e. (Facoltativo) Lascia Notification metadata (Metadati delle notifiche) vuoto. I dati degli eventi a cui passiamo EventBridge contengono tutte le informazioni necessarie per richiamare la funzione Lambda.
5. Scegli Crea.

Passaggio 5: test e verifica dell'evento

Per effettuare il test dell'evento, aggiorna il gruppo con scalabilità automatica aumentando di 1 la capacità desiderata del gruppo con scalabilità automatica. La funzione Lambda viene richiamata pochi secondi dopo aver aumentato la capacità desiderata.

Per aumentare le dimensioni del gruppo con dimensionamento automatico

1. Apri la [pagina Gruppi con dimensionamento automatico](#) nella console Amazon EC2.

2. Seleziona la casella di controllo accanto al gruppo con scalabilità automatica per visualizzare i dettagli in un riquadro inferiore e visualizzare ancora le righe superiori del riquadro superiore.
3. Nel riquadro inferiore, nella scheda Details (Dettagli) scegli Group details (Dettagli gruppo), Edit (Modifica).
4. Per Desired capacity (Capacità desiderata), aumentare il valore corrente di 1.
5. Scegli Aggiorna. Durante l'avvio o la terminazione di un'istanza, la colonna Status (Stato) visualizza lo stato Updating capacity (Aggiornamento della capacità).

Dopo avere aumentato la capacità desiderata, potrai verificare che la funzione Lambda è stata invocata.

Per visualizzare l'output della funzione Lambda

1. Apri la [pagina Log groups](#) della CloudWatch console.
2. Seleziona il nome del gruppo di log per la funzione Lambda (/aws/lambda/LogAutoScalingEvent).
3. Seleziona il nome del flusso di log per visualizzare i dati forniti dalla funzione per l'istanza avviata.

Successivamente, è possibile verificare che l'istanza sia stata avviata correttamente dalla descrizione delle attività di dimensionamento.

Come visualizzare le attività di dimensionamento

1. Dalla pagina Auto Scaling groups (Gruppi Auto Scaling), seleziona il gruppo.
2. Nella scheda Activity (Attività) in Activity history (Cronologia attività), la colonna Status (Stato) indica se il gruppo con scalabilità automatica ha avviato correttamente le istanze.
 - Se l'azione ha avuto esito positivo, l'attività di dimensionamento avrà lo stato "Successful" (Riuscito).
 - In caso di errore, dopo aver atteso alcuni minuti, verrà visualizzata un'attività di dimensionamento con lo stato "Cancelled" (Annullato) e un messaggio di stato "Instance failed to complete user's Lifecycle Action: Lifecycle Action with token e85eb647-4fe0-4909-b341-a6c42EXAMPLE was abandoned: Lifecycle Action Completed with ABANDON Result". (Istanza non riuscita a completare l'azione ciclo di vita dell'utente: Azione ciclo di vita con token

e85eb647-4fe0-4909-b341-a6c42EXAMPLE abbandonato: Azione ciclo di vita completata con risultato ABANDON).

Come ridurre le dimensioni del gruppo con scalabilità automatica

Se l'istanza aggiuntiva avviata per questo test non è necessaria, è possibile aprire la scheda Details (Dettagli) e ridurre di 1 il valore di Desired capacity (Capacità desiderata).

Fase 6: pulizia

Se hai finito di utilizzare le risorse create appositamente per l'esercitazione, segui la procedura seguente per eliminarle.

Come eliminare l'hook del ciclo di vita

1. Apri la [pagina Gruppi con dimensionamento automatico](#) nella console Amazon EC2.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.
3. Nella scheda Instance management (Gestione istanze) in Lifecycle hooks (Hook del ciclo di vita), scegli Create lifecycle hook (Crea hook del ciclo di vita)(LogAutoScalingEvent-hook).
4. Scegli Actions (Operazioni), Delete (Elimina).
5. Per confermare, scegli di nuovo Delete (Elimina).

Per eliminare la EventBridge regola Amazon

1. Apri la [pagina delle regole](#) nella EventBridge console Amazon.
2. In Event bus (Bus di eventi), seleziona il bus di eventi associato alla regola (Default).
3. Seleziona la casella di controllo accanto alla regola (LogAutoScalingEvent-rule).
4. Scegli Elimina.
5. Quando viene richiesta la conferma, digita il nome della regola e scegli Delete (Elimina).

Se hai finito di lavorare con la funzione di esempio, eliminala. È inoltre possibile eliminare il gruppo di log che memorizza i registri della funzione e il ruolo di esecuzione e la policy delle autorizzazioni creati dalla console.

Come eliminare una funzione Lambda

1. Aprire la [pagina Functions \(Funzioni\)](#) nella console Lambda.

2. Scegli la funzione (LogAutoScalingEvent).
3. Scegli Operazioni > Elimina.
4. Quando viene richiesta la conferma, digita **delete** per confermare l'eliminazione della funzione specificata e scegli Delete (Elimina).

Come eliminare il gruppo di log

1. Apri la [pagina Log groups](#) della CloudWatch console.
2. Scegli il gruppo di log della funzione (/aws/lambda/LogAutoScalingEvent).
3. Scegli Actions (Azioni), Delete log group(s) (Elimina gruppo/i di log).
4. Nella finestra di dialogo Delete log group(s) (Elimina gruppo/i di log) scegli Delete (Elimina).

Come eliminare il ruolo di esecuzione

1. Apri la [pagina Roles \(Ruoli\)](#) della console IAM.
2. Seleziona il ruolo della funzione (LogAutoScalingEvent-role).
3. Scegli Elimina.
4. Quando viene richiesta la conferma, digita il nome del ruolo e scegli Delete (Elimina).

Come eliminare la policy IAM

1. Apri la pagina [Policies \(Policy\)](#) nella console IAM.
2. Seleziona la policy creata (LogAutoScalingEvent-policy).
3. Scegli Operazioni > Elimina.
4. Quando viene richiesta la conferma, digita il nome della policy e scegli Delete (Elimina).

Risorse correlate

I seguenti argomenti correlati possono essere utili per creare EventBridge regole basate sugli eventi che si verificano alle istanze del gruppo Auto Scaling.

- [Utilizzato EventBridge per gestire gli eventi di Auto Scaling](#). Questa sezione mostra esempi di eventi per altri casi d'uso, inclusi gli eventi per il dimensionamento verticale.

- [Come aggiungere hook del ciclo di vita \(console\)](#). Questa procedura mostra come aggiungere lifecycle hook sia per il dimensionamento orizzontale (avvio delle istanze) che per il dimensionamento verticale (istanze che terminano o ritornano a un pool caldo).

Per un tutorial che illustra come utilizzare il servizio di metadati di istanza (IMDS) per richiamare un'azione dall'interno dell'istanza stessa, consulta [Tutorial: configurazione dei dati utente per recuperare lo stato del ciclo di vita di destinazione tramite i metadati dell'istanza](#).

Warm pool per Dimensionamento automatico Amazon EC2

Un warm pool offre la possibilità di ridurre la latenza per le applicazioni che hanno tempi di avvio eccezionalmente lunghi, ad esempio, perché le istanze devono scrivere enormi quantità di dati su disco. Con i warm pool, è possibile gestire la latenza per migliorare le prestazioni delle applicazioni evitando di eseguire un provisioning dei gruppi di Auto Scaling oltre il necessario. Per ulteriori informazioni, consulta il seguente post sul blog [Scalabilità più rapida delle applicazioni con i warm pool EC2 Auto Scaling](#).

Important

Creare un warm pool quando non è necessario può generare costi inutili. Se il primo avvio non causa problemi di latenza evidenti per l'applicazione, probabilmente non è necessario utilizzare un warm pool.

Argomenti

- [Concetti principali](#)
- [Prerequisiti](#)
- [Aggiornamento di istanze in un pool caldo](#)
- [Risorse correlate](#)
- [Limitazioni](#)
- [Utilizzare un hook del ciclo di vita con un warm pool](#)
- [Creazione di un pool caldo per un gruppo con dimensionamento automatico](#)
- [Visualizzare lo stato del controllo dell'integrità e il motivo degli errori relativi ad esso.](#)
- [Esempi di creazione e gestione di piscine calde con AWS CLI](#)

Concetti principali

Prima di iniziare, acquisisci familiarità con i concetti principali riportati di seguito:

Warm pool

Un warm pool è un pool di istanze EC2 pre-inizializzate che si trova accanto a un gruppo con scalabilità automatica. Ogni volta che l'applicazione deve essere dimensionata, il gruppo con scalabilità automatica può sviluppare sul warm pool per soddisfare la nuova capacità desiderata. L'obiettivo è garantire che le istanze siano pronte per iniziare rapidamente a servire il traffico delle applicazioni, accelerando la risposta a un evento di aumento orizzontale. Man mano che le istanze lasciano il warm pool, contano per la capacità desiderata del gruppo. Si tratta del cosiddetto avvio a caldo.

Mentre le istanze sono nel periodo di riscaldamento, le policy di scalabilità vengono dimensionate orizzontalmente solo se il valore del parametro delle istanze che si trovano nello stato `InService` è maggiore della soglia massima di allarme della policy di dimensionamento (che è uguale all'utilizzo di destinazione di una policy di dimensionamento con monitoraggio degli obiettivi).

Dimensione di un warm pool

Per impostazione predefinita, la dimensione del warm pool viene calcolata come la differenza tra la capacità massima del gruppo con scalabilità automatica e la capacità desiderata. Ad esempio, se la capacità desiderata del gruppo con scalabilità automatica è 6 e la capacità massima è 10, la dimensione del warm pool sarà 4 quando imposti il warm pool e questo viene inizializzato.

Per specificare separatamente la capacità massima del warm pool, impostare un valore per la capacità massima preparata superiore alla capacità corrente del gruppo. Quando imposti un valore per la capacità massima preparata, la dimensione del warm pool è calcolata come la differenza tra la capacità massima preparata e l'attuale capacità desiderata del gruppo. Ad esempio, se la capacità desiderata del gruppo con scalabilità automatica è 6, la capacità massima è 10 e la capacità massima preparata è 8, la dimensione del warm pool al momento dell'impostazione sarà 2 e verrà quindi inizializzato.

Potresti dover utilizzare l'opzione di capacità massima preparata solo quando lavori con gruppi con scalabilità automatica di grandi dimensioni per gestire i vantaggi in termini di costi derivanti dalla presenza di un warm pool. Ad esempio, un gruppo con scalabilità automatica con 1000 istanze, una capacità massima di 1500 (per fornire capacità extra per picchi di traffico di

emergenza) e un warm pool di 100 istanze potrebbero aiutarti a raggiungere gli obiettivi meglio che mantenere 500 istanze riservate per un uso futuro all'interno del warm pool.

Dimensione minima del warm pool

Valuta la possibilità di utilizzare l'impostazione della dimensione minima per impostare staticamente il numero minimo di istanze da conservare nel warm pool. Per impostazione predefinita, non è impostata alcuna dimensione minima.

Stato delle istanze del warm pool

È possibile mantenere istanze nel warm pool in uno dei tre stati: `Stopped`, `Running` o `Hibernated`. Mantenere le istanze nello stato `Stopped` è un modo efficace per ridurre al minimo i costi. Con le istanze arrestate, paghi solo per i volumi utilizzati e gli indirizzi IP elastici assegnati alle istanze.

In alternativa, si possono mantenere le istanze nello stato `Hibernated` per arrestarle senza eliminare i contenuti della memoria (RAM). Quando un'istanza viene ibernata, al sistema operativo viene segnalato di salvare i contenuti della RAM nel volume root Amazon EBS. Quando l'istanza viene riavviata, il volume root viene ripristinato allo stato precedente e i contenuti RAM vengono ricaricati. Mentre le istanze sono in ibernazione, si paga solo per i volumi EBS, inclusa l'archiviazione dei contenuti RAM e gli indirizzi IP elastici associati alle istanze.

Inoltre, è possibile mantenere delle istanze nello stato `Running` all'interno del warm pool, ma è vivamente sconsigliato, per evitare di incorrere in spese inutili. Quando le istanze vengono interrotte o ibernare, si risparmia il costo delle istanze stesse. Paghi per le istanze solo quando sono in esecuzione.

Hook del ciclo di vita

Gli [hook del ciclo di vita](#) consentono di mettere le istanze in uno stato di attesa in modo da poter eseguire operazioni personalizzate sulle istanze. Le operazioni personalizzate vengono eseguite all'avvio o prima della fine delle istanze.

In una configurazione di pool caldo, gli hook del ciclo di vita ritardano l'arresto o l'ibernazione delle istanze ed evitano che vengano messe in esecuzione durante un evento di aumento orizzontale fino a quando non hanno terminato l'inizializzazione. Se si aggiunge un warm pool al gruppo con scalabilità automatica senza un hook del ciclo di vita, le istanze che richiedono molto tempo per terminare l'inizializzazione potrebbero essere arrestate o ibernare e quindi messe in esecuzione durante un evento di aumento orizzontale prima che siano pronte.

Policy per il riutilizzo delle istanze

Per impostazione predefinita, Dimensionamento automatico Amazon EC2 termina le istanze quando il gruppo con scalabilità automatica si dimensiona. Quindi, avvia nuove istanze nel warm pool per sostituire quelle che sono state terminate.

Se invece si desidera restituire le istanze al warm pool, è possibile specificare una policy per il riutilizzo delle istanze. Ciò consente di riutilizzare le istanze già configurate per servire il traffico delle applicazioni. Per assicurarsi che sul warm pool non venga effettuato l'over-provisioning, Dimensionamento automatico Amazon EC2 può terminare le istanze nel warm pool per ridurre le dimensioni quando sono più grandi del necessario in base alle impostazioni. Quando vengono terminate le istanze nel warm pool, si utilizza la [policy di terminazione di default](#) per scegliere quali istanze terminare per prime.

Important

Se si desidera ibernare le istanze durante la riduzione orizzontale e ci sono istanze esistenti nel gruppo con scalabilità automatica, devono soddisfare i requisiti, ad esempio l'ibernazione. In caso contrario, quando le istanze ritornano nel warm pool, vengono nuovamente arrestate e non ibernare.

Note

Attualmente, si può specificare una policy per il riutilizzo delle istanze solo utilizzando la AWS CLI o un SDK. Questa funzionalità non è disponibile dalla console.

Prerequisiti

Prima di creare un pool caldo per il gruppo con dimensionamento automatico, decidi come utilizzare gli hook del ciclo di vita per inizializzare nuove istanze con uno stato iniziale appropriato.

Per eseguire azioni personalizzate sulle istanze mentre sono in stato di attesa a causa di un hook del ciclo di vita, hai due opzioni:

- Per scenari semplici in cui si desidera eseguire comandi sulle istanze all'avvio, è possibile includere uno script di dati utente quando si crea un modello di avvio o una configurazione di avvio

per il gruppo con scalabilità automatica. Gli script di dati utente sono solo normali script della shell o direttive cloud-init gestiti da [cloud-init](#) all'avvio delle istanze. Lo script può controllare anche quando le istanze passano allo stato successivo utilizzando l'ID dell'istanza su cui viene eseguito. Se non lo stai già facendo, aggiorna lo script per recuperare l'ID istanza dai metadati dell'istanza. Per ulteriori informazioni, consulta la sezione [Recupero dei metadati dell'istanza](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Tip

Per eseguire script di dati utente al riavvio di un'istanza, i dati utente devono essere nel formato multipart MIME e specificare quanto segue nella sezione `#cloud-config` dei dati utente:

```
#cloud-config
cloud_final_modules:
- [scripts-user, always]
```

- Per scenari avanzati in cui è necessario un servizio, ad esempio per eseguire operazioni mentre le istanze entrano o escono dal pool caldo, è possibile creare un lifecycle hook per il gruppo Auto Scaling e configurare il servizio di destinazione per eseguire azioni personalizzate basate sulle notifiche del ciclo di vita. AWS Lambda Per ulteriori informazioni, consulta [Destinazioni di notifica supportate](#).

Preparazione delle istanze all'ibernazione

Per preparare le istanze Auto Scaling a utilizzare lo stato del pool Hibernated, crea un nuovo modello di avvio o una configurazione di avvio impostata correttamente per supportare l'ibernazione delle istanze, come descritto nell'argomento [Hibernation prerequisites](#) (Prerequisiti dell'ibernazione) nella Guida per l'utente di Amazon EC2 per le istanze Linux. Quindi, associa il nuovo modello di avvio o la configurazione di avvio al gruppo con scalabilità automatica e avvia un aggiornamento delle istanze per sostituire quelle associate a un modello di avvio o alla configurazione di avvio precedenti. Per ulteriori informazioni, consulta [Usa un aggiornamento dell'istanza per aggiornare le istanze in un gruppo di Auto Scaling](#).

Aggiornamento di istanze in un pool caldo

Per aggiornare le istanze in un pool caldo, devi creare un nuovo modello o una nuova configurazione di avvio e associarla al gruppo con dimensionamento automatico. Tutte le nuove istanze vengono

avviate utilizzando la nuova AMI e gli altri aggiornamenti specificati nel modello o nella configurazione di avvio, ma le istanze esistenti non sono interessate.

Per forzare l'avvio di istanze del pool caldo sostitutive che utilizzano il nuovo modello o la nuova configurazione di avvio puoi avviare un aggiornamento in sequenza del tuo gruppo. Un aggiornamento delle istanze sostituisce le istanze InService per prime. Quindi sostituisce le istanze nel warm pool. Per ulteriori informazioni, consulta [Usa un aggiornamento dell'istanza per aggiornare le istanze in un gruppo di Auto Scaling](#).

Risorse correlate

Puoi visitare il nostro [GitHub repository per esempi di lifecycle hook](#) per piscine calde.

Limitazioni

- [I gruppi di istanze miste](#) non sono supportati. Non puoi aggiungere un pool caldo ai gruppi con dimensionamento automatico che sostituiscono il tipo di istanza specificato in un modello di avvio o che sono configurati per l'avvio di istanze Spot.
- Dimensionamento automatico Amazon EC2 può inserire un'istanza in stato Stopped Hibernated solo se ha un volume Amazon EBS come root. Le istanze che utilizzano archivi di istanze per il dispositivo root non possono essere arrestate o ibernare.
- Dimensionamento automatico Amazon EC2 può inserire un'istanza in stato Hibernated solo se soddisfa tutti i requisiti elencati nell'argomento [Hibernation prerequisites](#) (Prerequisiti dell'ibernazione) nella Guida per l'utente di Amazon EC2 per le istanze Linux.
- Se il warm pool è esaurito quando c'è un evento di aumento orizzontale, le istanze verranno avviate direttamente nel gruppo con scalabilità automatica (avvio a freddo). Gli avvii a freddo possono verificarsi anche nel caso di inaccessibilità di una zona di disponibilità.
- Se un'istanza all'interno del pool caldo riscontra un problema durante il processo di avvio che ne impedisce il raggiungimento InService dello stato, l'istanza verrà considerata un avvio fallito e terminata. Ciò vale indipendentemente dalla causa sottostante, ad esempio un errore di capacità insufficiente o qualsiasi altro fattore.
- Se provi a utilizzare un warm pool con un gruppo di nodi gestiti Amazon Elastic Kubernetes Service (Amazon EKS), le istanze ancora in fase di inizializzazione potrebbero essere registrate con il cluster Amazon EKS. Di conseguenza, il cluster potrebbe pianificare i lavori su un'istanza che si sta preparando per essere arrestate o ibernata.
- Allo stesso modo, se provi a utilizzare un warm pool con un cluster Amazon ECS, le istanze potrebbero registrarsi con il cluster prima che completino l'inizializzazione. Per risolvere il

problema, devi configurare un modello di avvio o una configurazione di avvio che includa una variabile di configurazione di agente speciale nei dati utente. Per ulteriori informazioni, consulta la sezione [Utilizzo di un warm pool per un gruppo con scalabilità automatica](#) nella Guida per gli sviluppatori di Amazon Elastic Container Service.

- Il supporto per l'ibernazione per le pool calde è disponibile in tutte le pubblicità in Regioni AWS cui sono disponibili Amazon EC2 Auto Scaling e ibernazione, ad eccezione di quanto segue:
 - Asia Pacific (Hyderabad)
 - Asia Pacifico (Melbourne)
 - Canada occidentale (Calgary)
 - Regione Cina (Pechino)
 - Regione Cina (Ningxia)
 - Europa (Spagna)
 - Israele (Tel Aviv)

Utilizzare un hook del ciclo di vita con un warm pool

Le istanze di un warm pool mantengono il proprio ciclo di vita indipendente per facilitare la creazione delle operazioni personalizzate appropriate per ogni transizione. Questo ciclo di vita è progettato per aiutarti a richiamare operazioni in un servizio di destinazione (ad esempio, una funzione Lambda) mentre un'istanza viene inizializzata e prima che venga messa in esecuzione.

Note

Le operazioni API utilizzate per aggiungere e gestire gli hook del ciclo di vita e le operazioni complete del ciclo di vita non vengono modificate. Viene modificato solo il ciclo di vita dell'istanza.

Per ulteriori informazioni sull'aggiunta di un hook del ciclo di vita, consulta [Aggiungere un hook del ciclo di vita](#). Per ulteriori informazioni sul completamento di un'operazione del ciclo di vita, consulta [Completare un'operazione del ciclo di vita](#).

Per le istanze che entrano nel warm pool, potrebbe essere necessario un hook del ciclo di vita per uno dei seguenti motivi:

- È consigliabile avviare istanze EC2 da un'AMI che richiede molto tempo per terminare l'inizializzazione.
- È consigliabile eseguire script di dati utente per avviare le istanze EC2.

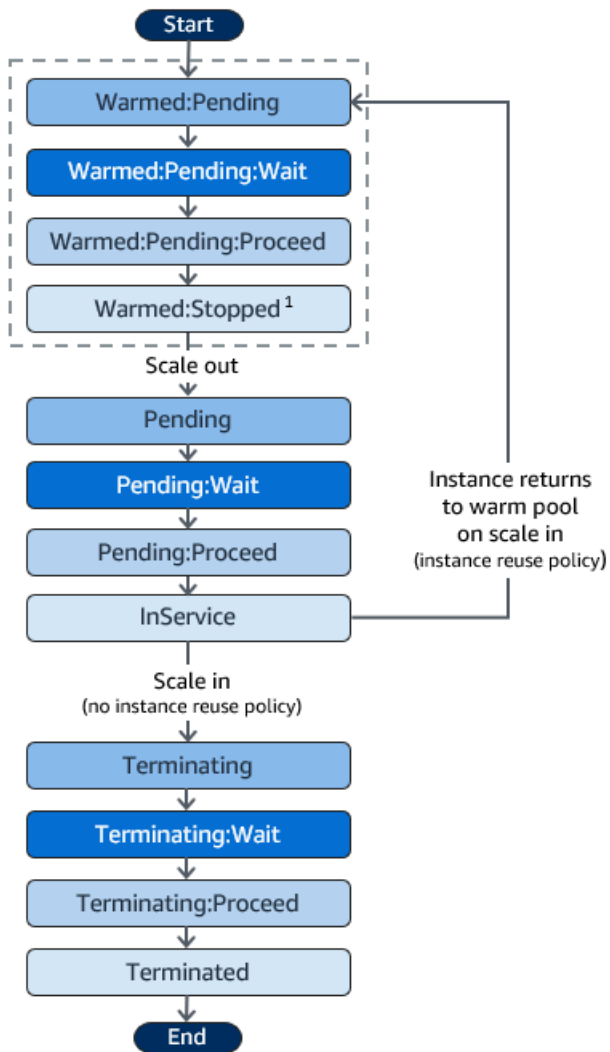
Per le istanze che escono dal warm pool, potrebbe essere necessario un hook del ciclo di vita per uno dei seguenti motivi:

- Puoi utilizzare un periodo di tempo aggiuntivo per preparare le istanze EC2 per l'uso. Ad esempio, puoi disporre di servizi da eseguire al riavvio di un'istanza prima che l'applicazione funzioni correttamente.
- È consigliabile compilare precedentemente i dati della cache, in modo che un nuovo server non venga avviato con una cache vuota.
- È consigliabile registrare le nuove istanze come istanze gestite con il servizio di gestione della configurazione.

Transizioni dello stato del ciclo di vita delle istanze in un warm pool

Un'istanza Auto Scaling può passare attraverso molti stati nel corso del ciclo di vita.

Il diagramma seguente mostra la transizione tra gli stati Auto Scaling quando utilizzi un warm pool:



¹ Questo stato varia in base all'impostazione dello stato del warm pool. Se lo stato del pool è impostato su Running, in tal caso è `Warmed:Running`. Se lo stato del pool è impostato su Hibernated, in tal caso è `Warmed:Hibernated`.

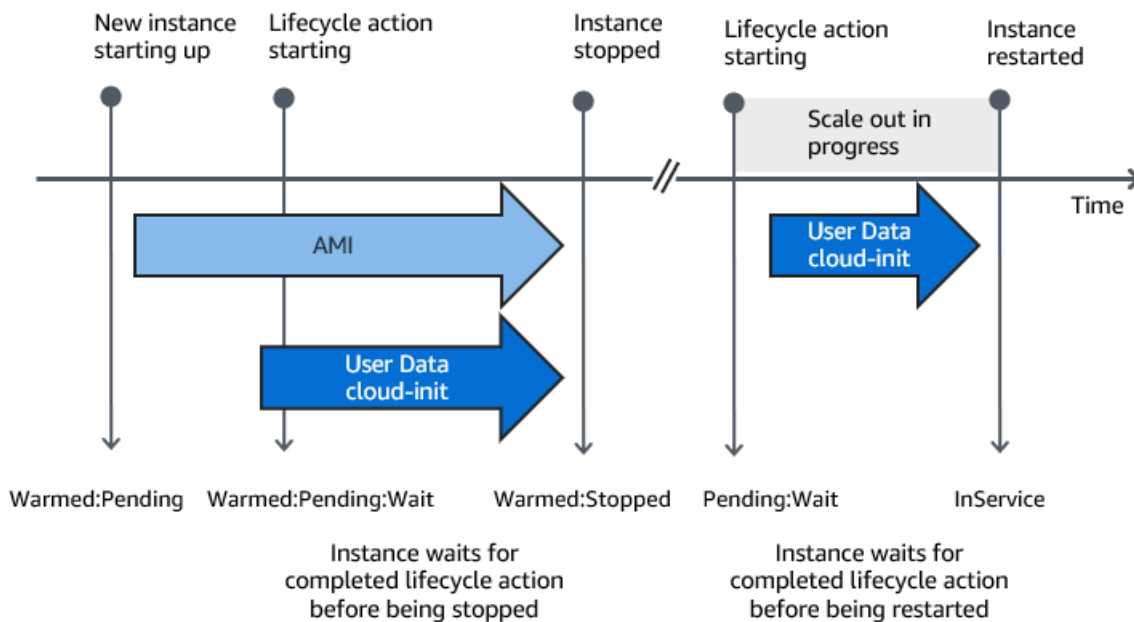
Quando aggiungi gli hook del ciclo di vita, considera quanto segue:

- Quando un hook del ciclo di vita è configurato per l'azione `autoscaling:EC2_INSTANCE_LAUNCHING` del ciclo di vita, un'istanza appena avviata viene sospesa per eseguire un'azione personalizzata quando raggiunge lo stato `Warmed:Pending:Wait`, quindi nuovamente quando viene riavviata e raggiunge lo stato `Pending:Wait`.
- Quando un hook del ciclo di vita è configurato per l'opzione `EC2_INSTANCE_TERMINATING`, un'istanza in fase di terminazione viene sospesa per eseguire un'azione personalizzata quando raggiunge lo stato `Terminating:Wait`. Tuttavia, se specifichi una policy di riutilizzo delle

istanze per restituire le istanze al warm pool su scala interna anziché terminarle, un'istanza che sta tornando al warm pool viene sospesa per eseguire un'azione personalizzata nello stato `Warmmed:Pending:Wait` per l'operazione `EC2_INSTANCE_TERMINATING` del ciclo di vita.

- Se la domanda sulla tua applicazione esaurisce il warm pool, Dimensionamento automatico Amazon EC2 può avviare le istanze direttamente nel gruppo con scalabilità automatica finché il gruppo non raggiunge la sua capacità massima. Se le istanze vengono avviate direttamente nel gruppo, vengono sospese solo per eseguire un'operazione personalizzata nello stato `Pending:Wait`.
- Per controllare per quanto tempo un'istanza rimane in uno stato di attesa prima di passare allo stato successivo, configura l'azione personalizzata per utilizzare il comando `complete-lifecycle-action`. Con gli hook del ciclo di vita, le istanze rimangono in stato di attesa finché non si notifica a Dimensionamento automatico Amazon EC2 che l'operazione del ciclo di vita specificata è completa o fino alla fine del periodo di timeout (un'ora per impostazione predefinita).

Di seguito viene riepilogato il flusso per un evento di aumento orizzontale.



Quando le istanze raggiungono uno stato di attesa, Dimensionamento automatico Amazon EC2 invia una notifica. Esempi di queste notifiche sono disponibili nella [EventBridge](#) sezione di questa guida. Per ulteriori informazioni, consulta [Esempi di eventi e modelli relativi al pool caldo](#).

Destinazioni di notifica supportate

Dimensionamento automatico Amazon EC2 fornisce supporto per la definizione di una delle seguenti destinazioni di notifica per le notifiche del ciclo di vita:

- EventBridge regole
- Argomenti di Amazon SNS
- Code Amazon SQS

Important

Ricorda che, se disponi di uno script di dati utente (cloud-init) nel modello di avvio o nella configurazione di avvio che configura le istanze al momento dell'avvio, non è necessario ricevere una notifica per eseguire operazioni personalizzate sulle istanze in fase di avvio o riavvio.

Le sezioni seguenti contengono collegamenti alla documentazione che descrive come configurare le destinazioni di notifica:

EventBridge regole: per eseguire il codice quando Amazon EC2 Auto Scaling mette un'istanza in uno stato di attesa, puoi creare una regola e specificare EventBridge una funzione Lambda come destinazione. Per richiamare diverse funzioni Lambda basate su notifiche del ciclo di vita diverse, puoi creare più regole e associare ciascuna regola a uno specifico modello di eventi e a una funzione Lambda. Per ulteriori informazioni, consulta [Crea EventBridge regole per eventi in piscina riscaldata](#).

Argomenti di Amazon SNS: per ricevere una notifica quando un'istanza viene inserita in uno stato di attesa, crei un argomento Amazon SNS e quindi imposti il filtro messaggi Amazon SNS per inviare le notifiche del ciclo di vita in modo diverso in base a un attributo del messaggio. Per ulteriori informazioni, consulta [Ricezione di notifiche tramite Amazon SNS](#).

Code Amazon SQS: per impostare un punto di consegna per le notifiche del ciclo di vita in cui un consumer pertinente può prenderle ed elaborarle, puoi creare una coda Amazon SQS e una coda di tipo consumer che elabora i messaggi dalla coda SQS. Se desideri che la coda di tipo consumer elabori le notifiche del ciclo di vita in modo diverso in base a un attributo del messaggio, devi impostarla in modo che analizzi e agisca sul messaggio quando un attributo specifico corrisponde al valore desiderato. Per ulteriori informazioni, consulta [Ricezione di notifiche tramite Amazon SQS](#).

Creazione di un pool caldo per un gruppo con dimensionamento automatico

In questo argomento viene descritto come creare un pool caldo per il gruppo con dimensionamento automatico.

Important

Prima di continuare, completa i [prerequisiti](#) per la creazione di un pool caldo e conferma di aver creato un hook del ciclo di vita per il tuo gruppo con dimensionamento automatico.

Creazione di un warm pool

Per creare un pool caldo per il gruppo con dimensionamento automatico, utilizza la procedura seguente.

Creazione di un warm pool (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto a un gruppo esistente.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Seleziona la scheda Instance management (Gestione delle istanze).
4. In Warm pool (warm pool), scegli Create warm pool (Crea warm pool).
5. Per configurare un warm pool, procedi come indicato di seguito:
 - a. In Warm pool instance state (Stato dell'istanza del warm pool), scegli a quale stato desideri passare le istanze quando entrano nel warm pool. Il valore predefinito è Stopped.
 - b. In Minimum warm pool size (Dimensione minima del warm pool), inserisci il numero minimo di istanze da mantenere nel pool.
 - c. Per Riutilizzo delle istanze, seleziona la casella di controllo Riutilizza su scala in per consentire alle istanze del gruppo Auto Scaling di tornare al pool caldo su scala in.
 - d. Per le dimensioni della piscina Warm, scegli una delle opzioni disponibili:
 - Specificazione predefinita: la dimensione della piscina calda è determinata dalla differenza tra la capacità massima e quella desiderata del gruppo Auto Scaling. Questa opzione semplifica la gestione della piscina calda. Dopo aver creato la piscina calda, le sue

dimensioni possono essere facilmente aggiornate semplicemente regolando la capacità massima del gruppo.

- Specifiche personalizzate: la dimensione della piscina calda è determinata dalla differenza tra un valore personalizzato e la capacità desiderata del gruppo Auto Scaling. Questa opzione offre la flessibilità necessaria per gestire le dimensioni della piscina calda indipendentemente dalla capacità massima del gruppo.
6. Visualizza la sezione Dimensioni stimate della piscina calda in base alle impostazioni correnti per confermare in che modo le specifiche predefinite o personalizzate si applicano alle dimensioni della piscina calda. Ricorda che la dimensione della piscina calda dipende dalla capacità desiderata del gruppo Auto Scaling, che cambierà se il gruppo si ridimensiona.
 7. Scegli Crea.

Eliminazione di un warm pool

Quando il warm pool non ti serve più, utilizza la seguente procedura per eliminarlo.

Come eliminare il warm pool (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto a un gruppo esistente.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Seleziona la scheda Instance management (Gestione delle istanze).
4. Per Warm pool (Pool caldo), scegli Actions (Operazioni, quindi Delete (Elimina)).
5. Quando viene richiesta la conferma, seleziona Delete (Elimina).

Visualizzare lo stato del controllo dell'integrità e il motivo degli errori relativi ad esso.

I controlli dell'integrità permettono a Dimensionamento automatico Amazon EC2 di determinare quando un'istanza non è integra e deve essere terminata. Per le istanze di warm pool mantenute nello stato Stopped, si avvale della conoscenza che Amazon EBS ha di un'istanza Stopped, per identificare le istanze non integre. Lo fa chiamando il metodo API DescribeVolumeStatus per determinare lo stato del volume EBS collegato all'istanza. Per le istanze di warm pool mantenute

nello stato `Running`, per determinare l'integrità dell'istanza si basa sui controlli di stato EC2. Sebbene non vi sia alcun periodo di tolleranza del controllo dell'integrità per le istanze del warm pool, Dimensionamento automatico Amazon EC2 non avvia il controllo fino al termine dell'hook del ciclo di vita.

Quando un'istanza viene rilevata non integra, Dimensionamento automatico Amazon EC2 elimina automaticamente l'istanza non integra e ne crea una nuova per sostituirla. Le istanze sono di solito terminate entro pochi minuti dopo il fallimento del controllo dell'integrità. Per ulteriori informazioni, consulta [Come visualizzare il motivo dei fallimenti dei controlli dell'integrità](#).

Sono supportati anche controlli di integrità personalizzati. Ciò può essere utile se disponi di un tuo sistema di controllo dell'integrità in grado di rilevare l'integrità di un'istanza e inviare queste informazioni a Dimensionamento automatico Amazon EC2. Per ulteriori informazioni, consulta [Controlli dell'integrità personalizzati](#).

Nella console Dimensionamento automatico Amazon EC2 è possibile visualizzare lo stato (integro o non integro) delle istanze del warm pool. Puoi anche visualizzare il loro stato di salute utilizzando AWS CLI o uno degli SDK.

Come visualizzare lo stato delle istanze del pool attivo (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina Gruppi Auto Scaling.

3. Nella scheda Instance management (Gestione istanze) in Warm pool instances (Istanze warm pool), la colonna Lifecycle (Ciclo di vita) contiene lo stato delle istanze.

La colonna Health status (Stato di integrità) mostra la valutazione che Dimensionamento automatico Amazon EC2 ha fatto dello stato dell'istanza.

Note

Le nuove istanze inizialmente sono integre. Fino al termine dell'hook del ciclo di vita, l'integrità di un'istanza non viene verificata.

Come visualizzare il motivo dei fallimenti dei controlli dell'integrità (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina Gruppi Auto Scaling.

3. Nella scheda Activity (Attività) in Activity history (Cronologia attività), la colonna Status (Stato) indica se il gruppo con scalabilità automatica ha avviato o terminato correttamente le istanze.

Se ha terminato eventuali istanze non integre, il campo Cause (Causa) mostra la data e l'ora della terminazione e il motivo dell'errore di controllo dell'integrità. Ad esempio, "At 2021-04-01T21:48:35Z an instance was taken out of service in response to EBS volume health check failure" (Al 2021-04-01T21:48:35Z un'istanza è stata messa fuori servizio in risposta al fallimento del controllo dell'integrità del volume EBS).

Come visualizzare lo stato delle istanze del warm pool (AWS CLI)

Visualizzate la piscina calda per un gruppo di Auto Scaling utilizzando il seguente [describe-warm-pool](#) comando.

```
aws autoscaling describe-warm-pool --auto-scaling-group-name my-asg
```

Output di esempio:

```
{
  "WarmPoolConfiguration": {
    "MinSize": 0,
    "PoolState": "Stopped"
  },
  "Instances": [
    {
      "InstanceId": "i-0b5e5e7521cfaa46c",
      "InstanceType": "t2.micro",
      "AvailabilityZone": "us-west-2a",
      "LifecycleState": "Warmed:Stopped",
      "HealthStatus": "Healthy",
      "LaunchTemplate": {
        "LaunchTemplateId": "lt-08c4cd42f320d5dcd",
        "LaunchTemplateName": "my-template-for-auto-scaling",
```



```

        "Version": "1"
    }
},
{
    "InstanceId": "i-0e21af9dcfb7aa6bf",
    "InstanceType": "t2.micro",
    "AvailabilityZone": "us-west-2a",
    "LifecycleState": "Warmup:Stopped",
    "HealthStatus": "Healthy",
    "LaunchTemplate": {
        "LaunchTemplateId": "lt-08c4cd42f320d5dcd",
        "LaunchTemplateName": "my-template-for-auto-scaling",
        "Version": "1"
    }
}
]
}

```

Come visualizzare il motivo dei fallimenti dei controlli dell'integrità (AWS CLI)

Utilizza il seguente comando [describe-scaling-activities](#).

```
aws autoscaling describe-scaling-activities --auto-scaling-group-name my-asg
```

Di seguito è riportata una risposta di esempio, dove Description indica che il gruppo con scalabilità automatica ha terminato un'istanza e Cause indica il motivo del fallimento del controllo dell'integrità.

Le attività di dimensionamento vengono ordinate in base all'ora di inizio. Le attività ancora in corso vengono descritte per prime.

```

{
  "Activities": [
    {
      "ActivityId": "4c65e23d-a35a-4e7d-b6e4-2eaa8753dc12",
      "AutoScalingGroupName": "my-asg",
      "Description": "Terminating EC2 instance: i-04925c838b6438f14",
      "Cause": "At 2021-04-01T21:48:35Z an instance was taken out of service in response to EBS volume health check failure.",
      "StartTime": "2021-04-01T21:48:35.859Z",
      "EndTime": "2021-04-01T21:49:18Z",
      "StatusCode": "Successful",
      "Progress": 100,
    }
  ]
}

```

```
    "Details": "{ \"Subnet ID\": \"subnet-5ea0c127\", \"Availability Zone\": \"us-west-2a\n\"...}\",\n    \"AutoScalingGroupARN\": \"arn:aws:autoscaling:us-\nwest-2:123456789012:autoScalingGroup:283179a2-\nf3ce-423d-93f6-66bb518232f7:autoScalingGroupName/my-asg\"\n  },\n  ...]\n}
```

Esempi di creazione e gestione di piscine calde con AWS CLI

È possibile creare e gestire piscine calde utilizzando AWS Management Console, AWS Command Line Interface (AWS CLI) o SDK.

Negli esempi seguenti viene illustrato come creare e gestire warm pool utilizzando la AWS CLI.

Indice

- [Esempio 1: mantenimento delle istanze nello stato Stopped](#)
- [Esempio 2: mantenimento delle istanze nello stato Running](#)
- [Esempio 3: mantenimento delle istanze nello stato Hibernated](#)
- [Esempio 4: restituzione delle istanze al warm pool durante la riduzione orizzontale](#)
- [Esempio 5: specificazione del numero minimo di istanze nel warm pool](#)
- [Esempio 6: definire la dimensione della piscina calda utilizzando una specifica personalizzata](#)
- [Esempio 7: definizione di una dimensione assoluta del warm pool](#)
- [Esempio 8: eliminazione di un warm pool](#)

Esempio 1: mantenimento delle istanze nello stato **Stopped**

L'[put-warm-pool](#) esempio seguente crea un pool caldo che mantiene le istanze in uno Stopped stato.

```
aws autoscaling put-warm-pool --auto-scaling-group-name my-asg /\n--pool-state Stopped
```

Esempio 2: mantenimento delle istanze nello stato **Running**

L'[put-warm-pool](#) esempio seguente crea un pool caldo che mantiene le istanze in uno Running stato anziché in uno stato Stopped

```
aws autoscaling put-warm-pool --auto-scaling-group-name my-asg /  
--pool-state Running
```

Esempio 3: mantenimento delle istanze nello stato **Hibernated**

L'[put-warm-pool](#) esempio seguente crea un pool caldo che mantiene le istanze in uno Hibernated stato anziché in uno stato Stopped. Ciò consente di arrestare le istanze senza eliminare i contenuti della memoria (RAM).

```
aws autoscaling put-warm-pool --auto-scaling-group-name my-asg /  
--pool-state Hibernated
```

Esempio 4: restituzione delle istanze al warm pool durante la riduzione orizzontale

L'[put-warm-pool](#) esempio seguente crea un pool caldo che mantiene le istanze in uno Stopped stato e include l'`--instance-reuse-policy` opzione. Il valore della policy per il riutilizzo delle istanze `'{"ReuseOnScaleIn": true}'` indica a Dimensionamento automatico Amazon EC2 di restituire le istanze al warm pool quando il gruppo con scalabilità automatica si dimensiona.

```
aws autoscaling put-warm-pool --auto-scaling-group-name my-asg /  
--pool-state Stopped --instance-reuse-policy '{"ReuseOnScaleIn": true}'
```

Esempio 5: specificazione del numero minimo di istanze nel warm pool

L'[put-warm-pool](#) esempio seguente crea un pool caldo che mantiene un minimo di 4 istanze, in modo che siano disponibili almeno 4 istanze per gestire i picchi di traffico.

```
aws autoscaling put-warm-pool --auto-scaling-group-name my-asg /  
--pool-state Stopped --min-size 4
```

Esempio 6: definire la dimensione della piscina calda utilizzando una specifica personalizzata

Per impostazione predefinita, Amazon EC2 Auto Scaling gestisce la dimensione del pool caldo come differenza tra la capacità massima e quella desiderata del gruppo Auto Scaling. Tuttavia, puoi gestire le dimensioni della piscina calda indipendentemente dalla capacità massima del gruppo utilizzando l'opzione `--max-group-prepared-capacity`

L'[put-warm-pool](#) esempio seguente crea un pool caldo e imposta il numero massimo di istanze che possono esistere contemporaneamente sia nel pool caldo che nel gruppo Auto Scaling. Se il gruppo ha una capacità desiderata di 800, il pool caldo avrà inizialmente una dimensione di 100 quando verrà inizializzato dopo l'esecuzione di questo comando.

```
aws autoscaling put-warm-pool --auto-scaling-group-name my-asg /  
--pool-state Stopped --max-group-prepared-capacity 900
```

Per mantenere un numero minimo di istanze nel warm pool, includi l'opzione `--min-size` con il comando, come segue.

```
aws autoscaling put-warm-pool --auto-scaling-group-name my-asg /  
--pool-state Stopped --max-group-prepared-capacity 900 --min-size 25
```

Esempio 7: definizione di una dimensione assoluta del warm pool

Se imposti i valori per le opzioni `--max-group-prepared-capacity` e `--min-size` allo stesso valore, il warm pool avrà una dimensione assoluta. L'[put-warm-pool](#) esempio seguente crea un pool caldo che mantiene una dimensione costante del pool caldo di 10 istanze.

```
aws autoscaling put-warm-pool --auto-scaling-group-name my-asg /  
--pool-state Stopped --min-size 10 --max-group-prepared-capacity 10
```

Esempio 8: eliminazione di un warm pool

Utilizzate il seguente [delete-warm-pool](#) comando per eliminare una piscina calda.

```
aws autoscaling delete-warm-pool --auto-scaling-group-name my-asg
```

Se ci sono istanze nella piscina calda o se sono in corso attività di ridimensionamento, usa il [delete-warm-pool](#) comando con l'`--force-delete` opzione. Questa opzione termina anche le istanze Amazon EC2 e le eventuali operazioni del ciclo di vita in sospeso.

```
aws autoscaling delete-warm-pool --auto-scaling-group-name my-asg --force-delete
```

Scollegare o collegare istanze

Puoi scollegare le istanze dal tuo gruppo Auto Scaling. Dopo il distacco, l'istanza diventa indipendente e può essere gestita da sola o collegata a un gruppo di Auto Scaling diverso, separato

dal gruppo originale a cui apparteneva. Ciò può essere utile, ad esempio, quando si desidera eseguire test utilizzando istanze esistenti che già eseguono l'applicazione.

Questo argomento fornisce istruzioni su come scollegare e collegare le istanze. Quando si collegano istanze, è possibile utilizzare anche un'istanza esistente anziché una distaccata.

Invece di scollegare e ricollegare un'istanza allo stesso gruppo, consigliamo di utilizzare la procedura di standby per rimuovere temporaneamente l'istanza dal gruppo. Per ulteriori informazioni, consulta [Rimuovere temporaneamente le istanze dal gruppo con scalabilità automatica](#).

Indice

- [Considerazioni sul distacco delle istanze](#)
- [Considerazioni per il collegamento delle istanze](#)
- [Sposta un'istanza in un gruppo diverso utilizzando detach e attach](#)

Considerazioni sul distacco delle istanze

Quando scollegate le istanze, tenete a mente questi punti:

- Puoi scollegare un'istanza solo quando si trova nello stato. `InService`
- Dopo aver scollegato un'istanza, questa continua a funzionare e comporta dei costi. Per evitare addebiti inutili, assicurati di ricollegare o terminare le istanze separate quando non sono più necessarie.
- Puoi scegliere di diminuire la capacità desiderata in base al numero di istanze da scollegare. Se scegli di non ridurre la capacità, Amazon EC2 Auto Scaling lancia nuove istanze per sostituire quelle separate per mantenere la capacità desiderata.
- Se il numero di istanze che state scollegando porterà il gruppo Auto Scaling al di sotto della sua capacità minima, dovete diminuire la capacità minima.
- Se si scollegano più istanze dalla stessa zona di disponibilità senza diminuire la capacità desiderata, il gruppo si ribilancerà da solo a meno che non si sospenda il processo. `AZRebalance`. Per ulteriori informazioni, consulta [Sospendi e riprendi i processi di Amazon EC2 Auto Scaling](#).
- Se alleggi un'istanza a un gruppo con scalabilità automatica a cui è allegato a un gruppo di destinazione del load balancer o un Classic Load Balancer, l'istanza viene deregistrata dal load balancer. Se lo svuotamento della connessione (ritardo nell'annullamento della registrazione) è abilitato per il load balancer, Dimensionamento automatico Amazon EC2 attende il completamento delle richieste in transito.

Note

Se stai scollegando istanze che si trovano nello stato Standby, fai attenzione. Il tentativo di scollegare le istanze dopo averle inserite nello stato Standby può causare la terminazione imprevista di altre istanze.

Considerazioni per il collegamento delle istanze

Quando allegati le istanze, tenete presente quanto segue:

- Amazon EC2 Auto Scaling tratta le istanze collegate allo stesso modo delle istanze lanciate dal gruppo stesso. Ciò significa che le istanze collegate possono essere terminate durante gli eventi di scalabilità, se selezionate.
- Quando alleghi le istanze, la capacità desiderata del gruppo aumenta per il numero di istanze che vengono allegati. Se la capacità desiderata dopo l'aggiunta delle nuove istanze supera la dimensione massima del gruppo, la richiesta di collegare più istanze ha esito negativo.
- Se aggiungi istanze al tuo gruppo causando una distribuzione non uniforme tra le zone di disponibilità, Amazon EC2 Auto Scaling ribilancia il gruppo per ristabilire una distribuzione uniforme a meno che non sospendi il processo. AZRebalance Per ulteriori informazioni, consulta [Sospendi e riprendi i processi di Amazon EC2 Auto Scaling](#).
- Se alleghi un'istanza a un gruppo con scalabilità automatica a cui è allegato a un gruppo di destinazione del load balancer o un sistema di Classic Load Balancer, l'istanza viene registrata con il load balancer.

L'istanza da allegare deve soddisfare i seguenti criteri:

- L'istanza è nello stato running con Amazon EC2.
- L'AMI utilizzata per avviare l'istanza deve essere tuttora esistente.
- L'istanza non è un membro di un altro gruppo con scalabilità automatica.
- L'istanza viene avviata in una delle zone di disponibilità definite nel gruppo Auto Scaling.
- Se al gruppo con scalabilità automatica è collegato un gruppo di destinazione del load balancer o un Classic Load Balancer, l'istanza e il load balancer devono trovarsi nello stesso VPC.

Sposta un'istanza in un gruppo diverso utilizzando detach e attach

Utilizzate una delle seguenti procedure per scollegare un'istanza dal gruppo Auto Scaling e collegarla a un gruppo Auto Scaling diverso.

Per creare un nuovo gruppo Auto Scaling da un'istanza distaccata, vedi [Creare un gruppo con dimensionamento automatico utilizzando parametri di un'istanza esistente](#) (non consigliato, crea una configurazione di avvio).

Console

Per scollegare un'istanza da un gruppo Auto Scaling

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Nella scheda Instance management (Gestione istanze) in Instances (Istanze), seleziona un'istanza e scegli Actions (Operazioni), quindi Detach (Distacca).
4. Nella finestra di dialogo Scollega istanza, mantieni selezionata la casella di controllo Sostituisci istanza per avviare un'istanza sostitutiva. Deseleziona la casella di controllo per diminuire la capacità desiderata.
5. Quando viene richiesta la conferma, digita **detach** per confermare la rimozione dell'istanza specificata dal gruppo con dimensionamento automatico e scegli Elimina istanza.

È ora possibile collegare l'istanza a un gruppo Auto Scaling diverso.

Come allegare un'istanza a un gruppo con scalabilità automatica

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. (Facoltativo) Nel pannello di navigazione, in Auto Scaling (Scalabilità automatica), scegli Auto Scaling Groups (Gruppi con scalabilità automatica). Seleziona il gruppo con scalabilità automatica e verifica che la sua dimensione massima sia sufficiente ad aggiungere un'altra istanza. In caso contrario, nella scheda Details (Dettagli) aumenta la capacità massima.
3. Nel pannello di navigazione, in Instances (Istanze) selezionare Instances (Istanze) e poi selezionarne una.

4. Scegli Actions (Operazioni), Instance settings (Impostazioni istanza), Attach to Auto Scaling Group (Allega al gruppo con scalabilità automatica).
5. Nella pagina Attach to Auto Scaling Group (Allega al gruppo con scalabilità automatica), in Auto Scaling Group (gruppo con scalabilità automatica), seleziona il gruppo e scegli Attach (Allega).
6. Se l'istanza non soddisfa i parametri, ti verrà mostrato un messaggio di errore con i dettagli. Ad esempio, l'istanza potrebbe non essere nella stessa zona di disponibilità del gruppo con scalabilità automatica. Scegli Chiudi e riprova con un gruppo Auto Scaling che soddisfa i criteri.

AWS CLI

Per scollegare e collegare un'istanza, utilizzate i seguenti comandi di esempio. Sostituisci ciascun *placeholder input dell'utente* con le tue informazioni.

Per scollegare un'istanza da un gruppo Auto Scaling

1. Per descrivere le istanze correnti, utilizzate il comando seguente. [describe-auto-scaling-instances](#)

```
aws autoscaling describe-auto-scaling-instances \  
  --query 'AutoScalingInstances[?AutoScalingGroupName==`my-asg`]'
```

L'esempio seguente mostra l'output prodotto quando si esegue questo comando.

Prendi nota dell'ID dell'istanza che intendi rimuovere dal gruppo. Questo ID ti servirà nel passaggio successivo.

```
{  
  "AutoScalingInstances": [  
    {  
      "ProtectedFromScaleIn": false,  
      "AvailabilityZone": "us-west-2a",  
      "LaunchTemplate": {  
        "LaunchTemplateName": "my-launch-template",  
        "Version": "1",  
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"  
      },  
      "InstanceId": "i-05b4f7d5be44822a6",
```



```
"InstanceType": "t3.micro",
"AutoScalingGroupName": "my-asg",
"HealthStatus": "HEALTHY",
"LifecycleState": "InService"
},
{
  "ProtectedFromScaleIn": false,
  "AvailabilityZone": "us-west-2a",
  "LaunchTemplate": {
    "LaunchTemplateName": "my-launch-template",
    "Version": "1",
    "LaunchTemplateId": "lt-050555ad16a3f9c7f"
  },
  "InstanceId": "i-0c20ac468fa3049e8",
  "InstanceType": "t3.micro",
  "AutoScalingGroupName": "my-asg",
  "HealthStatus": "HEALTHY",
  "LifecycleState": "InService"
},
{
  "ProtectedFromScaleIn": false,
  "AvailabilityZone": "us-west-2a",
  "LaunchTemplate": {
    "LaunchTemplateName": "my-launch-template",
    "Version": "1",
    "LaunchTemplateId": "lt-050555ad16a3f9c7f"
  },
  "InstanceId": "i-0787762faf1c28619",
  "InstanceType": "t3.micro",
  "AutoScalingGroupName": "my-asg",
  "HealthStatus": "HEALTHY",
  "LifecycleState": "InService"
},
{
  "ProtectedFromScaleIn": false,
  "AvailabilityZone": "us-west-2a",
  "LaunchTemplate": {
    "LaunchTemplateName": "my-launch-template",
    "Version": "1",
    "LaunchTemplateId": "lt-050555ad16a3f9c7f"
  },
  "InstanceId": "i-0f280a4c58d319a8a",
  "InstanceType": "t3.micro",
  "AutoScalingGroupName": "my-asg",
```

```

        "HealthStatus": "HEALTHY",
        "LifecycleState": "InService"
    }
]
}

```

2. [Per scollegare un'istanza senza diminuire la capacità desiderata, utilizzate il seguente comando `detach-instances`.](#)

```
aws autoscaling detach-instances --instance-ids i-05b4f7d5be44822a6 \
--auto-scaling-group-name my-asg
```

Per scollegare un'istanza e diminuire la capacità desiderata, includete l'opzione. `--should-decrement-desired-capacity`

```
aws autoscaling detach-instances --instance-ids i-05b4f7d5be44822a6 \
--auto-scaling-group-name my-asg --should-decrement-desired-capacity
```

È ora possibile collegare l'istanza a un gruppo Auto Scaling diverso.

Come allegare un'istanza a un gruppo con scalabilità automatica

1. Per collegare l'istanza a un gruppo Auto Scaling diverso, utilizzate il seguente comando [attach-instances](#).

```
aws autoscaling attach-instances --instance-ids i-05b4f7d5be44822a6 --auto-
scaling-group-name my-asg-for-testing
```

2. Per verificare la dimensione del gruppo Auto Scaling dopo aver collegato un'istanza, utilizzate il comando seguente. [describe-auto-scaling-groups](#)

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-names my-asg-
for-testing
```

La seguente risposta di esempio mostra che il gruppo ha due istanze in esecuzione, una delle quali è l'istanza allegata.

```
{
  "AutoScalingGroups": [
```

```
{
  "AutoScalingGroupName": "my-asg-for-testing",
  "AutoScalingGroupARN": "arn",
  "LaunchTemplate": {
    "LaunchTemplateName": "my-launch-template",
    "Version": "2",
    "LaunchTemplateId": "lt-050555ad16a3f9c7f"
  },
  "MinSize": 1,
  "MaxSize": 5,
  "DesiredCapacity": 2,
  ...
  "Instances": [
    {
      "ProtectedFromScaleIn": false,
      "AvailabilityZone": "us-west-2a",
      "LaunchTemplate": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "1",
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"
      },
      "InstanceId": "i-05b4f7d5be44822a6",
      "InstanceType": "t3.micro",
      "HealthStatus": "Healthy",
      "LifecycleState": "InService"
    },
    {
      "ProtectedFromScaleIn": false,
      "AvailabilityZone": "us-west-2a",
      "LaunchTemplate": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "2",
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"
      },
      "InstanceId": "i-00dcdfffd5175890",
      "InstanceType": "t3.micro",
      "HealthStatus": "Healthy",
      "LifecycleState": "InService"
    }
  ],
  ...
}
```

}

Rimuovere temporaneamente le istanze dal gruppo con scalabilità automatica

È possibile far passare un'istanza dallo stato InService allo stato Standby, aggiornare l'istanza o risolverne i problemi e riportare quindi l'istanza in servizio. Le istanze in standby fanno sempre parte del gruppo con scalabilità automatica, ma non gestiscono attivamente il traffico del load balancer.

Questa caratteristica consente di arrestare e avviare le istanze o di riavviarle senza preoccuparsi che Dimensionamento automatico Amazon EC2 termini le istanze durante i controlli di integrità o durante gli eventi di riduzione orizzontale.

Ad esempio, puoi modificare l'immagine Amazon Machine Image (AMI) per un gruppo con scalabilità automatica in qualsiasi momento modificando il modello o la configurazione di avvio. Tutte le istanze successive avviate dal gruppo con scalabilità automatica utilizzeranno questa AMI. Tuttavia, il gruppo con scalabilità automatica non aggiorna le istanze attualmente in servizio. Puoi terminare queste istanze e consentire a Dimensionamento automatico Amazon EC2 di sostituirle oppure utilizzare la funzione di aggiornamento delle istanze per terminare e sostituire le istanze. In alternativa, puoi mettere le istanze in standby, aggiornare il software e quindi riportare le istanze in servizio.

Il procedimento per distaccare le istanze da un gruppo con scalabilità automatica è simile a quello con cui vengono messe in standby. Il distacco delle istanze potrebbe essere utile se desideri collegarle a un gruppo diverso o gestire le istanze come istanze EC2 autonome ed eventualmente terminarle. Per ulteriori informazioni, consulta [Scollegare o collegare istanze](#).

Indice

- [Come funziona lo stato di standby](#)
- [Considerazioni](#)
- [Stato di integrità di un'istanza in standby](#)
- [Rimuovi temporaneamente un'istanza impostandola in standby](#)

Come funziona lo stato di standby

Lo stato di standby funziona come indicato di seguito per agevolare la rimozione temporanea di un'istanza dal gruppo con scalabilità automatica:

1. Tu metti un'istanza nello stato di standby. Questa vi rimane finché non la si fa uscire da questo stato.
2. Se c'è un gruppo di destinazione del load balancer o un Classic Load Balancer associato al gruppo con scalabilità automatica, l'istanza viene deregistrata dal load balancer. Se per il load balancer è abilitata la funzione di svuotamento connessione, Elastic Load Balancing attende, per impostazione predefinita, 300 secondi prima di completare il processo di deregistrazione, facilitando così il completamento delle richieste in transito.
3. È possibile aggiornare l'istanza o risolverne i problemi.
4. È possibile riportare l'istanza in servizio, facendola uscire dallo stato di standby.
5. Se c'è un gruppo di destinazione del load balancer o un Classic Load Balancer allegato al tuo gruppo con scalabilità automatica, l'istanza è registrata con il load balancer.

Per ulteriori informazioni sul ciclo di vita delle istanze in un gruppo con dimensionamento automatico, consulta [Ciclo di vita delle istanze di Amazon EC2 Auto Scaling](#).

Considerazioni

Di seguito sono riportate le considerazioni relative allo spostamento delle istanze dentro e fuori dallo stato di standby:

- Quando metti un'istanza in standby, puoi diminuire la capacità desiderata tramite questa operazione o mantenerla invariata.
 - Se scegli di diminuire la capacità desiderata del gruppo con dimensionamento automatico, Dimensionamento automatico Amazon EC2 avvia un'istanza per sostituire quella in standby. Lo scopo è di mantenere la capacità per l'applicazione mentre una o più istanze sono in standby.
 - Se scegli di diminuire la capacità desiderata del gruppo con dimensionamento automatico, impedisce l'avvio di un'istanza per sostituire quella in standby.
- Dopo aver riportato l'istanza in servizio, la capacità desiderata viene incrementata in modo da riflettere il numero di istanze presenti nel gruppo con dimensionamento automatico.
- Per effettuare l'aumento (e la diminuzione), la nuova capacità desiderata deve essere compresa tra la dimensione minima e massima del gruppo. In caso contrario, l'operazione non va a buon fine.
- Se in qualsiasi momento dopo aver messo un'istanza in standby o aver riportato l'istanza in servizio uscendo dallo stato di standby, il gruppo con dimensionamento automatico risulta non bilanciato tra le zone di disponibilità, Dimensionamento automatico Amazon EC2 compensa ribilanciando le zone

di disponibilità a meno che tu non sospenda il processo AZRebalance. Per ulteriori informazioni, consulta [Sospendi e riprendi i processi di Amazon EC2 Auto Scaling](#).

- Ti vengono addebitati i costi delle istanze in stato di standby.

Stato di integrità di un'istanza in standby

Dimensionamento automatico Amazon EC2 non esegue controlli dell'integrità su istanze in standby. Anche se l'istanza è in standby, il relativo stato di integrità riflette lo stato che aveva prima di entrare in standby. Dimensionamento automatico Amazon EC2 non esegue controlli l'integrità sull'istanza finché questa non viene rimessa in servizio.

Ad esempio, se metti un'istanza integra in standby e poi la termini, Dimensionamento automatico Amazon EC2 continua a segnalare l'istanza come integra. Se tenti di mettere in servizio un'istanza terminata che era in standby, Dimensionamento automatico Amazon EC2 esegue un controllo dell'integrità sull'istanza, determina che questa si sta terminando e non è integra e ne avvia una sostitutiva. Per ulteriori informazioni, consulta [Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico](#).

Rimuovi temporaneamente un'istanza impostandola in standby

Utilizzate una delle seguenti procedure per mettere temporaneamente fuori servizio un'istanza mettendola in stato di standby.

Console

Per rimuovere temporaneamente un'istanza

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Nella scheda Instance management (Gestione istanze) in Instances (Istanze), selezionare un'istanza.
4. Scegli Actions (Operazioni), Set to Standby (Imposta su Standby).
5. Nella finestra di dialogo Imposta su Standby, mantieni selezionata la casella di controllo Sostituisci istanza per avviare un'istanza sostitutiva. Deseleziona la casella di controllo per diminuire la capacità desiderata.

- Quando viene richiesta la conferma, digita **standby** per confermare lo stato Standby dell'istanza specificata, quindi sceglie Imposta in standby.
- È possibile aggiornare l'istanza o risolverne i problemi in base alle esigenze. Al termine, continuare con la fase successiva per riportare l'istanza in servizio.
- Seleziona l'istanza, scegli Azioni, Imposta su. InService Nella finestra di InService dialogo Imposta su, scegliete Imposta su InService.

AWS CLI

Per rimuovere temporaneamente un'istanza dal gruppo Auto Scaling, utilizzate i seguenti comandi di esempio. Sostituisci ciascun *placeholder input dell'utente* con le tue informazioni.

Per rimuovere temporaneamente un'istanza

- Utilizzate il [describe-auto-scaling-instances](#) comando seguente per identificare l'istanza da aggiornare.

```
aws autoscaling describe-auto-scaling-instances \  
  --query 'AutoScalingInstances[?AutoScalingGroupName==`my-asg`]'
```

L'esempio seguente mostra l'output prodotto quando si esegue questo comando.

Prendi nota dell'ID dell'istanza che intendi rimuovere dal gruppo. Questo ID ti servirà nel passaggio successivo.

```
{  
  "AutoScalingInstances": [  
    {  
      "ProtectedFromScaleIn": false,  
      "AvailabilityZone": "us-west-2a",  
      "LaunchTemplate": {  
        "LaunchTemplateName": "my-launch-template",  
        "Version": "1",  
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"  
      },  
      "InstanceId": "i-05b4f7d5be44822a6",  
      "InstanceType": "t3.micro",  
      "AutoScalingGroupName": "my-asg",  
      "HealthStatus": "HEALTHY",  
      "LifecycleState": "InService"  
    }  
  ]  
}
```

```

    },
    ...
  ]
}

```

2. Metti l'istanza in uno stato Standby utilizzando il seguente comando: [enter-standby](#). L'opzione `--should-decrement-desired-capacity` diminuisce la capacità desiderata in modo che il gruppo con scalabilità automatica non avvii un'istanza sostitutiva.

```

aws autoscaling enter-standby --instance-ids i-05b4f7d5be44822a6 \
  --auto-scaling-group-name my-asg --should-decrement-desired-capacity

```

Di seguito è riportata una risposta di esempio.

```

{
  "Activities": [
    {
      "ActivityId": "3b1839fe-24b0-40d9-80ae-bcd883c2be32",
      "AutoScalingGroupName": "my-asg",
      "Description": "Moving EC2 instance to Standby:
i-05b4f7d5be44822a6",
      "Cause": "At 2023-12-15T21:31:26Z instance i-05b4f7d5be44822a6 was
moved to standby
in response to a user request, shrinking the capacity from 4 to
3.",
      "StartTime": "2023-12-15T21:31:26.150Z",
      "StatusCode": "InProgress",
      "Progress": 50,
      "Details": "{\"Subnet ID\":\"subnet-c934b782\",\"Availability Zone
\":\"us-west-2a\"}"
    }
  ]
}

```

3. (Facoltativo) Verifica che l'istanza sia Standby attiva utilizzando il [describe-auto-scaling-instances](#) comando seguente.

```

aws autoscaling describe-auto-scaling-instances --instance-
ids i-05b4f7d5be44822a6

```

Di seguito è riportata una risposta di esempio. Nota che lo stato dell'istanza è ora Standby.


```
{
  "AutoScalingInstances": [
    {
      "ProtectedFromScaleIn": false,
      "AvailabilityZone": "us-west-2a",
      "LaunchTemplate": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "1",
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"
      },
      "InstanceId": "i-05b4f7d5be44822a6",
      "InstanceType": "t3.micro",
      "AutoScalingGroupName": "my-asg",
      "HealthStatus": "HEALTHY",
      "LifecycleState": "Standby"
    },
    ...
  ]
}
```

4. È possibile aggiornare l'istanza o risolverne i problemi in base alle esigenze. Al termine, continuare con la fase successiva per riportare l'istanza in servizio.
5. Riporta l'istanza in servizio utilizzando il seguente comando: [exit-standby](#).

```
aws autoscaling exit-standby --instance-ids i-05b4f7d5be44822a6 --auto-scaling-
group-name my-asg
```

Di seguito è riportata una risposta di esempio.

```
{
  "Activities": [
    {
      "ActivityId": "db12b166-cdcc-4c54-8aac-08c5935f8389",
      "AutoScalingGroupName": "my-asg",
      "Description": "Moving EC2 instance out of Standby:
i-05b4f7d5be44822a6",
      "Cause": "At 2023-12-15T21:46:14Z instance i-05b4f7d5be44822a6 was
moved out of standby in
      response to a user request, increasing the capacity from 3 to
4.",
      "StartTime": "2023-12-15T21:46:14.678Z",
```

```

        "StatusCode": "PreInService",
        "Progress": 30,
        "Details": "{\"Subnet ID\": \"subnet-c934b782\", \"Availability Zone
\": \"us-west-2a\"}"
    }
]
}

```

6. (Facoltativo) Verifica che l'istanza sia in di nuovo in servizio utilizzando il comando `describe-auto-scaling-instances` seguente.

```
aws autoscaling describe-auto-scaling-instances --instance-ids i-05b4f7d5be44822a6
```

Di seguito è riportata una risposta di esempio. Nota che lo stato dell'istanza è `InService`.

```

{
  "AutoScalingInstances": [
    {
      "ProtectedFromScaleIn": false,
      "AvailabilityZone": "us-west-2a",
      "LaunchTemplate": {
        "LaunchTemplateName": "my-launch-template",
        "Version": "1",
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"
      },
      "InstanceId": "i-05b4f7d5be44822a6",
      "InstanceType": "t3.micro",
      "AutoScalingGroupName": "my-asg",
      "HealthStatus": "HEALTHY",
      "LifecycleState": "InService"
    },
    ...
  ]
}

```

Eliminazione dell'infrastruttura Auto Scaling

Per eliminare completamente l'infrastruttura di dimensionamento, completa le seguenti operazioni.

Attività

- [Elimina il gruppo con scalabilità automatica](#)
- [\(Facoltativo\) Elimina la configurazione di avvio](#)
- [\(Facoltativo\) Elimina il modello di avvio](#)
- [\(Facoltativo\) Elimina il load balancer e i gruppi di destinazione](#)
- [\(Facoltativo\) Eliminare gli allarmi CloudWatch](#)

Elimina il gruppo con scalabilità automatica

Quando elimini un gruppo con scalabilità automatica, i valori desiderato, minimo e massimo sono impostati a 0. Di conseguenza, le istanze vengono arrestate. L'eliminazione di un'istanza comporta l'eliminazione anche dei registri o dei dati associati e di qualsiasi volume presente nell'istanza. Se ci sono istanze che non desideri terminare, puoi distaccarle, prima di eliminare il gruppo con scalabilità automatica. Se il gruppo dispone di policy di dimensionamento, l'eliminazione del gruppo elimina le policy, le operazioni di allarme sottostanti e qualsiasi allarme a cui non è più associata alcuna operazione.

Come eliminare il gruppo con dimensionamento automatico (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico, quindi scegli Azioni, Elimina.
3. Quando viene richiesta la conferma, digita **delete** per confermare l'eliminazione del gruppo con dimensionamento automatico specificato e scegli Elimina.

Un'icona di caricamento nella colonna Nome indica che il gruppo con dimensionamento automatico è in fase di eliminazione. Le colonne Desired (Desiderata), Min (Minima) e Max (Massima) mostrano 0 istanze per il gruppo con scalabilità automatica. Sono necessari alcuni minuti per terminare l'istanza ed eliminare il gruppo. Aggiorna l'elenco per visualizzare lo stato corrente.

Come eliminare il gruppo con dimensionamento automatico (AWS CLI)

Utilizzate il seguente [delete-auto-scaling-group](#) comando per eliminare il gruppo Auto Scaling. Questa operazione non funziona se il gruppo ha istanze EC2. È solo per gruppi senza istanze.

```
aws autoscaling delete-auto-scaling-group --auto-scaling-group-name my-asg
```

Se il gruppo ha istanze o attività di ridimensionamento in corso, utilizzate il [delete-auto-scaling-group](#) comando con l'opzione. `--force-delete` Ciò consente inoltre di terminare le istanze EC2. Quando elimini un gruppo con scalabilità automatica dalla console Dimensionamento automatico Amazon EC2, la console utilizza questa operazione per terminare le istanze EC2 ed eliminare il gruppo nello stesso momento.

```
aws autoscaling delete-auto-scaling-group --auto-scaling-group-name my-asg --force-delete
```

(Facoltativo) Elimina la configurazione di avvio

È possibile saltare questo passaggio per mantenere la configurazione di avvio per uso futuro.

Come eliminare la configurazione di avvio (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nel pannello sinistro di navigazione in Dimensionamento automatico, scegliere Gruppi con dimensionamento automatico.
3. Nella parte superiore della pagina, scegli Configurazioni di avvio. Quando viene richiesta la conferma, scegli Visualizza configurazioni di avvio per confermare che desideri visualizzare la pagina delle configurazioni di avvio.
4. Seleziona la configurazione di avvio e scegli Operazioni, Elimina configurazione di avvio.
5. Quando viene richiesta la conferma, seleziona Delete (Elimina).

Come eliminare la configurazione di avvio (AWS CLI)

Utilizza il seguente comando [delete-launch-configuration](#).

```
aws autoscaling delete-launch-configuration --launch-configuration-name my-launch-config
```

(Facoltativo) Elimina il modello di avvio

Puoi eliminare il modello di avvio o solo una delle sue versioni. Quando si elimina un modello di avvio, vengono eliminate tutte le relative versioni.

Se desideri mantenere il modello di avvio per utilizzarlo in futuro, ignora questo passaggio.

Come eliminare il modello di avvio (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nel pannello di navigazione, in Istanze, scegli Modelli di avvio.
3. Seleziona il modello di avvio e procedi in uno dei seguenti modi:
 - Scegliere Actions (Operazioni), Delete template (Elimina modello). Quando viene richiesta la conferma, digitare **Delete** per confermare l'eliminazione del modello di avvio specificato e scegliere Delete (Elimina).
 - Scegli Actions (Operazioni), Delete template version (Elimina la versione del modello). Selezionare la versione da eliminare e scegliere Delete (Elimina).

Come eliminare il modello di avvio (AWS CLI)

Utilizzate il seguente [delete-launch-template](#) comando per eliminare il modello e tutte le sue versioni.

```
aws ec2 delete-launch-template --launch-template-id lt-068f72b72934aff71
```

In alternativa, puoi usare il [delete-launch-template-versions](#) comando per eliminare una versione specifica di un modello di lancio.

```
aws ec2 delete-launch-template-versions --launch-template-id lt-068f72b72934aff71 --versions 1
```

(Facoltativo) Elimina il load balancer e i gruppi di destinazione

Se il gruppo con scalabilità automatica non è associato a un load balancer Elastic Load Balancing o se desideri mantenere il load balancer per uso futuro, salta questo passaggio.

Per eliminare il load balancer (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nel pannello di navigazione, sotto Bilanciamento del carico, scegli Sistemi di bilanciamento del carico.
3. Seleziona il load balancer e scegli Actions (Operazioni), Delete (Elimina).

4. Quando viene richiesta la conferma, seleziona Sì, elimina.

Come eliminare il gruppo di destinazione (console)

1. Nel pannello di navigazione, in Bilanciamento del carico scegli Gruppi di destinazione.
2. Selezionare il gruppo di destinazione e scegliere Operazioni, Elimina.
3. Quando viene richiesta la conferma, seleziona Sì, elimina.

Come eliminare il load balancer associato al gruppo con scalabilità automatica(AWS CLI)

Per Application Load Balancer e Network Load Balancer, utilizzate i seguenti comandi. [delete-load-balancerdelete-target-group](#)

```
aws elbv2 delete-load-balancer --load-balancer-arn my-load-balancer-arn
aws elbv2 delete-target-group --target-group-arn my-target-group-arn
```

Per Classic Load Balancers, utilizzate il seguente comando. [delete-load-balancer](#)

```
aws elb delete-load-balancer --load-balancer-name my-load-balancer
```

(Facoltativo) Eliminare gli allarmi CloudWatch

Per eliminare gli CloudWatch allarmi associati al tuo gruppo Auto Scaling, completa i seguenti passaggi. Ad esempio, potresti avere allarmi associati al dimensionamento in fasi o a semplici policy di dimensionamento.

Note

L'eliminazione di un gruppo Auto Scaling elimina CloudWatch automaticamente gli allarmi gestiti da Amazon EC2 Auto Scaling per una politica di scalabilità di tracciamento mirata.

Puoi saltare questo passaggio se il tuo gruppo Auto Scaling non è associato ad CloudWatch alcun allarme o se desideri conservare gli allarmi per usi futuri.

Per eliminare gli allarmi (console CloudWatch)

1. Apri la CloudWatch console all'indirizzo <https://console.aws.amazon.com/cloudwatch/>.

2. Nel pannello di navigazione, scegli Allarmi.
3. Scegli gli allarmi, quindi Action (Operazione), Delete (Elimina).
4. Quando viene richiesta la conferma, seleziona Delete (Elimina).

Per eliminare gli CloudWatch allarmi (AWS CLI)

Utilizza il comando [delete-alarms](#). Puoi eliminare uno o più allarmi alla volta. Ad esempio, utilizzare il seguente comando per eliminare gli allarmi Step-Scaling-AlarmHigh-AddCapacity e Step-Scaling-AlarmLow-RemoveCapacity.

```
aws cloudwatch delete-alarms --alarm-name Step-Scaling-AlarmHigh-AddCapacity Step-Scaling-AlarmLow-RemoveCapacity
```

Esempi di creazione e gestione di gruppi Auto Scaling con gli SDK AWS

È possibile creare un gruppo Auto Scaling utilizzando AWS Management Console, the AWS CLI, un AWS SDK e. AWS CloudFormation

I seguenti esempi di codice mostrano come creare, aggiornare, descrivere ed eliminare un gruppo Auto Scaling nel linguaggio di programmazione supportato preferito utilizzando gli AWS SDK.

Indice

- [Crea un gruppo Auto Scaling utilizzando un SDK AWS](#)
- [Aggiornare un gruppo Auto Scaling utilizzando un SDK AWS](#)
- [Descrivi un gruppo di Auto Scaling utilizzando un SDK AWS](#)
- [Eliminare un gruppo Auto Scaling utilizzando un SDK AWS](#)

Crea un gruppo Auto Scaling utilizzando un SDK AWS

I seguenti esempi di codice mostrano come utilizzare. `CreateAutoScalingGroup`

.NET

AWS SDK for .NET

Note

C'è altro da fare GitHub. Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
/// <summary>
/// Create a new Amazon EC2 Auto Scaling group.
/// </summary>
/// <param name="groupName">The name to use for the new Auto Scaling
/// group.</param>
/// <param name="launchTemplateName">The name of the Amazon EC2 Auto Scaling
/// launch template to use to create instances in the group.</param>
/// <returns>A Boolean value indicating the success of the action.</returns>
public async Task<bool> CreateAutoScalingGroupAsync(
    string groupName,
    string launchTemplateName,
    string availabilityZone)
{
    var templateSpecification = new LaunchTemplateSpecification
    {
        LaunchTemplateName = launchTemplateName,
    };

    var zoneList = new List<string>
    {
        availabilityZone,
    };

    var request = new CreateAutoScalingGroupRequest
    {
        AutoScalingGroupName = groupName,
        AvailabilityZones = zoneList,
        LaunchTemplate = templateSpecification,
        MaxSize = 6,
        MinSize = 1
    };
};
```



```
var response = await
_amazonAutoScaling.CreateAutoScalingGroupAsync(request);
Console.WriteLine($"{groupName} Auto Scaling Group created");
return response.HttpStatusCode == System.Net.HttpStatusCode.OK;
}
```

- Per i dettagli sull'API, [CreateAutoScalingGroup](#) consulta AWS SDK for .NET API Reference.

C++

SDK per C++

Note

C'è altro su GitHub. Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
Aws::Client::ClientConfiguration clientConfig;
// Optional: Set to the AWS Region (overrides config file).
// clientConfig.region = "us-east-1";

Aws::AutoScaling::AutoScalingClient autoScalingClient(clientConfig);

Aws::AutoScaling::Model::CreateAutoScalingGroupRequest request;
request.SetAutoScalingGroupName(groupName);
Aws::Vector<Aws::String> availabilityGroupZones;
availabilityGroupZones.push_back(
    availabilityZones[availabilityZoneChoice - 1].GetZoneName());
request.SetAvailabilityZones(availabilityGroupZones);
request.SetMaxSize(1);
request.SetMinSize(1);

Aws::AutoScaling::Model::LaunchTemplateSpecification
launchTemplateSpecification;
launchTemplateSpecification.SetLaunchTemplateName(templateName);
request.SetLaunchTemplate(launchTemplateSpecification);

Aws::AutoScaling::Model::CreateAutoScalingGroupOutcome outcome =
```

```
        autoScalingClient.CreateAutoScalingGroup(request);

    if (outcome.IsSuccess()) {
        std::cout << "Created Auto Scaling group '" << groupName << "'..."
                  << std::endl;
    }
    else if (outcome.GetError().GetErrorType() ==
             Aws::AutoScaling::AutoScalingErrors::ALREADY_EXISTS_FAULT) {
        std::cout << "Auto Scaling group '" << groupName << "' already
exists."
                  << std::endl;
    }
    else {
        std::cerr << "Error with AutoScaling::CreateAutoScalingGroup. "
                  << outcome.GetError().GetMessage()
                  << std::endl;
    }
}
```

- Per i dettagli sull'API, [CreateAutoScalingGroup](#) consulta AWS SDK for C++ API Reference.

CLI

AWS CLI

Esempio 1: creare un gruppo Auto Scaling

L'create-auto-scaling-groupesempio seguente crea un gruppo Auto Scaling in sottoreti in più zone di disponibilità all'interno di una regione. Le istanze vengono avviate con la versione predefinita del modello di avvio specificato. Tieni presente che le impostazioni predefinite vengono utilizzate per la maggior parte delle altre impostazioni, come le politiche di terminazione e la configurazione del controllo dello stato di salute.

```
aws autoscaling create-auto-scaling-group \
  --auto-scaling-group-name my-asg \
  --launch-template LaunchTemplateId=lt-1234567890abcde12 \
  --min-size 1 \
  --max-size 5 \
  --vpc-zone-identifier "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782"
```

Questo comando non produce alcun output.

Per ulteriori informazioni, consulta [Gruppi con scalabilità automatica](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

Esempio 2: per collegare un Application Load Balancer, un Network Load Balancer o un Gateway Load Balancer

Questo esempio specifica l'ARN di un gruppo target per un sistema di bilanciamento del carico che supporta il traffico previsto. Il tipo di controllo dello stato specifica ELB che quando Elastic Load Balancing segnala un'istanza come non integra, il gruppo Auto Scaling la sostituisce. Il comando definisce anche un periodo di tolleranza di secondi per il controllo dello stato di salute. 600 Il periodo di tolleranza aiuta a prevenire la chiusura anticipata delle istanze appena avviate.

```
aws autoscaling create-auto-scaling-group \  
  --auto-scaling-group-name my-asg \  
  --launch-template LaunchTemplateId=lt-1234567890abcde12 \  
  --target-group-arns arn:aws:elasticloadbalancing:us-  
west-2:123456789012:targetgroup/my-targets/943f017f100becff \  
  --health-check-type ELB \  
  --health-check-grace-period 600 \  
  --min-size 1 \  
  --max-size 5 \  
  --vpc-zone-identifier "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782"
```

Questo comando non produce alcun output.

Per maggiori informazioni, consulta [Elastic Load Balancing e Amazon EC2 Auto Scaling](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

Esempio 3: per specificare un gruppo di collocamenti e utilizzare la versione più recente del modello di lancio

Questo esempio avvia le istanze in un gruppo di posizionamento all'interno di una singola zona di disponibilità. Questo può essere utile per gruppi a bassa latenza con carichi di lavoro HPC. Questo esempio specifica anche la dimensione minima, la dimensione massima e la capacità desiderata del gruppo.

```
aws autoscaling create-auto-scaling-group \  
  --auto-scaling-group-name my-asg \  
  --launch-template LaunchTemplateId=lt-1234567890abcde12,Version='$Latest' \  
  --min-size 1 \  
  --max-size 5 \  
  --vpc-zone-identifier "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782"
```

```
--max-size 5 \  
--desired-capacity 3 \  
--placement-group my-placement-group \  
--vpc-zone-identifier "subnet-6194ea3b"
```

Questo comando non produce alcun output.

Per ulteriori informazioni, consulta [Gruppi di collocamento](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Esempio 4: Per specificare un gruppo Auto Scaling a istanza singola e utilizzare una versione specifica del modello di avvio

Questo esempio crea un gruppo Auto Scaling con capacità minima e massima impostata per 1 impone l'esecuzione di un'istanza. Il comando specifica anche la v1 di un modello di avvio in cui è specificato l'ID di un ENI esistente. Quando si utilizza un modello di avvio che specifica un ENI esistente per eth0, è necessario specificare una zona di disponibilità per il gruppo Auto Scaling che corrisponda all'interfaccia di rete, senza specificare anche un ID di sottorete nella richiesta.

```
aws autoscaling create-auto-scaling-group \  
  --auto-scaling-group-name my-asg-single-instance \  
  --launch-template LaunchTemplateName=my-template-for-auto-scaling,Version='1'  
 \  
  --min-size 1 \  
  --max-size 1 \  
  --availability-zones us-west-2a
```

Questo comando non produce alcun output.

Per ulteriori informazioni, consulta [Gruppi con scalabilità automatica](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

Esempio 5: specificare una politica di terminazione diversa

Questo esempio crea un gruppo Auto Scaling utilizzando una configurazione di avvio e imposta la politica di terminazione per terminare prima le istanze più vecchie. Il comando applica inoltre un tag al gruppo e alle relative istanze, con una chiave di e un valore diRole. WebServer

```
aws autoscaling create-auto-scaling-group \  
  --auto-scaling-group-name my-asg \  
  --tag Key=Role,Value=WebServer
```

```
--launch-configuration-name my-lc \  
--min-size 1 \  
--max-size 5 \  
--termination-policies "OldestInstance" \  
--tags "ResourceId=my-asg,ResourceType=auto-scaling-  
group,Key=Role,Value=WebServer,PropagateAtLaunch=true" \  
--vpc-zone-identifier "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782"
```

Questo comando non produce alcun output.

Per ulteriori informazioni, consulta [Working with Amazon EC2 Auto Scaling policy di terminazione nella Amazon EC2 Auto Scaling User Guide](#).

Esempio 6: specificare un hook per il ciclo di vita del lancio

Questo esempio crea un gruppo Auto Scaling con un lifecycle hook che supporta un'azione personalizzata all'avvio dell'istanza.

```
aws autoscaling create-auto-scaling-group \  
--cli-input-json file://~/config.json
```

Contenuto del file: config.json

```
{  
  "AutoScalingGroupName": "my-asg",  
  "LaunchTemplate": {  
    "LaunchTemplateId": "lt-1234567890abcde12"  
  },  
  "LifecycleHookSpecificationList": [{  
    "LifecycleHookName": "my-launch-hook",  
    "LifecycleTransition": "autoscaling:EC2_INSTANCE_LAUNCHING",  
    "NotificationTargetARN": "arn:aws:sqs:us-west-2:123456789012:my-sqs-  
queue",  
    "RoleARN": "arn:aws:iam::123456789012:role/my-notification-role",  
    "NotificationMetadata": "SQS message metadata",  
    "HeartbeatTimeout": 4800,  
    "DefaultResult": "ABANDON"  
  }],  
  "MinSize": 1,  
  "MaxSize": 5,  
  "VPCZoneIdentifier": "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782",  
  "Tags": [{  
    "ResourceType": "auto-scaling-group",
```

```
    "ResourceId": "my-asg",
    "PropagateAtLaunch": true,
    "Value": "test",
    "Key": "environment"
  ]
}
```

Questo comando non produce alcun output.

Per ulteriori informazioni, consulta [Hook del ciclo di vita di Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

Esempio 7: per specificare un hook del ciclo di vita della terminazione

Questo esempio crea un gruppo Auto Scaling con un lifecycle hook che supporta un'azione personalizzata alla chiusura dell'istanza.

```
aws autoscaling create-auto-scaling-group \
  --cli-input-json file://~/config.json
```

Contenuto di config.json.

```
{
  "AutoScalingGroupName": "my-asg",
  "LaunchTemplate": {
    "LaunchTemplateId": "lt-1234567890abcde12"
  },
  "LifecycleHookSpecificationList": [{
    "LifecycleHookName": "my-termination-hook",
    "LifecycleTransition": "autoscaling:EC2_INSTANCE_TERMINATING",
    "HeartbeatTimeout": 120,
    "DefaultResult": "CONTINUE"
  }],
  "MinSize": 1,
  "MaxSize": 5,
  "TargetGroupARNs": [
    "arn:aws:elasticloadbalancing:us-west-2:123456789012:targetgroup/my-
    targets/73e2d6bc24d8a067"
  ],
  "VPCZoneIdentifier": "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782"
}
```

Questo comando non produce alcun output.

Per ulteriori informazioni, consulta [Hook del ciclo di vita di Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

Esempio 8: Per specificare una politica di terminazione personalizzata

Questo esempio crea un gruppo Auto Scaling che specifica una policy di terminazione della funzione Lambda personalizzata che indica ad Amazon EC2 Auto Scaling su quali istanze è sicuro terminare in scalabilità.

```
aws autoscaling create-auto-scaling-group \  
  --auto-scaling-group-name my-asg-single-instance \  
  --launch-template LaunchTemplateName=my-template-for-auto-scaling \  
  --min-size 1 \  
  --max-size 5 \  
  --termination-policies "arn:aws:lambda:us-  
west-2:123456789012:function:HelloFunction:prod" \  
  --vpc-zone-identifier "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782"
```

Questo comando non produce alcun output.

Per ulteriori informazioni, consulta [Creazione di una politica di terminazione personalizzata con Lambda](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

- Per i dettagli sull'API, consulta [CreateAutoScalingGroup](#) Command Reference.AWS CLI

Java

SDK per Java 2.x

Note

C'è altro su GitHub. Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
import software.amazon.awssdk.core.waiters.WaiterResponse;  
import software.amazon.awssdk.regions.Region;  
import software.amazon.awssdk.services.autoscaling.AutoScalingClient;  
import software.amazon.awssdk.services.autoscaling.model.AutoScalingException;  
import  
  software.amazon.awssdk.services.autoscaling.model.CreateAutoScalingGroupRequest;
```

```
import
    software.amazon.awssdk.services.autoscaling.model.DescribeAutoScalingGroupsRequest;
import
    software.amazon.awssdk.services.autoscaling.model.DescribeAutoScalingGroupsResponse;
import
    software.amazon.awssdk.services.autoscaling.model.LaunchTemplateSpecification;
import software.amazon.awssdk.services.autoscaling.waiters.AutoScalingWaiter;

/**
 * Before running this SDK for Java (v2) code example, set up your development
 * environment, including your credentials.
 *
 * For more information, see the following documentation:
 *
 * https://docs.aws.amazon.com/sdk-for-java/latest/developer-guide/get-started.html
 */
public class CreateAutoScalingGroup {
    public static void main(String[] args) {
        final String usage = ""

            Usage:
                <groupName> <launchTemplateName> <serviceLinkedRoleARN>
<vpcZoneId>

            Where:
                groupName - The name of the Auto Scaling group.
                launchTemplateName - The name of the launch template.\s
                vpcZoneId - A subnet Id for a virtual private cloud (VPC)
where instances in the Auto Scaling group can be created.
            """;

        if (args.length != 3) {
            System.out.println(usage);
            System.exit(1);
        }

        String groupName = args[0];
        String launchTemplateName = args[1];
        String vpcZoneId = args[2];
        AutoScalingClient autoScalingClient = AutoScalingClient.builder()
            .region(Region.US_EAST_1)
            .build();
```



```
        createAutoScalingGroup(autoScalingClient, groupName, launchTemplateName,
vpcZoneId);
        autoScalingClient.close();
    }

    public static void createAutoScalingGroup(AutoScalingClient
autoScalingClient,
        String groupName,
        String launchTemplateName,
        String vpcZoneId) {

        try {
            AutoScalingWaiter waiter = autoScalingClient.waiter();
            LaunchTemplateSpecification templateSpecification =
LaunchTemplateSpecification.builder()
                .launchTemplateName(launchTemplateName)
                .build();

            CreateAutoScalingGroupRequest request =
CreateAutoScalingGroupRequest.builder()
                .autoScalingGroupName(groupName)
                .availabilityZones("us-east-1a")
                .launchTemplate(templateSpecification)
                .maxSize(1)
                .minSize(1)
                .vpcZoneIdentifier(vpcZoneId)
                .build();

            autoScalingClient.createAutoScalingGroup(request);
            DescribeAutoScalingGroupsRequest groupsRequest =
DescribeAutoScalingGroupsRequest.builder()
                .autoScalingGroupNames(groupName)
                .build();

            WaiterResponse<DescribeAutoScalingGroupsResponse> waiterResponse =
waiter
                .waitUntilGroupExists(groupsRequest);
            waiterResponse.matched().response().ifPresent(System.out::println);
            System.out.println("Auto Scaling Group created");

        } catch (AutoScalingException e) {
            System.err.println(e.awsErrorDetails().errorMessage());
            System.exit(1);
        }
    }
}
```

```
}  
}
```

- Per i dettagli sull'API, [CreateAutoScalingGroup](#) consulta AWS SDK for Java 2.x API Reference.

Kotlin

SDK per Kotlin

Note

C'è altro su GitHub. Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
suspend fun createAutoScalingGroup(groupName: String, launchTemplateNameVal: String, serviceLinkedRoleARNVal: String, vpcZoneIdVal: String) {  
    val templateSpecification = LaunchTemplateSpecification {  
        launchTemplateName = launchTemplateNameVal  
    }  
  
    val request = CreateAutoScalingGroupRequest {  
        autoScalingGroupName = groupName  
        availabilityZones = listOf("us-east-1a")  
        launchTemplate = templateSpecification  
        maxSize = 1  
        minSize = 1  
        vpcZoneIdentifier = vpcZoneIdVal  
        serviceLinkedRoleArn = serviceLinkedRoleARNVal  
    }  
  
    // This object is required for the waiter call.  
    val groupsRequestWaiter = DescribeAutoScalingGroupsRequest {  
        autoScalingGroupNames = listOf(groupName)  
    }  
  
    AutoScalingClient { region = "us-east-1" }.use { autoScalingClient ->  
        autoScalingClient.createAutoScalingGroup(request)  
        autoScalingClient.waitUntilGroupExists(groupsRequestWaiter)  
    }  
}
```

```
        println("$groupName was created!")
    }
}
```

- Per i dettagli sull'API, [CreateAutoScalingGroup](#) consulta AWS SDK for Kotlin API reference.

PHP

SDK per PHP

Note

C'è altro su [GitHub](#) Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
public function createAutoScalingGroup(
    $autoScalingGroupName,
    $availabilityZones,
    $minSize,
    $maxSize,
    $launchTemplateId
) {
    return $this->autoScalingClient->createAutoScalingGroup([
        'AutoScalingGroupName' => $autoScalingGroupName,
        'AvailabilityZones' => $availabilityZones,
        'MinSize' => $minSize,
        'MaxSize' => $maxSize,
        'LaunchTemplate' => [
            'LaunchTemplateId' => $launchTemplateId,
        ],
    ]);
}
```

- Per i dettagli sull'API, [CreateAutoScalingGroup](#) consulta AWS SDK for PHP API Reference.

PowerShell

Strumenti per PowerShell

Esempio 1: Questo esempio crea un gruppo Auto Scaling con il nome e gli attributi specificati. La capacità predefinita desiderata è la dimensione minima. Pertanto, questo gruppo Auto Scaling avvia due istanze, una in ciascuna delle due zone di disponibilità specificate.

```
New-ASAutoScalingGroup -AutoScalingGroupName my-asg -LaunchConfigurationName my-  
lc -MinSize 2 -MaxSize 6 -AvailabilityZone @("us-west-2a", "us-west-2b")
```

- Per i dettagli sull'API, vedere [CreateAutoScalingGroup](#) in AWS Tools for PowerShell Cmdlet Reference.

Python

SDK per Python (Boto3)

Note

C'è altro su [GitHub](#) Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
class AutoScalingWrapper:  
    """Encapsulates Amazon EC2 Auto Scaling actions."""  
  
    def __init__(self, autoscaling_client):  
        """  
        :param autoscaling_client: A Boto3 Amazon EC2 Auto Scaling client.  
        """  
        self.autoscaling_client = autoscaling_client  
  
    def create_group(  
        self, group_name, group_zones, launch_template_name, min_size, max_size  
    ):  
        """  
        Creates an Auto Scaling group.
```

```
        :param group_name: The name to give to the group.
        :param group_zones: The Availability Zones in which instances can be
        created.
        :param launch_template_name: The name of an existing Amazon EC2 launch
        template.

        The launch template specifies the
        configuration of
        instances that are created by auto scaling
        activities.
        :param min_size: The minimum number of active instances in the group.
        :param max_size: The maximum number of active instances in the group.
        """
    try:
        self.autoscaling_client.create_auto_scaling_group(
            AutoScalingGroupName=group_name,
            AvailabilityZones=group_zones,
            LaunchTemplate={
                "LaunchTemplateName": launch_template_name,
                "Version": "$Default",
            },
            MinSize=min_size,
            MaxSize=max_size,
        )
    except ClientError as err:
        logger.error(
            "Couldn't create group %s. Here's why: %s: %s",
            group_name,
            err.response["Error"]["Code"],
            err.response["Error"]["Message"],
        )
    raise
```

- Per i dettagli sull'API, consulta [CreateAutoScalingGroup AWSSDK for Python \(Boto3\) API Reference](#).

Rust

SDK per Rust

Note

C'è di più su. [GitHub](#) Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
async fn create_group(client: &Client, name: &str, id: &str) -> Result<(), Error>
{
    client
        .create_auto_scaling_group()
        .auto_scaling_group_name(name)
        .instance_id(id)
        .min_size(1)
        .max_size(5)
        .send()
        .await?;

    println!("Created AutoScaling group");

    Ok(())
}
```

- Per i dettagli sulle API, consulta il riferimento [CreateAutoScalingGroup](#) all'API AWS SDK for Rust.

Per esempi che puoi utilizzare per creare [gruppi di istanze miste](#), consulta le seguenti risorse.

- [AWS SDK per .NET](#)
- [AWS SDK for Go](#)
- [AWS SDK per JavaScript](#)
- [AWS SDK per PHP V3](#)
- [AWS SDK per Python](#)
- [AWS SDK per Ruby V3](#)

Aggiornare un gruppo Auto Scaling utilizzando un SDK AWS

I seguenti esempi di codice mostrano come utilizzare `UpdateAutoScalingGroup`

.NET

AWS SDK for .NET

Note

C'è altro da fare GitHub. Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
/// <summary>
/// Update the capacity of an Auto Scaling group.
/// </summary>
/// <param name="groupName">The name of the Auto Scaling group.</param>
/// <param name="launchTemplateName">The name of the EC2 launch template.</
param>
/// <param name="maxSize">The maximum number of instances that can be
/// created for the Auto Scaling group.</param>
/// <returns>A Boolean value indicating the success of the action.</returns>
public async Task<bool> UpdateAutoScalingGroupAsync(
    string groupName,
    string launchTemplateName,
    int maxSize)
{
    var templateSpecification = new LaunchTemplateSpecification
    {
        LaunchTemplateName = launchTemplateName,
    };

    var groupRequest = new UpdateAutoScalingGroupRequest
    {
        MaxSize = maxSize,
        AutoScalingGroupName = groupName,
        LaunchTemplate = templateSpecification,
    };

    var response = await
_amazonAutoScaling.UpdateAutoScalingGroupAsync(groupRequest);
```

```
        if (response.HttpStatusCode == System.Net.HttpStatusCode.OK)
        {
            Console.WriteLine($"You successfully updated the Auto Scaling group
{groupName}.");
            return true;
        }
        else
        {
            return false;
        }
    }
}
```

- Per i dettagli sull'API, [UpdateAutoScalingGroup](#) consulta AWS SDK for .NET API Reference.

C++

SDK per C++

Note

C'è altro su GitHub. Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
Aws::Client::ClientConfiguration clientConfig;
// Optional: Set to the AWS Region (overrides config file).
// clientConfig.region = "us-east-1";

Aws::AutoScaling::AutoScalingClient autoScalingClient(clientConfig);

Aws::AutoScaling::Model::UpdateAutoScalingGroupRequest request;
request.SetAutoScalingGroupName(groupName);
request.SetMaxSize(3);

Aws::AutoScaling::Model::UpdateAutoScalingGroupOutcome outcome =
    autoScalingClient.UpdateAutoScalingGroup(request);

if (!outcome.IsSuccess()) {
    std::cerr << "Error with AutoScaling::UpdateAutoScalingGroup. "
        << outcome.GetError().GetMessage()

```



```
        << std::endl;
    }
}
```

- Per i dettagli sull'API, [UpdateAutoScalingGroup](#) consulta AWS SDK for C++ API Reference.

CLI

AWS CLI

Esempio 1: Per aggiornare i limiti di dimensione di un gruppo di Auto Scaling

Questo esempio aggiorna il gruppo Auto Scaling specificato con una dimensione minima di 2 e una dimensione massima di 10.

```
aws autoscaling update-auto-scaling-group \
  --auto-scaling-group-name my-asg \
  --min-size 2 \
  --max-size 10
```

Questo comando non produce alcun output.

Per ulteriori informazioni, consulta [Impostazione dei limiti di capacità per il gruppo Auto Scaling](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

Esempio 2: aggiungere i controlli di integrità di Elastic Load Balancing e specificare quali zone di disponibilità e sottoreti utilizzare

Questo esempio aggiorna il gruppo Auto Scaling specificato per aggiungere i controlli di integrità Elastic Load Balancing. Questo comando aggiorna anche il valore di `--vpc-zone-identifier` con un elenco di ID di sottorete in più zone di disponibilità.

```
aws autoscaling update-auto-scaling-group \
  --auto-scaling-group-name my-asg \
  --health-check-type ELB \
  --health-check-grace-period 600 \
  --vpc-zone-identifier "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782"
```

Questo comando non produce alcun output.

Per maggiori informazioni, consulta [Elastic Load Balancing e Amazon EC2 Auto Scaling](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

Esempio 3: per aggiornare il gruppo di collocamento e la politica di cessazione

Questo esempio aggiorna il gruppo di collocamento e la politica di licenziamento da utilizzare.

```
aws autoscaling update-auto-scaling-group \  
  --auto-scaling-group-name my-asg \  
  --placement-group my-placement-group \  
  --termination-policies "OldestInstance"
```

Questo comando non produce alcun output.

Per ulteriori informazioni, consulta [Gruppi con scalabilità automatica](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

Esempio 4: per utilizzare la versione più recente del modello di lancio

Questo esempio aggiorna il gruppo Auto Scaling specificato per utilizzare la versione più recente del modello di avvio specificato.

```
aws autoscaling update-auto-scaling-group \  
  --auto-scaling-group-name my-asg \  
  --launch-template LaunchTemplateId=lt-1234567890abcde12,Version='$Latest'
```

Questo comando non produce alcun output.

Per ulteriori informazioni, consulta [Modelli di avvio](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

Esempio 5: per utilizzare una versione specifica del modello di lancio

Questo esempio aggiorna il gruppo Auto Scaling specificato per utilizzare una versione specifica di un modello di avvio anziché la versione più recente o predefinita.

```
aws autoscaling update-auto-scaling-group \  
  --auto-scaling-group-name my-asg \  
  --launch-template LaunchTemplateName=my-template-for-auto-scaling,Version='2'
```

Questo comando non produce alcun output.

Per ulteriori informazioni, consulta [Modelli di avvio](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

Esempio 6: definire una politica a istanze miste e abilitare il ribilanciamento della capacità

Questo esempio aggiorna il gruppo Auto Scaling specificato per utilizzare una politica di istanze miste e consente il ribilanciamento della capacità. Questa struttura consente di specificare gruppi con capacità Spot e On-Demand e utilizzare modelli di avvio diversi per architetture diverse.

```
aws autoscaling update-auto-scaling-group \  
  --cli-input-json file://~/config.json
```

Contenuto di config.json.

```
{  
  "AutoScalingGroupName": "my-asg",  
  "CapacityRebalance": true,  
  "MixedInstancesPolicy": {  
    "LaunchTemplate": {  
      "LaunchTemplateSpecification": {  
        "LaunchTemplateName": "my-launch-template-for-x86",  
        "Version": "$Latest"  
      },  
      "Overrides": [  
        {  
          "InstanceType": "c6g.large",  
          "LaunchTemplateSpecification": {  
            "LaunchTemplateName": "my-launch-template-for-arm",  
            "Version": "$Latest"  
          }  
        },  
        {  
          "InstanceType": "c5.large"  
        },  
        {  
          "InstanceType": "c5a.large"  
        }  
      ]  
    }  
  },  
}
```

```
    "InstancesDistribution": {
      "OnDemandPercentageAboveBaseCapacity": 50,
      "SpotAllocationStrategy": "capacity-optimized"
    }
  }
}
```

Questo comando non produce alcun output.

Per ulteriori informazioni, consultare la sezione relativa ai [Gruppi con dimensionamento automatico con più tipi di istanze e opzioni di acquisto](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

- Per i dettagli sull'API, consulta [UpdateAutoScalingGroup](#) Command Reference.AWS CLI

Java

SDK per Java 2.x

Note

C'è altro su GitHub. Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
public static void updateAutoScalingGroup(AutoScalingClient
autoScalingClient, String groupName,
    String launchTemplateName) {
    try {
        AutoScalingWaiter waiter = autoScalingClient.waiter();
        LaunchTemplateSpecification templateSpecification =
LaunchTemplateSpecification.builder()
            .launchTemplateName(launchTemplateName)
            .build();

        UpdateAutoScalingGroupRequest groupRequest =
UpdateAutoScalingGroupRequest.builder()
            .maxSize(3)
            .autoScalingGroupName(groupName)
            .launchTemplate(templateSpecification)
            .build();
```

```
        autoScalingClient.updateAutoScalingGroup(groupRequest);
        DescribeAutoScalingGroupsRequest groupsRequest =
DescribeAutoScalingGroupsRequest.builder()
        .autoScalingGroupNames(groupName)
        .build();

        WaiterResponse<DescribeAutoScalingGroupsResponse> waiterResponse =
waiter
        .waitUntilGroupInService(groupsRequest);
        waiterResponse.matched().response().ifPresent(System.out::println);
        System.out.println("You successfully updated the auto scaling group
" + groupName);

    } catch (AutoScalingException e) {
        System.err.println(e.awsErrorDetails().errorMessage());
        System.exit(1);
    }
}
```

- Per i dettagli sull'API, [UpdateAutoScalingGroup](#) consulta AWS SDK for Java 2.x API Reference.

Kotlin

SDK per Kotlin

Note

C'è altro su GitHub. Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
suspend fun updateAutoScalingGroup(groupName: String, launchTemplateNameVal:
String, serviceLinkedRoleARNVal: String) {
    val templateSpecification = LaunchTemplateSpecification {
        launchTemplateName = launchTemplateNameVal
    }

    val groupRequest = UpdateAutoScalingGroupRequest {
```

```

        maxSize = 3
        serviceLinkedRoleArn = serviceLinkedRoleARNVal
        autoScalingGroupName = groupName
        launchTemplate = templateSpecification
    }

    val groupsRequestWaiter = DescribeAutoScalingGroupsRequest {
        autoScalingGroupNames = listOf(groupName)
    }

    AutoScalingClient { region = "us-east-1" }.use { autoScalingClient ->
        autoScalingClient.updateAutoScalingGroup(groupRequest)
        autoScalingClient.waitUntilGroupExists(groupsRequestWaiter)
        println("You successfully updated the Auto Scaling group $groupName")
    }
}

```

- Per i dettagli sull'API, [UpdateAutoScalingGroup](#) consulta AWS SDK for Kotlin API reference.

PHP

SDK per PHP

Note

C'è altro su [GitHub](#) Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```

public function updateAutoScalingGroup($autoScalingGroupName, $args)
{
    if (array_key_exists('MaxSize', $args)) {
        $maxSize = ['MaxSize' => $args['MaxSize']];
    } else {
        $maxSize = [];
    }
    if (array_key_exists('MinSize', $args)) {
        $minSize = ['MinSize' => $args['MinSize']];
    } else {
        $minSize = [];
    }
}

```

```
$parameters = ['AutoScalingGroupName' => $autoScalingGroupName];  
$parameters = array_merge($parameters, $minSize, $maxSize);  
return $this->autoScalingClient->updateAutoScalingGroup($parameters);  
}
```

- Per i dettagli sull'API, [UpdateAutoScalingGroup](#) consulta AWS SDK for PHP API Reference.

PowerShell

Strumenti per PowerShell

Esempio 1: Questo esempio aggiorna la dimensione minima e massima del gruppo Auto Scaling specificato.

```
Update-ASAutoScalingGroup -AutoScalingGroupName my-asg -MaxSize 5 -MinSize 1
```

Esempio 2: questo esempio aggiorna il periodo di cooldown predefinito del gruppo Auto Scaling specificato.

```
Update-ASAutoScalingGroup -AutoScalingGroupName my-asg -DefaultCooldown 10
```

Esempio 3: questo esempio aggiorna le zone di disponibilità del gruppo Auto Scaling specificato.

```
Update-ASAutoScalingGroup -AutoScalingGroupName my-asg -AvailabilityZone @("us-west-2a", "us-west-2b")
```

Esempio 4: questo esempio aggiorna il gruppo Auto Scaling specificato per utilizzare i controlli di integrità Elastic Load Balancing.

```
Update-ASAutoScalingGroup -AutoScalingGroupName my-asg -HealthCheckType ELB -  
HealthCheckGracePeriod 60
```

- Per i dettagli sull'API, vedere [UpdateAutoScalingGroup](#) in AWS Tools for PowerShell Cmdlet Reference.

Python

SDK per Python (Boto3)

Note

C'è altro su GitHub Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
class AutoScalingWrapper:
    """Encapsulates Amazon EC2 Auto Scaling actions."""

    def __init__(self, autoscaling_client):
        """
        :param autoscaling_client: A Boto3 Amazon EC2 Auto Scaling client.
        """
        self.autoscaling_client = autoscaling_client

    def update_group(self, group_name, **kwargs):
        """
        Updates an Auto Scaling group.

        :param group_name: The name of the group to update.
        :param kwargs: Keyword arguments to pass through to the service.
        """
        try:
            self.autoscaling_client.update_auto_scaling_group(
                AutoScalingGroupName=group_name, **kwargs
            )
        except ClientError as err:
            logger.error(
                "Couldn't update group %s. Here's why: %s: %s",
                group_name,
                err.response["Error"]["Code"],
                err.response["Error"]["Message"],
            )
            raise
```


- Per i dettagli sull'API, consulta [UpdateAutoScalingGroup AWS SDK for Python \(Boto3\) API Reference](#).

Rust

SDK per Rust

Note

C'è di più su. [GitHub](#) Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
async fn update_group(client: &Client, name: &str, size: i32) -> Result<(),
Error> {
    client
        .update_auto_scaling_group()
        .auto_scaling_group_name(name)
        .max_size(size)
        .send()
        .await?;

    println!("Updated AutoScaling group");

    Ok(())
}
```

- Per i dettagli sulle API, consulta il riferimento [UpdateAutoScalingGroup](#) all'API AWS SDK for Rust.

Descrivi un gruppo di Auto Scaling utilizzando un SDK AWS

I seguenti esempi di codice mostrano come utilizzare `DescribeAutoScalingGroups`.

.NET

AWS SDK for .NET

Note

C'è altro da fare GitHub. Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
/// <summary>
/// Get data about the instances in an Amazon EC2 Auto Scaling group.
/// </summary>
/// <param name="groupName">The name of the Amazon EC2 Auto Scaling group.</
param>
/// <returns>A list of Amazon EC2 Auto Scaling details.</returns>
public async Task<List<AutoScalingInstanceDetails>>
DescribeAutoScalingInstancesAsync(
    string groupName)
{
    var groups = await DescribeAutoScalingGroupsAsync(groupName);
    var instanceIds = new List<string>();
    groups!.ForEach(group =>
    {
        if (group.AutoScalingGroupName == groupName)
        {
            group.Instances.ForEach(instance =>
            {
                instanceIds.Add(instance.InstanceId);
            });
        }
    });

    var scalingGroupsRequest = new DescribeAutoScalingInstancesRequest
    {
        MaxRecords = 10,
        InstanceIds = instanceIds,
    };

    var response = await
_amazonAutoScaling.DescribeAutoScalingInstancesAsync(scalingGroupsRequest);
```

```
    var instanceDetails = response.AutoScalingInstances;

    return instanceDetails;
}
```

- Per i dettagli sull'API, [DescribeAutoScalingGroups](#) consulta AWS SDK for .NET API Reference.

C++

SDK per C++

Note

C'è altro su GitHub. Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
Aws::Client::ClientConfiguration clientConfig;
// Optional: Set to the AWS Region (overrides config file).
// clientConfig.region = "us-east-1";

Aws::AutoScaling::AutoScalingClient autoScalingClient(clientConfig);

Aws::AutoScaling::Model::DescribeAutoScalingGroupsRequest request;
Aws::Vector<Aws::String> groupNames;
groupNames.push_back(groupName);
request.SetAutoScalingGroupNames(groupNames);

Aws::AutoScaling::Model::DescribeAutoScalingGroupsOutcome outcome =
    client.DescribeAutoScalingGroups(request);

if (outcome.IsSuccess()) {
    autoScalingGroup = outcome.GetResult().GetAutoScalingGroups();
}
else {
    std::cerr << "Error with AutoScaling::DescribeAutoScalingGroups. "
                << outcome.GetError().GetMessage()
                << std::endl;
```

```
}
```

- Per i dettagli sull'API, [DescribeAutoScalingGroups](#) consulta AWS SDK for C++ API Reference.

CLI

AWS CLI

Esempio 1: Per descrivere il gruppo Auto Scaling specificato

Questo esempio descrive il gruppo Auto Scaling specificato.

```
aws autoscaling describe-auto-scaling-groups \  
  --auto-scaling-group-name my-asg
```

Output:

```
{  
  "AutoScalingGroups": [  
    {  
      "AutoScalingGroupName": "my-asg",  
      "AutoScalingGroupARN": "arn:aws:autoscaling:us-  
west-2:123456789012:autoScalingGroup:930d940e-891e-4781-  
a11a-7b0acd480f03:autoScalingGroupName/my-asg",  
      "LaunchTemplate": {  
        "LaunchTemplateName": "my-launch-template",  
        "Version": "1",  
        "LaunchTemplateId": "lt-1234567890abcde12"  
      },  
      "MinSize": 0,  
      "MaxSize": 1,  
      "DesiredCapacity": 1,  
      "DefaultCooldown": 300,  
      "AvailabilityZones": [  
        "us-west-2a",  
        "us-west-2b",  
        "us-west-2c"  
      ],  
      "LoadBalancerNames": [],  
      "TargetGroupARNs": [],
```

```

    "HealthCheckType": "EC2",
    "HealthCheckGracePeriod": 0,
    "Instances": [
      {
        "InstanceId": "i-06905f55584de02da",
        "InstanceType": "t2.micro",
        "AvailabilityZone": "us-west-2a",
        "HealthStatus": "Healthy",
        "LifecycleState": "InService",
        "ProtectedFromScaleIn": false,
        "LaunchTemplate": {
          "LaunchTemplateName": "my-launch-template",
          "Version": "1",
          "LaunchTemplateId": "lt-1234567890abcde12"
        }
      }
    ],
    "CreatedTime": "2023-10-28T02:39:22.152Z",
    "SuspendedProcesses": [],
    "VPCZoneIdentifier": "subnet-5ea0c127,subnet-6194ea3b,subnet-
c934b782",
    "EnabledMetrics": [],
    "Tags": [],
    "TerminationPolicies": [
      "Default"
    ],
    "NewInstancesProtectedFromScaleIn": false,
    "ServiceLinkedRoleARN": "arn",
    "TrafficSources": []
  }
]
}

```

Esempio 2: Per descrivere i primi 100 gruppi di Auto Scaling specificati

Questo esempio descrive i gruppi di Auto Scaling specificati. Consente di specificare fino a 100 nomi di gruppi.

```

aws autoscaling describe-auto-scaling-groups \
  --max-items 100 \
  --auto-scaling-group-name "group1" "group2" "group3" "group4"

```

Vedi l'esempio 1 per un esempio di output.

Esempio 3: Per descrivere un gruppo Auto Scaling nella regione specificata

Questo esempio descrive i gruppi Auto Scaling nella regione specificata, fino a un massimo di 75 gruppi.

```
aws autoscaling describe-auto-scaling-groups \  
  --max-items 75 \  
  --region us-east-1
```

Vedi l'esempio 1 per un esempio di output.

Esempio 4: Per descrivere il numero specificato del gruppo Auto Scaling

Per restituire un numero specifico di gruppi Auto Scaling, utilizzate l'`--max-items` opzione.

```
aws autoscaling describe-auto-scaling-groups \  
  --max-items 1
```

Vedi l'esempio 1 per un esempio di output.

Se l'output include un `NextToken` campo, ci sono più gruppi. Per ottenere i gruppi aggiuntivi, utilizzate il valore di questo campo con l'`--starting-token` opzione in una chiamata successiva come segue.

```
aws autoscaling describe-auto-scaling-groups \  
  --starting-token Z3M3LMPEXAMPLE
```

Vedi l'esempio 1 per un esempio di output.

Esempio 5: Per descrivere i gruppi di Auto Scaling che utilizzano configurazioni di avvio

Questo esempio utilizza l'`--query` opzione per descrivere i gruppi di Auto Scaling che utilizzano configurazioni di avvio.

```
aws autoscaling describe-auto-scaling-groups \  
  --query 'AutoScalingGroups[?LaunchConfigurationName!=`null`]'
```

Output:

```
[  
  {  
    "AutoScalingGroupName": "my-asg",
```

```
    "AutoScalingGroupARN": "arn:aws:autoscaling:us-
west-2:123456789012:autoScalingGroup:930d940e-891e-4781-
a11a-7b0acd480f03:autoScalingGroupName/my-asg",
    "LaunchConfigurationName": "my-lc",
    "MinSize": 0,
    "MaxSize": 1,
    "DesiredCapacity": 1,
    "DefaultCooldown": 300,
    "AvailabilityZones": [
        "us-west-2a",
        "us-west-2b",
        "us-west-2c"
    ],
    "LoadBalancerNames": [],
    "TargetGroupARNs": [],
    "HealthCheckType": "EC2",
    "HealthCheckGracePeriod": 0,
    "Instances": [
        {
            "InstanceId": "i-088c57934a6449037",
            "InstanceType": "t2.micro",
            "AvailabilityZone": "us-west-2c",
            "HealthStatus": "Healthy",
            "LifecycleState": "InService",
            "LaunchConfigurationName": "my-lc",
            "ProtectedFromScaleIn": false
        }
    ],
    "CreatedTime": "2023-10-28T02:39:22.152Z",
    "SuspendedProcesses": [],
    "VPCZoneIdentifier": "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782",
    "EnabledMetrics": [],
    "Tags": [],
    "TerminationPolicies": [
        "Default"
    ],
    "NewInstancesProtectedFromScaleIn": false,
    "ServiceLinkedRoleARN": "arn",
    "TrafficSources": []
}
]
```

Per ulteriori informazioni, consulta [Filter AWS CLI output](#) nella AWS Command Line Interface User Guide.

- Per i dettagli sull'API, consulta [DescribeAutoScalingGroups AWS CLI Command Reference](#).

Java

SDK per Java 2.x

Note

C'è altro su GitHub. Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
import software.amazon.awssdk.regions.Region;
import software.amazon.awssdk.services.autoscaling.AutoScalingClient;
import software.amazon.awssdk.services.autoscaling.model.AutoScalingException;
import software.amazon.awssdk.services.autoscaling.model.AutoScalingGroup;
import
    software.amazon.awssdk.services.autoscaling.model.DescribeAutoScalingGroupsResponse;
import
    software.amazon.awssdk.services.autoscaling.model.DescribeAutoScalingGroupsRequest;
import software.amazon.awssdk.services.autoscaling.model.Instance;
import java.util.List;

/**
 * Before running this SDK for Java (v2) code example, set up your development
 * environment, including your credentials.
 *
 * For more information, see the following documentation:
 *
 * https://docs.aws.amazon.com/sdk-for-java/latest/developer-guide/get-
 * started.html
 */
public class DescribeAutoScalingInstances {
    public static void main(String[] args) {
        final String usage = ""

                Usage:
                <groupName>
```



```
        Where:
            groupName - The name of the Auto Scaling group.
            """;

    if (args.length != 1) {
        System.out.println(usage);
        System.exit(1);
    }

    String groupName = args[0];
    AutoScalingClient autoScalingClient = AutoScalingClient.builder()
        .region(Region.US_EAST_1)
        .build();

    String instanceId = getAutoScaling(autoScalingClient, groupName);
    System.out.println(instanceId);
    autoScalingClient.close();
}

public static String getAutoScaling(AutoScalingClient autoScalingClient,
String groupName) {
    try {
        String instanceId = "";
        DescribeAutoScalingGroupsRequest scalingGroupsRequest =
DescribeAutoScalingGroupsRequest.builder()
            .autoScalingGroupNames(groupName)
            .build();

        DescribeAutoScalingGroupsResponse response = autoScalingClient
            .describeAutoScalingGroups(scalingGroupsRequest);
        List<AutoScalingGroup> groups = response.autoScalingGroups();
        for (AutoScalingGroup group : groups) {
            System.out.println("The group name is " +
group.autoScalingGroupName());
            System.out.println("The group ARN is " +
group.autoScalingGroupARN());

            List<Instance> instances = group.instances();
            for (Instance instance : instances) {
                instanceId = instance.instanceId();
            }
        }
        return instanceId;
    } catch (AutoScalingException e) {
```

```
        System.err.println(e.awsErrorDetails().errorMessage());
        System.exit(1);
    }
    return "";
}
}
```

- Per i dettagli sull'API, [DescribeAutoScalingGroups](#) consulta AWS SDK for Java 2.x API Reference.

Kotlin

SDK per Kotlin

Note

C'è altro su GitHub. Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
suspend fun getAutoScalingGroups(groupName: String) {
    val scalingGroupsRequest = DescribeAutoScalingGroupsRequest {
        autoScalingGroupNames = listOf(groupName)
    }

    AutoScalingClient { region = "us-east-1" }.use { autoScalingClient ->
        val response =
        autoScalingClient.describeAutoScalingGroups(scalingGroupsRequest)
        response.autoScalingGroups?.forEach { group ->
            println("The group name is ${group.autoScalingGroupName}")
            println("The group ARN is ${group.autoScalingGroupArn}")
            group.instances?.forEach { instance ->
                println("The instance id is ${instance.instanceId}")
                println("The lifecycle state is " + instance.lifecycleState)
            }
        }
    }
}
```

- Per i dettagli sull'API, [DescribeAutoScalingGroups](#) consulta AWS SDK for Kotlin API reference.

PHP

SDK per PHP

Note

C'è altro su. GitHub Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
public function describeAutoScalingGroups($autoScalingGroupNames)
{
    return $this->autoScalingClient->describeAutoScalingGroups([
        'AutoScalingGroupNames' => $autoScalingGroupNames
    ]);
}
```

- Per i dettagli sull'API, [DescribeAutoScalingGroups](#) consulta AWS SDK for PHP API Reference.

PowerShell

Strumenti per PowerShell

Esempio 1: questo esempio elenca i nomi dei gruppi di Auto Scaling.

```
Get-ASAutoScalingGroup | format-table -property AutoScalingGroupName
```

Output:

```
AutoScalingGroupName
-----
my-asg-1
my-asg-2
my-asg-3
```

```
my-asg-4
my-asg-5
my-asg-6
```

Esempio 2: Questo esempio descrive il gruppo Auto Scaling specificato.

```
Get-ASAutoScalingGroup -AutoScalingGroupName my-asg-1
```

Output:

```
AutoScalingGroupARN      : arn:aws:autoscaling:us-
west-2:123456789012:autoScalingGroup:930d940e-891e-4781-a11a-7b0acd480
                          f03:autoScalingGroupName/my-asg-1
AutoScalingGroupName     : my-asg-1
AvailabilityZones        : {us-west-2b, us-west-2a}
CreatedTime              : 3/1/2015 9:05:31 AM
DefaultCooldown         : 300
DesiredCapacity          : 2
EnabledMetrics           : {}
HealthCheckGracePeriod  : 300
HealthCheckType         : EC2
Instances                : {my-1c}
LaunchConfigurationName  : my-1c
LoadBalancerNames       : {}
MaxSize                  : 0
MinSize                  : 0
PlacementGroup          :
Status                   :
SuspendedProcesses      : {}
Tags                    : {}
TerminationPolicies     : {Default}
VPCZoneIdentifier       : subnet-e4f33493,subnet-5264e837
```

Esempio 3: Questo esempio descrive i due gruppi di Auto Scaling specificati.

```
Get-ASAutoScalingGroup -AutoScalingGroupName @"my-asg-1", "my-asg-2"
```

Esempio 4: Questo esempio descrive le istanze Auto Scaling per il gruppo Auto Scaling specificato.

```
(Get-ASAutoScalingGroup -AutoScalingGroupName my-asg-1).Instances
```

Esempio 5: questo esempio descrive tutti i gruppi di Auto Scaling.

```
Get-ASAutoScalingGroup
```

Esempio 6: Questo esempio descrive tutti i gruppi di Auto Scaling, in batch da 10.

```
$nextToken = $null
do {
  Get-ASAutoScalingGroup -NextToken $nextToken -MaxRecord 10
  $nextToken = $AWSHistory.LastServiceResponse.NextToken
} while ($nextToken -ne $null)
```

Esempio 7: Questo LaunchTemplate esempio descrive il gruppo Auto Scaling specificato. Questo esempio presuppone che l'opzione «Instance Purchase options» sia impostata su «Aderisci al modello di lancio». Nel caso in cui questa opzione sia impostata su «Combina opzioni di acquisto e tipi di istanze», è LaunchTemplate possibile accedervi utilizzando ". MixedInstancesPolicy LaunchTemplate«proprietà.

```
(Get-ASAutoScalingGroup -AutoScalingGroupName my-ag-1).LaunchTemplate
```

Output:

LaunchTemplateId	LaunchTemplateName	Version
lt-06095fd619cb40371	test-launch-template	\$Default

- Per i dettagli sull'API, vedere [DescribeAutoScalingGroups](#) in AWS Tools for PowerShell Cmdlet Reference.

Python

SDK per Python (Boto3)

Note

C'è altro su. [GitHub](#) Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
class AutoScalingWrapper:
    """Encapsulates Amazon EC2 Auto Scaling actions."""

    def __init__(self, autoscaling_client):
        """
        :param autoscaling_client: A Boto3 Amazon EC2 Auto Scaling client.
        """
        self.autoscaling_client = autoscaling_client

    def describe_group(self, group_name):
        """
        Gets information about an Auto Scaling group.

        :param group_name: The name of the group to look up.
        :return: Information about the group, if found.
        """
        try:
            response = self.autoscaling_client.describe_auto_scaling_groups(
                AutoScalingGroupNames=[group_name]
            )
        except ClientError as err:
            logger.error(
                "Couldn't describe group %s. Here's why: %s: %s",
                group_name,
                err.response["Error"]["Code"],
                err.response["Error"]["Message"],
            )
            raise
        else:
            groups = response.get("AutoScalingGroups", [])
            return groups[0] if len(groups) > 0 else None
```

- Per i dettagli sull'API, consulta [DescribeAutoScalingGroups AWSSDK for Python \(Boto3\) API Reference](#).

Rust

SDK per Rust

Note

C'è di più su. [GitHub](#) Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
async fn list_groups(client: &Client) -> Result<(), Error> {
    let resp = client.describe_auto_scaling_groups().send().await?;

    println!("Groups:");

    let groups = resp.auto_scaling_groups();

    for group in groups {
        println!(
            "Name: {}",
            group.auto_scaling_group_name().unwrap_or("Unknown")
        );
        println!(
            "Arn: {}",
            group.auto_scaling_group_arn().unwrap_or("unknown"),
        );
        println!("Zones: {:?}", group.availability_zones(),);
        println!();
    }

    println!("Found {} group(s)", groups.len());

    Ok(())
}
```

- Per i dettagli sulle API, consulta il riferimento [DescribeAutoScalingGroups](#) all'API AWS SDK for Rust.

Eliminare un gruppo Auto Scaling utilizzando un SDK AWS

I seguenti esempi di codice mostrano come utilizzare `DeleteAutoScalingGroup`

.NET

AWS SDK for .NET

Note

C'è altro da fare GitHub. Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

Aggiorna la dimensione minima di un gruppo con dimensionamento automatico a zero, termina tutte le istanze del gruppo ed elimina il gruppo.

```
/// <summary>
/// Try to terminate an instance by its Id.
/// </summary>
/// <param name="instanceId">The Id of the instance to terminate.</param>
/// <returns>Async task.</returns>
public async Task TryTerminateInstanceById(string instanceId)
{
    var stopping = false;
    Console.WriteLine($"Stopping {instanceId}...");
    while (!stopping)
    {
        try
        {
            await
                _amazonAutoScaling.TerminateInstanceInAutoScalingGroupAsync(
                    new TerminateInstanceInAutoScalingGroupRequest()
                    {
                        InstanceId = instanceId,
                        ShouldDecrementDesiredCapacity = false
                    });
            stopping = true;
        }
        catch (ScalingActivityInProgressException)
        {
        }
    }
}
```



```
        Console.WriteLine($"Scaling activity in progress for
{instanceId}. Waiting...");
        Thread.Sleep(10000);
    }
}

/// <summary>
/// Tries to delete the EC2 Auto Scaling group. If the group is in use or in
progress,
/// waits and retries until the group is successfully deleted.
/// </summary>
/// <param name="groupName">The name of the group to try to delete.</param>
/// <returns>Async task.</returns>
public async Task TryDeleteGroupByName(string groupName)
{
    var stopped = false;
    while (!stopped)
    {
        try
        {
            await _amazonAutoScaling.DeleteAutoScalingGroupAsync(
                new DeleteAutoScalingGroupRequest()
                {
                    AutoScalingGroupName = groupName
                });
            stopped = true;
        }
        catch (Exception e)
            when ((e is ScalingActivityInProgressException)
                || (e is Amazon.AutoScaling.Model.ResourceInUseException))
        {
            Console.WriteLine($"Some instances are still running.
Waiting...");
            Thread.Sleep(10000);
        }
    }
}

/// <summary>
/// Terminate instances and delete the Auto Scaling group by name.
/// </summary>
/// <param name="groupName">The name of the group to delete.</param>
/// <returns>Async task.</returns>
```

```

public async Task TerminateAndDeleteAutoScalingGroupWithName(string
groupName)
{
    var describeGroupsResponse = await
    _amazonAutoScaling.DescribeAutoScalingGroupsAsync(
        new DescribeAutoScalingGroupsRequest()
        {
            AutoScalingGroupNames = new List<string>() { groupName }
        });
    if (describeGroupsResponse.AutoScalingGroups.Any())
    {
        // Update the size to 0.
        await _amazonAutoScaling.UpdateAutoScalingGroupAsync(
            new UpdateAutoScalingGroupRequest()
            {
                AutoScalingGroupName = groupName,
                MinSize = 0
            });
        var group = describeGroupsResponse.AutoScalingGroups[0];
        foreach (var instance in group.Instances)
        {
            await TryTerminateInstanceById(instance.InstanceId);
        }

        await TryDeleteGroupByName(groupName);
    }
    else
    {
        Console.WriteLine($"No groups found with name {groupName}.");
    }
}

```

```

/// <summary>
/// Delete an Auto Scaling group.
/// </summary>
/// <param name="groupName">The name of the Amazon EC2 Auto Scaling group.</
param>
/// <returns>A Boolean value indicating the success of the action.</returns>
public async Task<bool> DeleteAutoScalingGroupAsync(
    string groupName)
{

```

```
var deleteAutoScalingGroupRequest = new DeleteAutoScalingGroupRequest
{
    AutoScalingGroupName = groupName,
    ForceDelete = true,
};

var response = await
_amazonAutoScaling.DeleteAutoScalingGroupAsync(deleteAutoScalingGroupRequest);
if (response.HttpStatusCode == System.Net.HttpStatusCode.OK)
{
    Console.WriteLine($"You successfully deleted {groupName}");
    return true;
}

Console.WriteLine($"Couldn't delete {groupName}.");
return false;
}
```

- Per i dettagli sull'API, [DeleteAutoScalingGroup](#) consulta AWS SDK for .NET API Reference.

C++

SDK per C++

Note

C'è altro su GitHub. Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
Aws::Client::ClientConfiguration clientConfig;
// Optional: Set to the AWS Region (overrides config file).
// clientConfig.region = "us-east-1";

Aws::AutoScaling::AutoScalingClient autoScalingClient(clientConfig);

Aws::AutoScaling::Model::DeleteAutoScalingGroupRequest request;
request.SetAutoScalingGroupName(groupName);

Aws::AutoScaling::Model::DeleteAutoScalingGroupOutcome outcome =
```

```
        autoScalingClient.DeleteAutoScalingGroup(request);

        if (outcome.IsSuccess()) {
            std::cout << "Auto Scaling group '" << groupName << "' was
deleted."
                << std::endl;
        }
        else {
            std::cerr << "Error with AutoScaling::DeleteAutoScalingGroup. "
                << outcome.GetError().GetMessage()
                << std::endl;
            result = false;
        }
    }
}
```

- Per i dettagli sull'API, [DeleteAutoScalingGroup](#) consulta AWS SDK for C++ API Reference.

CLI

AWS CLI

Esempio 1: Per eliminare il gruppo Auto Scaling specificato

Questo esempio elimina il gruppo Auto Scaling specificato.

```
aws autoscaling delete-auto-scaling-group \
    --auto-scaling-group-name my-asg
```

Questo comando non produce alcun output.

Per ulteriori informazioni, consulta [Eliminazione dell'infrastruttura Auto Scaling](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

Esempio 2: Per forzare l'eliminazione del gruppo Auto Scaling specificato

Per eliminare il gruppo Auto Scaling senza attendere la chiusura delle istanze del gruppo, utilizzate l'opzione. `--force-delete`

```
aws autoscaling delete-auto-scaling-group \
    --auto-scaling-group-name my-asg \
    --force-delete
```

Questo comando non produce alcun output.

Per ulteriori informazioni, consulta [Eliminazione dell'infrastruttura Auto Scaling](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

- Per i dettagli sull'API, consulta [DeleteAutoScalingGroup](#) Command Reference.AWS CLI

Java

SDK per Java 2.x

Note

C'è altro su GitHub. Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
import software.amazon.awssdk.regions.Region;
import software.amazon.awssdk.services.autoscaling.AutoScalingClient;
import software.amazon.awssdk.services.autoscaling.model.AutoScalingException;
import
    software.amazon.awssdk.services.autoscaling.model.DeleteAutoScalingGroupRequest;

/**
 * Before running this SDK for Java (v2) code example, set up your development
 * environment, including your credentials.
 *
 * For more information, see the following documentation:
 *
 * https://docs.aws.amazon.com/sdk-for-java/latest/developer-guide/get-
 * started.html
 */
public class DeleteAutoScalingGroup {
    public static void main(String[] args) {
        final String usage = ""

            Usage:
                <groupName>

            Where:
                groupName - The name of the Auto Scaling group.
        """;
```

```
    if (args.length != 1) {
        System.out.println(usage);
        System.exit(1);
    }

    String groupName = args[0];
    AutoScalingClient autoScalingClient = AutoScalingClient.builder()
        .region(Region.US_EAST_1)
        .build();

    deleteAutoScalingGroup(autoScalingClient, groupName);
    autoScalingClient.close();
}

public static void deleteAutoScalingGroup(AutoScalingClient
autoScalingClient, String groupName) {
    try {
        DeleteAutoScalingGroupRequest deleteAutoScalingGroupRequest =
DeleteAutoScalingGroupRequest.builder()
            .autoScalingGroupName(groupName)
            .forceDelete(true)
            .build();

        autoScalingClient.deleteAutoScalingGroup(deleteAutoScalingGroupRequest);
        System.out.println("You successfully deleted " + groupName);

    } catch (AutoScalingException e) {
        System.err.println(e.awsErrorDetails().errorMessage());
        System.exit(1);
    }
}
}
```

- Per i dettagli sull'API, [DeleteAutoScalingGroup](#) consulta AWS SDK for Java 2.x API Reference.

Kotlin

SDK per Kotlin

Note

C'è altro su GitHub. Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
suspend fun deleteSpecificAutoScalingGroup(groupName: String) {
    val deleteAutoScalingGroupRequest = DeleteAutoScalingGroupRequest {
        autoScalingGroupName = groupName
        forceDelete = true
    }

    AutoScalingClient { region = "us-east-1" }.use { autoScalingClient ->
        autoScalingClient.deleteAutoScalingGroup(deleteAutoScalingGroupRequest)
        println("You successfully deleted $groupName")
    }
}
```

- Per i dettagli sull'API, [DeleteAutoScalingGroup](#) consulta AWS SDK for Kotlin API reference.

PHP

SDK per PHP

Note

C'è altro su GitHub. Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
public function deleteAutoScalingGroup($autoScalingGroupName)
{
    return $this->autoScalingClient->deleteAutoScalingGroup([
        'AutoScalingGroupName' => $autoScalingGroupName,
        'ForceDelete' => true,
    ]);
}
```

```
    ]);  
}
```

- Per i dettagli sull'API, [DeleteAutoScalingGroup](#) consulta AWS SDK for PHP API Reference.

PowerShell

Strumenti per PowerShell

Esempio 1: Questo esempio elimina il gruppo Auto Scaling specificato se non ha istanze in esecuzione. Prima di procedere con l'operazione, viene richiesta una conferma.

```
Remove-ASAutoScalingGroup -AutoScalingGroupName my-asg
```

Output:

```
Confirm  
Are you sure you want to perform this action?  
Performing operation "Remove-ASAutoScalingGroup (DeleteAutoScalingGroup)" on  
Target "my-asg".  
[Y] Yes [A] Yes to All [N] No [L] No to All [S] Suspend [?] Help (default is  
"Y"):
```

Esempio 2: se si specifica il parametro Force, non viene richiesta la conferma prima che l'operazione proceda.

```
Remove-ASAutoScalingGroup -AutoScalingGroupName my-asg -Force
```

Esempio 3: Questo esempio elimina il gruppo Auto Scaling specificato e termina tutte le istanze in esecuzione in esso contenute.

```
Remove-ASAutoScalingGroup -AutoScalingGroupName my-asg -ForceDelete $true -Force
```

- Per i dettagli sull'API, vedere [DeleteAutoScalingGroup](#) in Cmdlet Reference.AWS Tools for PowerShell

Python

SDK per Python (Boto3)

Note

C'è altro su GitHub Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

Aggiorna la dimensione minima di un gruppo con dimensionamento automatico a zero, termina tutte le istanze del gruppo ed elimina il gruppo.

```
class AutoScaler:
    """
    Encapsulates Amazon EC2 Auto Scaling and EC2 management actions.
    """

    def __init__(
        self,
        resource_prefix,
        inst_type,
        ami_param,
        autoscaling_client,
        ec2_client,
        ssm_client,
        iam_client,
    ):
        """
        :param resource_prefix: The prefix for naming AWS resources that are
        created by this class.
        :param inst_type: The type of EC2 instance to create, such as t3.micro.
        :param ami_param: The Systems Manager parameter used to look up the AMI
        that is
                created.
        :param autoscaling_client: A Boto3 EC2 Auto Scaling client.
        :param ec2_client: A Boto3 EC2 client.
        :param ssm_client: A Boto3 Systems Manager client.
        :param iam_client: A Boto3 IAM client.
        """
        self.inst_type = inst_type
        self.ami_param = ami_param
```

```

self.autoscaling_client = autoscaling_client
self.ec2_client = ec2_client
self.ssm_client = ssm_client
self.iam_client = iam_client
self.launch_template_name = f"{resource_prefix}-template"
self.group_name = f"{resource_prefix}-group"
self.instance_policy_name = f"{resource_prefix}-pol"
self.instance_role_name = f"{resource_prefix}-role"
self.instance_profile_name = f"{resource_prefix}-prof"
self.bad_creds_policy_name = f"{resource_prefix}-bc-pol"
self.bad_creds_role_name = f"{resource_prefix}-bc-role"
self.bad_creds_profile_name = f"{resource_prefix}-bc-prof"
self.key_pair_name = f"{resource_prefix}-key-pair"

def _try_terminate_instance(self, inst_id):
    stopping = False
    log.info(f"Stopping {inst_id}.")
    while not stopping:
        try:
            self.autoscaling_client.terminate_instance_in_auto_scaling_group(
                InstanceId=inst_id, ShouldDecrementDesiredCapacity=True
            )
            stopping = True
        except ClientError as err:
            if err.response["Error"]["Code"] == "ScalingActivityInProgress":
                log.info("Scaling activity in progress for %s. Waiting...",
inst_id)
                time.sleep(10)
            else:
                raise AutoScalerError(f"Couldn't stop instance {inst_id}:
{err}.")

def _try_delete_group(self):
    """
    Tries to delete the EC2 Auto Scaling group. If the group is in use or in
progress,
the function waits and retries until the group is successfully deleted.
    """
    stopped = False
    while not stopped:
        try:
            self.autoscaling_client.delete_auto_scaling_group(
                AutoScalingGroupName=self.group_name

```

```

        )
        stopped = True
        log.info("Deleted EC2 Auto Scaling group %s.", self.group_name)
    except ClientError as err:
        if (
            err.response["Error"]["Code"] == "ResourceInUse"
            or err.response["Error"]["Code"] ==
"ScalingActivityInProgress"
        ):
            log.info(
                "Some instances are still running. Waiting for them to
stop..."
            )
            time.sleep(10)
        else:
            raise AutoScalerError(
                f"Couldn't delete group {self.group_name}: {err}."
            )

    def delete_group(self):
        """
        Terminates all instances in the group, deletes the EC2 Auto Scaling
group.
        """
        try:
            response = self.autoscaling_client.describe_auto_scaling_groups(
                AutoScalingGroupNames=[self.group_name]
            )
            groups = response.get("AutoScalingGroups", [])
            if len(groups) > 0:
                self.autoscaling_client.update_auto_scaling_group(
                    AutoScalingGroupName=self.group_name, MinSize=0
                )
                instance_ids = [inst["InstanceId"] for inst in groups[0]
["Instances"]]
                for inst_id in instance_ids:
                    self._try_terminate_instance(inst_id)
                    self._try_delete_group()
            else:
                log.info("No groups found named %s, nothing to do.",
self.group_name)
        except ClientError as err:
            raise AutoScalerError(f"Couldn't delete group {self.group_name}:
{err}.")

```

- Per i dettagli sull'API, consulta [DeleteAutoScalingGroup AWS SDK for Python \(Boto3\) API Reference](#).

Rust

SDK per Rust

Note

C'è di più su. [GitHub](#) Trova l'esempio completo e scopri di più sulla configurazione e l'esecuzione nel [Repository di esempi di codice AWS](#).

```
async fn delete_group(client: &Client, name: &str, force: bool) -> Result<(),
Error> {
    client
        .delete_auto_scaling_group()
        .auto_scaling_group_name(name)
        .set_force_delete(if force { Some(true) } else { None })
        .send()
        .await?;

    println!("Deleted Auto Scaling group");

    Ok(())
}
```

- Per i dettagli sulle API, consulta il riferimento [DeleteAutoScalingGroup](#) all'API AWS SDK for Rust.

Ricicla le istanze del gruppo con dimensionamento automatico

Amazon EC2 Auto Scaling offre funzionalità che consentono di sostituire le istanze Amazon EC2 nel gruppo Auto Scaling dopo aver effettuato aggiornamenti, come l'aggiunta di un nuovo modello di lancio con una nuova Amazon Machine Image (AMI) o l'aggiunta di nuovi tipi di istanze. Inoltre, ti aiuta a semplificare gli aggiornamenti dandoti la possibilità di includerli nella stessa operazione che sostituisce le istanze.

In questa sezione sono incluse informazioni che consentono di completare le seguenti operazioni:

- Avviare un aggiornamento istanza per sostituire le istanze nel gruppo con dimensionamento automatico.
- Dichiarare aggiornamenti specifici che descrivono una configurazione desiderata e aggiornare il gruppo con scalabilità automatica alla configurazione desiderata.
- Saltare la sostituzione delle istanze già aggiornate.
- Utilizza i checkpoint per aggiornare le istanze in fasi ed eseguire verifiche sulle istanze in punti specifici.
- Ricevere notifiche via e-mail quando viene raggiunto un punto di controllo.
- Usare un rollback per ripristinare il gruppo con scalabilità automatica alla configurazione utilizzata in precedenza.
- Esegui automaticamente il rollback se l'aggiornamento dell'istanza fallisce per qualche motivo o se gli CloudWatch allarmi Amazon specificati entrano nello ALARM stato.
- Limitare la durata delle istanze per fornire versioni software coerenti e configurazioni delle istanze nel gruppo con scalabilità automatica.

Indice

- [Usa un aggiornamento dell'istanza per aggiornare le istanze in un gruppo di Auto Scaling](#)
- [Sostituire le istanze Auto Scaling basate sulla durata massima dell'istanza](#)

Usa un aggiornamento dell'istanza per aggiornare le istanze in un gruppo di Auto Scaling

È possibile utilizzare un aggiornamento dell'istanza per aggiornare le istanze nel gruppo Auto Scaling. Questa funzionalità può essere utile quando una modifica alla configurazione richiede la sostituzione delle istanze, specialmente se il gruppo Auto Scaling contiene un numero elevato di istanze.

Alcune situazioni in cui un aggiornamento dell'istanza può essere utile includono:

- Distribuzione di una nuova Amazon Machine Image (AMI) o di uno script di dati utente nel tuo gruppo di Auto Scaling. Puoi creare un nuovo modello di lancio con le modifiche e quindi utilizzare un aggiornamento dell'istanza per implementare immediatamente gli aggiornamenti.
- Migrazione delle istanze verso nuovi tipi di istanze per sfruttare i miglioramenti e le ottimizzazioni più recenti.
- Passaggio dai gruppi di Auto Scaling dall'utilizzo di una configurazione di avvio all'utilizzo di un modello di avvio. Puoi copiare le configurazioni di avvio nei modelli di avvio e quindi utilizzare un aggiornamento dell'istanza per aggiornare le istanze ai nuovi modelli. Per ulteriori informazioni sulla migrazione dei modelli di avvio, consulta [Migra i tuoi gruppi di Auto Scaling ai modelli di avvio](#).

Indice

- [Come funziona l'aggiornamento dell'istanza](#)
- [Informazioni sui valori predefiniti per l'aggiornamento di un'istanza](#)
- [Avvio dell'aggiornamento di un'istanza](#)
- [Monitora l'aggiornamento di un'istanza](#)
- [Annullamento dell'aggiornamento di un'istanza](#)
- [Annullamento delle modifiche tramite un rollback](#)
- [Utilizzo di un aggiornamento dell'istanza con l'opzione Ignora la corrispondenza](#)
- [Aggiungere punti di controllo a un aggiornamento delle istanze](#)

Come funziona l'aggiornamento dell'istanza

Questo argomento descrive come funziona l'aggiornamento di un'istanza e introduce i concetti chiave da comprendere per utilizzarlo in modo efficace.

Indice

- [Come funziona](#)
- [Concetti principali](#)
- [Periodo di tolleranza dei controlli dell'integrità](#)
- [Compatibilità del tipo di istanza](#)
- [Limitazioni](#)

Come funziona

Per aggiornare le istanze in un gruppo Auto Scaling, puoi definire una nuova configurazione che contenga la versione più recente dell'applicazione e tutti gli altri aggiornamenti che desideri apportare. Quindi, avvia un aggiornamento dell'istanza per sostituire le istanze esistenti con nuove in base a tale configurazione.

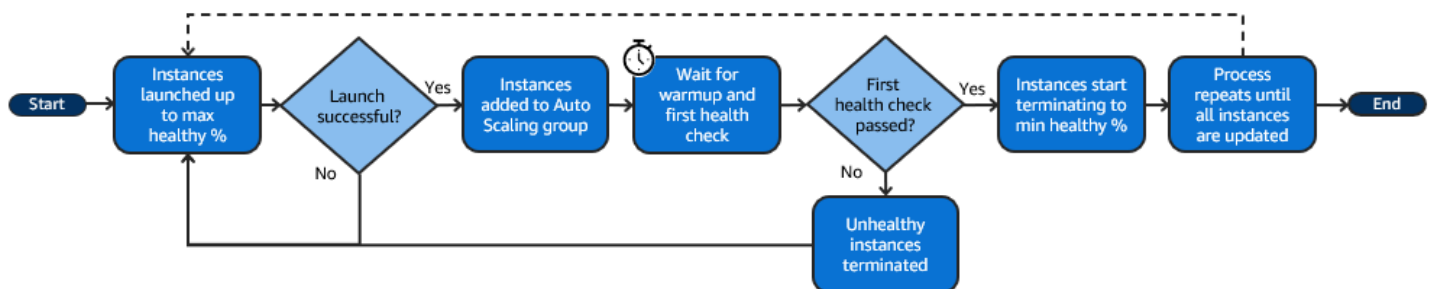
Per eseguire un aggiornamento dell'istanza:

1. Crea un nuovo modello di lancio o aggiorna il modello esistente con le modifiche di configurazione desiderate, ad esempio una nuova Amazon Machine Image (AMI). Per ulteriori informazioni, consulta [Creare un modello di avvio per un gruppo con dimensionamento automatico](#).
2. Avvia l'aggiornamento dell'istanza utilizzando la console AWS CLI Amazon EC2 Auto Scaling o l'SDK:
 - Specificate il nuovo modello di lancio o la nuova versione del modello di lancio che avete creato. Verrà utilizzato per avviare nuove istanze.
 - Imposta la percentuale di salute minima e massima preferita. Questo controlla quante istanze vengono sostituite contemporaneamente e se vengono lanciate nuove istanze prima di terminare quelle vecchie.
 - Configura tutte le impostazioni opzionali, ad esempio:
 - Checkpoint: sospende l'aggiornamento dell'istanza dopo una certa percentuale di sostituzioni per verificare l'avanzamento.
 - Ignora la corrispondenza: confronta le vecchie istanze con la nuova configurazione e sostituisci solo quelle che non corrispondono. Quando avvii l'aggiornamento di un'istanza dalla console, l'opzione Skip Matching è attiva per impostazione predefinita.
 - Tipi di istanze multiple: applica una [politica per le istanze miste](#) nuova o aggiornata come parte della configurazione desiderata.

Una volta avviato l'aggiornamento dell'istanza, Amazon EC2 Auto Scaling:

- Sostituisci le istanze in batch in base alle percentuali di integrità minima e massima.
- Avvia le nuove istanze prima di terminare quelle vecchie se la percentuale minima di integrità è impostata al 100%. Ciò garantisce che la capacità desiderata venga mantenuta in ogni momento.
- Controlla lo stato di salute delle istanze e concedi loro il tempo di riscaldarsi prima che vengano sostituite altre istanze.
- Termina e sostituisci le istanze che risultano non integre.
- Aggiorna automaticamente le impostazioni del gruppo Auto Scaling con le nuove modifiche alla configurazione dopo il completamento dell'aggiornamento dell'istanza.
- Se il tuo gruppo ha un pool caldo, Amazon EC2 Auto Scaling sostituisce innanzitutto le istanze. InService Quindi sostituisce le istanze nel warm pool.

Il seguente diagramma di flusso illustra il comportamento di avvio prima del termine quando si imposta la percentuale minima di integrità al 100%.



Note

Le percentuali di integrità minima e massima per l'aggiornamento di un'istanza devono essere specificate solo se non è stata impostata una politica di manutenzione dell'istanza o se è necessario sostituire la politica esistente. Per ulteriori informazioni, consulta [Policy di manutenzione delle istanze](#).

Analogamente, è necessario specificare il periodo di riscaldamento dell'istanza per un aggiornamento dell'istanza solo se non è stato abilitato il riscaldamento predefinito o se è necessario sovrascrivere l'impostazione predefinita. Per ulteriori informazioni, consulta [Impostazione della preparazione di default dell'istanza per un gruppo con scalabilità automatica](#).

Concetti principali

Prima di iniziare, acquisisci familiarità con i concetti principali di base dell'aggiornamento delle istanze, riportati di seguito:

Percentuale minima di integrità

La percentuale minima di integrità è la percentuale della capacità desiderata da mantenere in servizio, integra e pronta all'uso durante l'aggiornamento dell'istanza in modo che l'aggiornamento possa continuare. Ad esempio, se la percentuale minima di integrità è del 90% e la percentuale di integrità massima è del 100%, la percentuale di capacità che verrà sostituita sarà del 10%. Se le nuove istanze non superano i controlli dell'integrità, Dimensionamento automatico Amazon EC2 le termina e le sostituisce. Se l'aggiornamento dell'istanza non riesce ad avviare istanze integre, esso avrà esito negativo, lasciando intatto il restante 90% del gruppo. Se le nuove istanze rimangono integre e terminano il periodo di riscaldamento, Amazon EC2 Auto Scaling può continuare a sostituire altre istanze.

L'aggiornamento di un'istanza può sostituire le istanze una alla volta, più alla volta o tutte contemporaneamente. Per sostituire un'istanza alla volta, imposta la percentuale minima e massima di integrità al 100%. Ciò modifica il comportamento dell'aggiornamento dell'istanza da avviare prima del termine, il che impedisce alla capacità del gruppo di scendere al di sotto del 100% della capacità desiderata. Per sostituire tutte le istanze contemporaneamente, imposta tale percentuale allo 0%.

Percentuale massima di integrità

La percentuale massima integra è la percentuale della capacità desiderata che il gruppo con dimensionamento automatico può raggiungere quando si sostituiscono le istanze. La differenza tra il valore minimo e massimo non può essere maggiore di 100. Un intervallo più ampio aumenta il numero di istanze che possono essere sostituite contemporaneamente.

Preparazione dell'istanza

La preparazione dell'istanza è il periodo di tempo da quando lo stato di una nuova istanza cambia in `InService` fino a quando la sua inizializzazione è considerata terminata. Dopo aver determinato che un'istanza appena avviata è integra, se le istanze superano i controlli dell'integrità, Dimensionamento automatico Amazon EC2 non passa immediatamente alla sostituzione successiva. Attende il periodo di riscaldamento prima di passare alla sostituzione dell'istanza successiva. Questo può essere utile quando l'applicazione necessita ancora di un po' di tempo di inizializzazione prima di rispondere alle richieste.

Il funzionamento è identico alla preparazione dell'istanza predefinita. Pertanto, valgono le stesse considerazioni sul dimensionamento. Per ulteriori informazioni, consulta [Impostazione della preparazione di default dell'istanza per un gruppo con scalabilità automatica](#).

Configurazione desiderata

La configurazione desiderata è la nuova configurazione che desideri che Dimensionamento automatico Amazon EC2 implementi nel tuo gruppo con scalabilità automatica. Ad esempio, puoi specificare un nuovo modello di avvio e nuovi tipi di istanza per le istanze. Durante un aggiornamento delle istanze, Amazon EC2 Auto Scaling aggiorna il gruppo con scalabilità automatica alla configurazione desiderata. Se durante l'aggiornamento delle istanze si verifica un evento di aumento orizzontale, Amazon EC2 Auto Scaling avvia nuove istanze con la configurazione desiderata, anziché le impostazioni correnti del gruppo. Dopo aver completato l'aggiornamento delle istanze, Dimensionamento automatico Amazon EC2 aggiorna le impostazioni del gruppo con scalabilità automatica in modo da riflettere la nuova configurazione desiderata specificata nell'ambito dell'aggiornamento.

Salta corrispondenza

Ignora la corrispondenza indica a Dimensionamento automatico Amazon EC2 di ignorare le istanze che dispongono già degli aggiornamenti più recenti, in modo da non sostituire più istanze del necessario. Questa opzione è utile quando si desidera assicurarsi che il gruppo con scalabilità automatica utilizzi una particolare versione del modello di avvio e sostituisca solo le istanze che non la utilizzano.

Punti di controllo

Un punto di controllo è un punto nel tempo in cui l'aggiornamento dell'istanza si interrompe per un periodo di tempo specificato. Un aggiornamento delle istanze può avere più punti di controllo. Amazon EC2 Auto Scaling emette eventi per ogni punto di controllo. Pertanto, puoi aggiungere una EventBridge regola per inviare gli eventi a una destinazione, come Amazon SNS, per ricevere una notifica quando viene raggiunto un checkpoint. Dopo aver raggiunto un punto di controllo, è possibile verificare l'implementazione. Se vengono identificati problemi, è possibile annullare l'aggiornamento delle istanze o eseguirne il rollback. La possibilità di implementare gli aggiornamenti in fasi è un vantaggio fondamentale dei punti di controllo. Se non si utilizzano i punti di controllo, le sostituzioni a rotazione vengono eseguite in modo continuo.

Per ulteriori informazioni su tutte le impostazioni predefinite che puoi configurare all'avvio di un aggiornamento dell'istanza, consulta [Informazioni sui valori predefiniti per l'aggiornamento di un'istanza](#).

Periodo di tolleranza dei controlli dell'integrità

Dimensionamento automatico Amazon EC2 determina se un'istanza è integra in base allo stato dei controlli dell'integrità utilizzati dal gruppo con dimensionamento automatico. Per ulteriori informazioni, consulta [Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico](#).

Per assicurarsi che questi controlli dell'integrità vengano avviati il prima possibile, non impostare un periodo di tolleranza del controllo dell'integrità del gruppo troppo alto, ma abbastanza alto da consentire ai controlli dell'integrità di Elastic Load Balancing di determinare se una destinazione è disponibile per gestire le richieste. Per ulteriori informazioni, consulta [Impostazione del periodo di tolleranza del controllo dell'integrità per un gruppo con scalabilità automatica](#).

Compatibilità del tipo di istanza

Prima di modificare il tipo di istanza, è consigliabile verificarne il funzionamento con il modello di avvio. Questa operazione consente di confermare la compatibilità con l'AMI specificata. Ad esempio, se sono state avviate le istanze originali da un'AMI paravirtuale (PV) ma si desidera passare a un tipo di istanza di generazione corrente supportato solo da un'AMI di una macchina virtuale hardware (HVM), è necessario utilizzare un'AMI HVM nel modello di avvio.

Per verificare la compatibilità del tipo di istanza senza avviare le istanze, utilizza il comando [run-instances](#) con l'opzione `--dry-run`, come mostrato nell'esempio seguente.

```
aws ec2 run-instances --launch-template LaunchTemplateName=my-template,Version='1' --dry-run
```

Per informazioni sul modo in cui viene determinata la compatibilità, consulta [Compatibilità per cambiare il tipo di istanza](#) nella Guida per l'utente di Amazon EC2 per istanze Linux.

Limitazioni

- Durata totale: il tempo massimo durante il quale un aggiornamento delle istanze può continuare a sostituire attivamente le istanze è di 14 giorni.
- Differenza di comportamento specifica dei gruppi ponderati: se un gruppo di istanze miste è configurato con un peso dell'istanza maggiore o uguale alla capacità desiderata del gruppo, Dimensionamento automatico Amazon EC2 può sostituire tutte le istanze InService in una volta. Per evitare questa situazione, segui i suggerimenti riportati nell'argomento [Configurare un gruppo Auto Scaling per utilizzare i pesi delle istanze](#). Quando utilizzi pesi con il gruppo con scalabilità automatica, specifica una capacità desiderata maggiore del peso massimo.

- **Timeout di un'ora:** quando un aggiornamento delle istanze non è in grado di continuare a effettuare sostituzioni perché Dimensionamento automatico Amazon EC2 è in attesa di sostituire le istanze in standby o protette dalla riduzione orizzontale o perché le nuove istanze non superano i controlli dell'integrità, continua a riprovare per un'ora. Fornisce inoltre un messaggio di stato per aiutarti a risolvere il problema. Se trascorsa un'ora il problema persiste, l'operazione ha esito negativo. Lo scopo è lasciare tempo per il recupero in caso di problemi temporanei.
- **Distribuzione del codice tramite i dati utente:** Skip matching non verifica le modifiche al codice distribuite da uno script di dati utente. Se utilizzi i dati utente per estrarre nuovo codice e installare questi aggiornamenti su nuove istanze, ti consigliamo di disattivare lo skip matching per assicurarti che tutte le istanze ricevano il codice più recente, anche senza un aggiornamento della versione del modello di lancio.

Informazioni sui valori predefiniti per l'aggiornamento di un'istanza

Prima di avviare l'aggiornamento di un'istanza, puoi personalizzare varie preferenze che influiscono sull'aggiornamento dell'istanza. Alcune preferenze predefinite sono diverse a seconda che utilizzi la console o la riga di comando (o SDK).AWS CLI AWS

La tabella seguente elenca i valori predefiniti per le impostazioni di aggiornamento dell'istanza.

Impostazione	AWS CLI o SDK AWS	Console dimensionamento automatico Amazon EC2
CloudWatch allarme	Disabilitato (null)	Disabilitato
Rollback automatico	Disabilitato (false)	Disabilitato
Punti di controllo	Disabilitato (false)	Disabilitato
Ritardo del checkpoint	1 ora (3600 secondi)	1 ora
Preparazione dell'istanza	La preparazione dell'istanza di default , se definita, o in caso contrario il periodo di tolleranza a dei controlli dell'integrità .	La preparazione dell'istanza di default , se definita, o in caso contrario il periodo di tolleranza a dei controlli dell'integrità .
Percentuale massima di integrità	Varia in base alla policy di manutenzione dell'istanza.	Varia in base alla policy di manutenzione dell'istanza.

Impostazione	AWS CLI o SDK AWS	Console dimensionamento automatico Amazon EC2
	Se non esiste una policy di manutenzione dell'istanza, il valore predefinito è il 100% (nullo).	Se non esiste una policy di manutenzione dell'istanza, il valore predefinito è il 100% (nullo).
Percentuale minima di integrità	Varia in base alla policy di manutenzione dell'istanza. Se non esiste una policy di manutenzione dell'istanza, il valore predefinito è il 90%.	Varia in base alla policy di manutenzione dell'istanza. Se non esiste una policy di manutenzione dell'istanza, il valore predefinito è il 90%.
Istanze protette dalla riduzione orizzontale	Attendi	Ignorare
Salta corrispondenza	Disabilitato (false)	Abilitato
Istanze in standby	Attendi	Ignorare

Di seguito viene fornita una descrizione di ciascuna impostazione:

CloudWatch allarme (**AlarmSpecification**)

Le specifiche CloudWatch dell'allarme. CloudWatch gli allarmi possono essere utilizzati per identificare eventuali problemi e fallire l'operazione se un allarme entra in ALARM stato. Per ulteriori informazioni, consulta [Avvio dell'aggiornamento di un'istanza con rollback automatico](#).

Rollback automatico (**AutoRollback**)

Controlla se il Dimensionamento automatico Amazon EC2 esegue il rollback del gruppo con dimensionamento automatico alla configurazione precedente in caso di esito negativo dell'aggiornamento dell'istanza. Per ulteriori informazioni, consulta [Annullamento delle modifiche tramite un rollback](#).

Checkpoint (**CheckpointPercentages**)

Controlla se Dimensionamento automatico Amazon EC2 sostituisce le istanze in fasi. È utile se devi eseguire verifiche sulle istanze prima di sostituirle tutte. Per ulteriori informazioni, consulta [Aggiungere punti di controllo a un aggiornamento delle istanze](#).

Ritardo del checkpoint (**CheckpointDelay**)

Il tempo di attesa, espresso in secondi, necessario al raggiungimento di un checkpoint prima di continuare. Per ulteriori informazioni, consulta [Aggiungere punti di controllo a un aggiornamento delle istanze](#).

Preparazione dell'istanza (**InstanceWarmup**)

Intervallo di tempo, espresso in secondi, durante il quale Dimensionamento automatico Amazon EC2 attende che una nuova istanza sia considerata terminata prima di passare alla sostituzione dell'istanza successiva. Se hai già definito correttamente una preparazione dell'istanza di default per il gruppo con scalabilità automatica, non devi modificare la preparazione dell'istanza, a meno che non desideri ignorare l'impostazione predefinita. Per ulteriori informazioni, consulta [Impostazione della preparazione di default dell'istanza per un gruppo con scalabilità automatica](#).

Percentuale massima di integrità (**MaxHealthyPercentage**)

La percentuale della capacità desiderata che il gruppo con dimensionamento automatico può raggiungere quando si sostituiscono le istanze.

Percentuale minima di integrità (**MinHealthyPercentage**)

La percentuale della capacità desiderata del gruppo con dimensionamento automatico che deve essere in servizio, integra e pronta all'uso prima che l'operazione possa continuare.

Istanze protette dalla riduzione orizzontale (**ScaleInProtectedInstances**)

Determina il comportamento di Dimensionamento automatico Amazon EC2 nel caso in cui siano presenti istanze protette dalla riduzione orizzontale. Per ulteriori informazioni su tali istanze, consulta [Utilizzare la protezione di riduzione orizzontale dell'istanza](#).

Dimensionamento automatico Amazon EC2 offre le seguenti opzioni:

- **Replace (Refresh)**: sostituisce le istanze protette dallo scale-in.
- **Ignore (Ignore)**: ignora le istanze protette da scalabilità in e continua a sostituire le istanze non protette.
- **Wait (Wait)**: attende un'ora prima che l'utente rimuova la protezione scale-in. In caso contrario, l'aggiornamento dell'istanza ha esito negativo.

Ignora la corrispondenza (**SkipMatching**)

Determina se Dimensionamento automatico Amazon EC2 ignora la sostituzione delle istanze che corrispondono alla configurazione desiderata. Se non viene specificata alcuna configurazione desiderata, viene ignorata la sostituzione delle istanze con lo stesso modello di avvio e gli stessi

tipi di istanze utilizzati dal gruppo con scalabilità automatica prima dell'avvio dell'aggiornamento dell'istanza. Per ulteriori informazioni, consulta [Utilizzo di un aggiornamento dell'istanza con l'opzione Ignora la corrispondenza](#).

Istanze in standby (**StandbyInstances**)

Determina il comportamento di Dimensionamento automatico Amazon EC2 nel caso in cui siano presenti istanze con stato Standby. Per ulteriori informazioni su tali istanze, consulta [Rimuovere temporaneamente le istanze dal gruppo con scalabilità automatica](#).

Dimensionamento automatico Amazon EC2 offre le seguenti opzioni:

- **Terminate (Terminate)**: termina le istanze presenti in Standby
- **Ignore (Ignore)**: ignora le istanze presenti nello stato Standby e continua a sostituire le istanze che si trovano nello stato InService
- **Wait (Wait)** — Attende un'ora prima che l'utente restituisca le istanze al servizio. In caso contrario, l'aggiornamento dell'istanza ha esito negativo.

Avvio dell'aggiornamento di un'istanza

Important

Puoi eseguire il rollback dell'aggiornamento di un'istanza in corso per annullare eventuali modifiche. Per poter utilizzare questa opzione, il gruppo con scalabilità automatica deve soddisfare i prerequisiti per l'utilizzo dei rollback prima di avviare l'aggiornamento dell'istanza. Per ulteriori informazioni, consulta [Annullamento delle modifiche tramite un rollback](#).

Le seguenti procedure consentono di avviare un aggiornamento dell'istanza utilizzando o. AWS Management Console AWS CLI

Avvio dell'aggiornamento di un'istanza (console)

Se è la prima volta che avvii un aggiornamento delle istanze, utilizzare la console ti sarà utile per scoprire le caratteristiche e le opzioni disponibili.

Avviare un aggiornamento delle istanze nella console (procedura di base)

Se non hai definito in precedenza una [policy di istanze miste](#) per il gruppo con scalabilità automatica, utilizza la seguente procedura. Se in precedenza hai definito una policy di istanze miste, consulta

[Avviare un aggiornamento delle istanze nella console \(gruppo di istanze miste\)](#) per avviare l'aggiornamento di un'istanza.

Come avviare l'aggiornamento di un'istanza

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina Auto Scaling groups (Gruppi Auto Scaling).

3. Nella scheda Aggiornamento istanza in Active instance refresh (Aggiornamento attivo istanza), scegli Start instance refresh (Avvia aggiornamento istanza).
4. In Impostazioni di disponibilità, esegui la seguente operazione:
 - a. Per il metodo di sostituzione dell'istanza:
 - Se non hai impostato una policy di manutenzione delle istanze nel gruppo con dimensionamento automatico, l'impostazione predefinita per il metodo di sostituzione dell'istanza è Termina e avvia. Questo è il comportamento predefinito precedente per l'aggiornamento di un'istanza.
 - Se imposti una policy di manutenzione dell'istanza nel gruppo con dimensionamento automatico, vengono forniti valori predefiniti per il metodo di sostituzione dell'istanza. Per sovrascrivere la policy di manutenzione dell'istanza, scegli Sostituisci. La sostituzione si applica solo all'aggiornamento dell'istanza corrente. Al successivo avvio di un aggiornamento dell'istanza, questi valori vengono ripristinati ai valori predefiniti della policy di manutenzione dell'istanza.

Nella procedura seguente viene descritto come aggiornare il metodo di sostituzione delle istanze.

- i. Seleziona uno dei seguenti metodi di sostituzione delle istanze:
 - Avvio prima del termine: è necessario eseguire il provisioning di una nuova istanza prima di poter chiudere un'istanza esistente. Si tratta di una buona scelta per le applicazioni che privilegiano la disponibilità rispetto al risparmio.

- Termina e avvia: le nuove istanze vengono fornite contemporaneamente alla chiusura delle istanze esistenti. Si tratta di una buona scelta per le applicazioni che privilegiano il risparmio rispetto alla disponibilità. È anche una buona scelta per le applicazioni che non dovrebbero avviare una capacità superiore a quella attualmente disponibile.
 - Comportamento personalizzato: questa opzione consente di impostare un intervallo minimo e massimo personalizzato per la quantità di capacità che desideri rendere disponibile quando si sostituiscono le istanze. Questo approccio può aiutarti a raggiungere il giusto equilibrio tra costi e disponibilità.
- ii. Per Imposta una percentuale di integrità adeguata, inserisci i valori per uno o entrambi i campi seguenti. I campi di attivazione variano a seconda dell'opzione scelta per il metodo di sostituzione dell'istanza.
- Min: imposta la percentuale minima di integrità necessaria per procedere con l'aggiornamento dell'istanza.
 - Max: imposta la percentuale massima di integrità possibile durante l'aggiornamento dell'istanza.
- iii. Espandi la sezione Visualizza la capacità temporanea stimata durante le sostituzioni in base alla capacità desiderata per confermare in che modo i valori di Min e Max si applicano al gruppo. I valori esatti utilizzati dipendono dal valore di capacità desiderato, che cambierà se il gruppo cresce.
- iv. Espandi la sezione Imposta il comportamento di fallback per le dimensioni sostitutive non valide, quindi scegli se Viola la percentuale massima di integrità per dare priorità alla disponibilità o Viola la percentuale minima di integrità.

Non è consigliabile mantenere l'opzione Viola la percentuale minima di integrità predefinita per gruppi molto piccoli. Quando nel gruppo con dimensionamento automatico è presente una sola istanza, l'avvio di un aggiornamento delle istanze può causare un'interruzione.

Questo passaggio configura il comportamento di fallback se utilizzi un gruppo con dimensionamento automatico che non dispone ancora di una policy di manutenzione dell'istanza. Questa opzione non è disponibile e non viene visualizzata quando il gruppo ha una policy di manutenzione delle istanze. Questa opzione è disponibile anche solo per il metodo di sostituzione Termina e avvia. Gli altri metodi di sostituzione violeranno la percentuale massima di integrità per dare priorità alla disponibilità.

- b. In Preparazione dell'istanza, inserisci il numero di secondi da quando lo stato di una nuova istanza cambia in InService a quando termina l'inizializzazione. Dimensionamento automatico Amazon EC2 attende il tempo prestabilito prima di passare alla sostituzione dell'istanza successiva.

Durante la preparazione, un'istanza appena avviata non viene conteggiata per i parametri aggregati dell'istanza del gruppo con scalabilità automatica (come CPUUtilization, NetworkIn e NetworkOut). Se sono state aggiunte policy di dimensionamento al gruppo con scalabilità automatica, le attività di dimensionamento vengono eseguite in parallelo. Se imposti un intervallo lungo per il periodo di aggiornamento dell'istanza, ci vorrà più tempo prima che le istanze appena lanciate vengano visualizzate nelle metriche. Pertanto, un periodo di riscaldamento adeguato impedisce ad Amazon EC2 Auto Scaling di scalare su dati metrici obsoleti.


Se hai già definito correttamente una preparazione dell'istanza di default per il gruppo con scalabilità automatica, non devi modificare la preparazione dell'istanza. Puoi tuttavia impostare un valore per questa opzione se desideri sovrascrivere l'impostazione predefinita. Per ulteriori informazioni sull'impostazione della preparazione dell'istanza di default, consulta [Impostazione della preparazione di default dell'istanza per un gruppo con scalabilità automatica](#).

5. In Impostazioni di aggiornamento, esegui la seguente operazione:
 - a. (Facoltativo) Per Checkpoints (Punti di controllo), scegli Enable checkpoints (Abilita punti di controllo) per sostituire le istanze utilizzando un approccio incrementale o graduale all'aggiornamento delle istanze. Ciò consente di avere tempo aggiuntivo per la verifica tra i set delle sostituzioni. Se scegli di non abilitare i punti di controllo, le istanze vengono sostituite in un'operazione quasi continua.

Se abiliti i punti di controllo, consulta [Abilitazione dei punti di controllo \(console\)](#) per ulteriori passaggi.

- b. Attiva o disattiva Skip matching (Salta corrispondenza):
 - Per ignorare la sostituzione delle istanze che corrispondono già al modello di avvio, mantieni selezionata la casella di controllo Abilita Ignora la corrispondenza.
 - Se disattivi l'opzione Salta corrispondenza deselegionando questa casella di controllo, tutte le istanze possono essere sostituite.

Quando abiliti la funzione Ignora la corrispondenza, invece di utilizzare il modello esistente, puoi impostare un nuovo modello di avvio o una nuova versione del modello di avvio. Puoi eseguire questa operazione nella sezione Configurazione desiderata della pagina Avvio dell'aggiornamento dell'istanza.

 Note

Per utilizzare la funzionalità Salta corrispondenza per aggiornare un gruppo con scalabilità automatica che attualmente utilizza una configurazione di avvio, devi selezionare un modello di avvio in Configurazione desiderata. L'utilizzo della funzionalità Salta corrispondenza con una configurazione di avvio non è supportato.

- c. In Istanze in standby, scegli Ignora, Termina o Attendi. Ciò determina cosa succede se vengono individuate istanze il cui stato è Standby. Per ulteriori informazioni, consulta [Rimuovere temporaneamente le istanze dal gruppo con scalabilità automatica](#).

Se scegli Attendi, devi eseguire ulteriori operazioni per ripristinare l'utilizzo di queste istanze. In caso contrario, l'aggiornamento dell'istanza sostituisce tutte le istanze InService e attende un'ora. Nel caso in cui siano presenti ancora istanze in stato Standby, l'aggiornamento dell'istanza ha esito negativo. Per evitare questa situazione, scegli l'opzione Ignora o Termina per tali istanze.

- d. In Istanze protette dalla riduzione orizzontale, scegli Ignora, Sostituisci o Attendi. Ciò determina cosa succede se vengono individuate istanze protette dalla riduzione orizzontale. Per ulteriori informazioni, consulta [Utilizzare la protezione di riduzione orizzontale dell'istanza](#).

Se scegli Attendi, devi eseguire ulteriori operazioni per rimuovere la protezione dalla riduzione orizzontale da tali istanze. In caso contrario, l'aggiornamento dell'istanza sostituisce tutte le istanze non protette e attende un'ora. Nel caso in cui siano ancora presenti istanze protette dalla riduzione orizzontale, l'aggiornamento dell'istanza ha esito negativo. Per evitare questa situazione, scegli l'opzione Ignora o Sostituisci per tali istanze.

6. (Facoltativo) Per gli CloudWatch allarmi, scegli Abilita allarmi, quindi scegli uno o più CloudWatch allarmi. CloudWatch gli allarmi possono essere utilizzati per identificare eventuali problemi e fallire l'operazione se un allarme entra in stato. ALARM Per ulteriori informazioni, consulta [Avvio dell'aggiornamento di un'istanza con rollback automatico](#).

7. (Facoltativo) Espandi la sezione Configurazione desiderata per specificare gli aggiornamenti che desideri apportare al gruppo con scalabilità automatica.

Per questo passaggio, puoi scegliere di utilizzare la sintassi JSON o YAML per modificare i valori dei parametri anziché effettuare selezioni nell'interfaccia della console. Per farlo, scegli Use code editor (Usa editor di codice) e non Use console interface (Usa interfaccia console). La procedura seguente spiega come effettuare selezioni utilizzando l'interfaccia della console.


- a. Per Update launch template (Aggiornamento del modello di avvio):

- Se non hai creato un nuovo modello di avvio o una nuova versione del modello di avvio per il gruppo con scalabilità automatica, lascia questa casella di controllo deselezionata.
- In caso contrario, selezionala. Quando selezioni questa opzione, Dimensionamento automatico Amazon EC2 mostra il modello di avvio corrente e la relativa versione ed elenca tutte le altre versioni disponibili. Scegli il modello di avvio e la versione.

Una volta scelta la versione, puoi visualizzare le relative informazioni. Questa è la versione del modello di avvio che verrà utilizzata per sostituire le istanze nell'ambito di un aggiornamento delle istanze. Se l'aggiornamento viene eseguito correttamente, questa versione del modello di avvio verrà utilizzata anche ogni volta che vengono avviate nuove istanze, ad esempio quando il gruppo viene dimensionato.


- b. Per Choose a set of instance types and purchase options to override the instance type in the launch template (Scelta di un set di tipi di istanza e opzioni di acquisto per sovrascrivere il tipo di istanza nel modello di avvio):

- Non selezionare questa casella di controllo se vuoi utilizzare il tipo di istanza e l'opzione di acquisto specificati nel modello di avvio.
- Seleziona questa casella di controllo se desideri sovrascrivere il tipo di istanza nel modello di avvio o eseguire istanze spot. Puoi aggiungere manualmente ogni tipo di istanza oppure scegliere un tipo di istanza principale e un'opzione di raccomandazione che recupera eventuali tipi di istanza corrispondenti aggiuntivi. Se prevedi di avviare istanze spot, ti consigliamo di aggiungere alcuni tipi di istanza diversi. In questo modo, Dimensionamento automatico Amazon EC2 può avviare un altro tipo di istanza se la capacità dell'istanza non è sufficiente nelle zone di disponibilità scelte. Per ulteriori informazioni, consulta [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#).

 Warning

Le istanze spot non devono essere utilizzate con applicazioni che non sono in grado di gestire un'eventuale interruzione dell'istanza che potrebbe verificarsi se il servizio Spot di Amazon EC2 deve recuperare capacità.

Se selezioni questa casella di controllo, assicurati che il modello di avvio non richieda già istanze Spot. Non puoi utilizzare un modello di avvio che richiede alle istanze spot di creare un gruppo con scalabilità automatica che utilizzi più tipi di istanza e che avvii istanze spot e on demand.

 Note

Per utilizzare queste opzioni per un gruppo con scalabilità automatica che attualmente utilizza una configurazione di avvio, devi selezionare un modello di avvio in Update launch template (Aggiorna modello di avvio). La sovrascrittura del tipo di istanza nella configurazione di avvio non è supportata.

8. (Facoltativo) In Impostazioni di rollback, scegli Abilita il rollback automatico per eseguire automaticamente il rollback dell'aggiornamento dell'istanza in caso di errore.

Questa impostazione può essere abilitata solo se il gruppo con dimensionamento automatico soddisfa i prerequisiti per l'utilizzo dei rollback.

Per ulteriori informazioni, consulta [Annullamento delle modifiche tramite un rollback](#).

9. Rivedi tutte le selezioni per confermare che siano impostate correttamente.

A questo punto, è consigliabile verificare che le differenze tra le modifiche correnti e quelle proposte non influiscano sull'applicazione in modo imprevisto o indesiderato. Per verificare che il tipo di istanza è compatibile con il modello di avvio, consulta [Compatibilità del tipo di istanza](#).

10. Se le selezioni per l'aggiornamento delle istanze sono quelle desiderate, scegli Avvia aggiornamento dell'istanza.

Avviare un aggiornamento delle istanze nella console (gruppo di istanze miste)

Se hai creato un gruppo con scalabilità automatica con una [policy di istanze miste](#), utilizza la seguente procedura. Se in precedenza non hai definito una policy di istanze miste per il gruppo, consulta [Avviare un aggiornamento delle istanze nella console \(procedura di base\)](#) per avviare un aggiornamento delle istanze.

Come avviare l'aggiornamento di un'istanza

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina Auto Scaling groups (Gruppi Auto Scaling).

3. Nella scheda Aggiornamento istanza in Active instance refresh (Aggiornamento attivo istanza), scegli Start instance refresh (Avvia aggiornamento istanza).
4. In Impostazioni di disponibilità, esegui la seguente operazione:
 - a. Per il metodo di sostituzione dell'istanza:
 - Se non hai impostato una policy di manutenzione delle istanze nel gruppo con dimensionamento automatico, l'impostazione predefinita per il metodo di sostituzione dell'istanza è Termina e avvia. Questo è il comportamento predefinito precedente per l'aggiornamento di un'istanza.
 - Se imposti una policy di manutenzione dell'istanza nel gruppo con dimensionamento automatico, vengono forniti valori predefiniti per il metodo di sostituzione dell'istanza. Per sovrascrivere la policy di manutenzione dell'istanza, scegli Sostituisci. La sostituzione si applica solo all'aggiornamento dell'istanza corrente. Al successivo avvio di un aggiornamento dell'istanza, questi valori vengono ripristinati ai valori predefiniti della policy di manutenzione dell'istanza.

Nella procedura seguente viene descritto come aggiornare il metodo di sostituzione delle istanze.

- i. Seleziona uno dei seguenti metodi di sostituzione delle istanze:

- **Avvio prima del termine:** è necessario eseguire il provisioning di una nuova istanza prima di poter chiudere un'istanza esistente. Si tratta di una buona scelta per le applicazioni che privilegiano la disponibilità rispetto al risparmio.
 - **Termina e avvia:** le nuove istanze vengono fornite contemporaneamente alla chiusura delle istanze esistenti. Si tratta di una buona scelta per le applicazioni che privilegiano il risparmio rispetto alla disponibilità. È anche una buona scelta per le applicazioni che non dovrebbero avviare una capacità superiore a quella attualmente disponibile.
 - **Comportamento personalizzato:** questa opzione consente di impostare un intervallo minimo e massimo personalizzato per la quantità di capacità che desideri rendere disponibile quando si sostituiscono le istanze. Questo approccio può aiutarti a raggiungere il giusto equilibrio tra costi e disponibilità.
- ii. **Per Imposta una percentuale di integrità adeguata,** inserisci i valori per uno o entrambi i campi seguenti. I campi di attivazione variano a seconda dell'opzione scelta per il metodo di sostituzione dell'istanza.
- **Min:** imposta la percentuale minima di integrità necessaria per procedere con l'aggiornamento dell'istanza.
 - **Max:** imposta la percentuale massima di integrità possibile durante l'aggiornamento dell'istanza.
- iii. **Espandi la sezione Visualizza la capacità temporanea stimata durante le sostituzioni** in base alla capacità desiderata per confermare in che modo i valori di Min e Max si applicano al gruppo. I valori esatti utilizzati dipendono dal valore di capacità desiderato, che cambierà se il gruppo cresce.
- iv. **Espandi la sezione Imposta il comportamento di fallback per le dimensioni sostitutive non valide,** quindi scegli se Viola la percentuale massima di integrità per dare priorità alla disponibilità o Viola la percentuale minima di integrità.

Non è consigliabile mantenere l'opzione Viola la percentuale minima di integrità predefinita per gruppi molto piccoli. Quando nel gruppo con dimensionamento automatico è presente una sola istanza, l'avvio di un aggiornamento delle istanze può causare un'interruzione.

Questo passaggio configura il comportamento di fallback se utilizzi un gruppo con dimensionamento automatico che non dispone ancora di una policy di manutenzione dell'istanza. Questa opzione non è disponibile e non viene visualizzata quando il gruppo ha una policy di manutenzione delle istanze. Questa opzione è disponibile anche solo

per il metodo di sostituzione Termina e avvia. Gli altri metodi di sostituzione violeranno la percentuale massima di integrità per dare priorità alla disponibilità.

- b. In Preparazione dell'istanza, inserisci il numero di secondi da quando lo stato di una nuova istanza cambia in InService a quando termina l'inizializzazione. Dimensionamento automatico Amazon EC2 attende il tempo prestabilito prima di passare alla sostituzione dell'istanza successiva.

Durante la preparazione, un'istanza appena avviata non viene conteggiata per i parametri aggregati dell'istanza del gruppo con scalabilità automatica (come CPUUtilization, NetworkIn e NetworkOut). Se sono state aggiunte policy di dimensionamento al gruppo con scalabilità automatica, le attività di dimensionamento vengono eseguite in parallelo. Se imposti un intervallo lungo per il periodo di aggiornamento dell'istanza, ci vorrà più tempo prima che le istanze appena lanciate vengano visualizzate nelle metriche. Pertanto, un periodo di riscaldamento adeguato impedisce ad Amazon EC2 Auto Scaling di scalare su dati metrici obsoleti.

Se hai già definito correttamente una preparazione dell'istanza di default per il gruppo con scalabilità automatica, non devi modificare la preparazione dell'istanza. Puoi tuttavia impostare un valore per questa opzione se desideri sovrascrivere l'impostazione predefinita. Per ulteriori informazioni sull'impostazione della preparazione dell'istanza di default, consulta [Impostazione della preparazione di default dell'istanza per un gruppo con scalabilità automatica](#).

5. In Impostazioni di aggiornamento, esegui la seguente operazione:
 - a. (Facoltativo) Per Checkpoints (Punti di controllo), scegli Enable checkpoints (Abilita punti di controllo) per sostituire le istanze utilizzando un approccio incrementale o graduale all'aggiornamento delle istanze. Ciò consente di avere tempo aggiuntivo per la verifica tra i set delle sostituzioni. Se scegli di non abilitare i punti di controllo, le istanze vengono sostituite in un'operazione quasi continua.

Se abiliti i punti di controllo, consulta [Abilitazione dei punti di controllo \(console\)](#) per ulteriori passaggi.

- b. Attiva o disattiva Skip matching (Salta corrispondenza):
 - Per ignorare la sostituzione delle istanze che già corrispondono al modello di avvio e la sovrascrittura di qualsiasi tipo di istanza, mantieni selezionata la casella di controllo Abilita Ignora la corrispondenza.

- Se scegli di disattivare l'opzione Salta corrispondenza deselegionando questa casella di controllo, tutte le istanze possono essere sostituite.

Quando abiliti la funzione Ignora la corrispondenza, invece di utilizzare il modello esistente, puoi impostare un nuovo modello di avvio o una nuova versione del modello di avvio. Puoi eseguire questa operazione nella sezione Configurazione desiderata della pagina Avvio dell'aggiornamento dell'istanza. Puoi anche aggiornare le sovrascritture del tipo di istanza in Configurazione desiderata.

- c. In Istanze in standby, scegli Ignora, Termina o Attendi. Ciò determina cosa succede se vengono individuate istanze il cui stato è Standby. Per ulteriori informazioni, consulta [Rimuovere temporaneamente le istanze dal gruppo con scalabilità automatica](#).

Se scegli Attendi, devi eseguire ulteriori operazioni per ripristinare l'utilizzo di queste istanze. In caso contrario, l'aggiornamento dell'istanza sostituisce tutte le istanze InService e attende un'ora. Nel caso in cui siano presenti ancora istanze in stato Standby, l'aggiornamento dell'istanza ha esito negativo. Per evitare questa situazione, scegli l'opzione Ignora o Termina per tali istanze.

- d. In Istanze protette dalla riduzione orizzontale, scegli Ignora, Sostituisci o Attendi. Ciò determina cosa succede se vengono individuate istanze protette dalla riduzione orizzontale. Per ulteriori informazioni, consulta [Utilizzare la protezione di riduzione orizzontale dell'istanza](#).

Se scegli Attendi, devi eseguire ulteriori operazioni per rimuovere la protezione dalla riduzione orizzontale da tali istanze. In caso contrario, l'aggiornamento dell'istanza sostituisce tutte le istanze non protette e attende un'ora. Nel caso in cui siano ancora presenti istanze protette dalla riduzione orizzontale, l'aggiornamento dell'istanza ha esito negativo. Per evitare questa situazione, scegli l'opzione Ignora o Sostituisci per tali istanze.

6. (Facoltativo) Per gli CloudWatch allarmi, scegli Abilita allarmi, quindi scegli uno o più CloudWatch allarmi. CloudWatch gli allarmi possono essere utilizzati per identificare eventuali problemi e fallire l'operazione se un allarme entra in stato. ALARM Per ulteriori informazioni, consulta [Avvio dell'aggiornamento di un'istanza con rollback automatico](#).
7. Nella sezione Configurazione desiderata, procedi come segue.

Per questo passaggio, puoi scegliere di utilizzare la sintassi JSON o YAML per modificare i valori dei parametri anziché effettuare selezioni nell'interfaccia della console. Per farlo, scegli Use code

editor (Usa editor di codice) e non Use console interface (Usa interfaccia console). La procedura seguente spiega come effettuare selezioni utilizzando l'interfaccia della console.

a. Per Update launch template (Aggiornamento del modello di avvio):

- Se non hai creato un nuovo modello di avvio o una nuova versione del modello di avvio per il gruppo con scalabilità automatica, lascia questa casella di controllo deselezionata.
- In caso contrario, selezionala. Quando selezioni questa opzione, Dimensionamento automatico Amazon EC2 mostra il modello di avvio corrente e la relativa versione ed elenca tutte le altre versioni disponibili. Scegli il modello di avvio e la versione.

Una volta scelta la versione, puoi visualizzare le relative informazioni. Questa è la versione del modello di avvio che verrà utilizzata per sostituire le istanze nell'ambito di un aggiornamento delle istanze. Se l'aggiornamento viene eseguito correttamente, questa versione del modello di avvio verrà utilizzata anche ogni volta che vengono avviate nuove istanze, ad esempio quando il gruppo viene dimensionato.

b. Per Use these settings to override the instance type and purchase option defined in the launch template (Utilizzo di queste impostazioni per sovrascrivere il tipo di istanza e l'opzione di acquisto definite nel modello di avvio):

Per impostazione predefinita, questa casella di controllo è selezionata. Amazon EC2 Auto Scaling popola ciascun parametro con il valore attualmente impostato nella policy di istanze miste per il gruppo con scalabilità automatica. Aggiorna solo i valori per i parametri che desideri modificare. Per informazioni su queste impostazioni, consulta [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#).

Warning

Consigliamo di non deselezionare questa casella di controllo. Deselezionala solo se desideri interrompere l'utilizzo di una policy di istanze miste. Dopo un aggiornamento delle istanze, Amazon EC2 Auto Scaling aggiorna il gruppo affinché corrisponda alla configurazione desiderata. Se non include più una policy di istanze miste, Amazon EC2 Auto Scaling termina gradualmente tutte le istanze Spot attualmente in esecuzione e le sostituisce con istanze on demand. In alternativa, se il modello di avvio richiede istanze Spot, Amazon EC2 Auto Scaling termina gradualmente le istanze on demand attualmente in esecuzione e le sostituisce con istanze Spot.

8. (Facoltativo) In Impostazioni di rollback, scegli Abilita il rollback automatico per eseguire automaticamente il rollback dell'aggiornamento dell'istanza in caso di errore per qualsiasi motivo.

Questa impostazione può essere abilitata solo se il gruppo con dimensionamento automatico soddisfa i prerequisiti per l'utilizzo dei rollback.

Per ulteriori informazioni, consulta [Annullamento delle modifiche tramite un rollback](#).

9. Rivedi tutte le selezioni per confermare che siano impostate correttamente.

A questo punto, è consigliabile verificare che le differenze tra le modifiche correnti e quelle proposte non influiscano sull'applicazione in modo imprevisto o indesiderato. Per verificare che il tipo di istanza è compatibile con il modello di avvio, consulta [Compatibilità del tipo di istanza](#).

Se le selezioni per l'aggiornamento delle istanze sono quelle desiderate, scegli Avvia aggiornamento dell'istanza.

Avvio dell'aggiornamento di un'istanza (AWS CLI)

Come avviare l'aggiornamento di un'istanza

Utilizzate il [start-instance-refresh](#) comando seguente per avviare l'aggiornamento di un'istanza da AWS CLI. Puoi specificare le preferenze che desideri modificare in un file di configurazione JSON. Quando fai riferimento al file di configurazione, fornisci il percorso e il nome del file come illustrato nell'esempio seguente.

```
aws autoscaling start-instance-refresh --cli-input-json file://config.json
```

Contenuto di config.json.

```
{
  "AutoScalingGroupName": "my-asg",
  "Preferences": {
    "InstanceWarmup": 60,
    "MinHealthyPercentage": 50,
    "AutoRollback": true,
    "ScaleInProtectedInstances": Ignore,
    "StandbyInstances": Terminate
  }
}
```

Se le preferenze non vengono fornite, vengono utilizzati i valori predefiniti. Per ulteriori informazioni, consulta [Informazioni sui valori predefiniti per l'aggiornamento di un'istanza](#).

Output di esempio:

```
{
  "InstanceRefreshId": "08b91cf7-8fa6-48af-b6a6-d227f40f1b9b"
}
```

Monitora l'aggiornamento di un'istanza

Puoi monitorare un aggiornamento in corso dell'istanza o cercare lo stato degli aggiornamenti precedenti delle istanze delle ultime sei settimane utilizzando l'opzione o. AWS Management Console
AWS CLI

Monitora e controlla lo stato dell'aggiornamento di un'istanza

Per monitorare e controllare lo stato di un aggiornamento dell'istanza, utilizzate uno dei seguenti metodi:

Console

Tip

In questa procedura, le colonne denominate dovrebbero già essere visualizzate. Per visualizzare le colonne nascoste o modificare il numero di righe visualizzate, scegli l'icona a forma di ingranaggio nell'angolo in alto a destra della sezione per aprire la finestra delle preferenze. Aggiorna le impostazioni secondo necessità e scegli Conferma.

Per monitorare e controllare lo stato di un aggiornamento dell'istanza (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Nella scheda Instance refresh (Aggiornamento istanze) in Instance refresh history (Cronologia aggiornamento istanze), puoi determinare lo stato della richiesta esaminando

la colonna Stato. L'operazione passa Pending allo stato durante l'inizializzazione. Lo stato dovrebbe quindi passare rapidamente a InProgress. Quando tutte le istanze sono state aggiornate, lo stato cambia in Successful.

4. È possibile monitorare ulteriormente il successo o il fallimento delle attività in corso visualizzando le attività di ridimensionamento del gruppo. Nella scheda Activity (Attività) in Activity history (Cronologia attività), quando viene avviato l'aggiornamento delle istanze, vengono mostrate le voci al momento della terminazione delle istanze e un altro set di voci all'avvio delle istanze. Se hai numerose attività di scalabilità, puoi visualizzarne altre scegliendo l'icona > nella parte superiore della cronologia delle attività. Per informazioni sulla risoluzione dei problemi che potrebbero causare il fallimento delle attività, consulta [Risoluzione dei problemi di Amazon EC2 Auto Scaling](#).
5. (Facoltativo) Nella scheda Gestione delle istanze, in Istanze, è possibile esaminare lo stato di avanzamento di istanze specifiche in base alle esigenze.

AWS CLI

Per monitorare e controllare lo stato di un aggiornamento dell'istanza ()AWS CLI

Utilizza il seguente comando [describe-instance-refreshes](#).

```
aws autoscaling describe-instance-refreshes --auto-scaling-group-name my-asg
```

Di seguito è riportato un output di esempio.

Gli aggiornamenti delle istanze sono ordinati in base all'ora di inizio. Gli aggiornamenti delle istanze ancora in corso vengono descritti per primi.

```
{
  "InstanceRefreshes": [
    {
      "InstanceRefreshId": "08b91cf7-8fa6-48af-b6a6-d227f40f1b9b",
      "AutoScalingGroupName": "my-asg",
      "Status": "InProgress",
      "StatusReason": "Waiting for instances to warm up before continuing. For
example: i-0645704820a8e83ff is warming up.",
      "StartTime": "2023-11-24T16:46:52+00:00",
      "PercentageComplete": 50,
      "InstancesToUpdate": 0,
      "Preferences": {
        "MaxHealthyPercentage": 120,

```

```

        "MinHealthyPercentage":90,
        "InstanceWarmup":60,
        "SkipMatching":false,
        "AutoRollback":true,
        "ScaleInProtectedInstances":"Ignore",
        "StandbyInstances":"Ignore"
    }
},
{
    "InstanceRefreshId":"0e151305-1e57-4a32-a256-1fd14157c5ec",
    "AutoScalingGroupName":"my-asg",
    "Status":"Successful",
    "StartTime":"2023-11-22T13:53:37+00:00",
    "EndTime":"2023-11-22T13:59:45+00:00",
    "PercentageComplete":100,
    "InstancesToUpdate":0,
    "Preferences":{
        "MaxHealthyPercentage":120,
        "MinHealthyPercentage":90,
        "InstanceWarmup":60,
        "SkipMatching":false,
        "AutoRollback":true,
        "ScaleInProtectedInstances":"Ignore",
        "StandbyInstances":"Ignore"
    }
}
]
}

```

È possibile monitorare ulteriormente il successo o il fallimento delle attività in corso visualizzando le attività di scalabilità del gruppo. Le attività di scalabilità consentono inoltre di approfondire la ricerca di maggiori dettagli per la risoluzione dei problemi relativi all'aggiornamento dell'istanza. Per ulteriori informazioni, consulta [Risoluzione dei problemi di Amazon EC2 Auto Scaling](#).

Stati di aggiornamento dell'istanza

Quando avvii l'aggiornamento di un'istanza, quest'ultima passa allo stato In attesa. Passa da In sospeso a InProgress finché non raggiunge i valori Riuscito, Non riuscito, Annullato o RollbackSuccessfulRollbackFailed

Un aggiornamento dell'istanza può presentare gli stati seguenti:

Stato	Descrizione
Pending (In attesa)	La richiesta è stata creata, ma l'aggiornamento dell'istanza non è stato avviato.
InProgress	È in corso un aggiornamento dell'istanza.
Riuscito	Aggiornamento istanza completato correttamente.
Failed (Non riuscito)	Impossibile completare l'aggiornamento dell'istanza. È possibile risolvere i problemi utilizzando il motivo dello stato e le attività di dimensionamento.
Annullamento in corso	Un aggiornamento dell'istanza in corso viene annullato.
Annullato	L'aggiornamento dell'istanza viene annullato.
RollbackInProgress	È in corso il rollback di un aggiornamento dell'istanza.
RollbackFailed	Il rollback non è stato completato. È possibile risolvere i problemi utilizzando il motivo dello stato e le attività di dimensionamento.
RollbackSuccessful	Il rollback è stato completato correttamente.

Annullamento dell'aggiornamento di un'istanza

Puoi annullare l'aggiornamento di un'istanza quando l'operazione è ancora in corso, non quando è completata.

L'annullamento dell'aggiornamento di un'istanza non comporta il rollback di tutte le istanze già sostituite. Per ripristinare le modifiche apportate alle istanze, esegui un rollback. Per ulteriori informazioni, consulta [Annullamento delle modifiche tramite un rollback](#).

Argomenti

- [Annullamento dell'aggiornamento di un'istanza \(console\)](#)
- [Annullamento dell'aggiornamento di un'istanza \(AWS CLI\)](#)

Annullamento dell'aggiornamento di un'istanza (console)

Annullamento dell'aggiornamento delle istanze

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.
3. Nella scheda Aggiornamento istanza, in Aggiornamento dell'istanza attivo, scegli Operazioni, quindi Annulla.
4. Quando viene richiesta la conferma, seleziona Confirm (Conferma).

Lo stato dell'aggiornamento dell'istanza è impostato su Annullamento in corso. Al termine dell'annullamento, lo stato dell'aggiornamento dell'istanza è impostato su Annullato.

Annullamento dell'aggiornamento di un'istanza (AWS CLI)

Annullamento dell'aggiornamento delle istanze

Utilizzate il [cancel-instance-refresh](#) comando di AWS CLI e fornite il nome del gruppo Auto Scaling.

```
aws autoscaling cancel-instance-refresh --auto-scaling-group-name my-asg
```

Output di esempio:

```
{
  "InstanceRefreshId": "08b91cf7-8fa6-48af-b6a6-d227f40f1b9b"
}
```

Annullamento delle modifiche tramite un rollback

Puoi eseguire il rollback di un aggiornamento dell'istanza quando l'operazione è ancora in corso, non quando è stata completata. Tuttavia, puoi aggiornare nuovamente il gruppo con scalabilità automatica avviando un nuovo aggiornamento dell'istanza.

Durante il rollback, Dimensionamento automatico Amazon EC2 sostituisce le istanze implementate fino a quel momento. Le nuove istanze corrispondono all'ultima configurazione salvata nel gruppo con scalabilità automatica prima di avviare l'aggiornamento dell'istanza.

Dimensionamento automatico Amazon EC2 offre i seguenti metodi per eseguire il rollback:

- Rollback manuale: l'avvio manuale di un rollback consente di rimuovere gli elementi implementati fino al punto di rollback.
- Rollback automatico: Amazon EC2 Auto Scaling inverte automaticamente ciò che è stato distribuito se l'aggiornamento dell'istanza non riesce per qualche motivo o se CloudWatch vengono attivati degli allarmi specificati. ALARM

Indice

- [Considerazioni](#)
- [Avvio manuale di un rollback](#)
- [Avvio dell'aggiornamento di un'istanza con rollback automatico](#)

Considerazioni

Le seguenti considerazioni si applicano quando si utilizza il rollback:

- L'opzione rollback è disponibile solo se specifichi una configurazione desiderata come parte dell'avvio di un aggiornamento dell'istanza.
- Puoi ripristinare una versione precedente di un modello di avvio solo se la versione è una versione numerata specifica. L'opzione di rollback non è disponibile se il gruppo con dimensionamento automatico è configurato per utilizzare la versione del modello di lancio `$Latest` o `$Default`.
- Inoltre, non è possibile tornare a un modello di avvio configurato per utilizzare un alias AMI dal AWS Systems Manager Parameter Store.
- L'ultima configurazione salvata nel gruppo con dimensionamento automatico deve essere stabile. In caso contrario, il flusso di lavoro del rollback proseguirà ma alla fine l'operazione avrà esito negativo. Finché non si risolve il problema, il gruppo con scalabilità automatica potrebbe trovarsi in uno stato non riuscito, in cui non è in grado di avviare correttamente le istanze. Ciò potrebbe influire sulla disponibilità del servizio o dell'applicazione.

Avvio manuale di un rollback

Console

Per avviare manualmente un rollback dell'aggiornamento dell'istanza (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.

2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.
3. Nella scheda Aggiornamento istanza, in Aggiornamento dell'istanza attivo, scegli Operazioni, quindi Avvia rollback.
4. Quando viene richiesta la conferma, seleziona Confirm (Conferma).

AWS CLI

Per avviare manualmente un rollback dell'aggiornamento dell'istanza (AWS CLI)

Utilizzate il [rollback-instance-refresh](#) comando di AWS CLI e fornite il nome del gruppo Auto Scaling.

```
aws autoscaling rollback-instance-refresh --auto-scaling-group-name my-asg
```

Output di esempio:

```
{
  "InstanceRefreshId": "08b91cf7-8fa6-48af-b6a6-d227f40f1b9b"
}
```

Tip

Se questo comando genera un errore, assicurati di averlo aggiornato AWS CLI localmente alla versione più recente.

Avvio dell'aggiornamento di un'istanza con rollback automatico

Utilizzando la funzione di rollback automatico, puoi ripristinare automaticamente l'aggiornamento dell'istanza in caso di errore, ad esempio quando si verificano errori o viene attivato uno CloudWatch specifico allarme Amazon. ALARM

Se abiliti il rollback automatico e si verificano errori durante la sostituzione delle istanze, l'aggiornamento dell'istanza tenta di completare tutte le sostituzioni per un'ora prima che abbia esito negativo ed esegua il rollback. Questi errori sono in genere causati da errori di avvio di EC2, controlli dell'integrità configurati in modo errato o dalla mancata ignorazione o dalla mancata autorizzazione alla chiusura delle istanze nello stato Standby o protette dal dimensionamento verticale.

La specificazione degli CloudWatch allarmi è facoltativa. Per specificare un allarme, devi prima crearlo. Puoi specificare sia allarmi dei parametri che allarmi composti in. Per informazioni sulla creazione dell'allarme, consulta la [Amazon CloudWatch User Guide](#). Utilizzando i parametri di Elastic Load Balancing come esempio, se utilizzi un Application Load Balancer, puoi utilizzare i parametri HTTPCode_ELB_5XX_Count e HTTPCode_ELB_4XX_Count.

Considerazioni

- Se si specifica un CloudWatch avviso ma non si abilita il rollback automatico e lo stato dell'allarme passa aALARM, l'aggiornamento dell'istanza fallisce senza il rollback.
- Quando avvii l'aggiornamento di un'istanza puoi scegliere un massimo di 10 allarmi.
- Quando si sceglie un CloudWatch allarme, quest'ultimo deve essere in uno stato compatibile. Se lo stato dell'allarme è INSUFFICIENT_DATA o ALARM, ricevi un errore quando tenti di avviare l'aggiornamento dell'istanza.
- Quando crei un allarme da utilizzare per Dimensionamento automatico Amazon EC2, l'allarme deve includere come trattare i punti dati mancanti. Se un parametro è spesso privo di punti di dati in base per progettazione, lo stato dell'allarme è INSUFFICIENT_DATA durante tali periodi. In questo caso, Dimensionamento automatico Amazon EC2 non può sostituire le istanze finché non vengono trovati nuovi punti dati. Per forzare l'allarme a mantenere il precedente stato ALARM o OK, puoi scegliere di ignorare i dati mancanti. Per ulteriori informazioni, consulta [Configurazione del modo in cui gli allarmi trattano i dati mancanti](#) nella Amazon CloudWatch User Guide.

Console

Per avviare l'aggiornamento dell'istanza con il rollback automatico (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.
3. Nella scheda Aggiornamento istanza in Active instance refresh (Aggiornamento attivo istanza), scegli Start instance refresh (Avvia aggiornamento istanza).
4. Segui la procedura [Avvio dell'aggiornamento di un'istanza \(console\)](#) e configura le impostazioni di aggiornamento dell'istanza in base alle esigenze.
5. (Facoltativo) In Aggiorna impostazioni, per CloudWatch allarmi, scegli Abilita CloudWatch allarmi, quindi scegli uno o più allarmi per identificare eventuali problemi e fallire l'operazione se un allarme entra nello stato. ALARM

6. In Impostazioni di rollback, seleziona Abilita rollback automatico per eseguire il rollback automatico di un aggiornamento dell'istanza non riuscito alla configurazione salvata l'ultima volta nel gruppo con dimensionamento automatico prima di avviare l'aggiornamento dell'istanza.
7. Controlla le selezioni, quindi scegli Avvia aggiornamento dell'istanza.

AWS CLI

Per avviare l'aggiornamento dell'istanza con il rollback automatico (AWS CLI)

Utilizzate il [start-instance-refresh](#) comando e specificate `true` l'AutoRollbackopzione in Preferences

L'esempio seguente mostra come avviare un aggiornamento dell'istanza che verrà ripristinato automaticamente in caso di errore. Sostituisci i valori del parametro *italicized* con i tuoi valori.

```
aws autoscaling start-instance-refresh --cli-input-json file://config.json
```

Contenuto di config.json.

```
{
  "AutoScalingGroupName": "my-asg",
  "DesiredConfiguration": {
    "LaunchTemplate": {
      "LaunchTemplateName": "my-launch-template",
      "Version": "1"
    }
  },
  "Preferences": {
    "AutoRollback": true
  }
}
```

In alternativa, per eseguire automaticamente il rollback quando l'aggiornamento dell'istanza fallisce o quando uno specifico CloudWatch allarme è nello ALARM stato, specificate l'AlarmSpecificationopzione in Preferences e fornite il nome dell'avviso, come nell'esempio seguente. Sostituisci i valori del parametro *italicized* con i tuoi valori.

```
{
  "AutoScalingGroupName": "my-asg",
```

```
"DesiredConfiguration": {
  "LaunchTemplate": {
    "LaunchTemplateName": "my-launch-template",
    "Version": "1"
  }
},
"Preferences": {
  "AutoRollback": true,
  "AlarmSpecification": { "Alarms": [ "my-alarm" ] }
}
}
```

Se il comando viene eseguito correttamente, verrà visualizzato un output simile al seguente:

```
{
  "InstanceRefreshId": "08b91cf7-8fa6-48af-b6a6-d227f40f1b9b"
}
```

Tip

Se questo comando genera un errore, assicurati di averlo aggiornato AWS CLI localmente alla versione più recente.

Utilizzo di un aggiornamento dell'istanza con l'opzione Ignora la corrispondenza

Ignora la corrispondenza indica a Dimensionamento automatico Amazon EC2 di ignorare le istanze che dispongono già degli aggiornamenti più recenti, in modo da non sostituire più istanze del necessario. Ciò è utile quando desideri assicurarti che il gruppo con scalabilità automatica utilizzi una particolare versione del modello di avvio e sostituisca solo le istanze che ne utilizzano una diversa.

All'opzione Salta corrispondenza si applicano le seguenti considerazioni:

- Se avvii un aggiornamento dell'istanza con l'opzione Ignora la corrispondenza e specifichi una configurazione desiderata, Dimensionamento automatico Amazon EC2 verifica la corrispondenza di determinate istanze con la configurazione desiderata. Quindi, sostituisce solo le istanze che non corrispondono a tale configurazione. Al termine dell'aggiornamento dell'istanza, Dimensionamento automatico Amazon EC2 aggiorna il gruppo affinché corrisponda alla configurazione desiderata.

- Se avvii un aggiornamento dell'istanza con l'opzione Ignora la corrispondenza ma non specifichi una configurazione desiderata, Dimensionamento automatico Amazon EC2 verifica la corrispondenza di determinate istanze all'ultima configurazione salvata nel gruppo con scalabilità automatica. Quindi, sostituisce solo le istanze non corrispondenti.
- Puoi usare l'opzione Ignora la corrispondenza con un nuovo modello di avvio, una nuova versione del modello di avvio o un set di tipi di istanze. Se abiliti l'opzione Salta corrispondenza, ma nessuno di questi parametri viene modificato, l'aggiornamento dell'istanza avrà esito positivo senza sostituirla alcuna. Se hai apportato altre modifiche alla configurazione desiderata (ad esempio la modifica della strategia di allocazione Spot), Dimensionamento automatico Amazon EC2 attende che l'aggiornamento dell'istanza abbia esito positivo. Quindi, aggiorna le impostazioni del gruppo con scalabilità automatica in base alla nuova configurazione desiderata.
- Non puoi utilizzare l'opzione Salta corrispondenza con una nuova configurazione di avvio.
- Quando avvii un aggiornamento dell'istanza e fornisci la configurazione desiderata, Amazon EC2 Auto Scaling assicura che tutte le istanze utilizzino la configurazione desiderata. Pertanto, quando specifichi una `$Default` o `$Latest` come versione desiderata per il modello di lancio e poi crei una nuova versione del modello di lancio mentre è in corso un aggiornamento dell'istanza, tutte le istanze già sostituite verranno nuovamente sostituite.
- Skip matching non sa se uno script di dati utente nel modello di avvio recupererà il codice aggiornato e lo installerà su nuove istanze. Di conseguenza, skip matching potrebbe saltare la sostituzione delle istanze in cui è installato codice obsoleto. In questo caso, dovresti disattivare lo skip matching per assicurarti che tutte le istanze ricevano il codice più recente, anche senza un aggiornamento della versione del modello di lancio.

Questa sezione include AWS CLI le istruzioni per avviare un aggiornamento dell'istanza con skip matching abilitato. Per istruzioni su come usare la console, consulta [Avvio dell'aggiornamento di un'istanza \(console\)](#).

Ignora la corrispondenza (procedura di base)

Segui la procedura descritta in questa sezione per AWS CLI eseguire le seguenti operazioni:

- Creare il modello di avvio da applicare alle istanze.
- Avviare un aggiornamento dell'istanza per applicare il modello di avvio al gruppo con scalabilità automatica. Se non abiliti l'opzione Ignora la corrispondenza, tutte le istanze verranno sostituite. Ciò vale anche se il modello di avvio utilizzato per eseguire il provisioning dell'istanza è uguale a quello specificato per la configurazione desiderata.

Per utilizzare l'opzione Ignora la corrispondenza con un nuovo modello di avvio

1. Utilizzate il [create-launch-template](#) comando per creare un nuovo modello di avvio per il gruppo Auto Scaling. Includi l'opzione `--launch-template-data` e l'input JSON che definisce i dettagli delle istanze create per il gruppo con scalabilità automatica.

Ad esempio, utilizza il comando seguente per creare un modello di avvio di base con l'ID AMI `ami-0123456789abcdef0` e il tipo di istanza `t2.micro`.

```
aws ec2 create-launch-template --launch-template-name my-template-for-auto-scaling
--version-description version1 \
--launch-template-data
'{"ImageId": "ami-0123456789abcdef0", "InstanceType": "t2.micro"}'
```

Se il comando viene eseguito correttamente, verrà visualizzato un output simile al seguente:

```
{
  "LaunchTemplate": {
    "LaunchTemplateId": "lt-068f72b729example",
    "LaunchTemplateName": "my-template-for-auto-scaling",
    "CreatedBy": "arn:aws:iam::123456789012:user/Bob",
    "CreateTime": "2023-01-30T18:16:06.000Z",
    "DefaultVersionNumber": 1,
    "LatestVersionNumber": 1
  }
}
```

Per ulteriori informazioni, consulta [Esempi di creazione e gestione di modelli di avvio con \(\) AWS Command Line Interface AWS CLI](#).

2. Utilizzate il [start-instance-refresh](#) comando per avviare il flusso di lavoro di sostituzione delle istanze e applicate il nuovo modello di lancio con l'ID. `lt-068f72b729example` Poiché il modello di avvio è nuovo, ha una sola versione. Ciò significa che la versione 1 del modello di avvio è l'obiettivo di tale aggiornamento dell'istanza. Se durante l'aggiornamento dell'istanza si verifica un aumento orizzontale e Dimensionamento automatico Amazon EC2 effettua il provisioning di nuove istanze utilizzando la versione 1 di questo modello di avvio, le istanze non verranno sostituite. Al termine dell'operazione, il nuovo modello di avvio viene applicato correttamente al gruppo con scalabilità automatica.

```
aws autoscaling start-instance-refresh --cli-input-json file://config.json
```

Contenuto di config.json.

```
{
  "AutoScalingGroupName": "my-asg",
  "DesiredConfiguration": {
    "LaunchTemplate": {
      "LaunchTemplateId": "lt-068f72b729example",
      "Version": "$Default"
    }
  },
  "Preferences": {
    "SkipMatching": true
  }
}
```

Se il comando viene eseguito correttamente, verrà visualizzato un output simile al seguente:

```
{
  "InstanceRefreshId": "08b91cf7-8fa6-48af-b6a6-d227f40f1b9b"
}
```

Ignora la corrispondenza (gruppo di istanze miste)

Se disponi di un gruppo Auto Scaling con una [politica di istanze miste](#), segui i passaggi di questa sezione AWS CLI per avviare un aggiornamento dell'istanza con skip matching. Sono disponibili le seguenti opzioni:

- Fornire un nuovo modello di avvio da applicare a tutti i tipi di istanze specificati nella policy.
- Fornire un set aggiornato di tipi di istanze con o senza modifica del modello di avvio nella policy. Supponiamo, ad esempio, che tu voglia eseguire la migrazione da tipi di istanza indesiderati. In questo caso, dovresti utilizzare il modello di avvio così com'è, senza modificare l'AMI, i gruppi di sicurezza o altre specifiche delle istanze da sostituire.

Segui i passaggi indicati in una delle sezioni seguenti, a seconda dell'opzione più adatta alle tue esigenze.

Per utilizzare l'opzione Ignora la corrispondenza con un nuovo modello di avvio

1. Utilizzate il [create-launch-template](#) comando per creare un nuovo modello di avvio per il gruppo Auto Scaling. Includi l'opzione `--launch-template-data` e l'input JSON che definisce i dettagli delle istanze create per il gruppo con scalabilità automatica.

Ad esempio, utilizza il comando seguente per creare un modello di avvio con l'ID AMI *ami-0123456789abcdef0*.

```
aws ec2 create-launch-template --launch-template-name my-new-template --version-  
description version1 \  
--launch-template-data '{"ImageId": "ami-0123456789abcdef0"}'
```

Se il comando viene eseguito correttamente, verrà visualizzato un output simile al seguente:

```
{  
  "LaunchTemplate": {  
    "LaunchTemplateId": "lt-04d5cc9b88example",  
    "LaunchTemplateName": "my-new-template",  
    "CreatedBy": "arn:aws:iam::123456789012:user/Bob",  
    "CreateTime": "2023-01-31T15:56:02.000Z",  
    "DefaultVersionNumber": 1,  
    "LatestVersionNumber": 1  
  }  
}
```

Per ulteriori informazioni, consulta [Esempi di creazione e gestione di modelli di avvio con \(\) AWS Command Line Interface AWS CLI](#).

2. Per visualizzare la politica esistente per le istanze miste per il gruppo Auto Scaling, [describe-auto-scaling-groups](#) esegui il comando. Queste informazioni serviranno per avviare l'aggiornamento dell'istanza, come descritto nel passaggio successivo.

Il seguente comando di esempio restituisce la policy di istanze miste configurata per il gruppo con scalabilità automatica denominato *my-asg*.

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

Se il comando viene eseguito correttamente, verrà visualizzato un output simile al seguente:

```
{
  "AutoScalingGroups":[
    {
      "AutoScalingGroupName":"my-asg",
      "AutoScalingGroupARN":"arn",
      "MixedInstancesPolicy":{"
        "LaunchTemplate":{"
          "LaunchTemplateSpecification":{"
            "LaunchTemplateId":"lt-073693ed27example",
            "LaunchTemplateName":"my-old-template",
            "Version":"$Default"
          },
          "Overrides":[
            {
              "InstanceType":"c5.large"
            },
            {
              "InstanceType":"c5a.large"
            },
            {
              "InstanceType":"m5.large"
            },
            {
              "InstanceType":"m5a.large"
            }
          ]
        },
        "InstancesDistribution":{"
          "OnDemandAllocationStrategy":"prioritized",
          "OnDemandBaseCapacity":1,
          "OnDemandPercentageAboveBaseCapacity":50,
          "SpotAllocationStrategy":"price-capacity-optimized"
        }
      },
      "MinSize":1,
      "MaxSize":5,
      "DesiredCapacity":4,
      ...
    }
  ]
}
```

- Utilizzate il [start-instance-refresh](#) comando per avviare il flusso di lavoro di sostituzione delle istanze e applicate il nuovo modello di lancio con l'ID. *lt-04d5cc9b88example* Poiché il modello di avvio è nuovo, ha una sola versione. Ciò significa che la versione 1 del modello di avvio è l'obiettivo di tale aggiornamento dell'istanza. Se durante l'aggiornamento dell'istanza si verifica un aumento orizzontale e Dimensionamento automatico Amazon EC2 effettua il provisioning di nuove istanze utilizzando la versione 1 di questo modello di avvio, le istanze non verranno sostituite. Al termine dell'operazione, la policy di istanze miste aggiornata viene applicata correttamente al gruppo con scalabilità automatica.

```
aws autoscaling start-instance-refresh --cli-input-json file://config.json
```

Contenuto di config.json.

```
{
  "AutoScalingGroupName": "my-asg",
  "DesiredConfiguration": {
    "MixedInstancesPolicy": {
      "LaunchTemplate": {
        "LaunchTemplateSpecification": {
          "LaunchTemplateId": "lt-04d5cc9b88example",
          "Version": "$Default"
        },
        "Overrides": [
          {
            "InstanceType": "c5.large"
          },
          {
            "InstanceType": "c5a.large"
          },
          {
            "InstanceType": "m5.large"
          },
          {
            "InstanceType": "m5a.large"
          }
        ]
      }
    }
  },
  "InstancesDistribution": {
    "OnDemandAllocationStrategy": "prioritized",
    "OnDemandBaseCapacity": 1,
    "OnDemandPercentageAboveBaseCapacity": 50,
  }
}
```

```
        "SpotAllocationStrategy": "price-capacity-optimized"
      }
    }
  },
  "Preferences": {
    "SkipMatching": true
  }
}
```

Se il comando viene eseguito correttamente, verrà visualizzato un output simile al seguente:

```
{
  "InstanceRefreshId": "08b91cf7-8fa6-48af-b6a6-d227f40f1b9b"
}
```

Nella procedura seguente, fornirai un set aggiornato di tipi di istanze senza modificare il modello di avvio.

Per utilizzare l'opzione Ignora la corrispondenza con un set aggiornato di tipi di istanza

1. Per visualizzare la politica esistente per le istanze miste per il gruppo Auto Scaling, [describe-auto-scaling-groups](#) esegui il comando. Queste informazioni serviranno per avviare l'aggiornamento dell'istanza, come descritto nel passaggio successivo.

Il seguente comando di esempio restituisce la policy di istanze miste configurata per il gruppo con scalabilità automatica denominato *my-asg*.

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

Se il comando viene eseguito correttamente, verrà visualizzato un output simile al seguente:

```
{
  "AutoScalingGroups": [
    {
      "AutoScalingGroupName": "my-asg",
      "AutoScalingGroupARN": "arn",
      "MixedInstancesPolicy": {
        "LaunchTemplate": {
          "LaunchTemplateSpecification": {
```

```

    "LaunchTemplateId":"lt-073693ed27example",
    "LaunchTemplateName":"my-template-for-auto-scaling",
    "Version":"$Default"
  },
  "Overrides":[
    {
      "InstanceType":"c5.large"
    },
    {
      "InstanceType":"c5a.large"
    },
    {
      "InstanceType":"m5.large"
    },
    {
      "InstanceType":"m5a.large"
    }
  ]
},
"InstancesDistribution":{
  "OnDemandAllocationStrategy":"prioritized",
  "OnDemandBaseCapacity":1,
  "OnDemandPercentageAboveBaseCapacity":50,
  "SpotAllocationStrategy":"price-capacity-optimized"
}
},
"MinSize":1,
"MaxSize":5,
"DesiredCapacity":4,
...
}
]
}

```

- Utilizzate il [start-instance-refresh](#) comando per avviare il flusso di lavoro di sostituzione delle istanze e applicare gli aggiornamenti. Se desideri sostituire le istanze che utilizzano tipi di istanza precisi, la configurazione desiderata deve specificare la policy di istanze miste con solo i tipi di istanza desiderati. Puoi scegliere se aggiungere nuovi tipi di istanza al loro posto.

Il seguente comando di esempio avvia un aggiornamento dell'istanza senza il tipo di istanza indesiderato *m5a.large*. Quando un tipo di istanza nel gruppo non corrisponde a uno dei tre tipi di istanza rimanenti, le istanze vengono sostituite. Tieni presente che un aggiornamento dell'istanza non sceglie i tipi di istanza da cui eseguire il provisioning delle nuove istanze, al

contrario delle [strategie di allocazione](#). Al termine dell'operazione, la policy di istanze miste aggiornata viene applicata correttamente al gruppo con scalabilità automatica.

```
aws autoscaling start-instance-refresh --cli-input-json file://config.json
```

Contenuto di config.json

```
{
  "AutoScalingGroupName": "my-asg",
  "DesiredConfiguration": {
    "MixedInstancesPolicy": {
      "LaunchTemplate": {
        "LaunchTemplateSpecification": {
          "LaunchTemplateId": "lt-073693ed27example",
          "Version": "$Default"
        },
        "Overrides": [
          {
            "InstanceType": "c5.large"
          },
          {
            "InstanceType": "c5a.large"
          },
          {
            "InstanceType": "m5.large"
          }
        ]
      },
      "InstancesDistribution": {
        "OnDemandAllocationStrategy": "prioritized",
        "OnDemandBaseCapacity": 1,
        "OnDemandPercentageAboveBaseCapacity": 50,
        "SpotAllocationStrategy": "price-capacity-optimized"
      }
    }
  },
  "Preferences": {
    "SkipMatching": true
  }
}
```

Aggiungere punti di controllo a un aggiornamento delle istanze

Quando utilizzi un aggiornamento dell'istanza, puoi scegliere di sostituire le istanze in fasi, in modo da poter eseguire delle verifiche sulle istanze man mano che procedi. Per eseguire una sostituzione graduale, aggiungi dei punti di controllo, ovvero punti nel tempo in cui l'aggiornamento dell'istanza si interrompe. L'utilizzo dei punti di controllo consente di avere un maggiore controllo su come scegliere di aggiornare il gruppo con scalabilità automatica. Ti aiuta a verificare che l'applicazione funzioni in modo affidabile e prevedibile.

Indice

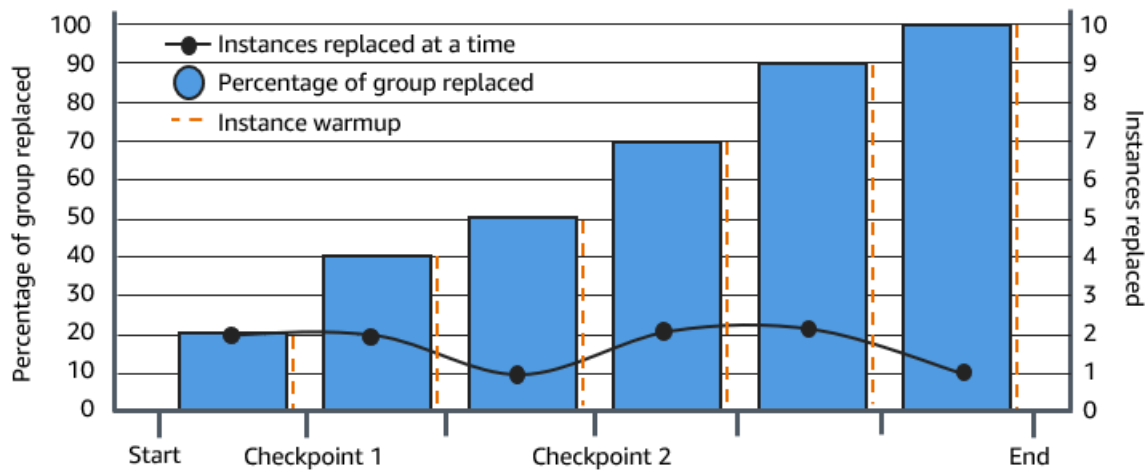
- [Come funziona](#)
- [Considerazioni](#)
- [Abilitazione dei punti di controllo \(console\)](#)
- [Abilitazione dei punti di controllo \(AWS CLI\)](#)

Come funziona

Quando si avvia un aggiornamento dell'istanza, si specificano i checkpoint come percentuali del numero totale di istanze nel gruppo Auto Scaling. Questi checkpoint indicano la percentuale minima di istanze nel gruppo Auto Scaling che devono essere nuove istanze prima che il checkpoint venga considerato raggiunto. Ad esempio, se i checkpoint sono nuovi[20, 50, 100], il primo checkpoint viene raggiunto quando il 20% delle istanze sono nuove, il secondo quando il 50% sono nuove e il checkpoint finale quando tutte le istanze sono nuove.

Amazon EC2 Auto Scaling accelera le sostituzioni delle istanze per rispettare le percentuali di checkpoint specificate, mantenendo al contempo la percentuale minima integra del gruppo. Per raggiungere una percentuale di checkpoint, Dimensionamento automatico Amazon EC2 a volte ne sostituisce una quantità inferiore, ma mai superiore a quella consentita dalla percentuale minima di integrità.

Considera il seguente gruppo con dimensionamento automatico che ha 10 istanze. Le percentuali del checkpoint sono [20, 50, 100]: la percentuale minima di integrità è dell'80% e la percentuale massima di integrità è del 100 percento. Per mantenere la percentuale minima di integrità, solo due istanze vengono sostituite contemporaneamente. Il diagramma seguente riepiloga le operazioni necessarie per la sostituzione delle istanze prima che venga raggiunto un punto di controllo.



Nell'esempio precedente, è previsto un periodo di riscaldamento dell'istanza per ogni nuova istanza che viene avviata. Potresti anche avere un hook del ciclo di vita che mette un'istanza in uno stato di attesa e poi esegue un'operazione personalizzata durante l'avvio o la chiusura.

Amazon EC2 Auto Scaling emette eventi per ogni checkpoint ad eccezione del checkpoint completo al 100%. Puoi aggiungere una EventBridge regola per inviare gli eventi a una destinazione come Amazon SNS. In questo modo, riceverai una notifica su quando puoi eseguire le verifiche richieste. Per ulteriori informazioni, consulta [Crea EventBridge regole, ad esempio eventi di aggiornamento..](#)

Considerazioni

Quando utilizzi i punti di controllo, tieni presente le considerazioni riportate di seguito:

- Poiché i punti di controllo sono basati su percentuali, il numero di istanze da sostituire cambia in base alla dimensione del gruppo. Quando si verifica un'attività di aumento orizzontale della dimensione del gruppo, un'operazione in corso potrebbe raggiungere nuovamente un punto di controllo. In tal caso, prima di continuare, Amazon EC2 Auto Scaling invia un'altra notifica e ripete il tempo di attesa tra i punti di controllo.
- In determinate circostanze, è possibile saltare un punto di controllo. Ad esempio, supponiamo che il gruppo con scalabilità automatica abbia due istanze e che le percentuali del punto di controllo siano [10, 40, 100]. Dopo la sostituzione della prima istanza, Amazon EC2 Auto Scaling calcola che il 50% del gruppo è stato sostituito. Poiché il 50% è superiore ai primi due punti di controllo, salta il primo (10) e invia una notifica per il secondo (40).
- L'annullamento dell'operazione interrompe eventuali ulteriori sostituzioni. Se l'operazione viene annullata o non riesce prima di raggiungere l'ultimo punto di controllo, le istanze già sostituite non vengono ripristinate alla configurazione precedente.

- In un aggiornamento parziale, quando riesegui l'operazione, Dimensionamento automatico Amazon EC2 non si riavvia dal punto dell'ultimo punto di controllo e si interrompe quando vengono sostituite solo le istanze precedenti. Tuttavia, punterà a sostituire le istanze precedenti prima di quelle nuove.
- La percentuale effettiva di completamento potrebbe essere superiore alla percentuale di quel checkpoint quando la percentuale del checkpoint è troppo bassa rispetto al numero di istanze del gruppo. Ad esempio, supponiamo che la percentuale del punto di controllo sia del 20% e che il gruppo abbia quattro istanze. Se Dimensionamento automatico Amazon EC2 sostituisce una delle quattro istanze, la percentuale effettiva sostituita (25 percento) sarà superiore alla percentuale del checkpoint (20 percento).
- Una volta raggiunto un checkpoint, la percentuale complessiva di completamento visualizzata viene aggiornata solo dopo il completamento del riscaldamento delle istanze. Ad esempio, le percentuali dei checkpoint corrispondono [20, 50] a un ritardo di 15 minuti e a una percentuale minima di salute dell'80%. Il gruppo Auto Scaling è composto da 10 istanze ed effettua le seguenti sostituzioni:
 - 0:00: due istanze precedenti vengono sostituite con istanze nuove.
 - 0:10: due nuove istanze terminano la preparazione.
 - 0:25: due istanze precedenti vengono sostituite con istanze nuove. (Per mantenere la percentuale minima di integrità, solo due istanze vengono sostituite).
 - 0:35: due nuove istanze terminano la preparazione.
 - 0:35: un'istanza precedente viene sostituita con una nuova.
 - 0:45: una nuova istanza termina la preparazione.

Alle 0:35, l'operazione interrompe l'avvio di nuove istanze. La percentuale di completamento non riflette in modo accurato il numero di sostituzioni completate (50%), perché la nuova istanza non ha terminato la preparazione. Dopo che la nuova istanza ha completato il periodo di riscaldamento alle 0:45, la percentuale di completamento mostra il 50%.

Abilitazione dei punti di controllo (console)

Puoi abilitare i punti di controllo prima di avviare l'aggiornamento delle istanze per sostituirle utilizzando un approccio incrementale o graduale. Ciò consente di avere tempo aggiuntivo per la verifica.

Come avviare un aggiornamento delle istanze che utilizza i punti di controllo

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina Auto Scaling groups (Gruppi Auto Scaling).

3. Nella scheda Aggiornamento istanza in Active instance refresh (Aggiornamento attivo istanza), scegli Start instance refresh (Avvia aggiornamento istanza).
4. Nella pagina Start instance refresh (Avvia aggiornamento istanza), inserisci i valori per Percentuale minima di integrità e Instance warmup (Preparazione istanza).
5. Seleziona la casella di controllo Enable checkpoints (Abilita punti di controllo).

Visualizzerai una casella in cui potrai definire la soglia percentuale per il primo punto di controllo.

6. Per Proceed until ____ % of the group is refreshed (Procedi fino al ____% dell'aggiornamento del gruppo), inserisci un numero da 1 a 100. Questa operazione imposta la percentuale per il primo punto di controllo.
7. Per aggiungere un altro punto di controllo, scegli Add checkpoint (Aggiungi punto di controllo) e quindi definisci la percentuale per il punto di controllo successivo.
8. Per specificare quanto tempo Amazon EC2 Auto Scaling attende dopo il raggiungimento di un punto di controllo, aggiorna i campi in Wait for **1 hour** between checkpoints (Attendi un'ora tra un punto di controllo e l'altro). L'unità di tempo può essere in ore, minuti o secondi.
9. Se hai effettuato tutte le selezioni per l'aggiornamento delle istanze, scegli Avvia aggiornamento dell'istanza

Abilitazione dei punti di controllo (AWS CLI)

Per avviare un aggiornamento dell'istanza con i checkpoint abilitati utilizzando il AWS CLI, è necessario un file di configurazione che definisca i seguenti parametri:

- `CheckpointPercentages`: specifica i valori di soglia per la percentuale di istanze da sostituire. Questi valori di soglia forniscono i punti di controllo. Quando la percentuale di istanze sostituite e preparate raggiunge una delle soglie specificate, l'operazione attende un periodo di tempo specificato. È possibile specificare il numero di secondi di attesa in `CheckpointDelay`. Una volta

trascorso il periodo di tempo specificato, l'aggiornamento delle istanze procede fino a raggiungere il punto di controllo successivo (se applicabile).

- `CheckpointDelay`: specifica il tempo di attesa, espresso in secondi, necessario al raggiungimento di un punto di controllo prima di continuare. Scegli un periodo di tempo sufficiente per eseguire le verifiche.

L'ultimo valore mostrato nell'array `CheckpointPercentages` descrive la percentuale del gruppo con scalabilità automatica che deve essere sostituito correttamente. Una volta che questa percentuale viene sostituita con successo e che ciascuna istanza ha completato la sua inizializzazione, l'operazione passa a `Successful`.

Come creare più punti di controllo

Per creare più checkpoint, utilizzate il seguente comando di esempio. [start-instance-refresh](#) In questo esempio viene configurato un aggiornamento delle istanze che inizialmente aggiorna l'1% del gruppo con scalabilità automatica. Dopo aver atteso 10 minuti, aggiorna il 19% successivo e attende altri 10 minuti. Infine, aggiorna il resto del gruppo prima di concludere l'operazione.

```
aws autoscaling start-instance-refresh --cli-input-json file://config.json
```

Contenuto di `config.json`.

```
{
  "AutoScalingGroupName": "my-asg",
  "Preferences": {
    "InstanceWarmup": 60,
    "MinHealthyPercentage": 80,
    "CheckpointPercentages": [1, 20, 100],
    "CheckpointDelay": 600
  }
}
```

Come creare un singolo punto di controllo

Per creare un singolo checkpoint, utilizzate il seguente comando di esempio [start-instance-refresh](#). In questo esempio viene configurato un aggiornamento delle istanze che, inizialmente, aggiorna il 20% del gruppo con scalabilità automatica. Dopo aver atteso 10 minuti, aggiorna il resto del gruppo prima di concludere l'operazione.

```
aws autoscaling start-instance-refresh --cli-input-json file://config.json
```

Contenuto di config.json.

```
{
  "AutoScalingGroupName": "my-asg",
  "Preferences": {
    "InstanceWarmup": 60,
    "MinHealthyPercentage": 80,
    "CheckpointPercentages": [20,100],
    "CheckpointDelay": 600
  }
}
```

Aggiornamento parziale del gruppo con scalabilità automatica

Per sostituire solo una parte del gruppo Auto Scaling e poi interromperlo completamente, utilizzate il seguente comando di esempio [start-instance-refresh](#). In questo esempio viene configurato un aggiornamento delle istanze che inizialmente aggiorna l'1% del gruppo con scalabilità automatica. Dopo aver atteso 10 minuti, aggiorna il successivo 19% prima di concludere l'operazione.

```
aws autoscaling start-instance-refresh --cli-input-json file://config.json
```

Contenuto di config.json.

```
{
  "AutoScalingGroupName": "my-asg",
  "Preferences": {
    "InstanceWarmup": 60,
    "MinHealthyPercentage": 80,
    "CheckpointPercentages": [1,20],
    "CheckpointDelay": 600
  }
}
```

Sostituire le istanze Auto Scaling basate sulla durata massima dell'istanza

La durata massima delle istanze specifica la quantità massima di tempo (in secondi) durante la quale un'istanza può restare in servizio prima che venga terminata e sostituita. Un caso d'uso comune può essere un requisito per sostituire le istanze in base a una pianificazione a causa di policy di sicurezza interne o controlli di conformità esterni.

È necessario specificare un valore di almeno 86.400 secondi (un giorno). Per cancellare un valore impostato in precedenza, specifica un nuovo valore pari a 0. Questa impostazione si applica a tutte le istanze attuali e future del gruppo con scalabilità automatica.

Indice

- [Considerazioni](#)
- [Impostazione della durata massima dell'istanza](#)
- [Limitazioni](#)

Considerazioni

Di seguito sono riportate le considerazioni relative all'utilizzo di questa funzionalità:

- Ogni volta che viene sostituita un'istanza precedente e viene avviata una nuova, quest'ultima utilizza il modello di avvio o la configurazione di avvio attualmente associata al gruppo con scalabilità automatica. Se il modello di avvio o la configurazione di avvio specifica l'ID Amazon Machine Image (AMI) di una versione diversa dell'applicazione, questa versione dell'applicazione verrà distribuita automaticamente.
- Se si imposta una durata massima dell'istanza su un valore troppo basso, è possibile che le istanze vengano sostituite più rapidamente del previsto. Amazon EC2 Auto Scaling di solito sostituisce le istanze una alla volta, con una pausa tra le sostituzioni. Tuttavia, se la durata massima specificata dell'istanza non fornisce il tempo sufficiente per sostituire ogni istanza singolarmente, Amazon EC2 Auto Scaling deve sostituire più di un'istanza alla volta. È possibile sostituire più istanze contemporaneamente, fino al 10% della capacità corrente del gruppo con scalabilità automatica. Per evitare di sostituire troppe istanze contemporaneamente, imposta una durata massima dell'istanza più lunga o utilizza la protezione scalabile dell'istanza per impedire temporaneamente la chiusura di singole istanze. Per ulteriori informazioni, consulta [Utilizzare la protezione di riduzione orizzontale dell'istanza](#).

- Per impostazione predefinita, Dimensionamento automatico Amazon EC2 crea una nuova attività di dimensionamento per terminare l'istanza, quindi la termina. Mentre l'istanza viene terminata, un'altra attività di dimensionamento avvia una nuova istanza. Puoi modificare questo comportamento in modo che venga avviato prima di terminare utilizzando una policy di manutenzione dell'istanza. Per ulteriori informazioni, consulta [Policy di manutenzione delle istanze](#).

Impostazione della durata massima dell'istanza

Quando si crea un gruppo con scalabilità automatica nella console, non è possibile impostare la durata massima dell'istanza. Tuttavia, dopo aver creato il gruppo, è possibile modificarlo per impostare la durata massima dell'istanza.

Impostazione della durata massima dell'istanza per un gruppo (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Visualizzerai un riquadro diviso nella parte inferiore della pagina Auto Scaling groups (Gruppi Auto Scaling) in cui vengono mostrate le informazioni sul gruppo selezionato.

3. Nella scheda Details (Dettagli) scegliere Advanced configurations (Configurazioni avanzate), Edit (Modifica).
4. Per Maximum instance lifetime (Durata massima dell'istanza), inserisci il numero massimo di secondi in cui un'istanza può essere in servizio.
5. Scegli Aggiorna.

Nella scheda Activity (Attività) in Activity history (Cronologia attività), è possibile visualizzare tutta la cronologia della sostituzione delle istanze per il gruppo.

Impostazione della durata massima dell'istanza per un gruppo (AWS CLI)

È inoltre possibile utilizzare AWS CLI per impostare la durata massima dell'istanza per i gruppi di Auto Scaling nuovi o esistenti.

Per i nuovi gruppi Auto Scaling, utilizzare il [create-auto-scaling-group](#) comando.

```
aws autoscaling create-auto-scaling-group --cli-input-json file://~/config.json
```

Di seguito è riportato un file `config.json` di esempio che mostra una durata massima dell'istanza di 2592000 secondi (30 giorni).

```
{
  "AutoScalingGroupName": "my-asg",
  "LaunchTemplate": {
    "LaunchTemplateName": "my-launch-template",
    "Version": "$Default"
  },
  "MinSize": 1,
  "MaxSize": 5,
  "MaxInstanceLifetime": 2592000,
  "VPCZoneIdentifier": "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782",
  "Tags": []
}
```

Per i gruppi Auto Scaling esistenti, utilizzare il [update-auto-scaling-group](#) comando.

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-existing-asg --
max-instance-lifetime 2592000
```

Come verificare la durata massima dell'istanza per un gruppo con scalabilità automatica

Utilizza il comando [describe-auto-scaling-groups](#).

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

Limitazioni

- Esattezza della durata massima non garantita per ogni istanza: non è garantita la sostituzione delle istanze alla fine della loro durata massima. In alcune situazioni, Amazon EC2 Auto Scaling potrebbe richiedere la sostituzione delle istanze dopo l'aggiornamento del parametro di durata massima dell'istanza. La ragione di questo comportamento è quella di evitare di sostituire tutte le istanze allo stesso tempo.
- Protezione scalabile in base alle istanze rispettata: Amazon EC2 Auto Scaling offre una protezione scalabile in base alle istanze per aiutarti a controllare quali istanze possono essere chiuse. Quando questa protezione è abilitata su un'istanza, Amazon EC2 Auto Scaling non interromperà l'istanza anche se ha raggiunto la durata massima dell'istanza.

- **Istanze terminate prima dell'avvio:** quando nel gruppo con dimensionamento automatico è presente una sola istanza, la caratteristica di durata massima dell'istanza può causare un'interruzione poiché Dimensionamento automatico Amazon EC2 termina un'istanza e quindi ne avvia una nuova come impostazione predefinita. Per modificare questo comportamento in modo che venga avviato prima della terminazione, consulta [Policy di manutenzione delle istanze](#).

Dimensionare le dimensioni del gruppo con scalabilità automatica

Il dimensionamento è la possibilità di aumentare o diminuire la capacità di elaborazione dell'applicazione. Il dimensionamento inizia con un evento, o operazione di dimensionamento, che indica a un gruppo con scalabilità automatica di avviare o terminare istanze Amazon EC2.

Dimensionamento automatico Amazon EC2 offre vari modi di regolare il dimensionamento per soddisfare al meglio le esigenze delle applicazioni. Di conseguenza, è importante conoscere bene l'applicazione in uso. Tieni a mente le seguenti considerazioni:

- Quale ruolo deve svolgere Dimensionamento automatico Amazon EC2 nell'architettura dell'applicazione? È comune pensare alla scalabilità automatica principalmente come a un modo di aumentare e diminuire la capacità, ma è utile anche per mantenere un numero costante di server.
- Quali sono i limiti di costo importanti? Poiché Amazon EC2 Auto Scaling utilizza istanze EC2, paghi solo per le risorse che utilizzi. Conoscere i limiti di costo ti aiuta a decidere quando dimensionare le applicazioni e in quale misura.
- Quali sono i parametri importanti per l'applicazione? Amazon CloudWatch supporta una serie di parametri diversi che puoi utilizzare con il tuo gruppo Auto Scaling.

Indice

- [Scegli il tuo metodo di dimensionamento](#)
- [Impostare i limiti di dimensionamento per il gruppo con dimensionamento automatico](#)
- [Impostazione della preparazione di default dell'istanza per un gruppo con scalabilità automatica](#)
- [Dimensionamento manuale per Dimensionamento automatico Amazon EC2](#)
- [Dimensionamento programmato per Dimensionamento automatico Amazon EC2](#)
- [Dimensionamento dinamico per Dimensionamento automatico Amazon EC2](#)
- [Dimensionamento predittivo per Dimensionamento automatico Amazon EC2](#)
- [Verificare quali istanze Auto Scaling vengono terminate durante la riduzione orizzontale](#)
- [Sospendi e riprendi i processi di Amazon EC2 Auto Scaling](#)

Scegli il tuo metodo di dimensionamento

Dimensionamento automatico Amazon EC2 offre vari modi per dimensionare il gruppo con scalabilità automatica.

Mantenere un numero fisso di istanze

L'impostazione predefinita per un gruppo con dimensionamento automatico è di non avere policy di ridimensionamento allegate o azioni pianificate, il che fa sì che mantenga una dimensione fissa. Dopo aver creato il gruppo con dimensionamento automatico, questo comincia avviando un numero di istanze sufficiente a soddisfare la capacità desiderata. Se non ci sono condizioni di dimensionamento collegate al gruppo, quest'ultimo continua a mantenere la capacità desiderata anche se un'istanza perde integrità. Dimensionamento automatico Amazon EC2 monitora l'integrità di ciascuna istanza nel gruppo con dimensionamento automatico. Quando scopre che un'istanza è diventata non integra, la sostituisce con una nuova istanza. Puoi leggere una descrizione più approfondita di questo processo in [Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico](#).

Dimensionamento manuale

Il dimensionamento manuale è il metodo più semplice per dimensionare il gruppo con dimensionamento automatico. È possibile aggiornare la capacità desiderata del gruppo Auto Scaling o terminare le istanze nel gruppo Auto Scaling. Per ulteriori informazioni, consulta [Dimensionamento manuale per Dimensionamento automatico Amazon EC2](#).

Dimensionamento in base a una pianificazione

Il ridimensionamento in base alla pianificazione significa che le azioni di ridimensionamento vengono eseguite automaticamente in funzione della data e dell'ora. Questo è utile quando si sa esattamente quando aumentare o diminuire il numero di istanze nel gruppo, semplicemente perché la necessità deriva da una pianificazione prevedibile. Per ulteriori informazioni, consulta [Dimensionamento programmato per Dimensionamento automatico Amazon EC2](#).

Scalabilità dinamica in base alla domanda

Un modo più avanzato di dimensionare le risorse, tramite il dimensionamento dinamico, consente di definire una policy di dimensionamento che ridimensiona dinamicamente il gruppo con scalabilità automatica per soddisfare le variazioni della domanda. Ad esempio, immaginiamo di avere un'applicazione Web attualmente eseguita su due istanze e di volere che l'utilizzo della CPU del gruppo con scalabilità automatica rimanga intorno al 50% quando il carico sull'applicazione varia. Questo metodo è utile per scalare man mano che si verificano cambiamenti di traffico, quando

non sai quando cambierà il traffico. È possibile configurare le policy di dimensionamento in modo che rispondano automaticamente. Esistono diversi tipi di policy (o una combinazione di essi) che puoi utilizzare per scalare in risposta alle variazioni del traffico. Per ulteriori informazioni, consulta [Dimensionamento dinamico per Dimensionamento automatico Amazon EC2](#).

Scala in modo proattivo

Per dimensionare la capacità EC2 più rapidamente, puoi anche combinare dimensionamento predittivo e dinamico (rispettivamente, approcci proattivi e reattivi). Utilizza il dimensionamento predittivo per aumentare il numero di istanze EC2 nel gruppo con scalabilità automatica in anticipo rispetto ai modelli giornalieri e settimanali nei flussi di traffico. Per ulteriori informazioni, consulta [Dimensionamento predittivo per Dimensionamento automatico Amazon EC2](#).

Impostare i limiti di dimensionamento per il gruppo con dimensionamento automatico

I limiti di dimensionamento rappresentano la dimensione minima e massima del gruppo desiderata per il tuo gruppo con dimensionamento automatico. I limiti per le dimensioni minima e massima vengono impostati separatamente.

La capacità desiderata per il gruppo può essere dimensionata in base ai limiti della dimensione minima e massima. La capacità desiderata deve essere uguale o superiore alla dimensione minima del gruppo e uguale o inferiore alla dimensione massima del gruppo.

- **Capacità desiderata:** rappresenta la capacità iniziale del gruppo con scalabilità automatica nel momento in cui viene creato. Un gruppo con scalabilità automatica tenta di mantenere la capacità desiderata. Inizia avviando il numero di istanze specificato per la capacità desiderata e manterrà questo numero di istanze finché al gruppo con scalabilità automatica non sono associate policy di dimensionamento o operazioni pianificate.
- **Capacità minima:** rappresenta la dimensione minima del gruppo. Quando vengono impostate le policy di dimensionamento, esse non possono ridurre la capacità desiderata del gruppo al di sotto del limite di capacità minima.
- **Capacità massima:** rappresenta la dimensione massima del gruppo. Quando vengono impostate le policy di dimensionamento, esse non possono aumentare la capacità desiderata oltre il limite di capacità massima.

I limiti di dimensione minima e massima si applicano anche nei seguenti scenari:

- Quando si dimensiona manualmente il gruppo con scalabilità automatica aggiornando la capacità desiderata.
- Quando vengono eseguite azioni pianificate che aggiornano la capacità desiderata. Se un'operazione pianificata viene eseguita senza specificare nuovi limiti di dimensione minima e massima per il gruppo, si applicano i limiti di dimensione minima e massima correnti del gruppo.

Un gruppo con scalabilità automatica tenta sempre di mantenere la capacità desiderata. Nei casi in cui un'istanza termina in modo imprevisto (ad esempio, a causa di un'interruzione di un'istanza spot, di un errore di controllo dell'integrità o di un'operazione umana), il gruppo avvia automaticamente una nuova istanza per mantenere la capacità desiderata.

Gestione di queste impostazioni nella console

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nel pannello di navigazione selezionare Auto Scaling, Gruppi Auto Scaling.
3. Dalla pagina Gruppo con dimensionamento automatico, seleziona la casella di controllo accanto al gruppo con dimensionamento automatico..

Si aprirà un riquadro diviso nella parte inferiore della pagina.

4. Nella scheda Dettagli, visualizza o modifica le impostazioni correnti per la capacità minima, massima e desiderata del gruppo. Per ulteriori informazioni, consulta [Modifica della capacità desiderata di un gruppo con dimensionamento automatico esistente](#).

Sopra il riquadro Dettagli, si trovano informazioni quali il numero attuale di istanze nel gruppo con dimensionamento automatico, la capacità minima e massima desiderata e una colonna di stato. Se il gruppo Auto Scaling utilizza i pesi delle istanze, puoi anche trovare il numero di unità di capacità che hanno contribuito alla capacità desiderata.

Per aggiungere o rimuovere colonne dall'elenco, scegli l'icona delle impostazioni nella parte superiore della pagina. Quindi, per Auto Scaling groups attributes (Attributi dei gruppi con scalabilità automatica), attiva o disattiva ciascuna colonna e in seguito scegli Confirm (Conferma).

Come verificare le dimensioni del gruppo con scalabilità automatica dopo aver apportato le modifiche

La colonna Instances (Istanze) mostra il numero di istanze attualmente in esecuzione. Durante l'avvio o la terminazione di un'istanza, la colonna Status (Stato) mostra lo stato Updating capacity (Aggiornamento della capacità in corso), come da immagine seguente.

<input checked="" type="checkbox"/>	Name	Launch template...	Instances	Status	Desired...	Min	Max
<input checked="" type="checkbox"/>	my-asg	my_template Version Def	0	Updating capacity	1	0	1

Attendi alcuni minuti e poi aggiorna la pagina per visualizzare lo stato più recente. Al termine di un'attività di dimensionamento, la colonna Instances (Istanze) mostra un valore aggiornato.

Puoi visualizzare il numero di istanze e lo stato delle istanze in esecuzione dalla scheda Instance management (Gestione istanze) sotto Instances (Istanze).

Impostazione della preparazione di default dell'istanza per un gruppo con scalabilità automatica

CloudWatch raccoglie e aggrega i dati di utilizzo, come CPU e I/O di rete, tra le istanze di Auto Scaling. Questi parametri vengono utilizzati per creare policy di dimensionamento che regolano il numero di istanze nel gruppo con scalabilità automatica man mano che il valore del parametro selezionato aumenta e diminuisce.

Puoi specificare quanto tempo dopo che un'istanza raggiunge InService lo stato in cui deve attendere prima di aggiungere i dati di utilizzo alle metriche aggregate. Questo periodo di tempo specificato viene chiamato riscaldamento predefinito dell'istanza. In questo modo si evita che la scalabilità dinamica venga influenzata dalle metriche relative a singole istanze che non gestiscono ancora il traffico delle applicazioni e che potrebbero registrare un utilizzo temporaneamente elevato delle risorse di elaborazione.

Per ottimizzare le prestazioni delle politiche di tracciamento degli obiettivi e di scalabilità dei passaggi, ti consigliamo vivamente di abilitare e configurare il riscaldamento predefinito delle istanze. Non è abilitato o configurato per impostazione predefinita.

Quando abiliti il riscaldamento predefinito dell'istanza, tieni presente che se il gruppo Auto Scaling è impostato per utilizzare una politica di manutenzione delle istanze o utilizzi un aggiornamento dell'istanza per sostituire le istanze, puoi impedire che le istanze vengano conteggiate nella percentuale minima di integrità prima che abbiano terminato l'inizializzazione.

Indice

- [Considerazioni sulle prestazioni di dimensionamento](#)
- [Scegli il tempo di riscaldamento predefinito dell'istanza](#)
- [Abilitazione della preparazione di default dell'istanza per un gruppo](#)
- [Verificare il riscaldamento predefinito dell'istanza per un gruppo](#)
- [Trova le politiche di scalabilità con un tempo di riscaldamento dell'istanza impostato in precedenza](#)
- [Cancella la preparazione dell'istanza precedentemente impostata per una policy di dimensionamento](#)

Considerazioni sulle prestazioni di dimensionamento

Per la maggior parte delle applicazioni è utile disporre di un tempo di riscaldamento dell'istanza predefinito che si applica a tutte le funzionalità, anziché tempi di riscaldamento diversi per funzionalità diverse. Ad esempio, se non si imposta un riscaldamento predefinito dell'istanza, la funzionalità di aggiornamento dell'istanza utilizza il periodo di tolleranza del controllo dello stato di salute come orario di riscaldamento predefinito. Se disponete di politiche di tracciamento degli obiettivi e di scalatura dei passaggi, queste utilizzano il valore impostato per il cooldown predefinito come tempo di riscaldamento predefinito. Se hai delle politiche di scalabilità predittiva, non hanno un tempo di riscaldamento predefinito.

Durante il riscaldamento delle istanze, le politiche di scalabilità dinamica vengono ridimensionate solo se il valore della metrica delle istanze che non si stanno riscaldando è superiore alla soglia massima di allarme della policy (o all'utilizzo previsto di una politica di scalabilità di tracciamento del target). Se la domanda diminuisce, la scalabilità dinamica diventa più conservativa per proteggere la disponibilità dell'applicazione. Ciò blocca le attività di scalabilità per la scalabilità dinamica fino al termine del riscaldamento delle nuove istanze.

Durante la scalabilità orizzontale, Amazon EC2 Auto Scaling considera le istanze in fase di riscaldamento come parte della capacità del gruppo quando decide quante istanze aggiungere al gruppo. Pertanto, più violazioni degli allarmi che richiedono l'aggiunta di una quantità simile di capacità si traducono in un'unica attività di scalabilità. L'intenzione è quella di ampliare continuamente, senza farlo eccessivamente.

Se il riscaldamento predefinito dell'istanza non è abilitato, il tempo di attesa di un'istanza prima di inviare le metriche CloudWatch e contarle ai fini della capacità corrente varierà da istanza a istanza. Pertanto, esiste la possibilità che le politiche di scalabilità funzionino in modo imprevedibile rispetto al carico di lavoro effettivo che si sta verificando.

Ad esempio, si consideri un'applicazione con uno schema di carico di lavoro ricorrente. on-and-off Una policy di dimensionamento predittivo viene utilizzata per prendere decisioni ricorrenti sull'opportunità di aumentare il numero di istanze. Poiché non è previsto un tempo di riscaldamento predefinito per le politiche di scalabilità predittiva, le istanze iniziano a contribuire immediatamente alle metriche aggregate. Se queste istanze hanno un utilizzo maggiore delle risorse all'avvio, l'aggiunta di istanze potrebbe causare un aumento delle metriche aggregate. A seconda del tempo necessario alla stabilizzazione dell'utilizzo, ciò potrebbe influire sulle policy di dimensionamento dinamico che utilizzano queste metriche. Se viene superata la soglia massima di allarme di una policy di dimensionamento dinamico, le dimensioni del gruppo aumentano nuovamente. Mentre le nuove istanze si stanno preparando, le attività su larga scala verranno bloccate.

Scegli il tempo di riscaldamento predefinito dell'istanza

La chiave per impostare la preparazione predefinita delle istanze è determinare il tempo necessario per completare l'inizializzazione delle istanze e per stabilizzare il consumo di risorse una volta raggiunto lo stato `InService`. Quando scegli il tempo di riscaldamento dell'istanza, cerca di mantenere un equilibrio ottimale tra la raccolta dei dati di utilizzo per il traffico legittimo e la riduzione al minimo della raccolta dei dati associata ai picchi di utilizzo temporanei all'avvio.

Supponiamo che tu abbia un gruppo con dimensionamento automatico collegato a un sistema di bilanciamento del carico Elastic Load Balancing. Una volta terminato l'avvio delle nuove istanze, esse vengono registrate nel sistema di bilanciamento del carico prima di entrare nello stato `InService`. Dopo che le istanze passano allo stato `InService`, il consumo di risorse può ancora subire picchi temporanei e richiedere tempo per stabilizzarsi. Ad esempio, il consumo di risorse per un server applicazioni che deve scaricare e memorizzare nella cache risorse di grandi dimensioni richiede più tempo per stabilizzarsi rispetto a un server Web leggero senza risorse di grandi dimensioni da scaricare. La preparazione dell'istanza fornisce il ritardo necessario per stabilizzare il consumo di risorse.

Important

Se non sei sicuro del tempo necessario per il riscaldamento, puoi iniziare con 300 secondi. Quindi diminuiscilo o aumentalo gradualmente fino a ottenere le migliori prestazioni di scalabilità per la tua applicazione. Potrebbe essere necessario eseguire questa operazione alcune volte per eseguire correttamente l'operazione. In alternativa, se disponi di politiche di scalabilità che hanno il proprio tempo di riscaldamento (`EstimatedInstanceWarmup`), puoi

utilizzare questo valore per iniziare. Per ulteriori informazioni, consulta [Trova le politiche di scalabilità con un tempo di riscaldamento dell'istanza impostato in precedenza](#).

Utilizza gli hook del ciclo di vita per i casi d'uso in cui sono presenti attività di configurazione o script da eseguire all'avvio. Gli hook del ciclo di vita possono ritardare la messa in servizio delle nuove istanze finché non hanno finito di inicializzarsi. Sono particolarmente utili se si dispone di script di bootstrap che richiedono tempo per essere completati. Se aggiungi un hook del ciclo di vita, puoi ridurre il valore della preparazione di default dell'istanza. Per ulteriori informazioni sull'utilizzo di Hook del ciclo di vita, consulta [Hook del ciclo di vita di Amazon EC2 Auto Scaling](#).

Abilitazione della preparazione di default dell'istanza per un gruppo

È possibile abilitare la preparazione di default dell'istanza durante la creazione di un gruppo con scalabilità automatica. È possibile abilitarla anche per i gruppi esistenti.

Abilitando la funzione di riscaldamento dell'istanza predefinita, non è più necessario specificare i valori per i parametri di riscaldamento per le seguenti funzionalità:

- [Aggiornamento istanza](#)
- [Dimensionamento con monitoraggio degli obiettivi](#)
- [Dimensionamento per fasi](#)

Console

Abilita il riscaldamento predefinito dell'istanza per un nuovo gruppo (console)

Quando crei il gruppo con scalabilità automatica, nella pagina Configurare le opzioni avanzate sotto Impostazioni aggiuntive, seleziona l'opzione Abilita il riscaldamento predefinito dell'istanza. Scegliete il tempo di riscaldamento necessario per la vostra applicazione.

AWS CLI

Abilita il riscaldamento predefinito dell'istanza per un nuovo gruppo (AWS CLI)

Per abilitare la preparazione di default dell'istanza per un gruppo con scalabilità automatica, aggiungi l'opzione `--default-instance-warmup` e specifica un valore da 0 a 3600, in secondi. Dopo averlo abilitato, un valore di `-1` disattiverà questa impostazione.

Il `create-auto-scaling-group` comando seguente crea un gruppo Auto Scaling con il nome `my-asg` e abilita il riscaldamento predefinito dell'istanza con un valore di 120 secondi.

```
aws autoscaling create-auto-scaling-group --auto-scaling-group-name my-asg --
default-instance-warmup 120 ...
```

 Tip

Se questo comando genera un errore, assicurati di averlo aggiornato localmente alla AWS CLI versione più recente.

Console

Abilita il riscaldamento predefinito dell'istanza per un gruppo esistente (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Sulla barra di navigazione nella parte superiore della schermata, seleziona la Regione AWS in cui è stato creato il gruppo con dimensionamento automatico.
3. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

4. Nella scheda Details (Dettagli) scegliere Advanced configurations (Configurazioni avanzate), Edit (Modifica).
5. Per Default Instance Warmup, scegliete il tempo di riscaldamento necessario per l'applicazione.
6. Scegli Aggiorna.

AWS CLI

Abilita il riscaldamento predefinito dell'istanza per un gruppo esistente (AWS CLI)

L'esempio seguente utilizza il `update-auto-scaling-group` comando per abilitare il riscaldamento predefinito dell'istanza con un valore di 120 secondi per un gruppo Auto Scaling esistente denominato `my-asg`.

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg --  
default-instance-warmup 120
```

i Tip

Se questo comando genera un errore, assicurati di averlo aggiornato localmente alla AWS CLI versione più recente.

Verificare il riscaldamento predefinito dell'istanza per un gruppo

Per verificare il riscaldamento predefinito dell'istanza per un gruppo con scalabilità automatica (AWS CLI)

Utilizza il seguente comando [describe-auto-scaling-groups](#). Sostituisci *my-asg* con il nome del gruppo con dimensionamento automatico.

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

Di seguito è riportata una risposta di esempio.

```
{  
  "AutoScalingGroups": [  
    {  
      "AutoScalingGroupName": "my-asg",  
      "AutoScalingGroupARN": "arn",  
      ...  
      "DefaultInstanceWarmup": 120  
    }  
  ]  
}
```

Trova le politiche di scalabilità con un tempo di riscaldamento dell'istanza impostato in precedenza

Per identificare se disponi di politiche che hanno un proprio tempo di riscaldamento `EstimatedInstanceWarmup`, esegui il seguente comando [describe-policies](#) utilizzando il. AWS CLI Sostituisci *my-asg* con il nome del gruppo con dimensionamento automatico.

```
aws autoscaling describe-policies --auto-scaling-group-name my-asg
--query 'ScalingPolicies[?EstimatedInstanceWarmup!=`null`]'
```

Di seguito è riportato un output di esempio.

```
[
  {
    "AutoScalingGroupName":"my-asg",
    "PolicyName":"cpu50-target-tracking-scaling-policy",
    "PolicyARN":"arn",
    "PolicyType":"TargetTrackingScaling",
    "StepAdjustments":[],
    "EstimatedInstanceWarmup":120,
    "Alarms":[{"
      "AlarmARN": "arn:aws:cloudwatch:us-west-2:123456789012:alarm:TargetTracking-my-
asg-AlarmHigh-fc0e4183-23ac-497e-9992-691c9980c38e",
      "AlarmName": "TargetTracking-my-asg-AlarmHigh-
fc0e4183-23ac-497e-9992-691c9980c38e"
    },
    {
      "AlarmARN": "arn:aws:cloudwatch:us-
west-2:123456789012:alarm:TargetTracking-my-asg-AlarmLow-61a39305-ed0c-47af-
bd9e-471a352ee1a2",
      "AlarmName": "TargetTracking-my-asg-AlarmLow-61a39305-ed0c-47af-
bd9e-471a352ee1a2"
    }
  ],
  "TargetTrackingConfiguration":{
    "PredefinedMetricSpecification":{
      "PredefinedMetricType":"ASGAverageCPUUtilization"
    },
    "TargetValue":50.0,
    "DisableScaleIn":false
  },
  "Enabled":true
},
  ... additional policies ...
]
```

Cancella la preparazione dell'istanza precedentemente impostata per una policy di dimensionamento

Dopo aver abilitato il riscaldamento predefinito dell'istanza, aggiorna tutte le politiche di ridimensionamento che hanno ancora il proprio tempo di riscaldamento per cancellare il valore impostato in precedenza. Altrimenti, sostituirà la preparazione predefinita dell'istanza.

Puoi aggiornare le politiche di scalabilità utilizzando la console o gli SDK. AWS CLI AWS Questa sezione descrive i passaggi per la console. Se utilizzi gli AWS SDK AWS CLI o, assicurati di mantenere la configurazione delle policy esistente, ma rimuovi la proprietà. `EstimatedInstanceWarmup` Quando aggiorni una politica di scalabilità esistente, la politica verrà sostituita da quella specificata durante la chiamata a livello di codice. [PutScalingPolicy](#) I valori originali non vengono conservati.

Come cancellare la preparazione dell'istanza precedentemente impostata per una policy di dimensionamento (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Nella scheda Dimensionamento automatico, in Policy di dimensionamento dinamico, scegli la policy che ti interessa, quindi scegli Azioni, Modifica.
4. Ad esempio, riscaldamento dell'istanza, cancella il valore di riscaldamento dell'istanza per utilizzare invece il valore di riscaldamento dell'istanza predefinito.
5. Scegli Aggiorna.

Dimensionamento manuale per Dimensionamento automatico Amazon EC2

Puoi regolare manualmente il numero di istanze EC2 nel tuo gruppo Auto Scaling in qualsiasi momento. Questo processo di modifica manuale del conteggio delle istanze viene definito ridimensionamento manuale. Il ridimensionamento manuale è un'alternativa al ridimensionamento automatico, soprattutto se si desidera apportare modifiche alla capacità una tantum.

Dopo aver ridimensionato manualmente il gruppo, Amazon EC2 Auto Scaling riprende le normali attività di auto scaling in base alle politiche di scalabilità e alle azioni pianificate che hai definito. Per i gruppi con il riscaldamento predefinito delle istanze abilitato, tutte le nuove istanze passano attraverso un periodo di riscaldamento prima di iniziare a contribuire alle metriche utilizzate per la scalabilità automatica. Questo periodo di riscaldamento aiuta a stabilizzare il gruppo alla nuova capacità. Per ulteriori informazioni, consulta [Impostazione della preparazione di default dell'istanza per un gruppo con scalabilità automatica](#).

A volte, potresti voler disabilitare temporaneamente le politiche di ridimensionamento e le azioni pianificate prima di ridimensionare manualmente un gruppo. In questo modo si evitano conflitti tra le azioni di ridimensionamento manuali e le attività di ridimensionamento automatizzate. Per ulteriori informazioni, consulta [Disattiva le attività di scalabilità](#).

Indice

- [Modifica della capacità desiderata di un gruppo con dimensionamento automatico esistente](#)
- [Terminare un'istanza nel gruppo con dimensionamento automatico \(AWS CLI\)](#)

Modifica della capacità desiderata di un gruppo con dimensionamento automatico esistente

Quando modifichi la capacità desiderata del tuo gruppo Auto Scaling, Amazon EC2 Auto Scaling gestisce il processo di avvio e chiusura delle istanze per raggiungere la nuova dimensione desiderata.

Console

Come modificare le dimensioni del gruppo con scalabilità automatica

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Nella parte inferiore della pagina viene visualizzato un riquadro diviso.

3. Nella scheda Details (Dettagli) scegliere Group details (Dettagli gruppo), Edit (Modifica).
4. Per Capacità desiderata, aumentare o diminuire la capacità desiderata. Ad esempio, per aumentare di uno la dimensione del gruppo, se il valore corrente è 1, immettete 2.

Se il nuovo valore della capacità desiderata è maggiore della capacità minima desiderata e della capacità massima desiderata, la capacità massima desiderata viene automaticamente aumentata fino al nuovo valore di capacità desiderato.

5. Al termine, scegli Update (Aggiorna).

Verificate che la dimensione del gruppo specificata abbia comportato l'avvio della stessa quantità di istanze. Ad esempio, se hai aumentato la dimensione del gruppo di uno, verifica che il gruppo Auto Scaling abbia lanciato un'istanza aggiuntiva.

Come verificare che le dimensioni del gruppo con scalabilità automatica siano cambiate

1. Nella scheda Attività, nella Cronologia delle attività, è possibile visualizzare lo stato di avanzamento delle attività associate al gruppo Auto Scaling. La colonna Status (Stato) mostra lo stato corrente della tua istanza. Durante l'avvio dell'istanza, la colonna di stato mostra `Not yet in service`. Una volta avviata l'istanza, lo stato diventa `Successful`. Puoi anche utilizzare l'icona di aggiornamento per visualizzare lo stato corrente dell'istanza. Per ulteriori informazioni, consulta [Verificare un'attività di dimensionamento per un gruppo con scalabilità automatica](#).
2. Nella scheda Gestione dell'istanza, in Istanze, puoi visualizzare lo stato dell'istanza. L'avvio di un'istanza richiede pochi minuti.
 - La colonna Lifecycle (Ciclo di vita) mostra lo stato dell'istanza. Inizialmente, l'istanza è in stato `Pending`. Lo stato di un'istanza pronta a ricevere traffico è `InService`.
 - La colonna Health status mostra il risultato dei controlli di integrità di Amazon EC2 Auto Scaling sulla tua istanza.

AWS CLI

L'esempio seguente presuppone la creazione di un gruppo con scalabilità automatica con una dimensione minima di 1 e massima di 5. Pertanto, attualmente il gruppo ha un'istanza in esecuzione.

Come modificare le dimensioni del gruppo con scalabilità automatica

Utilizzate il [set-desired-capacity](#) comando per modificare le dimensioni del gruppo Auto Scaling, come illustrato nell'esempio seguente.

```
aws autoscaling set-desired-capacity --auto-scaling-group-name my-asg \  
  --desired-capacity 2
```

Se scegli di rispettare il tempo di raffreddamento di default per il gruppo con scalabilità automatica, dovrai specificare l'opzione `--honor-cooldown` come mostrato nel seguente esempio. Per ulteriori informazioni, consulta [Tempo di raffreddamento per il dimensionamento per Dimensionamento automatico Amazon EC2](#).

```
aws autoscaling set-desired-capacity --auto-scaling-group-name my-asg \  
  --desired-capacity 2 --honor-cooldown
```

Come verificare le dimensioni del gruppo con scalabilità automatica

Utilizzate il [describe-auto-scaling-groups](#) comando per confermare che la dimensione del gruppo Auto Scaling è cambiata, come nell'esempio seguente.

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

Di seguito è riportato un esempio di output, che fornisce dettagli sul gruppo e sulle istanze avviate.

```
{  
  "AutoScalingGroups": [  
    {  
      "AutoScalingGroupName": "my-asg",  
      "AutoScalingGroupARN": "arn",  
      "LaunchTemplate": {  
        "LaunchTemplateName": "my-launch-template",  
        "Version": "1",  
        "LaunchTemplateId": "lt-050555ad16a3f9c7f"  
      },  
      "MinSize": 1,  
      "MaxSize": 5,  
      "DesiredCapacity": 2,  
      "DefaultCooldown": 300,  
      "AvailabilityZones": [  
        "us-west-2a"  
      ],  
      "LoadBalancerNames": [],  
      "TargetGroupARNs": [],  
      "HealthCheckType": "EC2",  
      "HealthCheckGracePeriod": 300,  
    }  
  ]  
}
```

```

    "Instances": [
      {
        "ProtectedFromScaleIn": false,
        "AvailabilityZone": "us-west-2a",
        "LaunchTemplate": {
          "LaunchTemplateName": "my-launch-template",
          "Version": "1",
          "LaunchTemplateId": "lt-050555ad16a3f9c7f"
        },
        "InstanceId": "i-05b4f7d5be44822a6",
        "InstanceType": "t3.micro",
        "HealthStatus": "Healthy",
        "LifecycleState": "Pending"
      },
      {
        "ProtectedFromScaleIn": false,
        "AvailabilityZone": "us-west-2a",
        "LaunchTemplate": {
          "LaunchTemplateName": "my-launch-template",
          "Version": "1",
          "LaunchTemplateId": "lt-050555ad16a3f9c7f"
        },
        "InstanceId": "i-0c20ac468fa3049e8",
        "InstanceType": "t3.micro",
        "HealthStatus": "Healthy",
        "LifecycleState": "InService"
      }
    ],
    "CreatedTime": "2019-03-18T23:30:42.611Z",
    "SuspendedProcesses": [],
    "VPCZoneIdentifier": "subnet-c87f2be0",
    "EnabledMetrics": [],
    "Tags": [],
    "TerminationPolicies": [
      "Default"
    ],
    "NewInstancesProtectedFromScaleIn": false,
    "ServiceLinkedRoleARN": "arn",
    "TrafficSources": []
  }
]
}

```


Nota che `DesiredCapacity` mostra il nuovo valore. Il gruppo con scalabilità automatica ha avviato un'altra istanza.

Terminare un'istanza nel gruppo con dimensionamento automatico (AWS CLI)

A volte potresti voler dimensionare manualmente nel tuo gruppo con dimensionamento automatico ma desideri terminare un'istanza specifica. È possibile ridimensionare manualmente il gruppo Auto Scaling utilizzando il comando [terminate-instance-in-auto-scaling-group](#) e specificando l'ID dell'istanza che si desidera terminare e l'opzione `--should-decrement-desired-capacity`, come mostrato nell'esempio seguente.

```
aws autoscaling terminate-instance-in-auto-scaling-group \  
  --instance-id i-026e4c9f62c3e448c --should-decrement-desired-capacity
```

Di seguito è riportato un esempio di output, che fornisce dettagli sull'attività di scaling.

```
{  
  "Activities": [  
    {  
      "ActivityId": "b8d62b03-10d8-9df4-7377-e464ab6bd0cb",  
      "AutoScalingGroupName": "my-asg",  
      "Description": "Terminating EC2 instance: i-026e4c9f62c3e448c",  
      "Cause": "At 2023-09-23T06:39:59Z instance i-026e4c9f62c3e448c was taken  
out of service in response to a user request, shrinking the capacity from 1 to 0.",  
      "StartTime": "2023-09-23T06:39:59.015000+00:00",  
      "StatusCode": "InProgress",  
      "Progress": 0,  
      "Details": "{\"Subnet ID\":\"subnet-6194ea3b\"},\"Availability Zone\":\"us-  
west-2c\"}"  
    }  
  ]  
}
```

Questa opzione non è disponibile dalla console. Tuttavia, puoi utilizzare la pagina Istanze della console Amazon EC2 per terminare un'istanza nel tuo gruppo Auto Scaling. In tal caso, Amazon EC2 Auto Scaling rileva che l'istanza non è più in esecuzione e la sostituisce automaticamente come parte del processo di controllo dello stato. Dopo la chiusura dell'istanza sono necessari uno o due minuti

prima che venga avviata una nuova istanza. Per informazioni su come terminare un'istanza, consulta [Terminare un'istanza](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Se si interrompono le istanze del gruppo e ciò causa una distribuzione non uniforme tra le zone di disponibilità, Amazon EC2 Auto Scaling ribilancia il gruppo per ristabilire una distribuzione uniforme a meno che non si sospenda il processo. AZRebalance Per ulteriori informazioni, consulta [Sospendi e riprendi i processi di Amazon EC2 Auto Scaling](#).

Dimensionamento programmato per Dimensionamento automatico Amazon EC2

Con il ridimensionamento programmato, puoi impostare il ridimensionamento automatico per la tua applicazione in base a variazioni di carico prevedibili. Create azioni pianificate che aumentano o diminuiscono la capacità desiderata del gruppo in momenti specifici.

Ad esempio, si verifica uno schema di traffico settimanale regolare in cui il carico aumenta a metà settimana e diminuisce verso la fine della settimana. Puoi configurare una pianificazione di scalabilità in Amazon EC2 Auto Scaling che si allinei a questo schema:

- Mercoledì mattina, un'azione programmata aumenta la capacità aumentando la capacità desiderata precedentemente impostata del gruppo Auto Scaling.
- Venerdì sera, un'altra azione pianificata riduce la capacità diminuendo la capacità desiderata precedentemente impostata del gruppo Auto Scaling.

Queste azioni di dimensionamento pianificate consentono di ottimizzare costi e prestazioni.

L'applicazione dispone di una capacità sufficiente per gestire il picco di traffico infrasettimanale, ma in altri momenti non fornisce una capacità non necessaria eccedente.

È possibile utilizzare contemporaneamente il ridimensionamento pianificato e le politiche di scalabilità per ottenere i vantaggi di entrambi gli approcci alla scalabilità. Dopo l'esecuzione di un'operazione pianificata di dimensionamento, la policy di dimensionamento può continuare a prendere decisioni sull'opportunità di dimensionare ulteriormente la capacità. In questo modo è possibile garantire di disporre di capacità sufficiente per la gestione dei carichi dell'applicazione. Sebbene l'applicazione si dimensiona per soddisfare la domanda, la capacità corrente deve rientrare nei valori di capacità minima e massima impostati dall'operazione pianificata.

Indice

- [Come funziona il dimensionamento programmato](#)
- [Pianificazioni ricorrenti](#)
- [Time zone \(Fuso orario\)](#)
- [Considerazioni](#)
- [Creazione di un'operazione pianificata](#)
- [Visualizza i dettagli delle azioni pianificate](#)
- [Verificare un'attività di dimensionamento](#)
- [Eliminazione di un'operazione pianificata](#)
- [Limitazioni](#)

Come funziona il dimensionamento programmato

Per utilizzare la scalabilità pianificata, crea azioni pianificate che indichino ad Amazon EC2 Auto Scaling di eseguire attività di scalabilità in momenti specifici. Quando si crea un'azione pianificata, si specifica il gruppo Auto Scaling, il momento in cui deve avvenire l'attività di scalabilità, la nuova capacità desiderata e, facoltativamente, una nuova capacità minima e una nuova capacità massima. È possibile creare operazioni pianificate sia una tantum che ricorrenti.

All'ora specificata, Amazon EC2 Auto Scaling si ridimensiona in base ai nuovi valori di capacità, confrontando la capacità attuale con la capacità desiderata specificata.

- Se la capacità attuale è inferiore alla capacità desiderata specificata, Amazon EC2 Auto Scaling ridimensiona o aggiunge istanze fino alla capacità desiderata specificata.
- Se la capacità attuale è superiore alla capacità desiderata specificata, Amazon EC2 Auto Scaling ridimensiona o rimuove le istanze fino alla capacità desiderata specificata.

Un'azione pianificata imposta la capacità desiderata, minima e massima del gruppo alla data e all'ora specificate. È possibile creare un'azione pianificata solo per una di queste capacità alla volta, ad esempio la capacità desiderata. Tuttavia, in alcuni casi è necessario includere la capacità minima e massima per garantire che la capacità desiderata specificata nell'azione non superi questi limiti.

Pianificazioni ricorrenti

Per creare una pianificazione ricorrente utilizzando AWS CLI o un SDK, specificate un'espressione cron e un fuso orario per descrivere quando l'azione pianificata deve ripetersi. È possibile specificare una data e un'ora per l'ora di avvio, per l'ora di fine o per entrambi i campi.

Per creare una pianificazione ricorrente utilizzando AWS Management Console, specifica il modello di ricorrenza, il fuso orario, l'ora di inizio e l'ora di fine facoltativa dell'azione pianificata. Tutte le opzioni del modello di ricorrenza sono basate su espressioni cron. In alternativa, puoi scrivere la tua espressione cron personalizzata.

Il formato dell'espressione cron supportato è costituito da cinque campi separati da spazi: [Minute] [Hour] [Day_of_Month] [Month_of_Year] [Day_of_Week]. Ad esempio, l'espressione cron `30 6 * * 2`, configura un'operazione pianificata che ricorre ogni martedì alle 6:30. L'asterisco viene utilizzato come carattere jolly per abbinare tutti i valori di un campo. Per altri esempi di espressioni cron, consulta <https://crontab.guru/examples.html>. Per informazioni su come scrivere le tue espressioni cron in questo formato, consulta [Crontab](#).

Scegli con attenzione l'ora di avvio e di fine. Ricorda quanto segue:

- Se si specifica un'ora di inizio, Dimensionamento automatico Amazon EC2 esegue l'operazione in quell'ora e quindi la riesegue in base alla pianificazione ricorrente specificata.
- Se si specifica un'ora di fine, l'iterazione dell'operazione si interrompe dopo tale orario. L'operazione pianificata non persiste nell'account una volta raggiunta l'ora di fine.
- L'ora di inizio e l'ora di fine devono essere impostate in UTC quando si utilizza o un SDK. AWS CLI

Time zone (Fuso orario)

Per impostazione predefinita, le pianificazioni ricorrenti impostate sono in formato UTC. Puoi modificare l'orario affinché corrisponda al fuso orario locale o a un fuso orario di un'altra parte della rete. Quando si specifica un fuso orario che osserva l'ora legale (DST), l'operazione vi si adatta automaticamente.

I valori validi sono i nomi canonici dei fusi orari del database dei fusi orari della Internet Assigned Numbers Authority (IANA). Ad esempio, l'ora orientale degli Stati Uniti è identificata canonicamente come `America/New_York` [Per ulteriori informazioni, vedere https://www.iana.org/time-zones](https://www.iana.org/time-zones).

Fusi orari basati sulla posizione, ad esempio regolati `America/New_York` automaticamente in base all'ora legale. Tuttavia, un fuso orario basato su UTC come `Etc/UTC`, è un tempo assoluto e non si regolerà sull'ora legale.

Ad esempio, disponiamo di una pianificazione ricorrente il cui fuso orario è `America/New_York`. La prima operazione di dimensionamento avviene nel fuso orario `America/New_York`, prima dell'inizio dell'ora legale. La successiva operazione di dimensionamento avviene nel fuso orario `America/`

New_York, dopo l'avvio dell'ora legale. La prima operazione si avvia alle 8:00 UTC-5 ora locale, mentre la seconda volta inizia alle 8:00 UTC-4 ora locale.

Se si crea un'azione pianificata utilizzando AWS Management Console e si specifica un fuso orario che rispetti l'ora legale, sia la pianificazione ricorrente che gli orari di inizio e fine si adattano automaticamente all'ora legale.

Considerazioni

Quando crei un'operazione pianificata, tieni presente quanto segue:

- L'ordine di esecuzione delle operazioni pianificate è garantito all'interno dello stesso gruppo, ma non per le operazioni pianificate tra i gruppi.
- Un'azione pianificata viene generalmente eseguita in pochi secondi. Tuttavia, l'operazione potrebbe essere ritardata fino a due minuti, rispetto all'ora di inizio pianificata. Poiché le operazioni all'interno di un gruppo con scalabilità automatica vengono eseguite nell'ordine in cui sono specificate, l'esecuzione delle operazioni pianificate con ore di inizio previste ravvicinate tra loro può richiedere più tempo.
- È possibile disattivare temporaneamente il dimensionamento programmato per un gruppo con scalabilità automatica sospendendo il processo di elaborazione `ScheduledActions`. Ciò permette di impedire che le operazioni pianificate siano attive senza doverle eliminare. Sarà poi quindi possibile riprendere il dimensionamento programmato quando si vorrà utilizzarlo nuovamente. Per ulteriori informazioni, consulta [Sospendi e riprendi i processi di Amazon EC2 Auto Scaling](#).
- Dopo aver creato un'operazione pianificata, è possibile aggiornare qualsiasi impostazione relativa ad essa, ad eccezione del nome.

Creazione di un'operazione pianificata

Per creare un'azione pianificata per il gruppo Auto Scaling, utilizzate uno dei seguenti metodi:

Console

Per creare un'operazione pianificata

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Nella scheda Automatic scaling (Dimensionamento automatico) scegliere Create scheduled action (Crea operazione pianificata) in Scheduled actions (Operazioni pianificate).
4. In Name (Nome) specifica un nome per l'operazione pianificata.
5. Per Capacità desiderata, Min, Max, scegli la nuova capacità desiderata del gruppo e i nuovi limiti di dimensioni minimi e massimi. La capacità desiderata deve essere uguale o superiore alla dimensione minima del gruppo e uguale o inferiore alla dimensione massima del gruppo.
6. In Compute (Calcolo), seleziona una delle opzioni disponibili.
 - Se desideri dimensionare in base a una pianificazione ricorrente, scegli la frequenza con cui Dimensionamento automatico Amazon EC2 deve eseguire l'operazione pianificata.
 - Se scegli un'operazione che inizia con Every (Ogni), viene creata l'espressione cron.
 - Se scegli Cron, inserisci un'espressione cron che specifichi quando eseguire l'operazione, in UTC.
 - Se desideri dimensionare una sola volta, scegli Once (Una volta).
7. In Time zone (Fuso orario), scegli un fuso orario. Il valore predefinito è Et c/UTC.

Tutti i fusi orari elencati provengono dal database del fuso orario IANA. Per ulteriori informazioni, consultate https://en.wikipedia.org/wiki/List_of_tz_database_time_zones.

8. Definisci una data e un'ora in Specific start time (Ora di inizio specifica).
 - Se scegli una pianificazione periodica, l'ora di avvio definisce quando verrà eseguita la prima operazione pianificata della serie ricorrente.
 - Se hai scelto l'opzione Once (Una volta) come ricorrenza, specifica la data e l'ora programmate per l'esecuzione dell'operazione.
9. (Facoltativo) Per le pianificazioni ricorrenti, è possibile specificare un'ora di termine scegliendo Set End Time (Imposta l'ora di fine) e quindi scegliendo una data e un'ora in End by (Termina entro).
10. Scegli Crea. La console visualizza le operazioni programmate per il gruppo con scalabilità automatica.

AWS CLI

Per creare un'azione pianificata, è possibile utilizzare uno dei seguenti comandi di esempio. Sostituisci ciascun *placeholder input dell'utente* con le tue informazioni.

Esempio: dimensionamento unico

Utilizzate il seguente comando [put-scheduled-update-group-action](#) con le `--desired-capacity` opzioni `--start-time "YYYY-MM-DDThh:mm:ssZ"` and.

```
aws autoscaling put-scheduled-update-group-action --scheduled-action-name my-one-time-action \  
  --auto-scaling-group-name my-asg --start-time "2021-03-31T08:00:00Z" --desired-capacity 3
```

Esempio: per pianificare il ridimensionamento in base a una pianificazione ricorrente

Utilizzate il seguente comando [put-scheduled-update-group-action](#) con le opzioni and. `--recurrence "cron expression"` `--desired-capacity`

```
aws autoscaling put-scheduled-update-group-action --scheduled-action-name my-recurring-action \  
  --auto-scaling-group-name my-asg --recurrence "0 9 * * *" --desired-capacity 3
```

Per impostazione predefinita, Amazon EC2 Auto Scaling esegue la pianificazione di ricorrenza specificata in base al fuso orario UTC. Per specificare un fuso orario diverso, includi l'`--time-zone`opzione e il nome del fuso orario IANA, come nell'esempio seguente.

```
--time-zone "America/New_York"
```

Per ulteriori informazioni, vedere https://en.wikipedia.org/wiki/List_of_tz_database_time_zones.

Visualizza i dettagli delle azioni pianificate

Per visualizzare i dettagli delle azioni pianificate imminenti per il gruppo Auto Scaling, utilizzate uno dei seguenti metodi:

Console

Per visualizzare i dettagli delle azioni pianificate

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona il gruppo con scalabilità automatica.

3. Nella scheda Ridimensionamento automatico, nella sezione Azioni pianificate, puoi visualizzare le azioni pianificate imminenti.

Tieni presente che la console mostra i valori per Ora di inizio e Ora di fine nell'ora locale con lo scostamento UTC in vigore alla data e all'ora specificate. Lo scostamento UTC è la differenza, espressa in ore e minuti, tra l'ora locale e l'UTC. Il valore per Time zone (Fuso orario) mostra il fuso orario richiesto, ad esempio `America/New_York`.

AWS CLI

Utilizza il seguente comando [describe-scheduled-actions](#).

```
aws autoscaling describe-scheduled-actions --auto-scaling-group-name my-asg
```

Se il comando viene eseguito correttamente, verrà visualizzato un output simile al seguente.

```
{
  "ScheduledUpdateGroupActions": [
    {
      "AutoScalingGroupName": "my-asg",
      "ScheduledActionName": "my-recurring-action",
      "Recurrence": "30 0 1 1,6,12 *",
      "ScheduledActionARN": "arn:aws:autoscaling:us-
west-2:123456789012:scheduledUpdateGroupAction:8e86b655-b2e6-4410-8f29-
b4f094d6871c:autoScalingGroupName/my-asg:scheduledActionName/my-recurring-action",
      "StartTime": "2020-12-01T00:30:00Z",
      "Time": "2020-12-01T00:30:00Z",
      "MinSize": 1,
      "MaxSize": 6,
      "DesiredCapacity": 4
    }
  ]
}
```

Verificare un'attività di dimensionamento

Per verificare le attività di dimensionamento associate alla dimensionamento pianificata, consulta [Verificare un'attività di dimensionamento per un gruppo con scalabilità automatica](#).

Eliminazione di un'operazione pianificata

Per eliminare un'azione pianificata, utilizzate uno dei seguenti metodi:

Console

Per eliminare un'operazione pianificata

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona il gruppo con scalabilità automatica.
3. Nella scheda Automatic scaling (Dimensionamento automatico) selezionare un'operazione pianificata in Scheduled actions (Operazioni pianificate).
4. Scegli Operazioni > Elimina.
5. Quando viene richiesta la conferma, seleziona Sì, elimina.

AWS CLI

Utilizza il seguente comando [delete-scheduled-action](#).

```
aws autoscaling delete-scheduled-action --auto-scaling-group-name my-asg \  
--scheduled-action-name my-recurring-action
```

Limitazioni

- I nomi delle azioni pianificate devono essere univoci per ciascun gruppo con scalabilità automatica.
- Un'operazione pianificata deve avere un valore ora univoco. Se tenti di pianificare un'attività in un momento in cui è già pianificata un'altra attività di dimensionamento, la richiesta viene respinta e restituisce un errore che indica che esiste già un'operazione pianificata con la stessa ora di avvio.
- È possibile creare un massimo di 125 operazioni pianificate per ciascun gruppo con scalabilità automatica.

Dimensionamento dinamico per Dimensionamento automatico Amazon EC2

Il dimensionamento dinamico consente di dimensionare la capacità del gruppo con scalabilità automatica in base alle modifiche del traffico.

Dimensionamento automatico Amazon EC2 supporta i seguenti tipi di policy di dimensionamento dinamico:

- **Target tracking scaling:** aumenta e diminuisce la capacità attuale del gruppo in base a una CloudWatch metrica Amazon e a un valore target. Il funzionamento è simile al modo in cui il termostato regola la temperatura di una casa: si seleziona la temperatura e il termostato fa il resto.
- **Dimensionamento per fasi:** aumenta e diminuisce la capacità attuale del gruppo in base a una serie di regolazioni del dimensionamento, chiamate regolazioni per fasi, che variano in base alla dimensione dell'utilizzo fuori limite segnalato dall'allarme.
- **Dimensionamento semplice:** aumenta e diminuisce la capacità attuale del gruppo in base a una singola regolazione del dimensionamento, con un tempo di raffreddamento tra ogni attività di dimensionamento..

Ti consigliamo vivamente di utilizzare le politiche di scalabilità di Target Tracking e di scegliere una metrica che cambi in modo inversamente proporzionale alla variazione della capacità del tuo gruppo Auto Scaling. Quindi, se raddoppi la dimensione del tuo gruppo Auto Scaling, la metrica diminuisce del 50 per cento. Ciò consente ai dati dei parametri di attivare con precisione gli eventi di scalabilità proporzionale. Sono incluse metriche come l'utilizzo medio della CPU o il numero medio di richieste per target.

Con il tracciamento del target, il gruppo Auto Scaling si ridimensiona in modo direttamente proporzionale al carico effettivo dell'applicazione. Ciò significa che oltre a soddisfare il fabbisogno immediato di capacità in risposta alle modifiche del carico, una policy di monitoraggio degli obiettivi può anche adattarsi alle modifiche del carico che si verificano nel tempo, ad esempio a causa di variazioni stagionali.

Le politiche di tracciamento di Target eliminano inoltre la necessità di definire manualmente gli CloudWatch allarmi e le regolazioni di ridimensionamento. Amazon EC2 Auto Scaling lo gestisce automaticamente in base all'obiettivo impostato.

Indice

- [Come funziona il dimensionamento dinamico](#)
- [Più policy di dimensionamento](#)
- [Policy di dimensionamento con monitoraggio degli obiettivi per Dimensionamento automatico Amazon EC2](#)
- [Policy di dimensionamento per fasi e semplice per Dimensionamento automatico Amazon EC2](#)
- [Tempo di raffreddamento per il dimensionamento per Dimensionamento automatico Amazon EC2](#)
- [Scalabilità basata su Amazon SQS](#)
- [Verificare un'attività di dimensionamento per un gruppo con scalabilità automatica](#)
- [Disabilitare una policy di dimensionamento per un gruppo con scalabilità automatica](#)
- [Eliminazione di una policy di dimensionamento](#)
- [Policy di dimensionamento di esempio per AWS Command Line Interface \(AWS CLI\)](#)

Come funziona il dimensionamento dinamico

Una politica di scalabilità dinamica indica ad Amazon EC2 Auto Scaling di tenere traccia di CloudWatch una metrica specifica e definisce l'azione da intraprendere quando CloudWatch l'allarme associato è in ALARM. I parametri utilizzati per attivare lo stato di allarme sono un'aggregazione dei parametri provenienti da tutte le istanze del gruppo con scalabilità automatica. (Ad esempio, supponiamo che tu abbia un gruppo con scalabilità automatica con due istanze, in cui un'istanza è al 60% della CPU e l'altra è al 40%; quindi, la CPU è a una media del 50%). Quando la policy è in vigore, Dimensionamento automatico Amazon EC2 regola la capacità desiderata del gruppo verso l'alto o verso il basso in caso di superamento della soglia di un allarme.

Quando viene richiamata una policy di dimensionamento, se il calcolo della capacità produce un numero al di fuori dell'intervallo delle dimensioni minima e massima del gruppo, Dimensionamento automatico Amazon EC2 assicura che la nuova capacità non superi mai i limiti delle dimensioni minima e massima. [La capacità viene misurata in due modi: utilizzando le stesse unità scelte quando si imposta la capacità desiderata in termini di istanze o utilizzando unità di capacità \(se vengono applicati i pesi delle istanze\).](#)

- Esempio 1: un gruppo con scalabilità automatica ha una capacità massima di 3, una capacità corrente di 2 e una policy di dimensionamento dinamico che aggiunge 3 istanze. Quando si esegue questa policy, Dimensionamento automatico Amazon EC2 aggiunge solo un'istanza al gruppo, per impedire che vengano superate le dimensioni massime.

- **Esempio 2:** un gruppo con scalabilità automatica ha una capacità minima di 2, una capacità corrente di 3 e una policy di dimensionamento dinamico che rimuove 2 istanze. Quando si invoca questa policy, Dimensionamento automatico Amazon EC2 rimuove solo un'istanza dal gruppo, per impedire che il gruppo scenda al di sotto delle dimensioni minime.

Quando la capacità desiderata raggiunge il limite massimo, l'aumento orizzontale si arresta. Se la domanda cala e la capacità diminuisce, Dimensionamento automatico Amazon EC2 può aumentare nuovamente le dimensioni.

L'eccezione si verifica quando si utilizzano i pesi delle istanze. In questo caso, Dimensionamento automatico Amazon EC2 può dimensionare oltre il limite massimo, ma solo fino al peso massimo delle istanze. Lo scopo è avvicinarsi il più possibile alla nuova capacità desiderata, pur rispettando le strategie di allocazione specificate per il gruppo. Le strategie di allocazione determinano quali tipi di istanza avviare. Il peso determina con quante unità ciascuna istanza contribuisce alla capacità desiderata del gruppo, in base al tipo di istanza.

- **Esempio 3:** un gruppo con scalabilità automatica ha una capacità massima di 12, una capacità corrente di 10 e una policy di dimensionamento dinamico che aggiunge 5 unità di capacità. Ai tipi di istanza è assegnato uno di questi tre pesi: 1, 4 o 6. Quando si richiama la policy di dimensionamento, Dimensionamento automatico Amazon EC2 sceglie di avviare un tipo di istanza con un peso pari a 6, in base alla strategia di allocazione. Come risultato, il gruppo avrà una capacità desiderata di 12 e una capacità corrente di 16.

Più policy di dimensionamento

Nella maggior parte dei casi, una policy di dimensionamento con monitoraggio degli obiettivi è sufficiente per configurare il gruppo con scalabilità automatica, affinché questo incrementi e decrementi automaticamente. Una policy di dimensionamento con monitoraggio degli obiettivi consente di selezionare un risultato desiderato e fare in modo che il gruppo con scalabilità automatica aggiunga e rimuova le istanze come necessario per ottenere questo risultato.

Per una configurazione di dimensionamento avanzata, il gruppo con scalabilità automatica può avere più di una policy di dimensionamento. Ad esempio, è possibile definire una o più policy di dimensionamento con monitoraggio degli obiettivi, una o più policy di dimensionamento per fasi o entrambe. Ciò offre una maggiore flessibilità, per affrontare diversi scenari.

Per illustrare il modo in cui più policy di dimensionamento operano insieme, consideriamo un'applicazione che utilizza un gruppo con scalabilità automatica e una coda Amazon SQS per

inviare richieste a una singola istanza EC2. Per garantire che l'applicazione offra prestazioni ottimali, vi sono due policy che controllano quando il gruppo con scalabilità automatica deve essere dimensionato orizzontalmente. Una è una policy con monitoraggio degli obiettivi che utilizza un parametro personalizzato per aumentare e ridurre la capacità in base al numero di messaggi SQS nella coda. L'altra è una politica di scalabilità graduale che utilizza il CloudWatch `CPUUtilization` parametro Amazon per aggiungere capacità quando l'istanza supera il 90% di utilizzo per un periodo di tempo specificato.

Quando ci sono più policy attive nello stesso momento, c'è la possibilità che ogni policy indichi al gruppo con scalabilità automatica di aumentare o ridurre orizzontalmente le risorse nello stesso momento. Ad esempio, è possibile che la `CPUUtilization` metrica aumenti e superi la soglia dell'CloudWatch allarme nello stesso momento in cui la metrica personalizzata SQS aumenta e supera la soglia dell'allarme metrico personalizzato.

In queste situazioni, Dimensionamento automatico Amazon EC2 sceglie la policy che offre la capacità maggiore per aumentare o ridurre orizzontalmente. Supponiamo ad esempio che la policy per `CPUUtilization` avvii un'istanza e che la policy per la coda SQS avvii due istanze. Se i parametri per il dimensionamento orizzontale per tutte le policy saranno soddisfatti nello stesso momento, Dimensionamento automatico Amazon EC2 privilegerà la policy della coda SQS. In questo modo il gruppo con scalabilità automatica avvierà due istanze.

L'approccio che prevede di dare precedenza alla policy che fornisce la capacità maggiore si applica anche quando le policy riducono orizzontalmente impiegando criteri diversi. Ad esempio, se una policy termina tre istanze, un'altra policy riduce il numero di istanze del 25% e il gruppo ha otto istanze al momento della riduzione orizzontale, Dimensionamento automatico Amazon EC2 privilegia la policy che offre il maggior numero di istanze per il gruppo. Di conseguenza, il gruppo con scalabilità automatica termina due istanze (25% di 8 = 2). Lo scopo è evitare che Dimensionamento automatico Amazon EC2 rimuova troppe istanze.

Suggeriamo, tuttavia, di prestare attenzione quando si utilizzano le policy di dimensionamento con monitoraggio degli obiettivi insieme alle policy di dimensionamento per fasi, per evitare che insorgano conflitti che possono causare comportamenti indesiderati. Ad esempio, se la policy di dimensionamento per fasi avvia un'attività di riduzione prima che la policy con monitoraggio degli obiettivi sia pronta per eseguirla, tale attività non verrà bloccata. Al termine dell'attività di riduzione, la policy con monitoraggio degli obiettivi potrebbe indicare al gruppo di aumentare di nuovo.

Policy di dimensionamento con monitoraggio degli obiettivi per Dimensionamento automatico Amazon EC2

Una politica di dimensionamento del tracciamento degli obiettivi ridimensiona automaticamente la capacità del gruppo Auto Scaling in base a un valore metrico target. Ciò consente all'applicazione di mantenere prestazioni ottimali ed efficienza in termini di costi senza intervento manuale.

Con il monitoraggio degli obiettivi, devi selezionare un parametro e un valore target che rappresenti il livello di utilizzo o velocità di trasmissione effettiva ideale per la tua applicazione. Amazon EC2 Auto Scaling crea e gestisce gli allarmi che richiamano eventi di scalabilità quando CloudWatch la metrica si discosta dall'obiettivo. Ad esempio, è simile al modo in cui un termostato mantiene una temperatura target.

Ad esempio, immaginiamo di avere un'applicazione Web attualmente eseguita su due istanze e di volere che l'utilizzo della CPU del gruppo con scalabilità automatica rimanga intorno al 50% quando il carico sull'applicazione varia. Questo ti offre la capacità aggiuntiva per gestire i picchi di traffico senza dover mantenere un numero eccessivo di risorse inattive.

È possibile soddisfare questa esigenza creando una policy di dimensionamento del monitoraggio degli obiettivi che si rivolge a un utilizzo medio della CPU del 50%. Quindi, il gruppo Auto Scaling eseguirà la scalabilità orizzontale o aumenterà la capacità quando la CPU supera il 50 per cento per gestire un carico maggiore. Aumenterà o ridurrà la capacità quando la CPU scende al di sotto del 50 per cento per ottimizzare i costi nei periodi di basso utilizzo.

Argomenti

- [Policy di dimensionamento con monitoraggio degli obiettivi multipli](#)
- [Selezionare i parametri.](#)
- [Definire il valore target](#)
- [Definisci il tempo di riscaldamento dell'istanza](#)
- [Considerazioni](#)
- [Creazione di una policy di dimensionamento con monitoraggio degli obiettivi](#)
- [Creazione di una policy di dimensionamento con monitoraggio degli obiettivi per Dimensionamento automatico Amazon EC2 tramite la matematica dei parametri](#)

Policy di dimensionamento con monitoraggio degli obiettivi multipli

Per ottimizzare le prestazioni di dimensionamento, è possibile utilizzare più policy di dimensionamento con monitoraggio degli obiettivi insieme, a condizione che ciascuna di esse utilizzi un parametro diverso. Ad esempio, l'utilizzo e la velocità di trasmissione effettiva possono influenzarsi a vicenda. Ogni volta che uno di questi parametri cambia, di solito implica che anche altri parametri saranno interessati. L'uso di più metriche fornisce quindi informazioni aggiuntive sul carico a cui è sottoposto il gruppo Auto Scaling. Questo può aiutare Amazon EC2 Auto Scaling a prendere decisioni più informate nel determinare la capacità da aggiungere al gruppo.

L'intenzione di Amazon EC2 Auto Scaling è quella di dare sempre la priorità alla disponibilità. Se una delle politiche di tracciamento di destinazione è pronta per essere scalata orizzontalmente, il gruppo Auto Scaling verrà ridimensionato. Sarà scalabile solo se tutte le politiche di tracciamento di Target (con la parte scalabile abilitata) sono pronte per essere ampliate.

Selezionare i parametri.

Puoi creare policy di dimensionamento con monitoraggio degli obiettivi sia con parametri predefiniti che personalizzati.

Quando crei una policy di dimensionamento con monitoraggio degli obiettivi con un tipo di parametro predefinito, scegli un parametro dalla seguente lista di parametri predefiniti:

- `ASGAverageCPUUtilization`: utilizzo medio CPU del gruppo con scalabilità automatica.
- `ASGAverageNetworkIn`: numero medio di byte ricevuti da una singola istanza su tutte le interfacce di rete.
- `ASGAverageNetworkOut`: numero medio di byte inviati da una singola istanza su tutte le interfacce di rete.
- `ALBRequestCountPerTarget`: numero medio di richieste Application Load Balancer per destinazione.

Important

Altre informazioni utili sulle metriche relative all'utilizzo della CPU, all'I/O di rete e al numero di richieste di Application Load Balancer per target sono disponibili rispettivamente nell'argomento [Elenco delle metriche CloudWatch disponibili per le istanze nella Guida utente](#)

[di Amazon EC2 per le istanze Linux e nelle CloudWatch metriche per l'Application Load Balancer nella User Guide for Application Load Balancers.](#)

Puoi scegliere altre CloudWatch metriche disponibili o le tue metriche CloudWatch specificando una metrica personalizzata. È necessario utilizzare AWS CLI o un SDK per creare una politica di tracciamento degli obiettivi con una specifica metrica personalizzata. Per un esempio che specifica una specifica metrica personalizzata per una politica di ridimensionamento del tracciamento degli obiettivi utilizzando il, vedi. [AWS CLI Policy di dimensionamento di esempio per AWS Command Line Interface \(AWS CLI\)](#)

Quando scegli un parametro, tieni presente quanto segue:

- Ti consigliamo di utilizzare parametri disponibili solo a intervalli di un minuto per aiutarti a dimensionare più rapidamente quando viene modificato l'utilizzo. Il monitoraggio degli obiettivi valuta i parametri aggregati con una granularità di un minuto per tutti i parametri predefiniti e personalizzati, ma il parametro sottostante potrebbe pubblicare i dati meno frequentemente. Ad esempio, tutti i parametri di Amazon EC2 vengono inviati a intervalli di cinque minuti per impostazione predefinita, ma sono configurabili fino a un minuto (noto come monitoraggio dettagliato). Questa scelta spetta ai singoli servizi. La maggior parte dei servizi cerca di utilizzare l'intervallo più breve possibile. Per ulteriori informazioni su come abilitare il monitoraggio dettagliato, consulta [Configurare il monitoraggio per le istanze Auto Scaling](#).
- Non tutti i parametri personalizzati funzionano per il monitoraggio degli obiettivi. Il parametro deve essere un parametro di utilizzo valido e deve descrivere quanto è impegnata un'istanza. Il valore del parametro deve aumentare e diminuire in proporzione al numero di istanze nel gruppo con scalabilità automatica. In questo modo i dati dei parametri possono essere utilizzati per il dimensionamento proporzionalmente al numero di istanze. Ad esempio, l'utilizzo della CPU di un gruppo con scalabilità automatica (ovvero il parametro Amazon EC2 CPUUtilization con la dimensione di parametro AutoScalingGroupName) funziona se il carico sul gruppo con scalabilità automatica viene distribuito tra le istanze.
- I seguenti parametri non funzionano per il monitoraggio degli obiettivi:
 - Il numero di richieste ricevute dal load balancer rivolto al gruppo con scalabilità automatica (ovvero, il parametro Elastic Load Balancing RequestCount). Il numero di richieste ricevute dal load balancer non cambia in base all'utilizzo del gruppo con scalabilità automatica.
 - La latenza della richiesta del load balancer (ovvero, il parametro Elastic Load Balancing Latency). La latenza della richiesta può aumentare in base all'aumento dell'utilizzo, ma non cambia necessariamente in modo proporzionale.

- La CloudWatch metrica della coda di Amazon SQS, `ApproximateNumberOfMessagesVisible` Il numero di messaggi in una coda potrebbe non cambiare in proporzione alla dimensione del gruppo con scalabilità automatica che elabora i messaggi provenienti dalla coda. Tuttavia, un parametro personalizzato che misuri il numero di messaggi nella coda per ogni istanza EC2 nel gruppo con scalabilità automatica può funzionare. Per ulteriori informazioni, consulta [Scalabilità basata su Amazon SQS](#).
- Per utilizzare il parametro `ALBRequestCountPerTarget`, è necessario specificare il parametro `ResourceLabel` affinché identifichi il gruppo di destinazione del load balancer associato al parametro. Per un esempio che specifica il `ResourceLabel` parametro per una politica di scalabilità di tracciamento degli obiettivi utilizzando il, vedi. AWS CLI [Policy di dimensionamento di esempio per AWS Command Line Interface \(AWS CLI\)](#)
- Quando una metrica emette valori 0 reali su CloudWatch (ad esempio `ALBRequestCountPerTarget`), un gruppo Auto Scaling può scalare fino a 0 quando non c'è traffico verso l'applicazione per un periodo di tempo prolungato. Per fare in modo che il gruppo con scalabilità automatica si riduca orizzontalmente a 0 istanze quando non vengono instradate richieste, la capacità minima del gruppo deve essere impostata a 0.
- Invece di pubblicare nuove metriche da utilizzare nella policy di scalabilità, è possibile utilizzare la matematica delle metriche per combinare quelle esistenti. Per ulteriori informazioni, consulta [Creazione di una policy di dimensionamento con monitoraggio degli obiettivi per Dimensionamento automatico Amazon EC2 tramite la matematica dei parametri](#).

Definire il valore target

Quando si crea una policy di dimensionamento del monitoraggio degli obiettivi, è necessario specificare un valore target. Il valore di destinazione rappresenta l'utilizzo medio ideale o la velocità di trasmissione effettiva per il gruppo con scalabilità automatica nel suo complesso. Per utilizzare le risorse in modo efficiente in termini di costi, impostare il valore target il più alto possibile con un buffer ragionevole per aumenti di traffico imprevisti. Quando l'applicazione viene aumentata orizzontalmente in modo ottimale per un normale flusso di traffico, il valore del parametro effettivo deve essere pari o appena inferiore al valore target.

Quando una policy di scalabilità si basa sulla velocità di trasmissione effettiva, ad esempio il conteggio delle richieste per destinazione per un Application Load Balancer, I/O di rete o altri parametri di conteggio, il valore di destinazione rappresenta la velocità di trasmissione effettiva media ottimale da una singola istanza per un periodo di un minuto.

Definisci il tempo di riscaldamento dell'istanza

Come opzione, è possibile specificare il numero di secondi necessari per la preparazione di un'istanza appena avviata. Fino alla scadenza del tempo di riscaldamento specificato, un'istanza non viene conteggiata nelle metriche aggregate delle istanze EC2 del gruppo Auto Scaling.

Durante il periodo di riscaldamento delle istanze, le politiche di scalabilità vengono ridimensionate solo se il valore metrico delle istanze che non si stanno riscaldando è superiore all'utilizzo previsto dalla policy.

Se il gruppo si dimensiona di nuovo, le istanze in fase di preparazione vengono conteggiate come parte della capacità desiderata per l'attività di dimensionamento successiva. L'intenzione è di aumentare di continuo (ma non in eccesso).

Mentre l'attività di dimensionamento orizzontale è in corso, tutte le attività di riduzione orizzontale avviate dalle policy di ridimensionamento vengono bloccate fino al termine della preparazione delle istanze. Al termine della preparazione delle istanze, se si verifica un evento di dimensionamento, tutte le istanze attualmente in fase di chiusura verranno conteggiate ai fini della capacità attuale del gruppo nel calcolo della nuova capacità desiderata. Pertanto, non rimuoviamo dal gruppo Auto Scaling più istanze del necessario.

Valore predefinito

Se non viene impostato alcun valore, la politica di scalabilità utilizzerà il valore predefinito, che è il valore per il riscaldamento dell'istanza [predefinito definito](#) per il gruppo. [Se il riscaldamento dell'istanza predefinito è nullo, torna al valore del cooldown predefinito](#). Si consiglia di utilizzare il riscaldamento predefinito dell'istanza per semplificare l'aggiornamento di tutte le politiche di ridimensionamento quando cambia il tempo di riscaldamento.

Considerazioni

Le seguenti considerazioni si applicano quando si usano le policy di dimensionamento con monitoraggio degli obiettivi:

- Non creare, modificare o eliminare gli CloudWatch allarmi utilizzati con una politica di ridimensionamento di Target Tracking. Amazon EC2 Auto Scaling crea e gestisce gli allarmi associati CloudWatch alle politiche di scalabilità di tracciamento di destinazione e li elimina quando non sono più necessari.
- Una policy di dimensionamento del monitoraggio degli obiettivi dà priorità alla disponibilità durante i periodi di fluttuazione dei livelli di traffico, riducendo orizzontalmente in modo più graduale

quando il traffico diminuisce. Se si desidera che il gruppo con scalabilità automatica venga ridotto orizzontale immediatamente al termine di un carico di lavoro, è possibile disabilitare la parte di riduzione orizzontale della policy. In questo modo si ha la flessibilità di utilizzare il metodo di riduzione orizzontale più adeguato alle proprie esigenze quando l'utilizzo è basso. Per garantire che la riduzione orizzontale avvenga il più rapidamente possibile, consigliamo di non utilizzare una semplice policy di dimensionamento per evitare che venga aggiunto un tempo di raffreddamento.

- Se nella metrica mancano punti dati, lo stato di allarme passa a `CloudWatch INSUFFICIENT_DATA`. In questo caso, Dimensionamento automatico Amazon EC2 non può dimensionare il gruppo finché non vengono trovati nuovi punti dati.
- Se la metrica viene riportata scarsamente in base alla progettazione, la matematica metrica può essere utile. Ad esempio, per utilizzare i valori più recenti, utilizzate la `FILL(m1, REPEAT)` funzione dove `m1` è la metrica.
- Potrebbero esserci delle differenze tra il valore di destinazione e i punti dati dei parametri reali. Ciò avviene perché agiamo sempre con prudenza, arrotondando per eccesso o per difetto, quando stabiliamo il numero di istanze da aggiungere o rimuovere. Ciò impedisce l'aggiunta di un numero insufficiente di istanze o la rimozione di troppe istanze. Tuttavia, per gruppi Auto Scaling più piccoli e con un numero inferiore di istanze, l'utilizzo del gruppo potrebbe sembrare lontano dal valore di destinazione. Ad esempio, supponiamo che tu imposti un valore di destinazione del 50% per l'utilizzo della CPU e il gruppo con scalabilità automatica lo superi. Potremmo stabilire che l'aggiunta di 1,5 istanze diminuirà l'utilizzo della CPU a quasi il 50%. Poiché non è possibile aggiungere 1,5 istanze, arrotondiamo e aggiungiamo due istanze. Questo potrebbe diminuire l'utilizzo della CPU a un valore al di sotto del 50%, ma garantisce che l'applicazione disponga di un numero sufficiente di risorse per essere supportata. Analogamente, se stabiliamo che la rimozione di 1,5 istanze aumenta l'utilizzo della CPU a oltre il 50%, rimuoviamo una sola istanza.

Per i gruppi Auto Scaling di grandi dimensioni e con più istanze, l'utilizzo viene distribuito su un maggior numero di istanze, nel qual caso l'aggiunta o la rimozione di istanze provoca un intervallo minore tra il valore di destinazione e i punti di dati dei parametri reali.

- Una policy di dimensionamento con monitoraggio degli obiettivi presuppone l'aumento orizzontale del gruppo con scalabilità automatica quando il parametro specificato supera il valore di destinazione. Quando il parametro specificato è inferiore al valore di destinazione, non è possibile utilizzare una policy di dimensionamento con monitoraggio degli obiettivi per aumentare orizzontalmente il gruppo con scalabilità automatica.

Creazione di una policy di dimensionamento con monitoraggio degli obiettivi

Per creare una politica di ridimensionamento del tracciamento degli obiettivi per il tuo gruppo Auto Scaling, utilizza uno dei seguenti metodi.

Prima di iniziare, verifica che la metrica preferita sia disponibile a intervalli di 1 minuto (rispetto all'intervallo predefinito di 5 minuti dei parametri di Amazon EC2).

Console

Creazione di una policy di dimensionamento con monitoraggio degli obiettivi per un nuovo gruppo con dimensionamento automatico

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Selezionare Create Auto Scaling group (Crea un gruppo con scalabilità automatica).
3. Nelle fasi 1, 2 e 3, scegli le opzioni desiderate e procedere a Step 4: Configure group size and scaling policies (Fase 4: configurazione delle dimensioni del gruppo e delle policy di dimensionamento).
4. Nella sezione Dimensionamento, specifica l'intervallo entro il quale desideri scalare, aggiornando la Capacità minima desiderata e la Capacità massima desiderata. Queste due impostazioni permettono al gruppo con scalabilità automatica di dimensionare dinamicamente. Per ulteriori informazioni, consulta [Impostare i limiti di dimensionamento per il gruppo con dimensionamento automatico](#).
5. In Policy di dimensionamento, scegli Policy di dimensionamento con monitoraggio degli obiettivi.
6. Per definire una policy, effettua le operazioni seguenti:
 - a. Specificare un nome per la policy.
 - b. Scegliere un parametro per Tipo parametro.

Se hai scelto Application Load Balancer request count per target (Conteggio delle richieste Application Load Balancer per destinazione), scegli un gruppo destinazione in Target group (Gruppo di destinazione).

- c. Specifica un Target value (Valore di destinazione) per il parametro.
- d. (Facoltativo) Ad esempio, riscalda l'istanza, aggiorna il valore di riscaldamento dell'istanza in base alle esigenze.

- e. (Facoltativo) Selezionare Disable scale in (Disabilita riduzione) per creare solo una policy di espansione. In questo modo, se desiderato, è possibile creare una policy di riduzione separata di tipo diverso.
7. Procedi e crea il gruppo con dimensionamento automatico. La policy di dimensionamento verrà creata dopo la creazione del gruppo con scalabilità automatica.

Creazione di una policy di dimensionamento con monitoraggio degli obiettivi per un gruppo con scalabilità automatica esistente

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Verifica che i limiti di dimensionamento siano impostati in modo appropriato. Ad esempio, se la capacità desiderata del gruppo è già al suo valore massimo, specifica un nuovo valore massimo per il dimensionamento orizzontale. Per ulteriori informazioni, consulta [Impostare i limiti di dimensionamento per il gruppo con dimensionamento automatico](#).
4. Nella scheda Automatic scaling (Scalabilità automatica) scegli Add policy (Aggiungi policy) in Dynamic scaling policies (Policy di dimensionamento dinamico).
5. Per definire una policy, effettua le operazioni seguenti:
 - a. Per Tipo di policy lascia l'impostazione di default per il Target tracking scaling (Dimensionamento con monitoraggio degli obiettivi).
 - b. Specificare un nome per la policy.
 - c. Scegliere un parametro per Tipo parametro. È possibile scegliere un solo tipo di parametro. Per utilizzare più di un parametro, crea più policy.

Se hai scelto Application Load Balancer request count per target (Conteggio delle richieste Application Load Balancer per destinazione), scegli un gruppo destinazione in Target group (Gruppo di destinazione).

- d. Specifica un Target value (Valore di destinazione) per il parametro.
- e. (Facoltativo) Ad esempio il riscaldamento dell'istanza, aggiorna il valore di riscaldamento dell'istanza secondo necessità.

- f. (Facoltativo) Selezionare Disable scale in (Disabilita riduzione) per creare solo una policy di espansione. In questo modo, se desiderato, è possibile creare una policy di riduzione separata di tipo diverso.
6. Scegli Crea.

AWS CLI

Per creare una politica di ridimensionamento del tracciamento degli obiettivi, puoi utilizzare il seguente esempio per iniziare. Sostituisci ciascun *placeholder input dell'utente* con le tue informazioni.

Note

Per ulteriori esempi, consulta [Policy di dimensionamento di esempio per AWS Command Line Interface \(AWS CLI\)](#).

Creazione di una policy di dimensionamento con monitoraggio degli obiettivi (AWS CLI)

1. Utilizzate il `cat` comando seguente per memorizzare un valore target per la vostra politica di scalabilità e una specifica metrica predefinita in un file JSON denominato `config.json` nella vostra home directory. Di seguito è riportato un esempio di configurazione di tracciamento degli obiettivi che mantiene l'utilizzo medio della CPU al 50%.

```
$ cat ~/config.json
{
  "TargetValue": 50.0,
  "PredefinedMetricSpecification":
  {
    "PredefinedMetricType": "ASGAverageCPUUtilization"
  }
}
```

Per ulteriori informazioni, consulta il riferimento [PredefinedMetricSpecification](#) all'API Amazon EC2 Auto Scaling.

2. Utilizza il `put-scaling-policy` comando, insieme al `config.json` file creato nel passaggio precedente, per creare la tua politica di scalabilità.

```
aws autoscaling put-scaling-policy --policy-name cpu50-target-tracking-scaling-policy \
  --auto-scaling-group-name my-asg --policy-type TargetTrackingScaling \
  --target-tracking-configuration file://config.json
```

In caso di successo, questo comando restituisce gli ARN e i nomi dei due CloudWatch allarmi creati per tuo conto.

```
{
  "PolicyARN": "arn:aws:autoscaling:us-west-2:123456789012:scalingPolicy:228f02c2-c665-4bfd-aaac-8b04080bea3c:autoScalingGroupName/my-asg:policyName/cpu50-target-tracking-scaling-policy",
  "Alarms": [
    {
      "AlarmARN": "arn:aws:cloudwatch:us-west-2:123456789012:alarm:TargetTracking-my-asg-AlarmHigh-fc0e4183-23ac-497e-9992-691c9980c38e",
      "AlarmName": "TargetTracking-my-asg-AlarmHigh-fc0e4183-23ac-497e-9992-691c9980c38e"
    },
    {
      "AlarmARN": "arn:aws:cloudwatch:us-west-2:123456789012:alarm:TargetTracking-my-asg-AlarmLow-61a39305-ed0c-47af-bd9e-471a352ee1a2",
      "AlarmName": "TargetTracking-my-asg-AlarmLow-61a39305-ed0c-47af-bd9e-471a352ee1a2"
    }
  ]
}
```

Creazione di una policy di dimensionamento con monitoraggio degli obiettivi per Dimensionamento automatico Amazon EC2 tramite la matematica dei parametri

Utilizzando la matematica metrica, puoi interrogare più CloudWatch metriche e utilizzare espressioni matematiche per creare nuove serie temporali basate su queste metriche. Puoi visualizzare le serie temporali risultanti nella CloudWatch console e aggiungerle ai dashboard. Per ulteriori informazioni sulla matematica dei parametri, consulta [Using metric Math nella Amazon User Guide](#). CloudWatch

Alle espressioni matematiche dei parametri si applicano le seguenti considerazioni:

- Puoi interrogare qualsiasi metrica disponibile. CloudWatch Ogni parametro è una combinazione univoca di nome del parametro, spazio dei nomi e nessuna o più dimensioni.
- È possibile utilizzare qualsiasi operatore aritmetico (+ - */^), funzione statistica (come AVG o SUM) o altra funzione che supporti. CloudWatch
- È possibile utilizzare i parametri e i risultati di altre espressioni matematiche nelle formule dell'espressione matematica.
- Qualsiasi espressione utilizzata in una specifica dei parametri deve restituire una singola serie temporale.
- È possibile verificare che un'espressione matematica metrica sia valida utilizzando la console o l'API. CloudWatch CloudWatch [GetMetricData](#)

Note

Puoi creare una politica di ridimensionamento del tracciamento degli obiettivi utilizzando la matematica metrica solo se utilizzi l'SDK o un SDK. AWS CLI Questa funzionalità non è ancora disponibile nella console e. AWS CloudFormation

Esempio: backlog della coda di Amazon SQS per istanza

Per calcolare il backlog della coda di Amazon SQS per istanza, prendi il numero approssimativo di messaggi disponibili per il recupero dalla coda e dividi tale numero per la capacità operativa del gruppo con scalabilità automatica, ossia il numero di istanze nello stato InService. Per ulteriori informazioni, consulta [Scalabilità basata su Amazon SQS](#).

La logica dell'espressione è questa:

$$\text{sum of (number of messages in the queue)/(number of InService instances)}$$

Quindi le informazioni sulle CloudWatch metriche sono le seguenti.

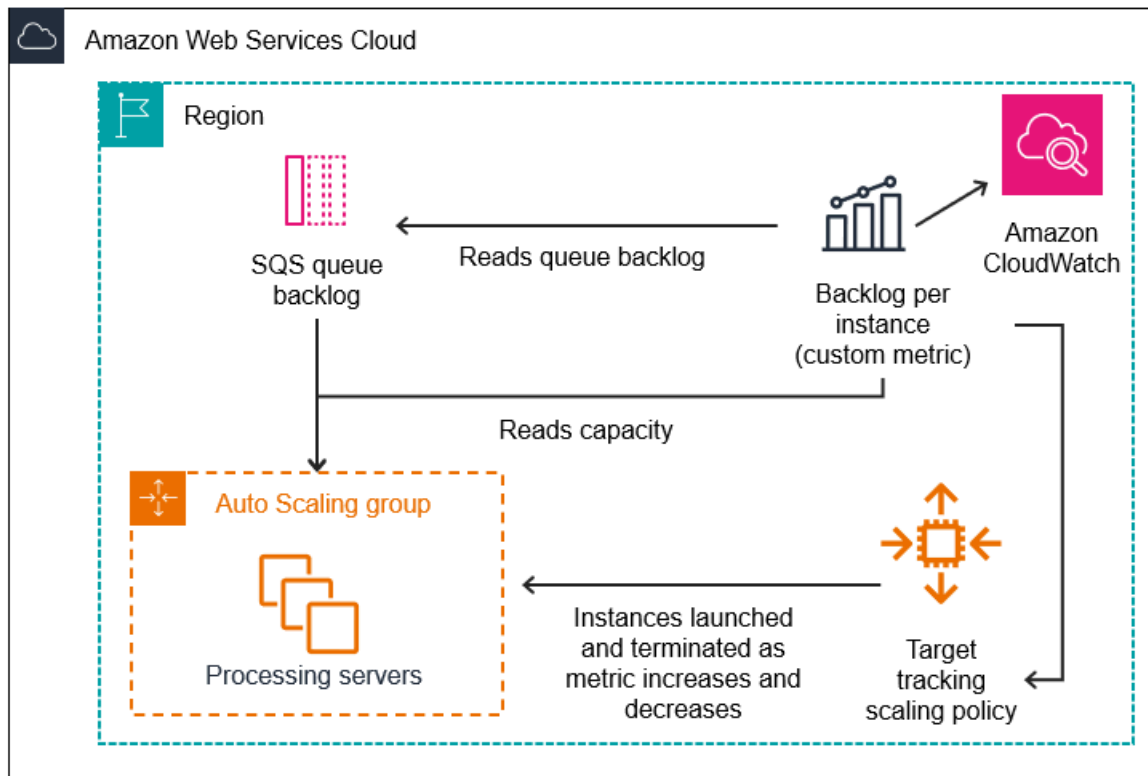
ID	CloudWatch metrico	Statistic	Periodo
m1	ApproximateNumberOfMessagesVisible	Somma	1 minuto

ID	CloudWatch metrico	Statistic	Periodo
m2	GroupInServiceInstances	Media	1 minuto

L'ID dell'operazione matematica sui parametri e l'espressione sono i seguenti.

ID	Expression
e1	$(m1)/(m2)$

Il diagramma seguente illustra l'architettura di questa metrica:



Per utilizzare questa matematica dei parametri al fine di creare una policy di dimensionamento con monitoraggio degli obiettivi (AWS CLI)

1. Memorizza l'espressione matematica dei parametri come parte di un specifico parametro personalizzato in un file JSON denominato `config.json`.

Utilizza la tabella seguente come guida. Sostituisci ciascun *placeholder input dell'utente* con le tue informazioni.

```
{
  "CustomizedMetricSpecification": {
    "Metrics": [
      {
        "Label": "Get the queue size (the number of messages waiting to be
processed)",
        "Id": "m1",
        "MetricStat": {
          "Metric": {
            "MetricName": "ApproximateNumberOfMessagesVisible",
            "Namespace": "AWS/SQS",
            "Dimensions": [
              {
                "Name": "QueueName",
                "Value": "my-queue"
              }
            ]
          },
          "Stat": "Sum"
        },
        "ReturnData": false
      },
      {
        "Label": "Get the group size (the number of InService instances)",
        "Id": "m2",
        "MetricStat": {
          "Metric": {
            "MetricName": "GroupInServiceInstances",
            "Namespace": "AWS/AutoScaling",
            "Dimensions": [
              {
                "Name": "AutoScalingGroupName",
                "Value": "my-asg"
              }
            ]
          },
          "Stat": "Average"
        },
        "ReturnData": false
      }
    ]
  }
}
```

```

    },
    {
      "Label": "Calculate the backlog per instance",
      "Id": "e1",
      "Expression": "m1 / m2",
      "ReturnData": true
    }
  ]
},
"TargetValue": 100
}

```

Per ulteriori informazioni, consulta il riferimento [TargetTrackingConfiguration](#) all'API Amazon EC2 Auto Scaling.

Note

Di seguito sono riportate alcune risorse aggiuntive che possono aiutarti a trovare nomi di metriche, namespace, dimensioni e statistiche relative alle metriche: CloudWatch

- Per informazioni sui parametri disponibili per AWS i servizi, consulta i [AWS servizi che pubblicano CloudWatch metriche](#) nella Amazon CloudWatch User Guide.
- [Per ottenere il nome esatto della metrica, lo spazio dei nomi e le dimensioni \(se applicabili\) di una CloudWatch metrica con, consulta list-metrics. AWS CLI](#)

2. Per creare questo criterio, esegui il [put-scaling-policy](#) comando utilizzando il file JSON come input, come illustrato nell'esempio seguente.

```

aws autoscaling put-scaling-policy --policy-name sqs-backlog-target-tracking-scaling-policy \
  --auto-scaling-group-name my-asg --policy-type TargetTrackingScaling \
  --target-tracking-configuration file://config.json

```

In caso di successo, questo comando restituisce l'Amazon Resource Name (ARN) della policy e gli ARN dei due CloudWatch allarmi creati per tuo conto.

```

{
  "PolicyARN": "arn:aws:autoscaling:us-
west-2:123456789012:scalingPolicy:228f02c2-c665-4bfd-

```

```
aaac-8b04080bea3c:autoScalingGroupName/my-asg:policyName/sqs-backlog-target-tracking-scaling-policy",
  "Alarms": [
    {
      "AlarmARN": "arn:aws:cloudwatch:us-west-2:123456789012:alarm:TargetTracking-my-asg-AlarmHigh-fc0e4183-23ac-497e-9992-691c9980c38e",
      "AlarmName": "TargetTracking-my-asg-AlarmHigh-fc0e4183-23ac-497e-9992-691c9980c38e"
    },
    {
      "AlarmARN": "arn:aws:cloudwatch:us-west-2:123456789012:alarm:TargetTracking-my-asg-AlarmLow-61a39305-ed0c-47af-bd9e-471a352ee1a2",
      "AlarmName": "TargetTracking-my-asg-AlarmLow-61a39305-ed0c-47af-bd9e-471a352ee1a2"
    }
  ]
}
```

Note

Se questo comando genera un errore, assicurati di averlo aggiornato AWS CLI localmente alla versione più recente.

Policy di dimensionamento per fasi e semplice per Dimensionamento automatico Amazon EC2

La scalabilità a fasi e le semplici politiche di scalabilità scalano la capacità del gruppo Auto Scaling in incrementi predefiniti in base agli allarmi. CloudWatch È possibile definire policy di dimensionamento separate per gestire il dimensionamento orizzontale (aumento della capacità) e il ridimensionamento (riduzione della capacità) in caso di superamento di una soglia di allarme.

Con Step Scaling e Simple Scaling, puoi creare e gestire gli allarmi che richiamano il processo di scalabilità. CloudWatch Quando viene violato un allarme, Amazon EC2 Auto Scaling avvia la politica di scalabilità associata a tale allarme.

Ti consigliamo vivamente di utilizzare le politiche di scalabilità di Target Tracking per adattare parametri come l'utilizzo medio della CPU o il numero medio di richieste per target. I parametri che

diminuiscono quando la capacità aumenta e aumentano quando la capacità diminuisce possono essere utilizzati per dimensionare proporzionalmente il numero di istanze mediante il monitoraggio degli obiettivi. Questo contribuisce a garantire che Dimensionamento automatico Amazon EC2 segua da vicino la curva della domanda per le applicazioni. Per ulteriori informazioni, consulta [Policy di dimensionamento con monitoraggio degli obiettivi](#).

Indice

- [Come funzionano le policy di dimensionamento a fasi](#)
- [Adeguaamenti delle fasi per il dimensionamento per fasi](#)
- [Tipi di regolazioni per il dimensionamento](#)
- [Preparazione dell'istanza](#)
- [Considerazioni](#)
- [Crea una politica di scalabilità graduale per la scalabilità orizzontale](#)
- [Crea una politica di scalabilità graduale per la scalabilità in](#)
- [Policy di dimensionamento semplice](#)

Come funzionano le policy di dimensionamento a fasi

Per utilizzare la scalatura a gradini, devi prima creare un CloudWatch allarme che monitora una metrica per il tuo gruppo Auto Scaling. Definisci il parametro, il valore di soglia e il numero di periodi di valutazione che determinano una violazione dell'allarme. Quindi, crea una politica di scalabilità dei passaggi che definisca come ridimensionare il gruppo quando viene superata la soglia di allarme.

Aggiungi le modifiche ai passaggi nella policy. È possibile definire diverse regolazioni delle fasi in base alla dimensione dell'allarme di violazione. Per esempio:

- Ridimensiona di 10 istanze se la metrica degli allarmi raggiunge il 60 per cento
- Scalabilità orizzontale di 30 istanze se la metrica degli allarmi raggiunge il 75 per cento
- Scalabilità orizzontale di 40 istanze se la metrica degli allarmi raggiunge l'85 per cento

Quando la soglia di allarme viene superata per il numero specificato di periodi di valutazione, Amazon EC2 Auto Scaling applicherà le modifiche ai passaggi definite nella policy. Le regolazioni possono continuare in caso di ulteriori violazioni degli allarmi fino al ripristino dello stato di allarme su OK.

Ogni istanza prevede un periodo di riscaldamento per evitare che le attività di scalabilità siano troppo reattive ai cambiamenti che si verificano in brevi periodi di tempo. Facoltativamente, puoi

configurare il periodo di riscaldamento per la tua politica di scalabilità. Tuttavia, consigliamo di utilizzare il riscaldamento predefinito dell'istanza per semplificare l'aggiornamento di tutte le politiche di ridimensionamento quando cambia il tempo di riscaldamento. Per ulteriori informazioni, consulta [Impostazione della preparazione di default dell'istanza per un gruppo con scalabilità automatica](#).

Le politiche di scalabilità semplici sono simili alle politiche di scalabilità graduale, tranne per il fatto che si basano su un singolo aggiustamento del ridimensionamento, con un periodo di tempo di recupero tra ogni attività di scalabilità. Per ulteriori informazioni, consulta [Policy di dimensionamento semplice](#).

Adeguamenti delle fasi per il dimensionamento per fasi

Quando si crea una policy di dimensionamento per fasi, è possibile specificare una o più regolazioni per fasi che dimensionano automaticamente il numero di istanze in modo dinamico in base all'utilizzo fuori limite segnalato dall'allarme. Ogni regolazione per fasi specifica quanto segue:

- Un limite inferiore per il valore del parametro
- Un limite superiore per il valore del parametro
- La quantità da dimensionare, in base al tipo di regolazione del dimensionamento

CloudWatch aggrega i punti dati metrici in base alla statistica della metrica associata all'allarme. CloudWatch Quando la soglia dell'allarme viene superata, viene richiamata la policy di dimensionamento appropriata. Amazon EC2 Auto Scaling applica il tipo di aggregazione ai punti CloudWatch dati delle metriche più recenti (anziché ai dati metrici grezzi). Confronta il valore dei parametri aggregati con i limiti superiore e inferiore definiti dagli adeguamenti delle fasi per determinare quali di queste eseguire.

Devi specificare i limiti superiore e inferiore relativi alla soglia dell'utilizzo fuori limite. Ad esempio, supponiamo che tu abbia creato un CloudWatch allarme e una politica di scalabilità orizzontale per quando la metrica è superiore al 50 per cento. Hai quindi creato un secondo allarme e una policy di riduzione orizzontale per quando il parametro è inferiore al 50 per cento. Hai apportato una serie di modifiche con un tipo di regolazione di `PercentChangeInCapacity` (o Percentuale di gruppo nella console) per ogni policy:

Esempio: Regolazioni per fasi per la policy di aumento orizzontale

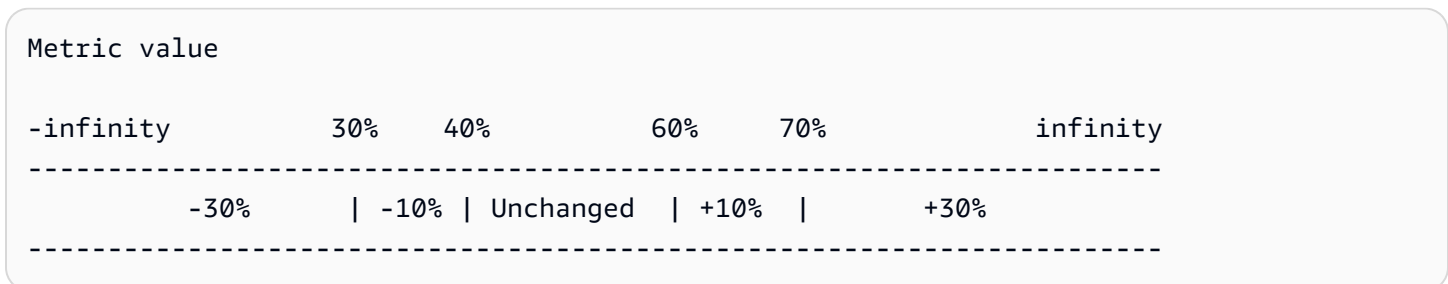
Limite inferiore	Limite superiore	Regolazione
0	10	0

Limite inferiore	Limite superiore	Regolazione
10	20	10
20	null	30

Esempio: Regolazioni per fasi della policy di riduzione orizzontale

Limite inferiore	Limite superiore	Regolazione
-10	0	0
-20	-10	-10
null	-20	-30

Questo crea la seguente configurazione di dimensionamento.



Supponiamo ora di utilizzare questa configurazione di scalabilità su un gruppo Auto Scaling che ha sia una capacità corrente che una capacità desiderata di 10. I seguenti punti riassumono il comportamento della configurazione di dimensionamento in relazione alla capacità desiderata e corrente del gruppo:

- La capacità corrente e desiderata vengono mantenute, mentre il valore del parametro aggregato è maggiore di 40 e minore di 60.
- Se il valore del parametro raggiunge 60, la capacità desiderata del gruppo aumenta di 1 istanza, fino a 11 istanze, in base alla seconda regolazione per fasi della policy di aumento orizzontale (aggiunta del 10% di 10 istanze). Dopo l'esecuzione della nuova istanza e la scadenza del tempo di riscaldamento specificato, la capacità attuale del gruppo aumenta a 11 istanze. Se il valore del parametro sale a 70 anche dopo questo aumento della capacità, la capacità desiderata del gruppo aumenta di altre 3 istanze, a 14. Questo si basa sulla modifica della terza fase della policy

di aumento orizzontale (aggiunta del 30% di 11 istanze, 3,3 istanze, arrotondato per difetto a 3 istanze).

- Se il valore del parametro arriva a 40, la capacità desiderata del gruppo diminuisce di 1 istanza, fino a 13 istanze, in base alla seconda regolazione per fasi della policy di riduzione orizzontale (rimozione del 10% di 14 istanze, ovvero 1,4 istanze, arrotondato per difetto a 1). Se il valore del parametro scende a 30 anche dopo questa diminuzione della capacità, la capacità desiderata del gruppo diminuisce di altre 3 istanze, a 10. Ciò si basa sulla regolazione della terza fase della policy di riduzione (rimuove il 30% delle 13 istanze, 3,9 istanze, arrotondate per difetto a 3).

Quando specifichi le regolazioni delle fasi per la policy di dimensionamento, tieni presente quanto segue:

- Se si utilizza il AWS Management Console, si specificano i limiti superiore e inferiore come valori assoluti. Se utilizzi AWS CLI o un SDK, specifichi i limiti superiore e inferiore relativi alla soglia di violazione.
- Gli intervalli delle regolazioni delle fasi non possono essere sovrapporsi o presentare scarti.
- Solo una regolazione delle fasi può avere un limite inferiore nullo (infinito negativo). Se una regolazione delle fasi presenta un limite inferiore negativo, allora deve esistere una regolazione con un limite inferiore nullo.
- Solo una regolazione delle fasi può avere un limite superiore nullo (infinito positivo). Se una regolazione delle fasi presenta un limite superiore positivo, deve esistere anche una regolazione con un limite superiore nullo.
- I limiti superiore e inferiore non possono essere nulli nella stessa regolazione delle fasi.
- Se il valore del parametro supera la soglia dell'utilizzo fuori limite, il limite inferiore è incluso e quello superiore è escluso. Se il valore del parametro è inferiore alla soglia dell'utilizzo fuori limite, il limite inferiore è escluso e il limite superiore è incluso.

Tipi di regolazioni per il dimensionamento

È possibile definire una policy di dimensionamento che esegue l'operazione di dimensionamento ottimale, in base al tipo di regolazione scelta. È possibile specificare il tipo di regolazione come percentuale della capacità corrente del gruppo con scalabilità automatica o in unità di capacità. Normalmente per unità di capacità si intende un'istanza, a meno che non si utilizzi la funzionalità di ponderazione delle istanze.

Dimensionamento automatico Amazon EC2 supporta i seguenti tipi di regolazioni per il dimensionamento per fasi e semplice:

- **ChangeInCapacity**: incrementa o decrementa la capacità corrente del gruppo del valore specificato. Un valore positivo incrementa la capacità, mentre un valore negativo della regolazione la decrementa. Ad esempio: se la capacità corrente del gruppo è 3 e la regolazione è 5, quando viene eseguita questa policy, aggiungiamo 5 unità di capacità, per un totale di 8 unità di capacità.
- **ExactCapacity**: modifica la capacità corrente del gruppo al valore specificato. Con questo tipo di adeguamento, specifica un valore non negativo. Ad esempio: se la capacità corrente del gruppo è 3 e la regolazione è 5, quando viene eseguita questa policy, la capacità viene modificata in 5 unità di capacità.
- **PercentChangeInCapacity**: incrementa o decrementa la capacità corrente del gruppo della percentuale specificata. Un valore positivo incrementa la capacità, mentre un valore negativo la decrementa. Ad esempio: se la capacità corrente è 10 e la regolazione è il 10%, quando viene eseguita questa policy, aggiungiamo 1 unità di capacità per un totale di 11 unità di capacità.

Note

Se il valore risultante non è intero, viene arrotondato come segue:

- I valori superiori a 1 vengono arrotondati per difetto. Ad esempio, 12.7 viene arrotondato in 12.
- I valori tra 0 e 1 vengono arrotondati a 1. Ad esempio, .67 viene arrotondato in 1.
- I valori tra 0 e -1 vengono arrotondati a -1. Ad esempio, -.58 viene arrotondato in -1.
- I valori inferiori a -1 vengono arrotondati per eccesso. Ad esempio, -6.67 viene arrotondato in -6.

Con **PercentChangeInCapacity**, puoi anche specificare il numero minimo di istanze da dimensionare con il parametro **MinAdjustmentMagnitude**. Ad esempio, supponi di creare una policy che aggiunge il 25% e di specificare un incremento minimo di 2 istanze. Se hai un gruppo con scalabilità automatica con 4 istanze e viene eseguita la policy di dimensionamento, il 25% di 4 è 1 istanza. Tuttavia, poiché hai specificato un incremento minimo di 2, vengono aggiunte 2 istanze.

Quando si utilizzano i [pesi delle istanze](#), l'effetto dell'impostazione del **MinAdjustmentMagnitude** parametro su un valore diverso da zero cambia. Il valore è espresso in unità di capacità. Per

impostare il numero minimo di istanze da dimensionare, scegli un valore per questo parametro pari almeno a quello del peso massimo dell'istanza.

Se utilizzate i pesi delle istanze, tenete presente che la capacità attuale del gruppo Auto Scaling può superare la capacità desiderata in base alle esigenze. Se il numero assoluto da decrementare, o la quantità che la percentuale indica di decrementare, è inferiore alla differenza tra la capacità corrente e quella desiderata, non viene eseguita alcuna operazione di dimensionamento. Nell'esaminare l'esito di una policy di dimensionamento quando viene superata la soglia di un allarme, dovrai tenere conto di questi comportamenti. Ad esempio, supponiamo che la capacità desiderata sia 30 e la capacità corrente sia 32. Quando la soglia dell'allarme viene superata, se la policy di dimensionamento decrementa la capacità desiderata di 1, non viene eseguita alcuna operazione di dimensionamento.

Preparazione dell'istanza

Nel caso di dimensionamento per fasi, è possibile specificare il numero di secondi necessari per la preparazione di un'istanza appena avviata. Fino alla scadenza del tempo di riscaldamento specificato, un'istanza non viene conteggiata nelle metriche aggregate delle istanze EC2 del gruppo Auto Scaling.

Durante il periodo di riscaldamento delle istanze, le politiche di scalabilità vengono ridimensionate solo se il valore della metrica delle istanze che non si stanno riscaldando è superiore alla soglia massima di allarme prevista dalla policy.

Se il gruppo si dimensiona di nuovo, le istanze in fase di preparazione vengono conteggiate come parte della capacità desiderata per l'attività di dimensionamento successiva. Pertanto, più utilizzi fuori limite degli allarmi che rientrano nella gamma della stessa regolazione per fasi determinano un'unica attività di dimensionamento. L'intenzione è di aumentare di continuo (ma non in eccesso).

Ad esempio, supponiamo che tu crei una policy in due passaggi. Il primo passaggio aggiunge il 10 per cento quando la metrica arriva al 60 e il secondo passaggio aggiunge il 30 per cento quando la metrica arriva al 70 per cento. Il tuo gruppo con dimensionamento automatico ha una capacità desiderata e corrente di 10. La capacità corrente e desiderata non cambiano, mentre il valore del parametro aggregato è minore di 60. Supponiamo che la metrica arrivi a 60, quindi venga aggiunta 1 istanza (il 10 per cento di 10 istanze). Quindi, la metrica arriva a 62 mentre la nuova istanza è ancora in fase di preparazione. La policy di dimensionamento calcola la nuova capacità desiderata in base alla capacità attuale, che è ancora pari a 10. Tuttavia, la capacità desiderata del gruppo è già aumentata a 11 istanze, perciò la policy di dimensionamento non la incrementa ulteriormente. Se il parametro raggiunge 70 mentre la nuova istanza è ancora in preparazione, dovremmo aggiungere 3

istanze (il 30% di 10 istanze). Tuttavia, la capacità desiderata del gruppo è già 11 perciò, per arrivare alla nuova capacità desiderata di 13 istanze, ne aggiungiamo solo 2.

Mentre l'attività di dimensionamento orizzontale è in corso, tutte le attività di riduzione orizzontale avviate dalle policy di ridimensionamento vengono bloccate fino al termine della preparazione delle istanze. Al termine della preparazione delle istanze, se si verifica un evento di dimensionamento, tutte le istanze attualmente in fase di chiusura verranno conteggiate ai fini della capacità attuale del gruppo nel calcolo della nuova capacità desiderata. Pertanto, non rimuoviamo dal gruppo Auto Scaling più istanze del necessario. Ad esempio, quando un'istanza è già terminata, se viene superata una soglia di regolazione dello stesso passaggio che ha diminuito la capacità desiderata di 1, non viene eseguita alcuna operazione di dimensionamento.

Valore predefinito

Se non viene impostato alcun valore, la politica di scalabilità utilizzerà il valore predefinito, che è il valore per il riscaldamento dell'[istanza predefinito definito](#) per il gruppo. [Se il riscaldamento dell'istanza predefinito è nullo, torna al valore del cooldown predefinito.](#)

Considerazioni

Le seguenti considerazioni si applicano quando si usano le policy di dimensionamento semplice a fasi:

- Valuta se riesci a prevedere le regolazioni dei passaggi sull'applicazione in modo sufficientemente accurato da utilizzare il dimensionamento a fasi. Se il parametro di dimensionamento aumenta o diminuisce in proporzione alla capacità della destinazione scalabile, ti consigliamo di utilizzare una policy di dimensionamento di monitoraggio della destinazione. È comunque possibile utilizzare il dimensionamento per fasi come policy aggiuntiva per una configurazione più avanzata. Ad esempio, puoi configurare una risposta più aggressiva quando l'utilizzo raggiunge un determinato livello.
- Assicurati di scegliere un margine adeguato tra le soglie di scalabilità orizzontale e ridimensionamento per evitare che si verifichino sbalzi. Flapping è un ciclo infinito di riduzione e aumento orizzontale. Ciò significa che se viene eseguita un'azione di dimensionamento, il valore della metrica cambierebbe per avviare un'altra azione di dimensionamento nella direzione opposta.

Crea una politica di scalabilità graduale per la scalabilità orizzontale

Per creare una politica di scalabilità in fasi per la scalabilità orizzontale per il gruppo Auto Scaling, utilizzate uno dei seguenti metodi:

Console

Fase 1: creare un CloudWatch allarme per la soglia metrica alta

1. Apri la CloudWatch console all'indirizzo <https://console.aws.amazon.com/cloudwatch/>.
2. Se necessario, modificare la regione . Nella barra di navigazione seleziona la regione in cui si trova il gruppo con scalabilità automatica.
3. Nel riquadro di navigazione, scegli Alarms (Allarmi), quindi Create alarm (Crea allarme).
4. Scegli Select Metric (Seleziona parametro).
5. Nella scheda All metrics (Tutti i parametri), scegli EC2, By Auto Scaling Group (Per gruppo con scalabilità automatica) e inserisci il nome del gruppo con scalabilità automatica nel campo di ricerca. Quindi, seleziona CPUUtilization e scegli Select metric (Seleziona parametro). Viene visualizzata la pagina Specify metric and conditions (Specifica parametro e condizioni) contenente un grafico e altre informazioni sul parametro.
6. In Period (Periodo), scegliere il periodo di valutazione per l'allarme, ad esempio 1 minuto. Durante la valutazione dell'allarme, ogni periodo è aggregato in un punto dati.

Note

Un periodo più breve crea un allarme più sensibile.

7. In Condizioni, effettuare le seguenti operazioni:
 - For Threshold type (Tipo di soglia), scegli Static (Statica).
 - Per Whenever **CPUUtilization** is, specifica se desideri che il valore della metrica sia maggiore o maggiore o uguale alla soglia per violare l'allarme. Quindi, per than (di), inserisci il valore di soglia che deve attivare l'allarme.

Important

Per utilizzare un allarme con una policy di aumento orizzontale (parametro superiore), assicurati di non scegliere minore di o minore di o uguale alla soglia.

8. In Configurazione aggiuntiva, eseguire le operazioni seguenti:
 - Per Datapoints to alarm (Punti dati per allarme), inserisci il numero di punti di dati (periodi di valutazione) durante i quali il valore del parametro deve corrispondere alle condizioni

di soglia per l'allarme. Ad esempio, due periodi consecutivi di 5 minuti fanno sì che trascorrono 10 minuti prima che lo stato di allarme venga richiamato.

- In Missing data treatment (Trattamento dei dati mancanti), scegli Treat missing data as bad (breaching threshold) (Considera i dati mancanti come non validi - superamento soglia).

Per ulteriori informazioni, consulta [Configurazione del modo in cui gli CloudWatch allarmi trattano i dati mancanti](#) nella Amazon CloudWatch User Guide.

9. Seleziona Successivo.

Viene visualizzata la pagina Configure actions (Configura operazioni).

10. In Notification (Notifica), seleziona un argomento Amazon SNS per segnalare quando l'allarme si trova nello stato ALARM, OK o INSUFFICIENT_DATA.

Per fare in modo che l'allarme invii più notifiche per lo stesso stato di allarme o per stati di allarme diversi, scegli Add notification (Aggiungi notifica).

Per fare in modo che l'allarme non invii notifiche, scegli Remove (Rimuovi).

11. Puoi lasciare vuote le altre sezioni della pagina Configure actions (Configura operazioni). Lasciando vuote le altre sezioni viene creato un allarme che però non viene associato a una policy di dimensionamento. Puoi associare l'allarme a una policy di dimensionamento dalla console Dimensionamento automatico Amazon EC2.

12. Seleziona Successivo.

13. Inserisci un nome (ad esempio, Step-Scaling-AlarmHigh-AddCapacity) e facoltativamente, una descrizione per l'allarme, quindi scegli Next (Successivo).

14. Scegli Crea allarme.

Usa la seguente procedura per continuare da dove avevi interrotto dopo aver creato il CloudWatch allarme.

Fase 2: Creare una politica di scalabilità graduale per la scalabilità orizzontale

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Verifica che i limiti di dimensionamento siano impostati in modo appropriato. Ad esempio, se la capacità desiderata del gruppo è già al suo valore massimo, specifica un nuovo valore massimo per il dimensionamento orizzontale. Per ulteriori informazioni, consulta [Impostare i limiti di dimensionamento per il gruppo con dimensionamento automatico](#).
4. Nella scheda Automatic scaling (Scalabilità automatica) scegli Add policy (Aggiungi policy) in Dynamic scaling policies (Policy di dimensionamento dinamico).
5. Per Tipo di policy, scegli Step scaling, quindi specifica un nome per la policy.
6. Per CloudWatch Allarme, scegli la tua sveglia. Se non hai ancora creato un allarme, scegli Crea un CloudWatch allarme e completa i passaggi da 4 a 14 della procedura precedente per creare un allarme.
7. Specificare la modifica della dimensione del gruppo corrente che questa policy apporterà quando viene eseguita con Take the action (Esegui l'operazione). È possibile aggiungere un numero specifico di istanze o una percentuale della dimensione del gruppo esistente oppure impostare il gruppo su una dimensione specifica.

Ad esempio, per creare una politica di scalabilità orizzontale che aumenti la capacità del gruppo del 30 per cento, scegli Add, inserisci 30 nel campo successivo e quindi scegli percent of group. Per impostazione predefinita, il limite inferiore di questa regolazione per fasi è la soglia di allarme e il limite superiore è infinito positivo (+).

8. Per aggiungere un'altra fase, scegliere Add step (Aggiungi fase), quindi definire la quantità di dimensionamento e i limiti inferiore e superiore della fase relativi alla soglia di allarme.
9. Per impostare un numero minimo di istanze da dimensionare, aggiorna il campo del numero in Add capacity units in increments of at least 1 capacity units (Aggiungi unità di capacità con incrementi di almeno 1 unità di capacità).
10. (Facoltativo) Ad esempio, aggiorna il valore di riscaldamento dell'istanza in base alle esigenze.
11. Scegli Crea.

AWS CLI

Per creare una politica di scalabilità graduale per la scalabilità orizzontale (aumento della capacità), è possibile utilizzare i seguenti comandi di esempio. Sostituisci ciascun *placeholder input dell'utente* con le tue informazioni.

Quando utilizzi AWS CLI, devi innanzitutto creare una politica di scalabilità graduale che fornisca istruzioni ad Amazon EC2 Auto Scaling su come eseguire la scalabilità orizzontale quando il

valore di una metrica aumenta. Quindi, crei l'allarme identificando la metrica da tenere d'occhio, definendo la soglia massima della metrica e altri dettagli per gli allarmi e associando l'allarme alla politica di scalabilità.

Fase 1: Creare una politica per la scalabilità orizzontale

Utilizzate il seguente [put-scaling-policy](#) comando per creare una politica di scalabilità denominata `my-step-scale-out-policy`, con un tipo di regolazione `PercentChangeInCapacity` che aumenti la capacità del gruppo in base alle seguenti regolazioni delle fasi (presupponendo una soglia di CloudWatch allarme del 60 per cento):

- Aumenta il numero delle istanze del 10% quando il valore del parametro è maggiore di o uguale al 60%, ma inferiore al 75%
- Aumenta il numero delle istanze del 20% quando il valore del parametro è maggiore di o uguale al 75%, ma inferiore all'85%
- Aumenta il numero delle istanze del 30% quando il valore del parametro è maggiore di o uguale all'85%

```
aws autoscaling put-scaling-policy \  
  --auto-scaling-group-name my-asg \  
  --policy-name my-step-scale-out-policy \  
  --policy-type StepScaling \  
  --adjustment-type PercentChangeInCapacity \  
  --metric-aggregation-type Average \  
  --step-adjustments  
MetricIntervalLowerBound=0.0,MetricIntervalUpperBound=15.0,ScalingAdjustment=10 \  
  
MetricIntervalLowerBound=15.0,MetricIntervalUpperBound=25.0,ScalingAdjustment=20 \  
  MetricIntervalLowerBound=25.0,ScalingAdjustment=30 \  
  --min-adjustment-magnitude 1
```

Registrare l'Amazon Resource Name (ARN) della policy. Ne hai bisogno per creare un CloudWatch allarme per la politica.

```
{  
  "PolicyARN":  
  "arn:aws:autoscaling:region:123456789012:scalingPolicy:4ee9e543-86b5-4121-b53b-aa4c23b5bbcc:autoScalingGroupName/my-asg:policyName/my-step-scale-in-policy  
}
```

Fase 2: Creare un CloudWatch allarme per la soglia metrica più alta

Utilizzate il seguente CloudWatch [put-metric-alarm](#) comando per creare un allarme che aumenti la dimensione del gruppo Auto Scaling in base a un valore di soglia medio della CPU del 60 per cento per almeno due periodi di valutazione consecutivi di due minuti. Per utilizzare un parametro personalizzato, specificane il nome in `--metric-name` e lo spazio dei nomi in `--namespace`.

```
aws cloudwatch put-metric-alarm --alarm-name Step-Scaling-AlarmHigh-AddCapacity \  
  --metric-name CPUUtilization --namespace AWS/EC2 --statistic Average \  
  --period 120 --evaluation-periods 2 --threshold 60 \  
  --comparison-operator GreaterThanOrEqualToThreshold \  
  --dimensions "Name=AutoScalingGroupName,Value=my-asg" \  
  --alarm-actions PolicyARN
```


Crea una politica di scalabilità graduale per la scalabilità in

Per creare una politica di scalabilità graduale per il gruppo Auto Scaling, utilizzate uno dei seguenti metodi:

Console

Fase 1: creare un CloudWatch allarme per la soglia metrica minima


1. Apri la CloudWatch console all'indirizzo <https://console.aws.amazon.com/cloudwatch/>.
2. Se necessario, modificare la regione . Nella barra di navigazione seleziona la regione in cui si trova il gruppo con scalabilità automatica.
3. Nel riquadro di navigazione, scegli Alarms (Allarmi), quindi Create alarm (Crea allarme).
4. Scegli Select Metric (Seleziona parametro).
5. Nella scheda All metrics (Tutti i parametri), scegli EC2, By Auto Scaling Group (Per gruppo con scalabilità automatica) e inserisci il nome del gruppo con scalabilità automatica nel campo di ricerca. Quindi, seleziona CPUUtilization e scegli Select metric (Seleziona parametro). Viene visualizzata la pagina Specify metric and conditions (Specifica parametro e condizioni) contenente un grafico e altre informazioni sul parametro.
6. In Period (Periodo), scegliere il periodo di valutazione per l'allarme, ad esempio 1 minuto. Durante la valutazione dell'allarme, ogni periodo è aggregato in un punto dati.

 Note

Un periodo più breve crea un allarme più sensibile.

7. In Condizioni, effettuare le seguenti operazioni:

- For Threshold type (Tipo di soglia), scegli Static (Statica).
- Per Whenever **CPUUtilization** is, specifica se desideri che il valore della metrica sia inferiore o inferiore o uguale alla soglia per violare l'allarme. Quindi, per than (di), inserisci il valore di soglia che deve attivare l'allarme.

 Important

Per utilizzare un allarme con una policy di riduzione orizzontale (parametro inferiore), assicurati di non scegliere maggiore di o maggiore di o uguale alla soglia.

8. In Configurazione aggiuntiva, eseguire le operazioni seguenti:

- Per Datapoints to alarm (Punti dati per allarme), inserisci il numero di punti di dati (periodi di valutazione) durante i quali il valore del parametro deve corrispondere alle condizioni di soglia per l'allarme. Ad esempio, due periodi consecutivi di 5 minuti fanno sì che trascorrono 10 minuti prima che lo stato di allarme venga richiamato.
- In Missing data treatment (Trattamento dei dati mancanti), scegli Treat missing data as bad (breaching threshold) (Considera i dati mancanti come non validi - superamento soglia). Per ulteriori informazioni, consulta [Configurazione del modo in cui gli CloudWatch allarmi trattano i dati mancanti](#) nella Amazon CloudWatch User Guide.

9. Seleziona Successivo.

Viene visualizzata la pagina Configure actions (Configura operazioni).

10. In Notification (Notifica), seleziona un argomento Amazon SNS per segnalare quando l'allarme si trova nello stato ALARM, OK o INSUFFICIENT_DATA.

Per fare in modo che l'allarme invii più notifiche per lo stesso stato di allarme o per stati di allarme diversi, scegli Add notification (Aggiungi notifica).

Per fare in modo che l'allarme non invii notifiche, scegli Remove (Rimuovi).

11. Puoi lasciare vuote le altre sezioni della pagina Configure actions (Configura operazioni). Lasciando vuote le altre sezioni viene creato un allarme che però non viene associato a una policy di dimensionamento. Puoi associare l'allarme a una policy di dimensionamento dalla console Dimensionamento automatico Amazon EC2.
12. Seleziona Successivo.
13. Inserisci un nome (ad esempio, Step-Scaling-AlarmLow-RemoveCapacity) e facoltativamente, una descrizione per l'allarme, quindi scegli Next (Successivo).
14. Scegli Crea allarme.

Usa la seguente procedura per continuare da dove avevi interrotto dopo aver creato l' CloudWatch allarme.

Fase 2: Creare una politica di scalabilità graduale per la scalabilità in

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.
3. Verifica che i limiti di dimensionamento siano impostati in modo appropriato. Ad esempio, se la capacità desiderata del gruppo è già al minimo, è necessario specificare un nuovo minimo per poter ampliare la capacità. Per ulteriori informazioni, consulta [Impostare i limiti di dimensionamento per il gruppo con dimensionamento automatico](#).
4. Nella scheda Automatic scaling (Scalabilità automatica) scegli Add policy (Aggiungi policy) in Dynamic scaling policies (Policy di dimensionamento dinamico).
5. Per Tipo di policy, scegliete Step scaling, quindi specificate un nome per la policy.
6. Per CloudWatch Allarme, scegli la tua sveglia. Se non hai ancora creato un allarme, scegli Crea un CloudWatch allarme e completa i passaggi da 4 a 14 della procedura precedente per creare un allarme.
7. Specificare la modifica della dimensione del gruppo corrente che questa policy apporterà quando viene eseguita con Take the action (Esegui l'operazione). È possibile rimuovere un numero specifico di istanze o una percentuale della dimensione del gruppo esistente, oppure impostare il gruppo su una dimensione specifica.

Ad esempio, per creare una politica scalabile che riduca la capacità del gruppo di due istanze, scegli Remove, inserisci 2 nel campo successivo, quindi scegli. `capacity units`

Per impostazione predefinita, il limite superiore di questa regolazione per fasi è la soglia di allarme e il limite inferiore è infinito negativo (-).

8. Per aggiungere un'altra fase, scegliere Add step (Aggiungi fase), quindi definire la quantità di dimensionamento e i limiti inferiore e superiore della fase relativi alla soglia di allarme.
9. Scegli Crea.

AWS CLI

Per creare una politica di scalabilità graduale per la scalabilità in (riduzione della capacità), è possibile utilizzare i seguenti comandi di esempio. Sostituisci ciascun *placeholder input dell'utente* con le tue informazioni.

Quando utilizzi AWS CLI, devi innanzitutto creare una politica di scalabilità graduale che fornisca istruzioni ad Amazon EC2 Auto Scaling su come eseguire la scalabilità quando il valore di una metrica diminuisce. Quindi, crei l'allarme identificando la metrica da tenere d'occhio, definendo la soglia minima del parametro e altri dettagli per gli allarmi e associando l'allarme alla politica di scalabilità.

Fase 1: Creare una politica per la scalabilità

Usa il seguente [put-scaling-policy](#) comando per creare una politica di scalabilità a fasi denominata `my-step-scale-in-policy`, con un tipo di regolazione `ChangeInCapacity` che riduca la capacità del gruppo di 2 istanze quando l' CloudWatch allarme associato supera il valore di soglia metrica più basso.

```
aws autoscaling put-scaling-policy \  
  --auto-scaling-group-name my-asg \  
  --policy-name my-step-scale-in-policy \  
  --policy-type StepScaling \  
  --adjustment-type ChangeInCapacity \  
  --step-adjustments MetricIntervalUpperBound=0.0,ScalingAdjustment=-2
```

Registrare l'Amazon Resource Name (ARN) della policy. Ne hai bisogno per creare l' CloudWatch allarme per la politica.

```
{  
  "PolicyARN": "arn:aws:autoscaling:region:123456789012:scalingPolicy:ac542982-cbeb-4294-891c-a5a941dfa787:autoScalingGroupName/my-asg:policyName/my-step-scale-out-policy
```

```
}
```

Fase 2: Creare un CloudWatch allarme per la soglia metrica minima

Utilizzate il seguente CloudWatch [put-metric-alarm](#) comando per creare un allarme che riduca la dimensione del gruppo Auto Scaling in base al valore medio della soglia della CPU del 40 per cento per almeno due periodi di valutazione consecutivi di due minuti. Per utilizzare un parametro personalizzato, specificane il nome in `--metric-name` e lo spazio dei nomi in `--namespace`.

```
aws cloudwatch put-metric-alarm --alarm-name Step-Scaling-AlarmLow-RemoveCapacity \  
  --metric-name CPUUtilization --namespace AWS/EC2 --statistic Average \  
  --period 120 --evaluation-periods 2 --threshold 40 \  
  --comparison-operator LessThanOrEqualToThreshold \  
  --dimensions "Name=AutoScalingGroupName,Value=my-asg" \  
  --alarm-actions PolicyARN
```

Policy di dimensionamento semplice

Gli esempi seguenti mostrano come utilizzare i comandi CLI per creare semplici politiche di scalabilità. Rimangono in questo documento come riferimento per tutti i clienti che desiderano utilizzarli, ma consigliamo di utilizzare invece le politiche di tracciamento degli obiettivi o di scalabilità dei passaggi.

Analogamente alle politiche di scalabilità graduale, le politiche di scalabilità semplici richiedono la creazione di CloudWatch allarmi per le politiche di scalabilità. Nelle politiche create, è inoltre necessario definire se aggiungere o rimuovere le istanze e quante, oppure impostare il gruppo su una dimensione esatta.

Una delle principali differenze tra le politiche di scalabilità graduale e le politiche di scalabilità semplici è rappresentata dalle regolazioni dei passaggi che si ottengono con le politiche di scalabilità graduale. Con la scalatura per fasi, puoi apportare modifiche più o meno grandi alla dimensione del gruppo in base alle regolazioni dei passaggi specificate.

Una semplice politica di scalabilità deve inoltre attendere il completamento di un'attività di scalabilità in corso o la sostituzione di un controllo dello stato di salute e il termine di un [periodo di recupero prima di rispondere](#) a ulteriori allarmi. Al contrario, con la scalabilità graduale, la policy continua a rispondere a ulteriori allarmi, anche mentre è in corso un'attività di scalabilità o la sostituzione di un

controllo dello stato di salute. Ciò significa che Amazon EC2 Auto Scaling valuta tutte le violazioni degli allarmi non appena riceve i messaggi di allarme. Per questo motivo, ti consigliamo di utilizzare invece politiche di scalabilità graduale, anche se hai solo una singola modifica della scalabilità.

Inizialmente, Dimensionamento automatico Amazon EC2 supportava solo policy di dimensionamento semplice. Se hai creato la tua politica di scalabilità prima dell'introduzione delle politiche di tracciamento degli obiettivi e di scalabilità per fasi, la tua politica viene considerata una semplice politica di scalabilità.

Crea una semplice politica di scalabilità per la scalabilità orizzontale

Utilizzate il seguente [put-scaling-policy](#) comando per creare una semplice politica di scalabilità denominata `my-simple-scale-out-policy`, con un tipo di regolazione `PercentChangeInCapacity` che aumenti la capacità del gruppo del 30 per cento quando l' CloudWatch allarme associato supera il valore di soglia metrico più elevato.

```
aws autoscaling put-scaling-policy --policy-name my-simple-scale-out-policy \  
  --auto-scaling-group-name my-asg --scaling-adjustment 30 \  
  --adjustment-type PercentChangeInCapacity
```

Registrare l'Amazon Resource Name (ARN) della policy. Ne hai bisogno per creare l' CloudWatch allarme per la politica.

Crea una semplice politica di scalabilità per scalare

Utilizzate il seguente [put-scaling-policy](#) comando per creare una semplice politica di scalabilità denominata `my-simple-scale-in-policy`, con un tipo di regolazione `ChangeInCapacity` che riduca la capacità del gruppo di un'istanza quando l' CloudWatch allarme associato supera il valore di soglia metrico inferiore.

```
aws autoscaling put-scaling-policy --policy-name my-simple-scale-in-policy \  
  --auto-scaling-group-name my-asg --scaling-adjustment -1 \  
  --adjustment-type ChangeInCapacity --cooldown 180
```

Registrare l'Amazon Resource Name (ARN) della policy. Ne hai bisogno per creare l' CloudWatch allarme per la politica.

Tempo di raffreddamento per il dimensionamento per Dimensionamento automatico Amazon EC2

Important

Come best practice, ti consigliamo di non utilizzare policy di dimensionamento semplice e tempi di raffreddamento di dimensionamento. È preferibile utilizzare una policy di dimensionamento con monitoraggio degli obiettivi o una policy di dimensionamento a fasi per le prestazioni di dimensionamento. Per una policy di dimensionamento che modifichi le dimensioni del gruppo con scalabilità automatica in modo proporzionale man mano che il valore del parametro di dimensionamento diminuisce o aumenta, consigliamo di effettuare il [monitoraggio degli obiettivi](#) mediante dimensionamento semplice o per fasi.

Quando crei policy di dimensionamento semplici per il tuo gruppo con dimensionamento automatico, ti consigliamo di configurare contemporaneamente il tempo di attesa del ridimensionamento.

Dopo che il gruppo con scalabilità automatica avvia o termina le istanze, il servizio attende il termine del tempo di raffreddamento prima dell'avvio di ulteriori attività di dimensionamento. Il tempo di raffreddamento aiuta a stabilizzare il gruppo con dimensionamento automatico e a impedirgli che avvii o termini altre istanze prima che gli effetti dell'operazione di dimensionamento precedente siano visibili.

Supponiamo, ad esempio, che una policy di dimensionamento semplice per l'utilizzo della CPU consigli l'avvio di due istanze. Dimensionamento automatico Amazon EC2 avvia due istanze e quindi sospende le attività di dimensionamento fino al termine del tempo di raffreddamento. Al termine del tempo di raffreddamento, tutte le attività di dimensionamento avviate dalle policy di dimensionamento semplice possono riprendere. Se l'utilizzo della CPU viola nuovamente la soglia alta dell'allarme, il gruppo con scalabilità automatica si dimensiona nuovamente e il tempo di raffreddamento diventa nuovamente effettivo. Tuttavia, se due istanze fossero sufficienti a ridurre il valore del parametro, il gruppo rimane alla sua dimensione attuale.

Indice

- [Considerazioni](#)
- [Ulteriori ritardi causati dagli hook del ciclo di vita](#)
- [Modifica del tempo di raffreddamento di default](#)
- [Impostazione di un tempo di raffreddamento per policy di scalabilità semplici e specifiche](#)

Considerazioni

Le seguenti considerazioni si applicano quando si utilizzano policy di scalabilità semplici e tempi di raffreddamento di scalabilità:

- Le policy di monitoraggio degli obiettivi e di dimensionamento a fasi possono avviare immediatamente un'attività di dimensionamento orizzontale senza attendere la fine del tempo di raffreddamento. Invece, ogni volta che il gruppo Auto Scaling avvia delle istanze, le singole istanze hanno un periodo di riscaldamento. Per ulteriori informazioni, consulta [Impostazione della preparazione di default dell'istanza per un gruppo con scalabilità automatica](#).
- Quando un'azione pianificata inizia all'ora prevista, può anche iniziare un'attività di dimensionamento immediatamente senza aspettare che il tempo di raffreddamento finisca.
- Se un'istanza diventa non integra, Dimensionamento automatico Amazon EC2 non attende il termine del tempo di raffreddamento, per sostituirla.
- Quando più istanze vengono avviate o terminate, il tempo di raffreddamento (sia il raffreddamento predefinito che quello specifico della policy di dimensionamento) ha effetto a partire da quando l'ultima istanza termina l'avvio o la terminazione.
- Quando si dimensiona manualmente il gruppo con scalabilità automatica, l'impostazione predefinita non prevede la fine del tempo di raffreddamento. Tuttavia, puoi ignorare questo comportamento e rispettare il tempo di recupero predefinito quando utilizzi l'SDK o un SDK per scalare AWS CLI manualmente.
- Di default, Elastic Load Balancing attende 300 secondi per completare il processo di annullamento della registrazione (svuotamento connessione). Se il gruppo è dietro a un load balancer Elastic Load Balancing, attenderà la cancellazione della registrazione delle istanze terminanti prima di iniziare il periodo di tempo di raffreddamento.

Ulteriori ritardi causati dagli hook del ciclo di vita

Se un [hook del ciclo di vita](#) viene richiamato, il tempo di raffreddamento inizia dopo aver completato l'operazione del ciclo di vita o al termine del periodo di timeout. Per esempio, consideriamo un gruppo con scalabilità automatica che ha un hook del ciclo di vita per l'avvio di un'istanza. Quando l'applicazione incorre in un aumento della domanda, il gruppo avvia istanze per aggiungere capacità. Poiché esiste un hook del ciclo di vita, l'istanza viene messa in stato di attesa e le attività di dimensionamento dovute a policy di dimensionamento semplice vengono sospese. Quando l'istanza entra nello stato `InService`, inizia il tempo di raffreddamento. Al termine del tempo di raffreddamento, vengono riprese le attività delle policy di dimensionamento semplice

Quando Elastic Load Balancing è abilitato, ai fini della scalabilità, il periodo di cooldown inizia quando l'istanza selezionata per la terminazione inizia il drenaggio della connessione (ritardo di annullamento della registrazione). Il periodo di cooldown non attende il termine del drenaggio della connessione o il completamento dell'operazione da parte del lifecycle hook. Ciò significa che qualsiasi attività di scalabilità dovuta a semplici policy di scalabilità può riprendere non appena il risultato dell'evento di riduzione orizzontale si riflette nella capacità del gruppo. In caso contrario, decidere di attendere il completamento di tutte e tre le attività, ovvero svuotamento connessione, hook del ciclo di vita e tempo di raffreddamento, richiede al gruppo con scalabilità automatica un tempo significativamente superiore, per la sospensione del dimensionamento.

Modifica del tempo di raffreddamento di default

Non è possibile impostare il tempo di raffreddamento predefinito al momento della creazione iniziale di un gruppo con scalabilità automatica nella console Dimensionamento automatico Amazon EC2. Di default, il tempo di raffreddamento è impostato su 300 secondi (5 minuti). Se necessario, è possibile aggiornarlo dopo la creazione del gruppo.

Modificare il tempo di raffreddamento predefinito (console)

Dopo aver creato il gruppo con scalabilità automatica, nella scheda Details (Dettagli), scegli Advanced configurations (Configurazioni avanzate), quindi Edit (Modifica). Per Tempo di raffreddamento predefinito, scegli l'intervallo di tempo desiderato in base al tempo di avvio dell'istanza o ad altre esigenze applicative.

Modificare il tempo di raffreddamento predefinito (AWS CLI)

Utilizzare i comandi seguenti per modificare il tempo di raffreddamento predefinito per gruppi Auto Scaling nuovi o già esistenti. Se il tempo di preparazione di default non è definito, viene utilizzato il valore di default di 300 secondi.

- [create-auto-scaling-group](#)
- [update-auto-scaling-group](#)

Per confermare il valore del cooldown predefinito, usa il comando. [describe-auto-scaling-groups](#)

Impostazione di un tempo di raffreddamento per policy di scalabilità semplici e specifiche

Per impostazione predefinita, tutte le policy di dimensionamento semplice utilizzano il tempo di raffreddamento predefinito definito per il gruppo con scalabilità automatica. Per indicare un tempo di

raffreddamento specifico per policy di dimensionamento semplici e specifiche, utilizza il parametro di raffreddamento facoltativo durante la creazione o l'aggiornamento della policy. Quando viene specificato un tempo di raffreddamento per una policy, questo sostituisce il tempo di raffreddamento predefinito.

Un uso comune per tempo di raffreddamento specifico per il dimensionamento è con una policy di dimensionamento. Poiché questa policy termina le istanze, Dimensionamento automatico Amazon EC2 necessita di meno tempo, per stabilire se terminare le istanze aggiuntive. La terminazione delle istanze dovrebbe essere un'operazione molto più rapida, rispetto all'avvio delle istanze. Il tempo di raffreddamento di default di 300 secondi è quindi troppo lungo. In questo caso, un tempo di raffreddamento specifico per la policy di dimensionamento con un valore inferiore per le policy di riduzione orizzontale può aiutare a ridurre i costi, permettendo al gruppo di eseguire il dimensionamento più velocemente.

Per creare o aggiornare policy di dimensionamento semplice nella console, scegli la scheda Automatic scaling (Dimensionamento automatico) dopo aver creato il gruppo. Per creare o aggiornare semplici politiche di ridimensionamento utilizzando il AWS CLI, utilizzare il [put-scaling-policy](#) comando. Per ulteriori informazioni, consulta [Policy di dimensionamento per fasi e di dimensionamento semplice](#).

Scalabilità basata su Amazon SQS

Important

Le seguenti informazioni e passaggi mostrano come calcolare il backlog della coda di Amazon SQS per istanza utilizzando l'attributo `ApproximateNumberOfMessages` queue prima di pubblicarlo come metrica personalizzata su CloudWatch. Tuttavia, ora puoi ridurre i costi e limitare il lavoro necessario per pubblicare il parametro personalizzato utilizzando la matematica dei parametri. Per ulteriori informazioni, consulta [Creazione di una policy di dimensionamento con monitoraggio degli obiettivi per Dimensionamento automatico Amazon EC2 tramite la matematica dei parametri](#).

In questa sezione viene illustrato come dimensionare il gruppo con scalabilità automatica in risposta alle modifiche apportate al carico di sistema in una coda Amazon Simple Queue Service (Amazon SQS). Per ulteriori informazioni su come utilizzare Amazon SQS, consulta [Guida per sviluppatori Amazon Simple Queue](#).

Vi sono alcuni scenari in cui si potrebbe pensare al dimensionamento in risposta a un'attività in una coda Amazon SQS. Ad esempio, supponi di avere un'app Web che permette agli utenti di caricare immagini e di utilizzarle online. In questo scenario, ogni immagine richiede il ridimensionamento e la codifica, prima di poter essere pubblicata. L'app viene eseguita su istanze EC2 in un gruppo con scalabilità automatica configurato per gestire le tipiche percentuali di caricamento. Le istanze non integre vengono terminate e sostituite per mantenere i livelli di istanze correnti in qualsiasi momento. L'app colloca i dati bitmap grezzi delle immagini in una coda SQS per l'elaborazione. Le elabora e pubblica le immagini elaborate dove possono essere visualizzate dagli utenti. L'architettura per questo scenario funziona bene se il numero di caricamenti di immagini non varia nel tempo. Tuttavia, se il numero di caricamenti cambia nel tempo, per scalare la capacità del gruppo con scalabilità automatica, potresti prendere in considerazione l'utilizzo del dimensionamento dinamico.

Indice

- [Utilizzare il monitoraggio degli obiettivi con il giusto parametro](#)
- [Limitazioni e prerequisiti](#)
- [Configurazione del dimensionamento in base ad Amazon SQS](#)
- [Amazon SQS e la protezione per la riduzione orizzontale delle istanze](#)

Utilizzare il monitoraggio degli obiettivi con il giusto parametro

Se utilizzi una policy di dimensionamento con monitoraggio degli obiettivi basato su un parametro personalizzato della coda Amazon SQS, il dimensionamento dinamico può adattarsi in modo più efficace alla curva di domanda dell'applicazione. Per ulteriori informazioni sulla scelta dei parametri per il monitoraggio degli obiettivi, consulta [Selezionare i parametri..](#)

Il problema legato all'utilizzo di una metrica di CloudWatch Amazon SQS come `ApproximateNumberOfMessagesVisible` per il tracciamento delle destinazioni è che il numero di messaggi nella coda potrebbe non cambiare proporzionalmente alla dimensione del gruppo Auto Scaling che elabora i messaggi dalla coda. Ciò accade perché il numero di messaggi nella coda SQS non definisce esclusivamente il numero di istanze necessarie. Il numero di istanze nel gruppo con scalabilità automatica può essere determinato da più fattori, tra cui la quantità di tempo necessaria per elaborare un messaggio e la quantità di latenza (ritardo della coda) accettabile.

La soluzione è utilizzare un parametro backlog per instance (backlog per istanza) dove il valore di destinazione è il backlog accettabile per istanza da mantenere. È possibile calcolare questi valori nel modo seguente:

- **Backlog per instance (Backlog per istanza):** per calcolare il backlog per istanza, occorre partire dall'attributo della coda `ApproximateNumberOfMessages` per determinare la lunghezza della coda SQS (numero di messaggi disponibili per il recupero dalla coda). Per ottenere il backlog per istanza, questo numero va diviso per la capacità in esecuzione del parco istanze, che per un gruppo con scalabilità automatica è il numero di istanze nello stato `InService`.
- **Acceptable backlog per instance (Backlog accettabile per istanza):** per calcolare il valore di destinazione, prima determina ciò che l'applicazione può accettare in termini di latenza. Quindi dividi il valore di latenza accettabile per il tempo medio che un'istanza EC2 richiede per elaborare un messaggio.

Ad esempio, supponiamo che al momento si disponga di un gruppo con scalabilità automatica con 10 istanze e il numero di messaggi visibili nella coda (`ApproximateNumberOfMessages`) è 1500. Se il tempo di elaborazione medio è 0,1 secondi per ogni messaggio e la latenza massima accettabile è 10 secondi, il backlog per istanza accettabile è $10/0,1$, ovvero 100 messaggi. Questo significa che 100 è il valore di destinazione per la policy con monitoraggio degli obiettivi. Quando il backlog per istanza raggiunge il valore target, si verificherà un evento di aumentazione orizzontale. Poiché il backlog per istanza è già di 150 messaggi (1500 messaggi / 10 istanze), il gruppo si aumenta orizzontalmente e si aumenta orizzontalmente di cinque istanze per mantenere la proporzione con il valore target.

Nelle procedure seguenti viene illustrato come pubblicare il parametro personalizzato e creare una policy di dimensionamento con monitoraggio degli obiettivi che permetta di configurare il gruppo con scalabilità automatica in modo che venga dimensionato in base a questi calcoli.

Important

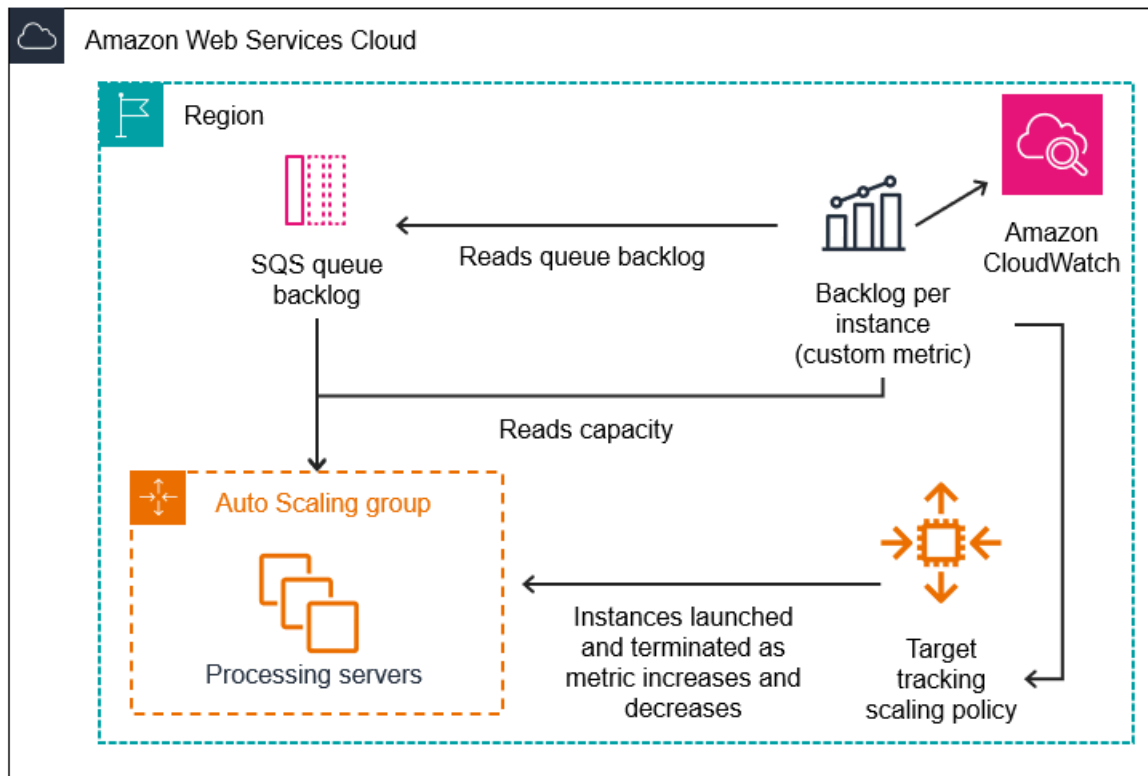
Ricorda di utilizzare la matematica dei parametri per ridurre i costi. Per ulteriori informazioni, consulta [Creazione di una policy di dimensionamento con monitoraggio degli obiettivi per Dimensionamento automatico Amazon EC2 tramite la matematica dei parametri](#).

Sono tre, le parti principali di questa configurazione:

- Un gruppo con scalabilità automatica per gestire le istanze EC2 per l'elaborazione dei messaggi provenienti dalla coda SQS.
- Una metrica personalizzata da inviare ad Amazon CloudWatch che misura il numero di messaggi in coda per istanza EC2 nel gruppo Auto Scaling.

- Una politica di tracciamento degli obiettivi che configura il gruppo Auto Scaling in modo che venga scalato in base alla metrica personalizzata e a un valore target impostato. CloudWatch gli allarmi richiamano la politica di scalabilità.

Nel diagramma seguente viene illustrata l'architettura di questa configurazione.



Limitazioni e prerequisiti

Per utilizzare questa configurazione, devi essere consapevole delle seguenti limitazioni:

- È necessario utilizzare AWS CLI o un SDK su cui pubblicare la metrica personalizzata. CloudWatch Puoi quindi monitorare la tua metrica con. AWS Management Console
- La console Dimensionamento automatico Amazon EC2 non supporta policy di dimensionamento con monitoraggio degli obiettivi che utilizzano parametri personalizzati. È necessario utilizzare AWS CLI o un SDK per specificare una metrica personalizzata per la politica di scalabilità.

Le seguenti sezioni ti spiegano come utilizzare i AWS CLI per le attività che devi eseguire. Ad esempio, per ottenere dati metrici che riflettano l'uso attuale della coda, si utilizza il comando [get-queue-attributes](#) SQS. Assicurati che la CLI sia [installata](#) e [configurata](#).

Prima di iniziare, dovrai disporre di una coda Amazon SQS da utilizzare. Le seguenti sezioni presuppongono che si disponga già di una coda (standard o FIFO), di un gruppo con scalabilità automatica e di istanze EC2 che eseguono l'applicazione che utilizza la coda. Per ulteriori informazioni su Amazon SQS, consulta la [Guida per gli sviluppatori di Amazon Simple Queue Service](#).

Configurazione del dimensionamento in base ad Amazon SQS

Attività

- [Passaggio 1: creare una metrica personalizzata CloudWatch](#)
- [Fase 2: creazione una policy di dimensionamento con monitoraggio degli obiettivi](#)
- [Fase 3: test delle policy di dimensionamento](#)

Passaggio 1: creare una metrica personalizzata CloudWatch

Un parametro personalizzato viene definito utilizzando uno spazio dei nomi e un nome parametro a scelta. Gli spazi dei nomi per i parametri personalizzati non possono iniziare con `AWS/`. Per ulteriori informazioni sulla pubblicazione di metriche personalizzate, consulta l'argomento [Pubblica metriche personalizzate](#) nella Amazon CloudWatch User Guide.

Segui questa procedura per creare la metrica personalizzata leggendo prima le informazioni dal tuo account. AWS Successivamente, calcola il parametro backlog per istanza, come suggerito in una sezione precedente. Infine, pubblica questo numero su con una granularità CloudWatch di 1 minuto. Quando possibile, consigliamo vivamente di dimensionare in base ai parametri con una granularità di 1 minuto, per garantire una risposta più rapida alle modifiche apportate al carico del sistema.

Per creare una metrica personalizzata () CloudWatch AWS CLI

1. Utilizzate il [get-queue-attributes](#) comando SQS per ottenere il numero di messaggi in attesa nella coda (). `ApproximateNumberOfMessages`

```
aws sqs get-queue-attributes --queue-url https://  
sqs.region.amazonaws.com/123456789/MyQueue \  
--attribute-names ApproximateNumberOfMessages
```

2. Utilizzate il [describe-auto-scaling-groups](#) comando per ottenere la capacità di esecuzione del gruppo, ovvero il numero di istanze nello stato del ciclo di vita. `InService` Questo comando restituisce le istanze di un gruppo con scalabilità automatica insieme allo stato del loro ciclo di vita.

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-names my-asg
```

3. Calcola il backlog per istanza dividendo il numero approssimativo di messaggi disponibili per il recupero dalla coda per la capacità operativa del gruppo.
4. Crea uno script che viene eseguito ogni minuto per recuperare il valore del backlog per istanza e pubblicarlo su una metrica personalizzata. CloudWatch Quando pubblichi una metrica personalizzata, specifichi il nome, lo spazio dei nomi, l'unità, il valore e nessuna o più dimensioni del parametro. Una dimensione è composta da un nome di dimensione e un valore di dimensione.

Per pubblicare la metrica personalizzata, sostituisci i valori segnaposto in *corsivo* con il nome della metrica preferita, il valore della metrica, uno spazio dei nomi (purché non inizi con "AWS«) e le dimensioni (opzionale), quindi esegui il comando seguente. [put-metric-data](#)

```
aws cloudwatch put-metric-data --metric-name MyBacklogPerInstance --
namespace MyNamespace \
  --unit None --value 20 --
dimensions MyOptionalMetricDimensionName=MyOptionalMetricDimensionValue
```

Dopo che l'applicazione ha emesso la metrica desiderata, i dati vengono inviati a CloudWatch. La metrica è visibile nella console. CloudWatch Puoi accedervi accedendo AWS Management Console e accedendo alla pagina. CloudWatch Puoi visualizzare il parametro passando alla pagina dei parametri oppure cercandolo nella casella di ricerca. Per informazioni sulla visualizzazione dei parametri, consulta [Visualizza i parametri disponibili](#) nella Amazon CloudWatch User Guide.

Fase 2: creazione una policy di dimensionamento con monitoraggio degli obiettivi

Il parametro creato può ora essere aggiunto a una policy di dimensionamento con monitoraggio degli obiettivi.

Creazione di una policy di dimensionamento con monitoraggio degli obiettivi (AWS CLI)

1. Utilizza il comando `cat` seguente per archiviare un valore di destinazione per la policy di dimensionamento e una specifica del parametro personalizzato in un file JSON con nome `config.json` nella directory principale. Sostituisci ciascun *placeholder input dell'utente* con le tue informazioni. Per `TargetValue`, calcola il backlog accettabile per accettabile e inseriscilo qui. Per calcolare questo numero, stabilisci un valore di latenza normale

e dividilo per il tempo medio necessario per elaborare un messaggio, come descritto in una precedente sezione.

Se non hai specificato alcuna dimensione per il parametro creato nel passaggio 1, non includere alcuna dimensione nel parametro specifico personalizzato.

```
$ cat ~/config.json
{
  "TargetValue":100,
  "CustomizedMetricSpecification":{
    "MetricName":"MyBacklogPerInstance",
    "Namespace":"MyNamespace",
    "Dimensions":[
      {
        "Name":"MyOptionalMetricDimensionName",
        "Value":"MyOptionalMetricDimensionValue"
      }
    ],
    "Statistic":"Average",
    "Unit":"None"
  }
}
```

2. Utilizza il [put-scaling-policy](#) comando, insieme al `config.json` file creato nel passaggio precedente, per creare la tua politica di scalabilità.

```
aws autoscaling put-scaling-policy --policy-name sqs100-target-tracking-scaling-policy \  
  --auto-scaling-group-name my-asg --policy-type TargetTrackingScaling \  
  --target-tracking-configuration file://~/config.json
```

Ciò crea due allarmi: uno per l'aumento e uno per la riduzione orizzontali. Restituisce inoltre l'Amazon Resource Name (ARN) della policy con cui è registrata CloudWatch, che CloudWatch utilizza per richiamare la scalabilità ogni volta che la soglia metrica viene violata.

Fase 3: test delle policy di dimensionamento

Una volta completata la configurazione, verifica che la policy di dimensionamento funzioni. Puoi testarla incrementando il numero di messaggi nella coda SQS e verificando che il gruppo con scalabilità automatica abbia avviato un'istanza EC2 aggiuntiva. Puoi testarla anche diminuendo il

numero di messaggi nella coda SQS e verificando che il gruppo con scalabilità automatica abbia terminato un'istanza EC2.

Per testare la funzione di aumento orizzontale

1. Segui i passaggi descritti in [Creazione di una coda Amazon SQS standard e invio di un messaggio](#) o [Creazione di una coda Amazon SQS FIFO e invio di un messaggio per aggiungere messaggi alla coda](#). Verifica di aver aumentato il numero di messaggi nella coda in modo che il parametro backlog per istanza superi il valore di destinazione.

Possono trascorrere alcuni minuti prima che le modifiche richiama l'allarme.

2. Usa il [describe-auto-scaling-groups](#) comando per verificare che il gruppo abbia avviato un'istanza.

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

Come testare la funzione di riduzione orizzontale

1. Segui la procedura descritta in [Ricevere ed eliminare un messaggio \(console\)](#) per eliminare i messaggi dalla coda. Verifica di aver ridotto il numero di messaggi nella coda, in modo che il parametro backlog per istanza sia inferiore al valore di destinazione.

Possono trascorrere alcuni minuti prima che le modifiche richiama l'allarme.

2. Usa il [describe-auto-scaling-groups](#) comando per verificare che il gruppo abbia terminato un'istanza.

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

Amazon SQS e la protezione per la riduzione orizzontale delle istanze

I messaggi che non sono stati elaborati nel momento in cui un'istanza viene terminata vengono restituiti alla coda SQS dove possono essere elaborati da un altro daemon su un'istanza ancora in esecuzione. Per le applicazioni in cui vengono eseguite attività di lunga durata, è possibile utilizzare facoltativamente la protezione per la riduzione orizzontale dell'istanza per avere il controllo su quali dipendenti della coda vengono terminati quando il gruppo con scalabilità automatica viene ridotto.

Lo pseudocodice seguente mostra un modo per proteggere i processi di lavoro a esecuzione prolungata e basati sulla coda dalla terminazione della riduzione orizzontale.

```
while (true)
{
    SetInstanceProtection(False);
    Work = GetNextWorkUnit();
    SetInstanceProtection(True);
    ProcessWorkUnit(Work);
    SetInstanceProtection(False);
}
```

Per ulteriori informazioni, consulta [Progetta le tue applicazioni su Dimensionamento automatico Amazon EC2 per gestire senza problemi la terminazione delle istanze.](#)

Verificare un'attività di dimensionamento per un gruppo con scalabilità automatica

Nella sezione Dimensionamento automatico Amazon EC2 della console Amazon EC2, la voce Activity history (Cronologia delle attività) per un gruppo con scalabilità automatica consente di visualizzare lo stato corrente di un'attività di dimensionamento attualmente in corso. Quando l'attività di dimensionamento è terminata, è possibile verificare se ha esito positivo o meno. Ciò è particolarmente utile quando si creano gruppi Auto Scaling o si aggiungono condizioni di dimensionamento ai gruppi esistenti.

Quando aggiungi un monitoraggio degli obiettivi, un passaggio o una semplice policy di dimensionamento al tuo gruppo con scalabilità automatica, Dimensionamento automatico Amazon EC2 inizia immediatamente a valutare la policy rispetto al parametro. L'allarme del parametro passa nello stato ALARM (ALLARME) quando viene verificato un utilizzo fuori limite per un numero specificato di periodi di valutazione. Ciò significa che una policy di dimensionamento potrebbe attivare l'attività di dimensionamento subito dopo la sua creazione. Dopo che Dimensionamento automatico Amazon EC2 ha regolato la capacità desiderata in risposta a una policy di dimensionamento, puoi verificare l'attività di dimensionamento nell'account. Per ricevere una notifica e-mail da Dimensionamento automatico Amazon EC2 che avvisi dell'avvenuta attivazione dell'attività di dimensionamento, segui le istruzioni riportate in [Opzioni di notifica Amazon SNS per Amazon EC2 Auto Scaling.](#)

 Tip

Nella seguente procedura, si guardano le sezioni Cronologia attività e Istanze del gruppo con scalabilità automatica. In entrambi, le colonne con nome dovrebbero essere già visualizzate. Per visualizzare le colonne nascoste o modificare il numero di righe visualizzate, scegli l'icona a forma di ingranaggio nell'angolo in alto a destra di ogni sezione per aprire le preferenze modali, aggiornare le impostazioni secondo necessità e scegli Conferma.

Come visualizzare le attività di dimensionamento per il gruppo con scalabilità automatica (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Nella barra di navigazione nella parte superiore della schermata, seleziona la Regione in cui si trova il gruppo con dimensionamento automatico.
3. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

4. Nella scheda Activity (Attività), in Activity history (Cronologia attività), la colonna Status (Stato) indica se il gruppo con dimensionamento automatico ha avviato o terminato correttamente le istanze o se l'attività di dimensionamento è ancora in corso.
5. (Facoltativo) Se hai molte attività di dimensionamento, puoi scegliere l'icona > sul bordo superiore della cronologia delle attività per visualizzare la pagina successiva delle attività di dimensionamento.
6. Nella scheda Instance management (Gestione istanze) in Instances (Istanze), la colonna Lifecycle (Ciclo di vita) contiene lo stato delle istanze. Dopo l'avvio dell'istanza e al termine degli hook del ciclo di vita, lo stato del ciclo di vita diventa InService. La colonna Health status (Stato d'integrità) mostra il risultato dei controlli dell'integrità della tua istanza EC2.

Come visualizzare le attività di dimensionamento per un gruppo con scalabilità automatica (AWS CLI)

Utilizza il seguente comando [describe-scaling-activities](#).

```
aws autoscaling describe-scaling-activities --auto-scaling-group-name my-asg
```

Di seguito è riportato un output di esempio.

Le attività di dimensionamento vengono ordinate in base all'ora di inizio. Le attività ancora in corso vengono descritte per prime.

```
{
  "Activities": [
    {
      "ActivityId": "5e3a1f47-2309-415c-bfd8-35aa06300799",
      "AutoScalingGroupName": "my-asg",
      "Description": "Terminating EC2 instance: i-06c4794c2499af1df",
      "Cause": "At 2020-02-11T18:34:10Z a monitor alarm TargetTracking-my-asg-AlarmLow-
b9376cab-18a7-4385-920c-dfa3f7783f82 in state ALARM triggered policy my-target-
tracking-policy changing the desired capacity from 3 to 2. At 2020-02-11T18:34:31Z
an instance was taken out of service in response to a difference between desired and
actual capacity, shrinking the capacity from 3 to 2. At 2020-02-11T18:34:31Z instance
i-06c4794c2499af1df was selected for termination.",
      "StartTime": "2020-02-11T18:34:31.268Z",
      "EndTime": "2020-02-11T18:34:53Z",
      "StatusCode": "Successful",
      "Progress": 100,
      "Details": "{\"Subnet ID\":\"subnet-5ea0c127\",\"Availability Zone\":\"us-west-2a
\"...}\",
      "AutoScalingGroupARN": "arn"
    },
    ...
  ]
}
```

Per una descrizione dei campi nell'output, consulta [Attività](#) nella Riferimenti generali alle API Amazon EC2 Auto Scaling.

Per informazioni su come recuperare attività di dimensionamento per un gruppo eliminato e per informazioni sui tipi di errori che potresti riscontrare e su come gestirli, consulta la sezione [Risoluzione dei problemi di Amazon EC2 Auto Scaling](#).

Disabilitare una policy di dimensionamento per un gruppo con scalabilità automatica

In questo argomento viene descritto come disabilitare temporaneamente una policy di dimensionamento in modo che non apporti modifiche al numero di istanze contenute nel gruppo con scalabilità automatica. Quando disabiliti una policy di dimensionamento, i dettagli della configurazione vengono conservati per consentirti di riabilitare rapidamente la policy. Si tratta di una soluzione più

semplice, rispetto all'eliminazione temporanea di una policy al momento non necessaria e alla sua successiva ricreazione.

Quando una policy di dimensionamento è disabilitata, il gruppo con scalabilità automatica non viene aumentato o ridotto in base agli allarmi dei parametri che segnalano l'utilizzo fuori limite dall'intervallo. Tuttavia, nessuna delle attività di dimensionamento ancora in corso viene arrestata.

Le policy di dimensionamento disabilitate vengono comunque contate nelle quote per il numero di policy di dimensionamento che puoi aggiungere a un gruppo con scalabilità automatica.

Come disabilitare una policy di dimensionamento (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Alla scheda Automatic scaling (Scalabilità automatica), sotto Dynamic scaling policies (Policy di dimensionamento dinamico), seleziona la casella di controllo nell'angolo in alto a destra della policy di dimensionamento desiderata.
4. Scorri fino alla parte superiore della sezione Dynamic scaling policies (Policy di dimensionamento dinamico) e scegli Actions (Operazioni), Disable (Disabilita).

Quando sei pronto ad abilitare nuovamente la policy di dimensionamento, ripeti questa procedura e scegli Actions (Operazioni), Enable (Abilita). Dopo aver riabilitato una policy di dimensionamento, il gruppo con scalabilità automatica può avviare immediatamente l'operazione di dimensionamento, se sono presenti allarmi in stato ALARM (ALLARME).

Come disabilitare una policy di dimensionamento (AWS CLI)

Utilizzate il [put-scaling-policy](#) comando con l'`--no-enabled` opzione come segue. Specifica tutte le opzioni nel comando come faresti per la creazione della policy.

```
aws autoscaling put-scaling-policy --auto-scaling-group-name my-asg \  
  --policy-name my-scaling-policy --policy-type TargetTrackingScaling \  
  --estimated-instance-warmup 360 \  
  --target-tracking-configuration '{ "TargetValue": 70,  
"PredefinedMetricSpecification": { "PredefinedMetricType":  
"ASGAverageCPUUtilization" } }' \  

```

```
--no-enabled
```

Come abilitare nuovamente una policy di dimensionamento (AWS CLI)

Utilizzate il [put-scaling-policy](#) comando con l'`--enabled` opzione come segue. Specifica tutte le opzioni nel comando come faresti per la creazione della policy.

```
aws autoscaling put-scaling-policy --auto-scaling-group-name my-asg \  
  --policy-name my-scaling-policy --policy-type TargetTrackingScaling \  
  --estimated-instance-warmup 360 \  
  --target-tracking-configuration '{ "TargetValue": 70,  
"PredefinedMetricSpecification": { "PredefinedMetricType":  
"ASGAverageCPUUtilization" } }' \  
  --enabled
```

Come descrivere una policy di dimensionamento (AWS CLI)

Utilizza il comando [describe-policies](#) per verificare che una policy di dimensionamento sia in stato “abilitata”.

```
aws autoscaling describe-policies --auto-scaling-group-name my-asg \  
  --policy-names my-scaling-policy
```

Di seguito è riportato un output di esempio.

```
{  
  "ScalingPolicies": [  
    {  
      "AutoScalingGroupName": "my-asg",  
      "PolicyName": "my-scaling-policy",  
      "PolicyARN": "arn:aws:autoscaling:us-  
west-2:123456789012:scalingPolicy:1d52783a-b03b-4710-  
bb0e-549fd64378cc:autoScalingGroupName/my-asg:policyName/my-scaling-policy",  
      "PolicyType": "TargetTrackingScaling",  
      "StepAdjustments": [],  
      "Alarms": [  
        {  
          "AlarmName": "TargetTracking-my-asg-  
AlarmHigh-9ca53fdd-7cf5-4223-938a-ae1199204502",  
          "AlarmARN": "arn:aws:cloudwatch:us-  
west-2:123456789012:alarm:TargetTracking-my-asg-AlarmHigh-9ca53fdd-7cf5-4223-938a-  
ae1199204502"        }  
      ]  
    }  
  ]  
}
```

```
    },
    {
      "AlarmName": "TargetTracking-my-asg-AlarmLow-7010c83d-d55a-4a7a-abe0-1cf8b9de6d6c",
      "AlarmARN": "arn:aws:cloudwatch:us-west-2:123456789012:alarm:TargetTracking-my-asg-AlarmLow-7010c83d-d55a-4a7a-abe0-1cf8b9de6d6c"
    }
  ],
  "TargetTrackingConfiguration": {
    "PredefinedMetricSpecification": {
      "PredefinedMetricType": "ASGAverageCPUUtilization"
    },
    "TargetValue": 70.0,
    "DisableScaleIn": false
  },
  "Enabled": true
}
]
```

Eliminazione di una policy di dimensionamento

Quando una policy di dimensionamento non è più necessaria, puoi eliminarla. A seconda del tipo di politica di scalabilità, potrebbe essere necessario eliminare anche gli CloudWatch allarmi. L'eliminazione di una politica di ridimensionamento di Target Tracking elimina anche tutti gli allarmi associati. CloudWatch L'eliminazione di una politica di ridimensionamento graduale o di una semplice politica di ridimensionamento elimina l'azione di allarme sottostante, ma non elimina l' CloudWatch allarme, anche se non ha più un'azione associata.

Come eliminare una policy di dimensionamento (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Alla scheda Automatic scaling (Scalabilità automatica), sotto Dynamic scaling policies (Policy di dimensionamento dinamico), seleziona la casella di controllo nell'angolo in alto a destra della policy di dimensionamento desiderata.

4. Scorri fino alla parte superiore della sezione Dynamic scaling policies (Policy di dimensionamento dinamico) e scegli Actions (Operazioni), Delete (Elimina).
5. Quando viene richiesta la conferma, seleziona Sì, elimina.
6. (Facoltativo) Se hai eliminato una politica di scalabilità a fasi o una semplice politica di scalabilità, procedi come segue per eliminare l' CloudWatch allarme associato alla politica. È possibile saltare questi sottopassaggi per mantenere l'allarme per un utilizzo futuro.
 - a. [Apri la CloudWatch console all'indirizzo https://console.aws.amazon.com/cloudwatch/](https://console.aws.amazon.com/cloudwatch/).
 - b. Nel pannello di navigazione, scegli Allarmi.
 - c. Scegli l'allarme, ad esempio Step-Scaling-AlarmHigh-AddCapacity, quindi selezionare Action (Operazione), Delete (Elimina).
 - d. Quando viene richiesta la conferma, seleziona Delete (Elimina).

Come ottenere le policy di dimensionamento per un gruppo con scalabilità automatica (AWS CLI)

Prima di eliminare una policy di dimensionamento, utilizzare il seguente comando, [describe-policies](#), per visualizzare le policy di dimensionamento create per il gruppo con scalabilità automatica. È possibile utilizzare l'output per eliminare la politica e gli CloudWatch allarmi.

```
aws autoscaling describe-policies --auto-scaling-group-name my-asg
```

Puoi filtrare i risultati in base al tipo di policy di dimensionamento utilizzando il parametro `--query`. Questa sintassi per `query` è applicabile su Linux o macOS. In Windows, modifica le virgolette singole in virgolette doppie.

```
aws autoscaling describe-policies --auto-scaling-group-name my-asg  
--query 'ScalingPolicies[?PolicyType==`TargetTrackingScaling`]'
```

Di seguito è riportato un output di esempio.

```
[  
  {  
    "AutoScalingGroupName": "my-asg",  
    "PolicyName": "cpu50-target-tracking-scaling-policy",  
    "PolicyARN": "PolicyARN",  
    "PolicyType": "TargetTrackingScaling",  
    "StepAdjustments": [],
```

```

    "Alarms": [
      {
        "AlarmARN": "arn:aws:cloudwatch:us-
west-2:123456789012:alarm:TargetTracking-my-asg-AlarmHigh-
fc0e4183-23ac-497e-9992-691c9980c38e",
        "AlarmName": "TargetTracking-my-asg-AlarmHigh-
fc0e4183-23ac-497e-9992-691c9980c38e"
      },
      {
        "AlarmARN": "arn:aws:cloudwatch:us-
west-2:123456789012:alarm:TargetTracking-my-asg-AlarmLow-61a39305-ed0c-47af-
bd9e-471a352ee1a2",
        "AlarmName": "TargetTracking-my-asg-AlarmLow-61a39305-ed0c-47af-
bd9e-471a352ee1a2"
      }
    ],
    "TargetTrackingConfiguration": {
      "PredefinedMetricSpecification": {
        "PredefinedMetricType": "ASGAverageCPUUtilization"
      },
      "TargetValue": 50.0,
      "DisableScaleIn": false
    },
    "Enabled": true
  }
]

```

Come eliminare la policy di dimensionamento (AWS CLI)

Utilizza il comando seguente: [delete-policy](#).

```

aws autoscaling delete-policy --auto-scaling-group-name my-asg \
  --policy-name cpu50-target-tracking-scaling-policy

```

Per eliminare la CloudWatch sveglia (AWS CLI)

Per politiche di scalabilità semplici e dettagliate, utilizzate il comando [delete-alarms](#) per eliminare l' CloudWatch allarme associato alla policy. Se desideri mantenere l'allarme per utilizzarlo in futuro, salta questo passaggio. Puoi eliminare uno o più allarmi alla volta. Ad esempio, utilizzare il seguente comando per eliminare gli allarmi Step-Scaling-AlarmHigh-AddCapacity e Step-Scaling-AlarmLow-RemoveCapacity.


```
aws cloudwatch delete-alarms --alarm-name Step-Scaling-AlarmHigh-AddCapacity Step-Scaling-AlarmLow-RemoveCapacity
```

Policy di dimensionamento di esempio per AWS Command Line Interface (AWS CLI)

Puoi creare politiche di scalabilità per Amazon EC2 Auto Scaling tramite AWS Management Console, o SDK. AWS CLI

Gli esempi seguenti mostrano come creare politiche di scalabilità per Amazon EC2 Auto Scaling con il comando. AWS CLI [put-scaling-policy](#) Sostituisci ciascun *placeholder input dell'utente* con le tue informazioni.

Per iniziare a scrivere politiche di scalabilità utilizzando AWS CLI, consulta gli esercizi introduttivi in and. [Policy di dimensionamento con monitoraggio degli obiettivi](#) [Policy di dimensionamento per fasi e di dimensionamento semplice](#)

Esempio 1: applicazione di una policy di dimensionamento con monitoraggio degli obiettivi con uno specifico parametro di default

```
aws autoscaling put-scaling-policy --policy-name cpu50-target-tracking-scaling-policy \  
  --auto-scaling-group-name my-asg --policy-type TargetTrackingScaling \  
  --target-tracking-configuration file://config.json  
{  
  "TargetValue": 50.0,  
  "PredefinedMetricSpecification": {  
    "PredefinedMetricType": "ASGAverageCPUUtilization"  
  }  
}
```

Per ulteriori informazioni, consulta il riferimento [PredefinedMetricSpecification](#) all'API Amazon EC2 Auto Scaling.

Note

Se il file non si trova nella directory corrente, digita il percorso completo del file. Per ulteriori informazioni sulla lettura AWS CLI dei valori dei parametri da un file, consultate [Caricamento AWS CLI dei parametri da un file](#) nella Guida per l' AWS Command Line Interface utente.

Esempio 2: applicazione di una policy di dimensionamento con monitoraggio degli obiettivi con uno specifico parametro personalizzato

```
aws autoscaling put-scaling-policy --policy-name sqs100-target-tracking-scaling-policy \
  \
  --auto-scaling-group-name my-asg --policy-type TargetTrackingScaling \
  --target-tracking-configuration file://config.json
{
  "TargetValue": 100.0,
  "CustomizedMetricSpecification": {
    "MetricName": "MyBacklogPerInstance",
    "Namespace": "MyNamespace",
    "Dimensions": [{
      "Name": "MyOptionalMetricDimensionName",
      "Value": "MyOptionalMetricDimensionValue"
    }],
    "Statistic": "Average",
    "Unit": "None"
  }
}
```

Per ulteriori informazioni, consulta il riferimento [CustomizedMetricSpecification](#) all'API Amazon EC2 Auto Scaling.

Esempio 3: applicazione di una policy di dimensionamento con monitoraggio degli obiettivi per il solo aumento orizzontale

```
aws autoscaling put-scaling-policy --policy-name alb1000-target-tracking-scaling-policy \
  \
  --auto-scaling-group-name my-asg --policy-type TargetTrackingScaling \
  --target-tracking-configuration file://config.json
{
  "TargetValue": 1000.0,
  "PredefinedMetricSpecification": {
    "PredefinedMetricType": "ALBRequestCountPerTarget",
    "ResourceLabel": "app/my-alb/778d41231b141a0f/targetgroup/my-alb-target-
group/943f017f100becff"
  },
  "DisableScaleIn": true
}
```

Esempio 4: creazione di una policy di dimensionamento per fasi per l'aumento orizzontale

```
aws autoscaling put-scaling-policy \  
  --auto-scaling-group-name my-asg \  
  --policy-name my-step-scale-out-policy \  
  --policy-type StepScaling \  
  --adjustment-type PercentChangeInCapacity \  
  --metric-aggregation-type Average \  
  --step-adjustments  
MetricIntervalLowerBound=10.0,MetricIntervalUpperBound=20.0,ScalingAdjustment=10 \  
  
MetricIntervalLowerBound=20.0,MetricIntervalUpperBound=30.0,ScalingAdjustment=20 \  
  MetricIntervalLowerBound=30.0,ScalingAdjustment=30 \  
  --min-adjustment-magnitude 1
```

Registrare l'Amazon Resource Name (ARN) della policy. È necessario l'ARN quando si crea l'CloudWatch allarme.

Esempio 5: creazione di una policy di dimensionamento per fasi per la riduzione orizzontale

```
aws autoscaling put-scaling-policy \  
  --auto-scaling-group-name my-asg \  
  --policy-name my-step-scale-in-policy \  
  --policy-type StepScaling \  
  --adjustment-type ChangeInCapacity \  
  --step-adjustments MetricIntervalUpperBound=0.0,ScalingAdjustment=-2
```

Registrare l'Amazon Resource Name (ARN) della policy. È necessario l'ARN quando si crea l'CloudWatch allarme.

Esempio 6: applicazione di policy di dimensionamento semplice per l'aumento orizzontale

```
aws autoscaling put-scaling-policy --policy-name my-simple-scale-out-policy \  
  --auto-scaling-group-name my-asg --scaling-adjustment 30 \  
  --adjustment-type PercentChangeInCapacity --min-adjustment-magnitude 2
```

Registrare l'Amazon Resource Name (ARN) della policy. È necessario l'ARN quando si crea l'CloudWatch allarme.

Esempio 7: applicazione di policy di dimensionamento semplice per la riduzione orizzontale

```
aws autoscaling put-scaling-policy --policy-name my-simple-scale-in-policy \  
  --auto-scaling-group-name my-asg --scaling-adjustment -2 \  
  --adjustment-type ChangeInCapacity --min-adjustment-magnitude 2
```

```
--auto-scaling-group-name my-asg --scaling-adjustment -1 \  
--adjustment-type ChangeInCapacity --cooldown 180
```

Registrare l'Amazon Resource Name (ARN) della policy. È necessario l'ARN quando si crea l'CloudWatch allarme.

Dimensionamento predittivo per Dimensionamento automatico Amazon EC2

La scalabilità predittiva funziona analizzando i dati storici di carico per rilevare modelli giornalieri o settimanali nei flussi di traffico. Utilizza queste informazioni per prevedere le esigenze di capacità future in modo che Amazon EC2 Auto Scaling possa aumentare in modo proattivo la capacità del gruppo Auto Scaling per far fronte al carico previsto.

Il dimensionamento predittivo è adatto per le situazioni in cui si hanno:

- Traffico ciclico, ad esempio, un elevato uso di risorse durante i normali orari di ufficio e un basso utilizzo di risorse durante la notte
- Modelli di on-and-off carico di lavoro ricorrenti, come elaborazione in batch, test o analisi periodica dei dati
- Applicazioni che richiedono molto tempo per l'inizializzazione, causando un notevole impatto della latenza sulle prestazioni delle applicazioni durante gli eventi di aumento orizzontale

In generale, se hai modelli regolari di aumento del traffico e applicazioni che richiedono molto tempo per l'inizializzazione, è consigliabile che utilizzi il dimensionamento predittivo. Il dimensionamento predittivo consente di dimensionare più velocemente avviando la capacità in anticipo rispetto al carico previsto, in confronto all'utilizzo del solo dimensionamento dinamico, che è di natura reattiva. La scalabilità predittiva può anche potenzialmente farti risparmiare sulla fattura EC2, aiutandoti a evitare la necessità di fornire una capacità eccessiva.

Ad esempio, consideriamo un'applicazione che ha un utilizzo elevato durante l'orario di lavoro e uno ridotto durante la notte. All'inizio di ogni giornata lavorativa, il dimensionamento predittivo può aggiungere capacità prima del primo afflusso di traffico. Ciò permette all'applicazione di mantenere elevata disponibilità e prestazioni quando si passa da un periodo di utilizzo inferiore a un periodo di utilizzo più elevato. Non è necessario attendere che il dimensionamento dinamico reagisca alla variazione di traffico. Inoltre, non dovrai dedicare tempo alla verifica dei modelli di carico

dell'applicazione e al tentativo di pianificare la giusta quantità di capacità con il dimensionamento programmato.

Argomenti

- [Funzionamento del dimensionamento predittivo](#)
- [Crea una politica di scalabilità predittiva](#)
- [Valutazione delle policy di dimensionamento predittivo](#)
- [Sovrascrivere i valori di previsione mediante operazioni pianificate](#)
- [Advanced predictive scaling policy configurations using custom metrics \(Configurazioni avanzate delle policy di dimensionamento predittivo utilizzando parametri personalizzati\)](#)

Funzionamento del dimensionamento predittivo

Questo argomento spiega come funziona la scalabilità predittiva e descrive cosa prendere in considerazione quando si crea una politica di scalabilità predittiva.

Argomenti

- [Come funziona](#)
- [Limite massimo di capacità](#)
- [Considerazioni](#)
- [Regioni supportate](#)

Come funziona

Per utilizzare la scalabilità predittiva, crea una politica di scalabilità predittiva che specifichi la metrica da monitorare e analizzare. CloudWatch Affinché la scalabilità predittiva inizi a prevedere i valori futuri, questa metrica deve contenere almeno 24 ore di dati.

Dopo aver creato la policy, la scalabilità predittiva inizia ad analizzare i dati metrici relativi agli ultimi 14 giorni per identificare i modelli. Utilizza questa analisi per generare una previsione oraria dei requisiti di capacità per le prossime 48 ore. La previsione viene aggiornata ogni 6 ore utilizzando i CloudWatch dati più recenti. Con l'arrivo di nuovi dati, la scalabilità predittiva è in grado di migliorare continuamente l'accuratezza delle previsioni future.

La prima volta che abiliti la scalabilità predittiva, questa viene eseguita in modalità di sola previsione. In questa modalità, genera previsioni sulla capacità ma non ridimensiona effettivamente

il gruppo Auto Scaling in base a tali previsioni. Ciò consente di valutare l'accuratezza e l'idoneità della previsione. È possibile visualizzare i dati di previsione utilizzando l'operazione `GetPredictiveScalingForecast` API o il AWS Management Console.

Dopo aver esaminato i dati di previsione e deciso di iniziare il ridimensionamento in base a tali dati, passa la politica di scalabilità alla modalità previsione e scala. In questa modalità:

- Se la previsione prevede un aumento del carico, Amazon EC2 Auto Scaling aumenterà la capacità mediante la scalabilità orizzontale.
- Se la previsione prevede una riduzione del carico, non verrà ridimensionata per rimuovere la capacità. Se si desidera rimuovere la capacità che non è più necessaria, è necessario creare politiche di scalabilità dinamiche.

Per impostazione predefinita, Amazon EC2 Auto Scaling ridimensiona il gruppo Auto Scaling all'inizio di ogni ora in base alla previsione per quell'ora. Facoltativamente, puoi specificare un'ora di inizio precedente utilizzando la `SchedulingBufferTime` proprietà nell'operazione `PutScalingPolicy` API o l'impostazione delle istanze pre-Launch in AWS Management Console. Ciò fa sì che Amazon EC2 Auto Scaling lanci nuove istanze prima della domanda prevista, dando loro il tempo di avviarsi e prepararsi a gestire il traffico.

Per supportare il lancio di nuove istanze prima della domanda prevista, ti consigliamo vivamente di abilitare il riscaldamento delle istanze predefinito per il tuo gruppo Auto Scaling. Questo specifica un periodo di tempo dopo un'attività di scalabilità orizzontale durante il quale Amazon EC2 Auto Scaling non sarà scalabile, anche se le politiche di scalabilità dinamica indicano che la capacità deve essere ridotta. Questo ti aiuta a garantire che le istanze appena lanciate abbiano il tempo sufficiente per iniziare a gestire l'aumento del traffico prima di essere prese in considerazione per operazioni di scalabilità. Per ulteriori informazioni, consulta [Impostazione della preparazione di default dell'istanza per un gruppo con scalabilità automatica](#).

Limite massimo di capacità

I gruppi Auto Scaling hanno un'impostazione di capacità massima che limita il numero massimo di istanze EC2 che possono essere avviate per il gruppo. Per impostazione predefinita, quando vengono impostate politiche di scalabilità, non possono aumentare la capacità oltre la sua capacità massima.

In alternativa, è possibile consentire l'aumento automatico della capacità massima del gruppo se la capacità prevista si avvicina o supera la capacità massima del gruppo Auto

Scaling. Per abilitare questo comportamento, utilizzate le `MaxCapacityBuffer` proprietà `MaxCapacityBreachBehavior` and nell'operazione `PutScalingPolicy` API o l'impostazione di comportamento `Max capacity` in. AWS Management Console

Warning

Prestate attenzione quando consentite l'aumento automatico della capacità massima. Ciò può comportare l'avvio di un numero di istanze superiore a quello previsto se la maggiore capacità massima non viene monitorata e gestita. L'aumento della capacità massima diventa quindi la nuova capacità massima normale per il gruppo Auto Scaling fino a quando non lo si aggiorna manualmente. La capacità massima non torna automaticamente al valore massimo originale.

Considerazioni

- Conferma che il dimensionamento predittivo sia adeguato al carico di lavoro. Un carico di lavoro è adeguato per il dimensionamento predittivo se presenta modelli di carico ricorrenti specifici per il giorno della settimana o l'ora del giorno. Per verificarlo, configura le policy di dimensionamento predittivo in modalità solo previsione e fai riferimento ai suggerimenti nella console. Dimensionamento automatico Amazon EC2 offre consigli basati su osservazioni relative alle potenziali prestazioni delle policy. Valuta la previsione e i suggerimenti prima di permettere al dimensionamento predittivo di dimensionare attivamente l'applicazione.
- Per avviare la previsione, il dimensionamento predittivo richiede almeno 24 ore di dati cronologici. Tuttavia, le previsioni sono più efficaci se i dati cronologici si estendono su due settimane intere. Se aggiorni l'applicazione creando un nuovo gruppo con scalabilità automatica ed eliminando quello precedente, il nuovo gruppo con scalabilità automatica richiede 24 ore di dati cronologici del carico, prima che il dimensionamento predittivo possa iniziare a generare nuovamente i forecast. Puoi utilizzare i parametri personalizzati per aggregare i parametri tra gruppi con scalabilità automatica vecchi e nuovi. In caso contrario, potrebbe essere necessario attendere alcuni giorni, al fine di avere una previsione più accurata.
- Scegliete una metrica di carico che rappresenti con precisione il carico completo dell'applicazione e che rappresenti l'aspetto dell'applicazione su cui è più importante basare la scalabilità.
- L'utilizzo della scalabilità dinamica con la scalabilità predittiva ti aiuta a seguire da vicino la curva di domanda per la tua applicazione, adattandola verso l'alto durante i periodi di traffico ridotto e la scalabilità orizzontale quando il traffico è superiore al previsto. Quando sono attive più policy di dimensionamento, ciascuna di esse determina la capacità desiderata in modo indipendente,

che viene impostata al massimo. Ad esempio, se sono necessarie 10 istanze per rimanere all'utilizzo di destinazione in una policy di dimensionamento con monitoraggio degli obiettivi e 8 istanze devono rimanere all'utilizzo di destinazione in una policy di dimensionamento predittivo, la capacità desiderata del gruppo è impostata su 10. Se non conosci il ridimensionamento dinamico, ti consigliamo di utilizzare le politiche di scalabilità di Target Tracking. Per ulteriori informazioni, consulta [Dimensionamento dinamico per Dimensionamento automatico Amazon EC2](#).

- Un presupposto fondamentale del dimensionamento predittivo è che il gruppo con scalabilità automatica deve essere omogeneo e tutte le istanze devono avere la stessa capacità. Se ciò non è vero per il tuo gruppo, la capacità prevista può essere imprecisa. Pertanto, fai attenzione quando crei politiche di scalabilità predittiva per [gruppi di istanze miste](#), poiché è possibile fornire istanze di tipi diversi con capacità diversa. Di seguito sono riportati alcuni esempi in cui la capacità prevista sarà imprecisa:
 - La policy di dimensionamento predittivo si basa sull'utilizzo della CPU, ma il numero di vCPU su ciascuna istanza Auto Scaling varia a seconda dei tipi di istanza.
 - La policy di dimensionamento predittivo si basa sulla rete in entrata o in uscita, ma la velocità effettiva della larghezza di banda di rete per ogni istanza Auto Scaling varia a seconda dei tipi di istanza. Ad esempio, i tipi di istanza M5 e M5n sono simili, ma il tipo di istanza M5n offre una velocità effettiva di rete significativamente più elevata.

Regioni supportate

Amazon EC2 Auto Scaling supporta politiche di scalabilità predittiva nei Regioni AWS seguenti paesi: Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (Oregon), Stati Uniti occidentali (California settentrionale), Africa (Città del Capo), Canada (Centrale), UE (Francoforte), UE (Irlanda), UE (Londra), UE (Milano), UE (Parigi), UE (Stoccolma), Asia Pacifico (Hong Kong), Asia Pacifico (Giacarta), Asia Pacifico (Mumbai), Asia Pacifico (Osaka), Asia Pacifico (Tokyo), Asia Pacifico (Singapore), Asia Pacifico (Seoul), Asia Pacifico (Sydney), Medio Oriente (Bahrain), Medio Oriente (Emirati Arabi Uniti), Sud America (San Paolo), Cina (Pechino), Cina (Ningxia), AWS GovCloud (Stati Uniti orientali) e AWS GovCloud (Stati Uniti occidentali).

Crea una politica di scalabilità predittiva

Le seguenti procedure consentono di creare una politica di scalabilità predittiva utilizzando o. AWS Management Console o AWS CLI

Se il gruppo con scalabilità automatica è nuovo, prima che Dimensionamento automatico Amazon EC2 possa generare una previsione deve fornire almeno 24 ore di dati.

Indice

- [Creazione di una policy di dimensionamento predittivo \(console\)](#)
- [Creazione di una policy di dimensionamento predittivo \(AWS CLI\)](#)

Creazione di una policy di dimensionamento predittivo (console)

Se è la prima volta che crei una policy di scalabilità predittiva, ti consigliamo di utilizzare la console per creare più policy di scalabilità predittiva in modalità solo previsionale. Ciò consente di testare i potenziali effetti di diverse metriche e valori obiettivo. È possibile creare più criteri di dimensionamento predittivo per ciascun gruppo con scalabilità automatica, ma solo una delle policy può essere utilizzata per il dimensionamento attivo.

Creazione di una policy di dimensionamento predittivo nella console (parametri predefiniti)

Utilizza la procedura seguente per creare una policy di dimensionamento predittivo utilizzando i parametri predefiniti (numero di richieste di CPU, I/O di rete o Application Load Balancer per destinazione). Il modo più semplice per creare una policy di dimensionamento predittivo è utilizzare i parametri predefiniti. Se preferisci utilizzare i parametri personalizzati, consulta [Creazione di una policy di dimensionamento predittivo nella console \(parametri predefiniti\)](#).

Creazione di una policy di dimensionamento predittivo

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Nella scheda Automatic scaling (Scalabilità automatica) scegli Create predictive scaling policy (Crea policy di dimensionamento predittivo) in Scaling policies (Policy di dimensionamento).
4. Inserisci un nome per la policy.
5. Attiva Scale based on forecast (Dimensiona in base al forecast) per dare a Dimensionamento automatico Amazon EC2 l'autorizzazione ad avviare subito il dimensionamento.

Per mantenere la policy in modalità forecast only (solo forecast), mantieni l'opzione Scale based on forecast (Dimensiona in base al forecast) disattivata.

6. In Metrics (Parametri), scegli i parametri nell'elenco di opzioni. Le opzioni includono: CPU, Network In (Rete in ingresso), Rete in uscita (Network Out), Application Load Balancer

request count (Conteggio richieste Application Load Balancer) e Custom metric pair (Coppia di parametri personalizzati).

Se hai scelto Application Load Balancer request count per target (Conteggio delle richieste Application Load Balancer per destinazione), scegli un gruppo di destinazione in Target group (Gruppo di destinazione). Application Load Balancer request count per target (Conteggio di richieste Application Load Balancer per destinazione) è supportato solo avrai allegato un gruppo di destinazione Application Load Balancer al gruppo con scalabilità automatica.

Se hai scelto Custom metric pair (Coppia di parametri personalizzati), scegli i singoli parametri dagli elenchi a discesa per Load metric (Parametro del carico) e Scaling metric (Parametro di dimensionamento).

7. In Target utilization (Utilizzo di destinazione), inserisci il valore di destinazione che Dimensionamento automatico Amazon EC2 deve mantenere. Dimensionamento automatico Amazon EC2 aumenta la capacità finché l'utilizzo medio arriva a quello di destinazione o finché non raggiunge il numero massimo di istanze specificato.

Se il parametro di dimensionamento è...	Allora l'utilizzo target rappresenta...
CPU	La percentuale di CPU che ogni istanza dovrebbe idealmente utilizzare.
Rete in ingresso	Il numero medio di byte al minuto che ogni istanza dovrebbe idealmente ricevere.
Rete in uscita	Il numero medio di byte al minuto che ogni istanza dovrebbe idealmente inviare.
Conteggio di richieste Application Load Balancer per destinazione	Il numero medio di richieste al minuto che ogni istanza dovrebbe idealmente ricevere.

8. (Facoltativo) In Pre-launch instances (Pre-avviare le istanze), scegli con quale anticipo desideri avviare le istanze prima che il forecast richieda l'aumento del carico.
9. (Facoltativo) In Max capacity behavior (Comportamento di capacità max), scegli se consentire a Dimensionamento automatico Amazon EC2 di aumentare la capacità massima del gruppo quando la capacità predittiva prevista supera la capacità massima definita. L'attivazione di

questa impostazione permette l'aumento orizzontale durante i periodi in cui si prevede che il traffico sia al massimo.

10. (Facoltativo) In Buffer maximum capacity above the forecasted capacity (Capacità massima del buffer superiore alla capacità prevista), scegli la capacità aggiuntiva da utilizzare quando quella prevista è prossima o superiore alla capacità massima. Il valore è specificato come una percentuale relativa alla capacità di previsione. Ad esempio, se il buffer è 10, indica un buffer del 10%. Pertanto, se la capacità di previsione è 50 e quella massima è 40, la capacità massima effettiva è 55.

Se impostato su 0, Dimensionamento automatico Amazon EC2 potrebbe essere in grado di dimensionare una capacità superiore alla capacità massima, in modo da uguagliare ma non superare quella prevista.

11. Scegli Create predictive scaling policy (Crea policy di dimensionamento predittivo).

Creazione di una policy di dimensionamento predittivo nella console (parametri predefiniti)

Utilizza la procedura seguente per creare una policy di dimensionamento predittivo utilizzando i parametri personalizzati. Le metriche personalizzate possono includere altre metriche fornite da CloudWatch o metriche su cui pubblici. CloudWatch Per utilizzare il numero di richieste di CPU, I/O di rete o Application Load Balancer per destinazione, consulta [Creazione di una policy di dimensionamento predittivo nella console \(parametri predefiniti\)](#).

Utilizza la procedura seguente per creare una policy di dimensionamento predittivo utilizzando i parametri personalizzati.

- È necessario fornire le query non elaborate che consentano ad Amazon EC2 Auto Scaling di interagire con le metriche. CloudWatch Per ulteriori informazioni, consulta [Advanced predictive scaling policy configurations using custom metrics \(Configurazioni avanzate delle policy di dimensionamento predittivo utilizzando parametri personalizzati\)](#). Per essere sicuri che Amazon EC2 Auto Scaling possa estrarre i dati CloudWatch delle metriche, verifica che ogni query restituisca punti dati. Confermalo utilizzando la CloudWatch console o il funzionamento dell' CloudWatch [GetMetricDataAPI](#).

Note

Forniamo dei payload JSON di esempio nell'editor JSON nella console Dimensionamento automatico Amazon EC2. Questi esempi forniscono un riferimento per le coppie chiave-valore necessarie per aggiungere altre CloudWatch metriche fornite da AWS o metriche

su cui hai pubblicato in precedenza. CloudWatch Puoi utilizzarli come punto di partenza, quindi personalizzarli in base alle tue esigenze.

- Se per i parametri usi una formula, devi costruire manualmente il JSON per adattarlo al tuo scenario univoco. Per ulteriori informazioni, consulta [Utilizzare le espressioni matematiche del parametro](#). Prima di utilizzare la formula del parametro nella tua policy, verifica che le query dei parametri basate su espressioni matematiche siano valide e restituiscano una singola serie temporale. Confermalo utilizzando la CloudWatch console o l'operazione API. CloudWatch [GetMetricData](#)

Se commetti un errore in una query fornendo dati errati, ad esempio il nome del gruppo con scalabilità automatica errato, la previsione non conterrà alcun dato. Per la risoluzione dei problemi relativi ai parametri personalizzati, consulta [Considerazioni e risoluzione dei problemi](#).

Creazione di una policy di dimensionamento predittivo

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Nella scheda Automatic scaling (Scalabilità automatica) scegli Create predictive scaling policy (Crea policy di dimensionamento predittivo) in Scaling policies (Policy di dimensionamento).
4. Inserisci un nome per la policy.
5. Attiva Scale based on forecast (Dimensiona in base al forecast) per dare a Dimensionamento automatico Amazon EC2 l'autorizzazione ad avviare subito il dimensionamento.

Per mantenere la policy in modalità forecast only (solo forecast), mantieni l'opzione Scale based on forecast (Dimensiona in base al forecast) disattivata.

6. Per Metrics (Parametri), scegli Custom metric pair (Coppia di parametri personalizzati).
 - a. Per Load metric, scegli CloudWatch Metrica personalizzata per utilizzare una metrica personalizzata. Crea il payload JSON che contiene la definizione dei parametri di carico per la policy e incollalo nella casella dell'editor JSON, sostituendo ciò che è già presente nella casella.
 - b. Per Scaling metric, scegli Metrica personalizzata per utilizzare una CloudWatch metrica personalizzata. Crea il payload JSON che contiene la definizione dei parametri di

dimensionamento per la policy e incollalo nella casella dell'editor JSON, sostituendo ciò che è già presente nella casella.

- c. (Facoltativo) Per aggiungere un parametro di capacità personalizzato, seleziona la casella di controllo Add custom capacity metric (Aggiungi parametro di capacità personalizzato). Crea il payload JSON che contiene la definizione dei parametri di capacità per la policy e incollalo nella casella dell'editor JSON, sostituendo ciò che è già presente nella casella.

È necessario abilitare questa opzione solo per creare una nuova serie temporale per la capacità solo se i dati dei parametri di capacità riguardano più gruppi con scalabilità automatica. In questo caso, dovrai utilizzare la formula dei parametri per aggregare i dati in un'unica serie temporale.

7. In Target utilization (Utilizzo di destinazione), inserisci il valore di destinazione che Dimensionamento automatico Amazon EC2 deve mantenere. Dimensionamento automatico Amazon EC2 aumenta la capacità finché l'utilizzo medio arriva a quello di destinazione o finché non raggiunge il numero massimo di istanze specificato.
8. (Facoltativo) In Pre-launch instances (Pre-avviare le istanze), scegli con quale anticipo desideri avviare le istanze prima che il forecast richieda l'aumento del carico.
9. (Facoltativo) In Max capacity behavior (Comportamento di capacità max), scegli se consentire a Dimensionamento automatico Amazon EC2 di aumentare la capacità massima del gruppo quando la capacità predittiva prevista supera la capacità massima definita. L'attivazione di questa impostazione permette l'aumento orizzontale durante i periodi in cui si prevede che il traffico sia al massimo.
10. (Facoltativo) In Buffer maximum capacity above the forecasted capacity (Capacità massima del buffer superiore alla capacità prevista), scegli la capacità aggiuntiva da utilizzare quando quella prevista è prossima o superiore alla capacità massima. Il valore è specificato come una percentuale relativa alla capacità di previsione. Ad esempio, se il buffer è 10, indica un buffer del 10%. Pertanto, se la capacità di previsione è 50 e quella massima è 40, la capacità massima effettiva è 55.

Se impostato su 0, Dimensionamento automatico Amazon EC2 potrebbe essere in grado di dimensionare una capacità superiore alla capacità massima, in modo da uguagliare ma non superare quella prevista.

11. Scegli Create predictive scaling policy (Crea policy di dimensionamento predittivo).

Creazione di una policy di dimensionamento predittivo (AWS CLI)

Utilizza AWS CLI quanto segue per configurare le politiche di scalabilità predittiva per il tuo gruppo Auto Scaling. Sostituisci ciascun *placeholder input dell'utente* con le tue informazioni.

Per ulteriori informazioni sui CloudWatch parametri che puoi specificare, consulta il riferimento alle API [PredictiveScalingMetricSpecification](#) di Amazon EC2 Auto Scaling.

Esempio 1: policy di dimensionamento predittivo in grado di creare forecast ma non di dimensionare

La policy di esempio seguente mostra una configurazione completa delle policy che utilizzano i parametri di utilizzo della CPU per il dimensionamento predittivo con un utilizzo di destinazione 40. ForecastOnly viene utilizzato per impostazione predefinita, a meno che non specifichi esplicitamente quale modalità utilizzare. Salva questa configurazione in un file denominato `config.json`.

```
{
  "MetricSpecifications": [
    {
      "TargetValue": 40,
      "PredefinedMetricPairSpecification": {
        "PredefinedMetricType": "ASGCPUUtilization"
      }
    }
  ]
}
```

Per creare la policy dalla riga di comando, esegui il [put-scaling-policy](#) comando con il file di configurazione specificato, come illustrato nell'esempio seguente.

```
aws autoscaling put-scaling-policy --policy-name cpu40-predictive-scaling-policy \
  --auto-scaling-group-name my-asg --policy-type PredictiveScaling \
  --predictive-scaling-configuration file://config.json
```

In caso di esito positivo, questo comando restituisce l'Amazon Resource Name (ARN) della policy.

```
{
  "PolicyARN": "arn:aws:autoscaling:region:account-id:scalingPolicy:2f4f5048-d8a8-4d14-b13a-d1905620f345:autoScalingGroupName/my-asg:policyName/cpu40-predictive-scaling-policy",
}
```

```
"Alarms": []
}
```

Esempio 2: policy di dimensionamento predittivo in grado di creare forecast e di dimensionare

Per una policy che permette a Dimensionamento automatico Amazon EC2 di prevedere e dimensionare, aggiungi la proprietà `Mode` con un valore di `ForecastAndScale`. Nell'esempio seguente viene illustrata una configurazione di policy che utilizza i parametri del conteggio delle richieste di Application Load Balancer. L'utilizzo di destinazione è `1000` e il dimensionamento predittivo è impostato sulla modalità `ForecastAndScale`.

```
{
  "MetricSpecifications": [
    {
      "TargetValue": 1000,
      "PredefinedMetricPairSpecification": {
        "PredefinedMetricType": "ALBRequestCount",
        "ResourceLabel": "app/my-alb/778d41231b141a0f/targetgroup/my-alb-
target-group/943f017f100becff"
      }
    }
  ],
  "Mode": "ForecastAndScale"
}
```

Per creare questa politica, esegui il [put-scaling-policy](#) comando con il file di configurazione specificato, come illustrato nell'esempio seguente.

```
aws autoscaling put-scaling-policy --policy-name alb1000-predictive-scaling-policy \
  --auto-scaling-group-name my-asg --policy-type PredictiveScaling \
  --predictive-scaling-configuration file://config.json
```

In caso di esito positivo, questo comando restituisce l'Amazon Resource Name (ARN) della policy.

```
{
  "PolicyARN": "arn:aws:autoscaling:region:account-
id:scalingPolicy:19556d63-7914-4997-8c81-d27ca5241386:autoScalingGroupName/my-
asg:policyName/alb1000-predictive-scaling-policy",
  "Alarms": []
}
```

Esempio 3: policy di dimensionamento predittivo in grado di dimensionare oltre la capacità massima

Nell'esempio seguente viene illustrato come creare una policy in grado di dimensionare oltre al limite di dimensione massima del gruppo quando è necessario per gestire un carico superiore al normale. Per impostazione predefinita, Dimensionamento automatico Amazon EC2 non dimensiona la capacità EC2 oltre la capacità massima definita. Tuttavia, potrebbe essere utile lasciarla dimensionare più in alto, con una capacità leggermente superiore, per evitare problemi di prestazioni o disponibilità.

Per fornire spazio a Dimensionamento automatico Amazon EC2 per il provisioning di capacità aggiuntiva quando prevedi che la capacità sarà pari o molto vicina alle dimensioni massime del tuo gruppo, specifica le proprietà `MaxCapacityBreachBehavior` `MaxCapacityBuffer`, come mostrato nell'esempio seguente. Devi specificare `MaxCapacityBreachBehavior` con un valore di `IncreaseMaxCapacity`. Il numero massimo di istanze che il gruppo può avere dipende dal valore di `MaxCapacityBuffer`.

```
{
  "MetricSpecifications": [
    {
      "TargetValue": 70,
      "PredefinedMetricPairSpecification": {
        "PredefinedMetricType": "ASGCPUUtilization"
      }
    }
  ],
  "MaxCapacityBreachBehavior": "IncreaseMaxCapacity",
  "MaxCapacityBuffer": 10
}
```

In questo esempio, la policy è configurata per usare un buffer del 10% (`"MaxCapacityBuffer": 10`), in modo che se la capacità prevista è 50 e quella massima è 40, la capacità massima effettiva è 55. Una policy in grado di dimensionare una capacità superiore alla capacità massima in modo da uguagliare ma non superare la capacità prevista avrebbe un buffer di 0 (`"MaxCapacityBuffer": 0`).

Per creare questa politica, esegui il [put-scaling-policy](#) comando con il file di configurazione specificato, come illustrato nell'esempio seguente.

```
aws autoscaling put-scaling-policy --policy-name cpu70-predictive-scaling-policy \
  --auto-scaling-group-name my-asg --policy-type PredictiveScaling \
  --predictive-scaling-configuration file://config.json
```


In caso di esito positivo, questo comando restituisce l'Amazon Resource Name (ARN) della policy.

```
{
  "PolicyARN": "arn:aws:autoscaling:region:account-id:scalingPolicy:d02ef525-8651-4314-
bf14-888331ebd04f:autoScalingGroupName/my-asg:policyName/cpu70-predictive-scaling-
policy",
  "Alarms": []
}
```

Valutazione delle policy di dimensionamento predittivo

Prima di dimensionare il gruppo con scalabilità automatica tramite una policy di dimensionamento predittivo, esamina i consigli e gli altri dati nella console Dimensionamento automatico Amazon EC2. È un'opzione importante per assicurarsi che le previsioni siano accurate prima di applicare una policy di dimensionamento predittivo che dimensiona la capacità effettiva.

Se il gruppo con scalabilità automatica è nuovo, attendere 24 ore per permettere a Dimensionamento automatico Amazon EC2 di creare la prima previsione.

Quando Dimensionamento automatico Amazon EC2 crea una previsione, utilizza dati cronologici. Se il gruppo con scalabilità automatica non dispone ancora di molti dati cronologici recenti, Dimensionamento automatico Amazon EC2 potrebbe temporaneamente riempire la previsione con aggregati creati dagli aggregati storici attualmente disponibili. Le previsioni vengono popolate per un massimo di due settimane prima della data di creazione di una policy.

Indice

- [Visualizzazione dei suggerimenti per il dimensionamento predittivo](#)
- [Analisi dei grafici di monitoraggio di dimensionamento predittivo](#)
- [Monitora le metriche di scalabilità predittiva con CloudWatch](#)

Visualizzazione dei suggerimenti per il dimensionamento predittivo

Per un'analisi efficace, Dimensionamento automatico Amazon EC2 dovrebbe avere almeno due policy di dimensionamento predittivo da confrontare. Tuttavia, è ancora possibile esaminare i risultati per una singola policy. Quando crei più policy, puoi valutare una policy che utilizza un parametro rispetto a un parametro che ne utilizza uno diverso. Puoi anche valutare l'impatto di diverse combinazioni di valori di destinazione e parametri. Dopo aver creato le policy di dimensionamento

predittivo, Dimensionamento automatico Amazon EC2 inizia immediatamente a valutare quale policy è in grado di dimensionare il gruppo in modo ottimale.

Per visualizzare i suggerimenti nella console Dimensionamento automatico Amazon EC2

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Nella scheda Auto scaling, in Policy di dimensionamento predittivo, puoi visualizzare i dettagli di una policy e i relativi suggerimenti. Il suggerimento indica se l'utilizzo della policy di dimensionamento predittivo garantisce risultati migliori rispetto al non utilizzo.

Se non sei sicuro che una policy di dimensionamento predittivo sia appropriata per il tuo gruppo, consulta le colonne Impatto sulla disponibilità e Impatto sui costi per scegliere quella giusta. Le informazioni di ogni colonna indicano l'impatto della policy.

- Impatto sulla disponibilità: indica se l'utilizzo della policy eviterebbe un impatto negativo sulla disponibilità eseguendo il provisioning di un numero sufficiente di istanze per gestire il carico di lavoro, rispetto al mancato utilizzo della policy.
- Impatto sui costi: indica se l'utilizzo della policy eviterebbe un impatto negativo sui costi non eseguendo un provisioning eccessivo delle istanze, rispetto al mancato utilizzo della policy. Se il provisioning è eccessivo, le istanze risultano sottoutilizzate o inattive, comportando un maggiore impatto sui costi.

Se disponi di più policy, accanto al nome della policy che offre i maggiori vantaggi in termini di disponibilità a un costo inferiore viene visualizzato il tag Previsione migliore. Viene attribuito un peso maggiore all'impatto sulla disponibilità.

4. (Facoltativo) Per selezionare il periodo di tempo desiderato per i risultati dei suggerimenti, scegli il valore preferito dal menu a discesa Periodo di valutazione: 2 giorni, 1 settimana, 2 settimane, 4 settimane, 6 settimane o 8 settimane. Per impostazione predefinita, il periodo di valutazione è rappresentato dalle ultime due settimane. Un periodo di valutazione maggiore fornisce più punti dati per i risultati del suggerimento. Tuttavia, l'aggiunta di più punti dati potrebbe non migliorare i risultati se i modelli di carico sono cambiati, ad esempio dopo un periodo di domanda eccezionalmente elevata. In questo caso, puoi ottenere un suggerimento più mirato esaminando i dati più recenti.

Note

I suggerimenti vengono generati solo per le policy in modalità solo previsione. Questa funzione restituisce i risultati migliori quando una policy è in modalità solo previsione per tutto il periodo di valutazione. Se avvii una policy in modalità Previsione e dimensionamento e in un secondo momento la modifichi in una modalità solo previsione, è probabile che i risultati siano falsati. Questo accade perché la policy ha già contribuito alla capacità effettiva.

Analisi dei grafici di monitoraggio di dimensionamento predittivo

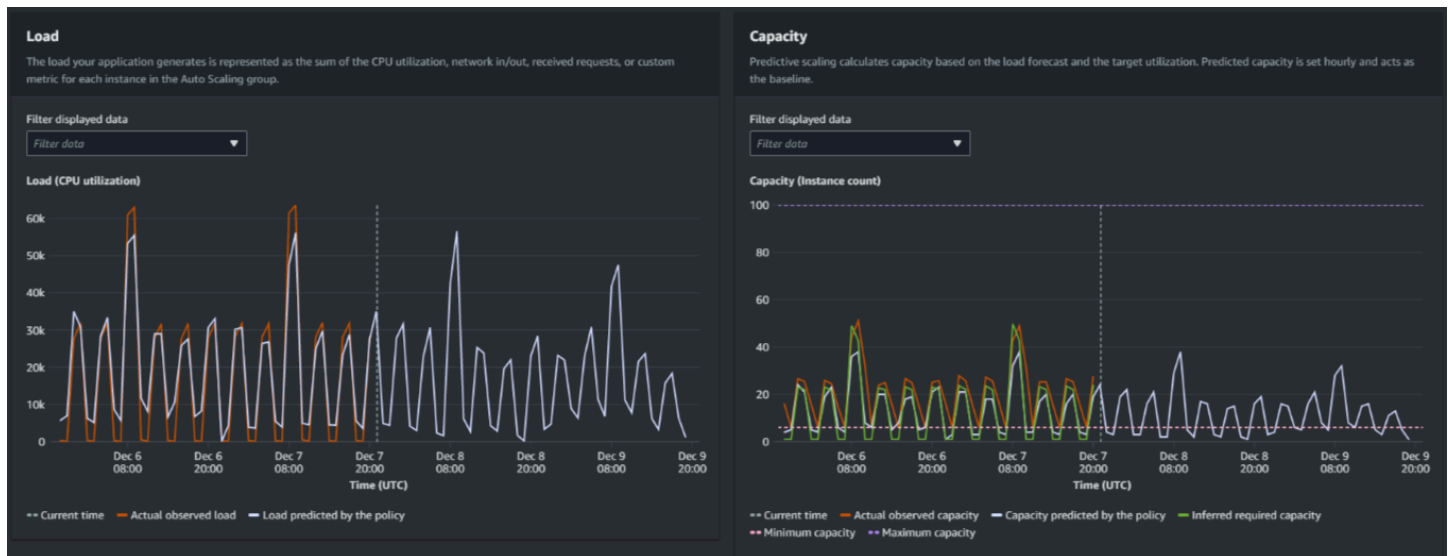
Nella console Dimensionamento automatico Amazon EC2, puoi esaminare le previsioni dei giorni, delle settimane o dei mesi precedenti per visualizzare le prestazioni della policy nel tempo. Puoi inoltre utilizzare queste informazioni per valutare l'accuratezza delle previsioni nel momento in cui decidi di applicare la policy per dimensionare la capacità effettiva.

Per esaminare i grafici di monitoraggio di dimensionamento predittivo nella console Dimensionamento automatico Amazon EC2

1. Scegli una policy dall'elenco Policy di dimensionamento predittivo.
2. Nella sezione Monitoraggio, puoi visualizzare le previsioni passate e future della policy in termini di carico e capacità rispetto ai valori effettivi. Il grafico Carico mostra la previsione di carico e i valori effettivi per il parametro di carico scelto. Il grafico Capacità mostra il numero di istanze previste dalla policy e il numero effettivo di istanze avviate. La linea verticale separa i valori storici dalle previsioni future. Questi grafici diventano disponibili poco dopo la creazione della policy.
3. (Facoltativo) Per modificare la quantità di dati cronologici mostrati nel grafico, scegli il valore desiderato dal menu a discesa Periodo di valutazione, nella parte superiore della pagina. Il periodo di valutazione non trasforma in alcun modo i dati di questa pagina, ma modifica soltanto la quantità di dati cronologici mostrati.

L'immagine seguente mostra i grafici Carico e Capacità quando le previsioni sono state applicate più volte. Le previsioni di dimensionamento predittivo si basano sui dati di carico storici. Il carico generato dall'applicazione è rappresentato come la somma dell'utilizzo della CPU, della rete in ingresso/uscita, delle richieste ricevute o del parametro personalizzato per ogni istanza del gruppo con scalabilità automatica. Il dimensionamento predittivo calcola le esigenze future in termini di capacità in base

alla previsione di carico e all'utilizzo di destinazione che si desidera ottenere per il parametro di dimensionamento.



Confronto dei dati nel grafico Carico

Ogni riga orizzontale rappresenta un diverso insieme di punti dati riportati a intervalli di un'ora:

1. Carico effettivo osservato utilizza la statistica SUM per il parametro di carico scelto per mostrare il carico orario totale in passato.
2. Carico previsto dalla policy mostra la previsione del carico orario. Questa previsione si basa sulle due settimane precedenti di osservazioni del carico effettivo.

Confronto dei dati nel grafico Capacità

Ogni riga orizzontale rappresenta un diverso insieme di punti dati riportati a intervalli di un'ora:

1. Capacità effettiva osservata mostra la capacità effettiva del gruppo con scalabilità automatica in passato, che dipende dalle altre policy di ridimensionamento e dalla dimensione minima del gruppo in vigore per il periodo di tempo selezionato.
2. Capacità prevista dalla policy mostra la capacità di base che si prevede di avere all'inizio di ogni ora, quando la policy è in modalità Previsione e dimensionamento.
3. Capacità richiesta differita mostra la capacità ideale per mantenere il parametro di dimensionamento al valore target scelto.
4. Capacità minima indica la capacità minima del gruppo con scalabilità automatica.
5. Capacità massima indica la capacità massima del gruppo con scalabilità automatica.

Per calcolare la capacità richiesta differita, partiamo dal presupposto che ogni istanza sia utilizzata allo stesso modo a un determinato valore target. In pratica, le istanze non vengono utilizzate allo stesso modo. Tuttavia, ipotizzando che l'utilizzo sia distribuito uniformemente tra le istanze, possiamo fare una stima verosimile della quantità di capacità necessaria. Il requisito di capacità viene quindi calcolato in modo che sia inversamente proporzionale al parametro di dimensionamento utilizzato per la relativa policy. In altre parole, all'aumentare della capacità, il parametro di dimensionamento diminuisce alla stessa velocità. Ad esempio, se la capacità raddoppia, il parametro di dimensionamento si dimezza.

La formula per la capacità richiesta differita è:

$$\text{sum of } (\text{actualCapacityUnits} * \text{scalingMetricValue}) / (\text{targetUtilization})$$

Ad esempio, prendiamo `actualCapacityUnits` (10) e `scalingMetricValue` (30) per una determinata ora. Prendiamo quindi il valore `targetUtilization` specificato nella policy di dimensionamento predittivo (60) e calcoliamo la capacità richiesta differita per la stessa ora. Viene restituito il valore 5. Ciò significa che cinque è la quantità di capacità richiesta necessaria per mantenere tale capacità in modo direttamente inversamente proporzionale al valore target del parametro di dimensionamento.

Note

Sono disponibili diversi fattori per regolare e migliorare i risparmi sui costi e la disponibilità dell'applicazione.

- Utilizza il dimensionamento predittivo per la capacità di base e il dimensionamento dinamico per gestire la capacità aggiuntiva. Il dimensionamento dinamico funziona indipendentemente dal dimensionamento predittivo, riducendo e aumentando orizzontalmente in base all'utilizzo corrente. Innanzitutto, Dimensionamento automatico Amazon EC2 calcola il numero consigliato di istanze per ogni policy di dimensionamento dinamico. Quindi, effettua il dimensionamento in base alla policy che fornisce il maggior numero di istanze.
- Per consentire la riduzione orizzontale quando il carico diminuisce, il gruppo con scalabilità automatica deve sempre disporre di almeno una policy di dimensionamento predittivo con la porzione di riduzione orizzontale abilitata.
- Puoi migliorare le prestazioni di dimensionamento assicurandoti che la capacità minima e massima non siano troppo restrittive. Una policy con un numero consigliato di istanze

che non rientra nell'intervallo di capacità minima e massima non sarà in grado di effettuare l'aumento o la riduzione orizzontale.

Monitora le metriche di scalabilità predittiva con CloudWatch

A seconda delle tue esigenze, potresti preferire accedere ai dati di monitoraggio per la scalabilità predittiva da Amazon CloudWatch anziché dalla console Amazon EC2 Auto Scaling. Dopo aver creato una policy di dimensionamento predittivo, la policy raccoglie i dati che vengono utilizzati per prevedere il carico e la capacità future. Una volta raccolti, questi dati vengono automaticamente archiviati a CloudWatch intervalli regolari. Quindi, è possibile utilizzarla CloudWatch per visualizzare le prestazioni della politica nel tempo. Puoi anche creare CloudWatch allarmi per avvisarti quando gli indicatori di performance superano i limiti definiti in. CloudWatch

Argomenti

- [Visualizzazione dei dati di previsione storici](#)
- [Creazione di parametri di precisione utilizzando la matematica dei parametri](#)

Visualizzazione dei dati di previsione storici

È possibile visualizzare i dati di previsione del carico e della capacità per una politica di scalabilità predittiva in. CloudWatch Ciò può essere utile quando si visualizzano le previsioni rispetto ad altre CloudWatch metriche in un unico grafico. Può essere utile anche quando si visualizza un intervallo di tempo più ampio in modo da poter vedere le tendenze nel tempo. Puoi accedere ai parametri cronologici fino a 15 mesi per avere una prospettiva migliore sulle performance di una policy.

Per ulteriori informazioni, consulta [Parametri e dimensioni di dimensionamento predittivo](#).

Per visualizzare i dati di previsione storici utilizzando la console CloudWatch

1. Apri la CloudWatch console all'[indirizzo https://console.aws.amazon.com/cloudwatch/](https://console.aws.amazon.com/cloudwatch/).
2. Nel pannello di navigazione, scegli Metrics (Parametri), quindi scegli All metrics (Tutti i parametri).
3. Scegli il parametro spazio dei nomi Auto Scaling (Dimensionamento automatico).
4. Scegli una delle seguenti opzioni per visualizzare i parametri relativi alla previsione del carico o alla previsione della capacità:
 - Previsioni del carico di dimensionamento predittivo

- Previsioni di capacità di dimensionamento predittivo
5. Nel campo di ricerca, immettere il nome della policy di dimensionamento predittivo o il nome del gruppo con scalabilità automatica, quindi premere Enter (Invio) per filtrare i risultati.
 6. Per creare il grafico di un parametro, seleziona la casella di controllo accanto al parametro. Per modificare il nome del grafico, seleziona l'icona a forma di matita. Per modificare l'intervallo di tempo, seleziona uno dei valori predefiniti o scegli custom (personalizzato). Per ulteriori informazioni, consulta la sezione [Rappresentazione grafica di una metrica](#) nella Amazon CloudWatch User Guide.
 7. Per modificare la statistica, seleziona la scheda Graphed metrics (Parametri nel grafico). Scegli l'intestazione di colonna o un valore singolo, quindi scegli un'altra statistica. Sebbene sia possibile scegliere qualsiasi statistica per ogni metrica, non tutte le statistiche sono utili per i parametri e le metriche. PredictiveScalingLoadForecastPredictiveScalingCapacityForecast Ad esempio, le statistiche media, minimo e massimo sono utili, ma la statistica somma non lo è.
 8. Per aggiungere un altro parametro al grafico, in Browse (Sfoggia), scegli All (Tutti), trova il parametro specifico, quindi seleziona la casella di controllo accanto a esso. Puoi aggiungere fino a 10 parametri.

Ad esempio, per aggiungere i valori effettivi per l'utilizzo della CPU al grafico, scegli il spazio dei nomi EC2 e poi scegli By Auto Scaling Group (Per gruppo con scalabilità automatica). Seleziona la casella di controllo per il parametro CPUUtilization e lo specifico gruppo con scalabilità automatica.

9. (Facoltativo) Per aggiungere il grafico a una CloudWatch dashboard, scegli Azioni, Aggiungi alla dashboard.

Creazione di parametri di precisione utilizzando la matematica dei parametri

Con la matematica metrica, puoi interrogare più CloudWatch metriche e utilizzare espressioni matematiche per creare nuove serie temporali basate su queste metriche. Puoi visualizzare le serie temporali risultanti sulla CloudWatch console e aggiungerle ai dashboard. Per ulteriori informazioni sulla matematica dei parametri, consulta [Using metric Math nella](#) Amazon User Guide. CloudWatch

Utilizzando la matematica dei parametri, puoi rappresentare graficamente i dati generati da Dimensionamento automatico Amazon EC2 per il dimensionamento predittivo in diversi modi. Questo ti aiuta a monitorare le prestazioni delle policy nel tempo e ti aiuta a capire se la tua combinazione di parametri può essere migliorata.

Ad esempio, è possibile utilizzare un'espressione matematica dei parametri per monitorare il [errore percentuale assoluto medio](#) (MAPE). Il parametro MAPE aiuta a monitorare la differenza tra i valori previsti e i valori effettivi osservati durante una determinata finestra di previsione. Le modifiche nel valore di MAPE possono indicare se le prestazioni della policy stanno peggiorando nel tempo in quanto la natura della tua applicazione cambia. Un aumento del MAPE segnala un divario più ampio tra i valori previsti e i valori effettivi.

Esempio: espressione matematica dei parametri

Per iniziare a utilizzare questo tipo di grafico, puoi creare un'espressione matematica dei parametri come quella mostrata nell'esempio seguente.

```
{
  "MetricDataQueries": [
    {
      "Expression": "TIME_SERIES(AVG(ABS(m1-m2)/m1))",
      "Id": "e1",
      "Period": 3600,
      "Label": "MeanAbsolutePercentageError",
      "ReturnData": true
    },
    {
      "Id": "m1",
      "Label": "ActualLoadValues",
      "MetricStat": {
        "Metric": {
          "Namespace": "AWS/EC2",
          "MetricName": "CPUUtilization",
          "Dimensions": [
            {
              "Name": "AutoScalingGroupName",
              "Value": "my-asg"
            }
          ]
        },
        "Period": 3600,
        "Stat": "Sum"
      },
      "ReturnData": false
    },
    {
      "Id": "m2",
      "Label": "ForecastedLoadValues",
```



```

"MetricStat": {
  "Metric": {
    "Namespace": "AWS/AutoScaling",
    "MetricName": "PredictiveScalingLoadForecast",
    "Dimensions": [
      {
        "Name": "AutoScalingGroupName",
        "Value": "my-asg"
      },
      {
        "Name": "PolicyName",
        "Value": "my-predictive-scaling-policy"
      },
      {
        "Name": "PairIndex",
        "Value": "0"
      }
    ]
  },
  "Period": 3600,
  "Stat": "Average"
},
"ReturnData": false
}
]
}

```

Invece di un singolo parametro, è disponibile una serie di strutture di query di dati dei parametri per `MetricDataQueries`. Ogni articolo in `MetricDataQueries` ottiene un parametro o esegue un'espressione matematica. Il primo articolo, `e1`, è l'espressione matematica. L'espressione designata imposta il parametro `ReturnData` a `true`, che alla fine produce una singola serie temporale. Per tutte le altre parametri, il valore `ReturnData` è `false`.

Nell'esempio, l'espressione designata utilizza i valori effettivi e previsti come input e restituisce la nuova metrica (MAPE). `m1` è la CloudWatch metrica che contiene i valori di carico effettivi (supponendo che l'utilizzo della CPU sia la metrica di carico originariamente specificata per la politica denominata). `my-predictive-scaling-policy` `m2` è la CloudWatch metrica che contiene i valori di carico previsti. La sintassi matematica per il parametro MAPE è la seguente:

Media di $(\text{abs}((\text{Effettivo} - \text{Previsto})/(\text{Effettivo})))$

Visualizza i parametri di precisione e imposta allarmi

Per visualizzare i dati delle metriche di precisione, seleziona la scheda Metriche nella console. CloudWatch È possibile rappresentare graficamente i dati da lì. Per ulteriori informazioni, consulta [Aggiungere un'espressione matematica a un CloudWatch grafico](#) nella Amazon CloudWatch User Guide.

Puoi anche impostare un allarme su un parametro che stai monitorando dalla sezione Metrics (Parametri). Nella scheda Graphed metrics (Parametri nel grafico), seleziona l'icona Create alarm (Crea allarme) nella colonna Actions (Operazioni). L'icona Create alarm (Crea allarme) è rappresentata come una piccola campana. Per ulteriori informazioni e opzioni di notifica, consulta [Creazione di un CloudWatch allarme basato su un'espressione matematica metrica](#) e [Notifica agli utenti delle modifiche agli allarmi](#) nella Amazon CloudWatch User Guide.

In alternativa, puoi utilizzare [GetMetricData](#) ed [PutMetricAlarm](#) eseguire calcoli utilizzando la matematica metrica e creare allarmi in base all'output.

Sovrascrivere i valori di previsione mediante operazioni pianificate

Talvolta, potrebbero essere disponibili ulteriori informazioni sui requisiti futuri dell'applicazione che il calcolo del forecast non è in grado di prendere in considerazione. Ad esempio, i calcoli del forecast potrebbero sottovalutare la capacità necessaria per un evento di marketing imminente. È possibile utilizzare le operazioni pianificate per sostituire temporaneamente il forecast nei periodi di tempo futuri. Le operazioni pianificate possono essere eseguite su base periodica o in una data e un'ora specifiche in cui si manifestino variazioni della domanda una tantum.

Ad esempio, è possibile creare un'operazione pianificata con una capacità minima superiore a quella prevista. In fase di runtime, Dimensionamento automatico Amazon EC2 aggiorna la capacità minima del gruppo con scalabilità automatica. Poiché il dimensionamento predittivo ottimizza la capacità, viene osservata un'azione pianificata con una capacità minima superiore ai valori di forecast. Ciò fa sì che la capacità non sia inferiore al previsto. Per interrompere la sostituzione del forecast, utilizza una seconda operazione pianificata per riportare la capacità minima all'impostazione originale.

La procedura seguente descrive le fasi per sostituire il forecast nei periodi di tempo futuri.

Indice

- [Fase 1: analizza i dati di serie temporali \(opzionale\)](#)
- [Fase 2: creazione di due operazioni pianificate](#)

Fase 1: analizza i dati di serie temporali (opzionale)

Inizia analizzando i dati delle serie temporali dei forecast. Si tratta di un passaggio facoltativo, ma è utile se desideri comprendere i dettagli del forecast.

1. Recupero del forecast

Dopo aver creato il forecast, puoi avviare una query per un periodo di tempo specifico nel forecast. L'obiettivo della query è ottenere lo scenario completo dei dati delle serie temporali per un periodo di tempo specifico.

La query può includere fino a due giorni di dati di forecast futuro. Se utilizzi il dimensionamento predittivo per un certo periodo di tempo, puoi anche accedere ai dati dei forecast precedenti. Tuttavia, la durata massima tra l'ora di inizio e di fine è 30 giorni.

Per ottenere la previsione utilizzando il [get-predictive-scaling-forecast](#) AWS CLI comando, fornite i seguenti parametri nel comando:

- Inserisci il nome del gruppo con scalabilità automatica nel parametro `--auto-scaling-group-name`.
- Inserisci il nome della policy nel parametro `--policy-name`.
- Inserisci l'ora di inizio nel parametro `--start-time` affinché restituisca solo i dati di forecast per il periodo di tempo o dopo l'intervallo di tempo specificato.
- Inserisci l'ora di fine nel parametro `--end-time` affinché restituisca solo i dati di forecast per il periodo di tempo precedente all'intervallo di tempo specificato.

```
aws autoscaling get-predictive-scaling-forecast --auto-scaling-group-name my-asg \  
--policy-name cpu40-predictive-scaling-policy \  
--start-time "2021-05-19T17:00:00Z" \  
--end-time "2021-05-19T23:00:00Z"
```

Se riuscito, il comando restituirà dati simili a quelli dell'esempio seguente:

```
{  
  "LoadForecast": [  
    {  
      "Timestamps": [  
        "2021-05-19T17:00:00+00:00",  
        "2021-05-19T18:00:00+00:00",
```

```
        "2021-05-19T19:00:00+00:00",
        "2021-05-19T20:00:00+00:00",
        "2021-05-19T21:00:00+00:00",
        "2021-05-19T22:00:00+00:00",
        "2021-05-19T23:00:00+00:00"
    ],
    "Values": [
        153.0655799339254,
        128.8288551285919,
        107.1179447150675,
        197.3601844551528,
        626.4039934516954,
        596.9441277518481,
        677.9675713779869
    ],
    "MetricSpecification": {
        "TargetValue": 40.0,
        "PredefinedMetricPairSpecification": {
            "PredefinedMetricType": "ASGCPUUtilization"
        }
    }
}
],
"CapacityForecast": {
    "Timestamps": [
        "2021-05-19T17:00:00+00:00",
        "2021-05-19T18:00:00+00:00",
        "2021-05-19T19:00:00+00:00",
        "2021-05-19T20:00:00+00:00",
        "2021-05-19T21:00:00+00:00",
        "2021-05-19T22:00:00+00:00",
        "2021-05-19T23:00:00+00:00"
    ],
    "Values": [
        2.0,
        2.0,
        2.0,
        2.0,
        4.0,
        4.0,
        4.0
    ]
}
},
"UpdateTime": "2021-05-19T01:52:50.118000+00:00"
```

}

La risposta include due forecast: `LoadForecast` e `CapacityForecast`. `LoadForecast` mostra il forecast del carico orario. `CapacityForecast` mostra i valori di forecast per la capacità necessaria su base oraria per gestire il carico previsto pur mantenendo un `TargetValue` di 40,0 (40% utilizzo medio della CPU).

2. Identificazione del periodo di tempo di destinazione

Identifica l'ora o le ore in cui deve avvenire la variazione della domanda una tantum. Ricorda che le date e le ore mostrate nel forecast sono in UTC.

Fase 2: creazione di due operazioni pianificate

Ora crea quindi due operazioni pianificate per un periodo di tempo specifico in cui l'applicazione avrà un carico superiore a quello previsto. Ad esempio, se è previsto un evento di marketing che genererà traffico nel tuo sito per un periodo di tempo limitato, puoi pianificare un'operazione singola per aggiornare la capacità minima all'ora di inizio prevista. Quindi, pianifica un'altra operazione per riportare la capacità minima all'impostazione originale al termine dell'evento.

Come creare due operazioni pianificate per eventi singoli (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Nella scheda Automatic scaling (Dimensionamento automatico) scegliere Create scheduled action (Crea operazione pianificata) in Scheduled actions (Operazioni pianificate).
4. Inserisci le seguenti impostazioni per le operazioni programmate:
 - a. In Name (Nome) specifica un nome per l'operazione pianificata.
 - b. In Min, inserisci la nuova capacità minima per il gruppo con scalabilità automatica. I Min devono essere minori o uguali alla dimensione massima del gruppo. Se il tuo valore per Min è superiore alla dimensione massima del gruppo, dovrai aggiornare Max.
 - c. In Recurrence (Ricorrenza), scegli Once (Una volta).
 - d. In Time zone (Fuso orario), scegli un fuso orario. Se non sceglierai alcun fuso orario, verrà utilizzato ETC/UTC per impostazione predefinita.

- e. Definisci il valore: Specific start time (Ora di inizio specifica).
5. Scegli Crea.

La console visualizza le operazioni programmate per il gruppo con scalabilità automatica.

6. Configura una seconda operazione programmata per ripristinare l'impostazione originale della capacità minima alla fine dell'evento. Il dimensionamento predittivo può dimensionare la capacità solo quando il valore impostato per Min è inferiore ai valori di forecast.

Come creare due operazioni pianificate per eventi singoli (AWS CLI)

Per utilizzare il AWS CLI per creare le azioni pianificate, utilizzare il comando [put-scheduled-update-group-action](#).

Ad esempio, definiamo una pianificazione che mantenga una capacità minima di tre istanze il 19 maggio alle 17:00 per otto ore. I comandi seguenti mostrano come implementare questo scenario.

Il primo comando [put-scheduled-update-group-action](#) indica ad Amazon EC2 Auto Scaling di aggiornare la capacità minima del gruppo Auto Scaling specificato alle 17:00 UTC del 19 maggio 2021.

```
aws autoscaling put-scheduled-update-group-action --scheduled-action-name my-event-start \  
  --auto-scaling-group-name my-asg --start-time "2021-05-19T17:00:00Z" --minimum-  
  capacity 3
```

Il secondo comando indica a Dimensionamento automatico Amazon EC2 di impostare la capacità minima del gruppo su una alle 1:00 UTC del 20 maggio 2021.

```
aws autoscaling put-scheduled-update-group-action --scheduled-action-name my-event-end \  
  --auto-scaling-group-name my-asg --start-time "2021-05-20T01:00:00Z" --minimum-  
  capacity 1
```

Dopo aver aggiunto queste operazioni pianificate al gruppo con scalabilità automatica, Dimensionamento automatico Amazon EC2 esegue le seguenti operazioni:

- Alle 17:00 UTC del 19 maggio 2021, viene eseguita la prima operazione pianificata. Se il gruppo include meno di tre istanze, il gruppo si dimensiona su tre istanze. Durante questo periodo e

per le otto ore successive, Dimensionamento automatico Amazon EC2 può continuare a essere aumentato orizzontalmente, se la capacità prevista è superiore alla capacità effettiva o se è in atto una policy di dimensionamento dinamico.

- All' 01:00 UTC del 20 maggio 2021, viene eseguita la seconda operazione pianificata. Questo restituisce la capacità minima all'impostazione originale alla fine dell'evento.

Dimensionamento in base a pianificazioni ricorrenti

Per sostituire il forecast per lo stesso periodo di tempo ogni settimana, crea due operazioni pianificate e fornisci la logica di data e ora utilizzando un'espressione cron.

Il formato dell'espressione cron è costituito da cinque campi separati da spazi: [Minute] [Hour] [Day_of_Month] [Month_of_Year] [Day_of_Week]. I campi possono contenere tutti i valori consentiti, inclusi i caratteri speciali.

Ad esempio, la seguente espressione cron campi esegue un'operazione ogni giorno alle 06:30. L'asterisco viene utilizzato come carattere jolly per abbinare tutti i valori di un campo.

```
30 6 * * 2
```

Consulta anche

Per ulteriori informazioni su come creare, elencare, modificare ed eliminare azioni programmate, vedi [Dimensionamento programmato per Dimensionamento automatico Amazon EC2](#).

Advanced predictive scaling policy configurations using custom metrics (Configurazioni avanzate delle policy di dimensionamento predittivo utilizzando parametri personalizzati)

In una policy di dimensionamento predittivo, puoi utilizzare parametri predefiniti o personalizzati. I parametri personalizzati sono utili quando i parametri predefiniti (CPU, I/O di rete e numero di richieste di Application Load Balancer) non descrivono sufficientemente il carico dell'applicazione.

Quando crei una politica di scalabilità predittiva con metriche personalizzate, puoi specificare altre CloudWatch metriche fornite da AWS oppure puoi specificare metriche che definisci e pubblichi tu stesso. Puoi anche utilizzare la matematica metrica per aggregare e trasformare le metriche esistenti in una nuova serie temporale che non viene tracciata automaticamente. AWS Quando si combinano

valori nei dati, ad esempio calcolando nuove somme o medie si parla di aggregazione. I dati risultanti sono chiamati aggregato.

La sezione seguente contiene le best practice e alcuni esempi di come costruire la struttura JSON per la policy.

Indice

- [Best practice](#)
- [Prerequisiti](#)
- [Costruzione di JSON per i parametri personalizzati](#)
- [Considerazioni e risoluzione dei problemi](#)
- [Limitazioni](#)

Best practice

Le seguenti best practice consentono di utilizzare i parametri personalizzati in modo più efficace:

- Per la specifica del parametro del carico, il parametro più utile è un parametro che rappresenta il carico su un gruppo con scalabilità automatica nel suo complesso, indipendentemente dalla capacità del gruppo.
- Per la specifica del parametro di dimensionamento, il parametro più utile da dimensionare è un parametro di velocità effettiva o utilizzo medio per istanza.
- Il parametro di dimensionamento deve essere inversamente proporzionale alla capacità. Cioè, se il numero di istanze nel gruppo con scalabilità automatica aumenta, il parametro di dimensionamento dovrebbe diminuire approssimativamente della stessa proporzione. Per garantire che il dimensionamento predittivo si comporti come previsto, anche il parametro del carico e il parametro di ridimensionamento devono essere fortemente correlati tra loro.
- L'utilizzo di destinazione deve corrispondere al tipo di parametro di dimensionamento. Per una configurazione di policy che utilizza l'utilizzo della CPU, questa è una percentuale di destinazione. Per una configurazione di policy che utilizza la velocità effettiva, ad esempio il numero di richieste o messaggi, si tratta del numero di richieste o messaggi di destinazione per istanza durante un intervallo di un minuto.
- Se questi suggerimenti non vengono seguiti, i valori futuri previsti della serie temporale probabilmente non saranno corretti. Per verificare che i dati siano corretti, puoi visualizzare i valori previsti nella console Dimensionamento automatico Amazon EC2. In alternativa, dopo aver

creato la politica di scalabilità predittiva, ispeziona `CapacityForecast` gli oggetti restituiti da una `LoadForecast` chiamata all'API. [GetPredictiveScalingForecast](#)

- Ti suggeriamo vivamente di configurare il dimensionamento predittivo in Forecast only (solo Forecast), in modo da poter valutare il Forecast prima che il dimensionamento predittivo inizi attivamente la capacità di dimensionamento.

Prerequisiti

Per aggiungere parametri personalizzati alla tua policy di dimensionamento predittivo, devi disporre delle autorizzazioni `cloudwatch:GetMetricData`.

Per specificare le tue metriche anziché le metriche AWS fornite da te, devi prima pubblicarle su CloudWatch. Per ulteriori informazioni, consulta [Pubblicazione di metriche personalizzate](#) nella Amazon CloudWatch User Guide.

Quando pubblichi i tuoi parametri, assicurati di pubblicare i punti dati con una frequenza minima di cinque minuti. Amazon EC2 Auto Scaling recupera i punti CloudWatch dati in base alla durata del periodo necessario. Ad esempio, la specifica dei parametri di carico utilizza parametri orari per misurare il carico dell'applicazione. CloudWatch utilizza i dati metrici pubblicati per fornire un unico valore di dati per ogni periodo di un'ora aggregando tutti i punti dati con timestamp che rientrano in ogni periodo di un'ora.

Costruzione di JSON per i parametri personalizzati

La sezione seguente contiene esempi su come configurare la scalabilità predittiva da cui interrogare i dati. CloudWatch. Esistono due metodi diversi per configurare questa opzione e il metodo scelto influisce sul formato utilizzato per costruire il JSON per la policy di scalabilità predittiva. Quando si utilizza la formula per i parametri, il formato del JSON varia ulteriormente in base alla formula eseguita.

1. Per creare una policy che ottenga i dati direttamente da altre CloudWatch metriche fornite da AWS o da metriche su cui pubblichi, consulta. CloudWatch [Esempio di policy di dimensionamento predittivo con parametri di carico e dimensionamento personalizzati \(AWS CLI\)](#)
2. Per creare una policy in grado di interrogare più CloudWatch metriche e utilizzare espressioni matematiche per creare nuove serie temporali basate su tali metriche, consulta. [Utilizzare le espressioni matematiche del parametro](#)

Esempio di policy di dimensionamento predittivo con parametri di carico e dimensionamento personalizzati (AWS CLI)

Per creare una politica di scalabilità predittiva con metriche di carico e scalabilità personalizzate con AWS CLI, memorizza gli argomenti per in un file JSON denominato. `--predictive-scaling-configuration config.json`

Inizi ad aggiungere parametri personalizzati sostituendo i valori sostituibili nell'esempio seguente con quelli dei tuoi parametri e del tuo obiettivo di destinazione.

```
{
  "MetricSpecifications": [
    {
      "TargetValue": 50,
      "CustomizedScalingMetricSpecification": {
        "MetricDataQueries": [
          {
            "Id": "scaling_metric",
            "MetricStat": {
              "Metric": {
                "MetricName": "MyUtilizationMetric",
                "Namespace": "MyNameSpace",
                "Dimensions": [
                  {
                    "Name": "MyOptionalMetricDimensionName",
                    "Value": "MyOptionalMetricDimensionValue"
                  }
                ]
              },
              "Stat": "Average"
            }
          ]
        },
        "CustomizedLoadMetricSpecification": {
          "MetricDataQueries": [
            {
              "Id": "load_metric",
              "MetricStat": {
                "Metric": {
                  "MetricName": "MyLoadMetric",
                  "Namespace": "MyNameSpace",
                  "Dimensions": [
```

```

        {
          "Name": "MyOptionalMetricDimensionName",
          "Value": "MyOptionalMetricDimensionValue"
        }
      ]
    },
    "Stat": "Sum"
  }
}

```

Per ulteriori informazioni, consulta il riferimento [MetricDataQuery](#) all'API Amazon EC2 Auto Scaling.

Note

Di seguito sono riportate alcune risorse aggiuntive che possono aiutarti a trovare nomi di metriche, namespace, dimensioni e statistiche relative alle metriche: CloudWatch

- Per informazioni sui parametri disponibili per AWS i servizi, consulta i [AWS servizi che pubblicano CloudWatch metriche](#) nella Amazon CloudWatch User Guide.
- [Per ottenere il nome esatto della metrica, lo spazio dei nomi e le dimensioni \(se applicabili\) di una CloudWatch metrica con, consulta list-metrics. AWS CLI](#)

Per creare questo criterio, esegui il [put-scaling-policy](#) comando utilizzando il file JSON come input, come illustrato nell'esempio seguente.

```

aws autoscaling put-scaling-policy --policy-name my-predictive-scaling-policy \
  --auto-scaling-group-name my-asg --policy-type PredictiveScaling \
  --predictive-scaling-configuration file://config.json

```

In caso di esito positivo, questo comando restituisce l'Amazon Resource Name (ARN) della policy.

```

{
  "PolicyARN": "arn:aws:autoscaling:region:account-id:scalingPolicy:2f4f5048-d8a8-4d14-b13a-d1905620f345:autoScalingGroupName/my-asg:policyName/my-predictive-scaling-policy",
  "Alarms": []
}

```

}

Utilizzare le espressioni matematiche del parametro

La sezione seguente fornisce informazioni ed esempi di policy di scalabilità predittiva che mostrano come utilizzare la formula dei parametri nella policy.

Indice

- [Comprendere la matematica dei parametri](#)
- [Esempio di policy di dimensionamento predittivo che combina parametri utilizzando la formula dei parametri \(AWS CLI\)](#)
- [Esempio di policy di dimensionamento predittivo da utilizzare in uno scenario di implementazione blu/verde \(AWS CLI\)](#)

Comprendere la matematica dei parametri

Se tutto ciò che vuoi fare è aggregare i dati metrici esistenti, la matematica metrica ti consente di risparmiare lo sforzo e il costo della pubblicazione di un'altra CloudWatch metrica su. CloudWatch Puoi utilizzare qualsiasi metrica AWS fornita e puoi anche utilizzare le metriche che definisci come parte delle tue applicazioni. Ad esempio, potresti calcolare il backlog della coda di Amazon SQS per istanza. È possibile eseguire questa operazione prendendo il numero approssimativo di messaggi disponibili per il recupero dalla coda e dividendo tale numero per la capacità operativa del gruppo con scalabilità automatica.

Per ulteriori informazioni, consulta [Using metric Math](#) nella Amazon CloudWatch User Guide.

Se scegli di utilizzare un'espressione matematica dei parametri nella policy di dimensionamento predittivo, considera i seguenti punti:

- Le operazioni di matematica dei parametri utilizzano i punti dati della combinazione univoca di nome del parametro, spazio dei nomi e coppie di parametri chiave/valore delle dimensioni.
- Puoi utilizzare qualsiasi operatore aritmetico (+ - */^), funzione statistica (come AVG o SUM) o altra funzione che supporti. CloudWatch
- È possibile utilizzare i parametri e i risultati di altre espressioni matematiche nelle formule dell'espressione matematica.
- Le espressioni matematiche dei parametri possono essere costituite da aggregazioni diverse. Tuttavia, per il risultato finale dell'aggregazione è una best practice utilizzare Average per il parametro di dimensionamento e Sum per il parametro del carico.

- Qualsiasi espressione utilizzata in una specifica dei parametri deve restituire una singola serie temporale.

Per utilizzare i parametri matematici, procedi come segue:

- Scegli una o più metriche. CloudWatch Quindi, crea l'espressione. Per ulteriori informazioni, consulta [Using metric Math](#) nella Amazon CloudWatch User Guide.
- Verifica che l'espressione matematica della metrica sia valida utilizzando la CloudWatch console o l'API. CloudWatch [GetMetricData](#)

Esempio di policy di dimensionamento predittivo che combina parametri utilizzando la formula dei parametri (AWS CLI)

A volte, invece di specificare direttamente il parametro, potrebbe essere necessario prima elaborarne i dati in qualche modo. Ad esempio, potresti avere un'applicazione che estrae il lavoro da una coda Amazon SQS e potresti utilizzare il numero di elementi nella coda come policy di dimensionamento predittivo. Il numero di messaggi nella coda non definisce esclusivamente il numero di istanze necessarie. Pertanto, è necessario più lavoro per creare un parametro che si può utilizzare per calcolare il backlog per istanza. Per ulteriori informazioni, consulta [Scalabilità basata su Amazon SQS](#).

Di seguito viene illustrato un esempio di policy di dimensionamento predittivo per questo scenario. Specifica i parametri di dimensionamento e del carico basati sul parametro `ApproximateNumberOfMessagesVisible` di Amazon SQS, ovvero il numero di messaggi disponibili per il recupero dalla coda. Utilizza anche il parametro `GroupInServiceInstances` di Dimensionamento automatico Amazon EC2 e un'espressione matematica per calcolare il backlog per istanza del parametro di dimensionamento.

```
aws autoscaling put-scaling-policy --policy-name my-sqs-custom-metrics-policy \  
  --auto-scaling-group-name my-asg --policy-type PredictiveScaling \  
  --predictive-scaling-configuration file://config.json  
{  
  "MetricSpecifications": [  
    {  
      "TargetValue": 100,  
      "CustomizedScalingMetricSpecification": {  
        "MetricDataQueries": [  
          {
```

```

    "Label": "Get the queue size (the number of messages waiting to be
processed)",
    "Id": "queue_size",
    "MetricStat": {
      "Metric": {
        "MetricName": "ApproximateNumberOfMessagesVisible",
        "Namespace": "AWS/SQS",
        "Dimensions": [
          {
            "Name": "QueueName",
            "Value": "my-queue"
          }
        ]
      },
      "Stat": "Sum"
    },
    "ReturnData": false
  },
  {
    "Label": "Get the group size (the number of running instances)",
    "Id": "running_capacity",
    "MetricStat": {
      "Metric": {
        "MetricName": "GroupInServiceInstances",
        "Namespace": "AWS/AutoScaling",
        "Dimensions": [
          {
            "Name": "AutoScalingGroupName",
            "Value": "my-asg"
          }
        ]
      },
      "Stat": "Sum"
    },
    "ReturnData": false
  },
  {
    "Label": "Calculate the backlog per instance",
    "Id": "scaling_metric",
    "Expression": "queue_size / running_capacity",
    "ReturnData": true
  }
]
},

```

```

"CustomizedLoadMetricSpecification": {
  "MetricDataQueries": [
    {
      "Id": "load_metric",
      "MetricStat": {
        "Metric": {
          "MetricName": "ApproximateNumberOfMessagesVisible",
          "Namespace": "AWS/SQS",
          "Dimensions": [
            {
              "Name": "QueueName",
              "Value": "my-queue"
            }
          ],
        },
        "Stat": "Sum"
      },
      "ReturnData": true
    }
  ]
}

```

L'esempio restituisce l'ARN della policy.

```

{
  "PolicyARN": "arn:aws:autoscaling:region:account-id:scalingPolicy:2f4f5048-d8a8-4d14-b13a-d1905620f345:autoScalingGroupName/my-asg:policyName/my-sqs-custom-metrics-policy",
  "Alarms": []
}

```

Esempio di policy di dimensionamento predittivo da utilizzare in uno scenario di implementazione blu/verde (AWS CLI)

Un'espressione di ricerca fornisce un'opzione avanzata in cui è possibile eseguire query per un parametro da più gruppi Auto Scaling ed eseguire espressioni matematiche su di essi. Ciò è particolarmente utile per implementazioni blu/verde.

Note

Un'implementazione blu/verde è un metodo di implementazione in cui si creano due gruppi Auto Scaling separati ma identici. Solo uno dei gruppi riceve il traffico di produzione. Il traffico utente viene inizialmente indirizzato al gruppo con scalabilità automatica precedente ("blu"), mentre un nuovo gruppo ("verde") viene utilizzato per testare e valutare una nuova versione di un'applicazione o di un servizio. Il traffico utente viene spostato sul gruppo con scalabilità automatica verde dopo che una nuova implementazione è stata testata e accettata. È quindi possibile eliminare il gruppo blu dopo che l'implementazione ha avuto esito positivo.

Quando vengono creati nuovi gruppi Auto Scaling nell'ambito di un'implementazione blu/verde, la cronologia dei parametri di ciascun gruppo può essere automaticamente inclusa nella policy di dimensionamento predittivo senza dover modificare le specifiche dei parametri. Per ulteriori informazioni, consulta [Utilizzo delle politiche di scalabilità predittiva di EC2 Auto Scaling con implementazioni Blue/Green sul](#) blog di Compute. AWS

Di seguito viene illustrato un esempio di policy che mostra come eseguire questa operazione. In questo esempio, la policy utilizza il parametro `CPUUtilization` emesso da Amazon EC2. Utilizza il parametro `GroupInServiceInstances` di Dimensionamento automatico Amazon EC2 e un'espressione matematica per calcolare il valore del parametro di dimensionamento per istanza. Definisce inoltre una specifica del parametro di capacità per ottenere il parametro `GroupInServiceInstances`.

L'espressione di ricerca trova `CPUUtilization` delle istanze in più gruppi Auto Scaling in base ai criteri di ricerca specificati. Se successivamente crei un nuovo gruppo con scalabilità automatica che soddisfi gli stessi criteri di ricerca, `CPUUtilization` delle istanze nel nuovo gruppo con scalabilità automatica viene automaticamente incluso.

```
aws autoscaling put-scaling-policy --policy-name my-blue-green-predictive-scaling-policy \  
  --auto-scaling-group-name my-asg --policy-type PredictiveScaling \  
  --predictive-scaling-configuration file://config.json  
{  
  "MetricSpecifications": [  
    {  
      "TargetValue": 25,  
      "CustomizedScalingMetricSpecification": {  
        "MetricDataQueries": [  

```



```

    {
      "Id": "load_sum",
      "Expression": "SUM(SEARCH('{AWS/EC2,AutoScalingGroupName} MetricName=
\"CPUUtilization\" ASG-myapp', 'Sum', 300))",
      "ReturnData": false
    },
    {
      "Id": "capacity_sum",
      "Expression": "SUM(SEARCH('{AWS/AutoScaling,AutoScalingGroupName}
MetricName=\"GroupInServiceInstances\" ASG-myapp', 'Average', 300))",
      "ReturnData": false
    },
    {
      "Id": "weighted_average",
      "Expression": "load_sum / capacity_sum",
      "ReturnData": true
    }
  ]
},
"CustomizedLoadMetricSpecification": {
  "MetricDataQueries": [
    {
      "Id": "load_sum",
      "Expression": "SUM(SEARCH('{AWS/EC2,AutoScalingGroupName} MetricName=
\"CPUUtilization\" ASG-myapp', 'Sum', 3600))"
    }
  ]
},
"CustomizedCapacityMetricSpecification": {
  "MetricDataQueries": [
    {
      "Id": "capacity_sum",
      "Expression": "SUM(SEARCH('{AWS/AutoScaling,AutoScalingGroupName}
MetricName=\"GroupInServiceInstances\" ASG-myapp', 'Average', 300))"
    }
  ]
}
}
]
}
}

```

L'esempio restituisce l'ARN della policy.

```
{
  "PolicyARN": "arn:aws:autoscaling:region:account-id:scalingPolicy:2f4f5048-d8a8-4d14-
b13a-d1905620f345:autoScalingGroupName/my-asg:policyName/my-blue-green-predictive-
scaling-policy",
  "Alarms": []
}
```

Considerazioni e risoluzione dei problemi

Se si verifica un problema durante l'utilizzo di parametri personalizzati, si consiglia di procedere come segue:

- Se viene restituito un messaggio di errore, leggi il messaggio e risolvi il problema segnalato, se possibile.
- Se si verifica un problema quando tenti di utilizzare un'espressione di ricerca in uno scenario di implementazione blu/verde, assicurati innanzitutto di capire come creare un'espressione di ricerca che cerca una corrispondenza parziale anziché una corrispondenza esatta. Inoltre, verifica che la query trovi solo i gruppi Auto Scaling che eseguono l'applicazione specifica. Per ulteriori informazioni sulla sintassi delle espressioni di ricerca, consulta la sintassi delle [espressioni di CloudWatch ricerca](#) nella Amazon CloudWatch User Guide.
- Se non hai convalidato un'espressione in anticipo, il [put-scaling-policy](#) comando la convalida quando crei la tua politica di scalabilità. Tuttavia, esiste la possibilità che questo comando non riesca a identificare la causa esatta degli errori rilevati. Per risolvere i problemi, risolvetevi gli errori che ricevete in risposta a una richiesta al comando. [get-metric-data](#) È inoltre possibile risolvere i problemi relativi all'espressione dalla console. CloudWatch
- Quando visualizzi i grafici Load (Carico) e Capacity (Capacità) nella console, il grafico Capacity (Capacità) potrebbe non mostrare alcun dato. Per garantire che i grafici dispongano di dati completi, assicurati di abilitare in modo coerente i parametri di gruppo per i gruppi Auto Scaling. Per ulteriori informazioni, consulta [Abilitazione dei parametri del gruppo con scalabilità automatica \(console\)](#).
- La specifica del parametro della capacità è utile solo per le implementazioni blu/verdi quando si dispone di applicazioni eseguite in diversi gruppi Auto Scaling nel corso del loro ciclo di vita. Questo parametro personalizzato consente di fornire la capacità totale di più gruppi Auto Scaling. Il dimensionamento predittivo utilizza questo strumento per mostrare i dati cronologici nel grafico Capacità nella console.
- Devi specificare `false` per `ReturnData` se `MetricDataQueries` specifica la funzione `SEARCH()` da sola senza una funzione matematica come `SUM()`. Questo perché le espressioni

di ricerca possono restituire più serie temporali e una specifica del parametro basata su un'espressione può restituire solo una serie temporale.

- Tutti i parametri coinvolti in un'espressione di ricerca devono avere la stessa risoluzione.

Limitazioni

- Puoi eseguire query su punti dati con un massimo di 10 parametri in un'unica specifica del parametro.
- Ai fini di questo limite, un'espressione conta come un parametro.

Verificare quali istanze Auto Scaling vengono terminate durante la riduzione orizzontale

Amazon EC2 Auto Scaling utilizza le politiche di terminazione per decidere l'ordine di chiusura delle istanze. Puoi utilizzare una politica predefinita o creare una politica personalizzata per soddisfare i tuoi requisiti specifici. Utilizzando una policy personalizzata o una protezione scalabile in base alle istanze, puoi anche impedire al gruppo Auto Scaling di terminare le istanze che non sono ancora pronte per la chiusura.

Indice

- [Quando Amazon EC2 Auto Scaling utilizza politiche di terminazione](#)
- [Configurazione delle politiche di terminazione per Amazon EC2 Auto Scaling](#)
- [Creare una policy di terminazione personalizzata con Lambda.](#)
- [Utilizzare la protezione di riduzione orizzontale dell'istanza](#)
- [Progetta le tue applicazioni su Dimensionamento automatico Amazon EC2 per gestire senza problemi la terminazione delle istanze](#)

Quando Amazon EC2 Auto Scaling utilizza politiche di terminazione

Nelle sezioni seguenti, vengono descritti gli scenari in cui Dimensionamento automatico Amazon EC2 utilizza le policy di terminazione.

Indice

- [Eventi di riduzione orizzontale](#)

- [Aggiornamento istanza](#)
- [Ribilanciamento della zona di disponibilità](#)

Eventi di riduzione orizzontale

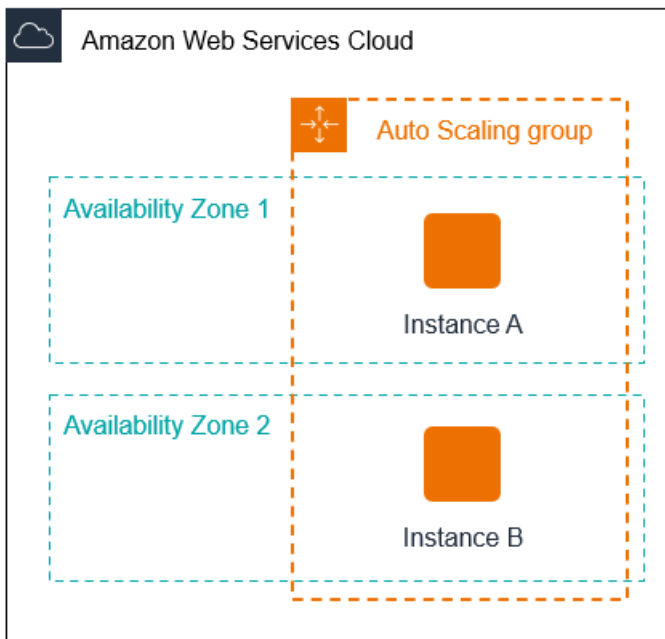
Un evento di riduzione orizzontale può verificarsi anche a causa di un'operazione pianificata o in cui un nuovo valore per la capacità desiderata di un gruppo con scalabilità automatica è inferiore alla capacità corrente del gruppo.

Gli eventi di riduzione orizzontale si verificano nei seguenti scenari:

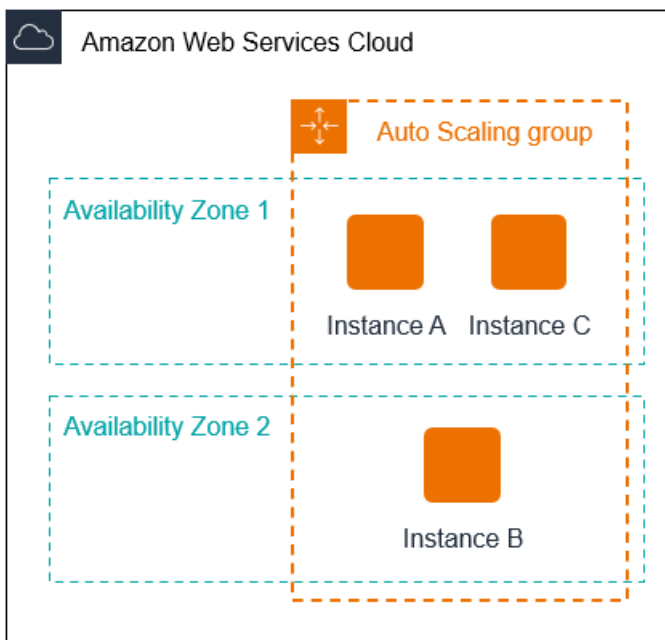
- Quando si utilizzano policy di dimensionamento dinamico e la dimensione del gruppo diminuisce a seguito di modifiche nel valore di un parametro
- Quando si utilizza il dimensionamento programmato e la dimensione del gruppo diminuisce a seguito di un'operazione pianificata
- Quando hai ridotto manualmente le dimensioni del gruppo

Nell'esempio seguente viene illustrato il funzionamento delle policy di terminazione quando si verifica un evento di dimensionamento.

1. Prendiamo l'esempio in cui si dispone di un gruppo con scalabilità automatica con un tipo di istanza, due zone di disponibilità e una capacità desiderata di due istanze. È inoltre dotato di una policy di dimensionamento dinamico che aggiunge e rimuove le istanze quando l'utilizzo delle risorse aumenta o diminuisce. Le due istanze di questo gruppo sono distribuite tra le due zone di disponibilità, come illustrato nel diagramma seguente.

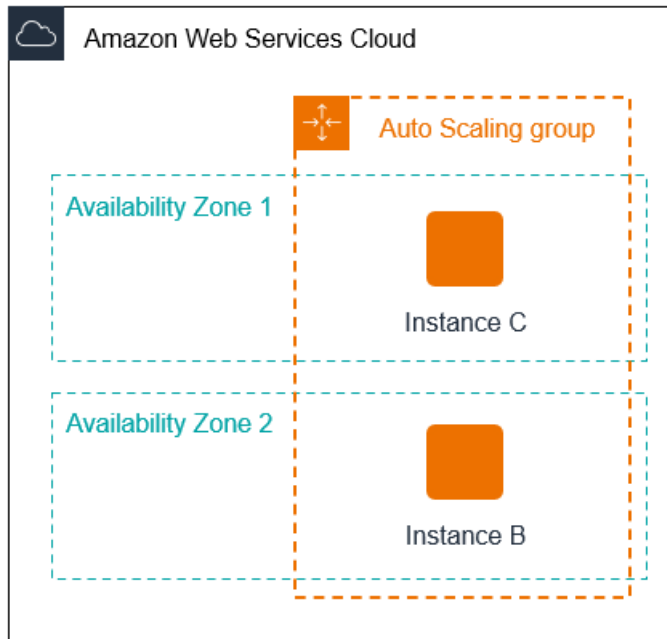


- Quando il gruppo con scalabilità automatica viene dimensionato, Dimensionamento automatico Amazon EC2 avvia una nuova istanza. Il gruppo con scalabilità automatica ora dispone di tre istanze, distribuite tra le due zone di disponibilità, come illustrato nel diagramma seguente.



- Quando il gruppo con scalabilità automatica viene ridotto, Dimensionamento automatico Amazon EC2 termina una delle istanze.
- Se non hai assegnato una policy di terminazione specifica al gruppo, Dimensionamento automatico Amazon EC2 utilizza quella di default. Seleziona la zona di disponibilità con due istanze e termina l'istanza che è stata avviata da una configurazione di avvio, da un modello di

avvio diverso o dalla versione più vecchia del modello di avvio corrente. Se le istanze sono state avviate dallo stesso modello e versione di lancio, Amazon EC2 Auto Scaling seleziona l'istanza più vicina all'ora di fatturazione successiva e la interrompe.



Aggiornamento istanza

Puoi avviare un aggiornamento dell'istanza per aggiornare le istanze nel tuo gruppo Auto Scaling. Durante l'aggiornamento di un'istanza, Dimensionamento automatico Amazon EC2 termina le istanze del gruppo e quindi avvia le sostituzioni per le istanze terminate. La policy di terminazione per il gruppo con scalabilità automatica controlla quali istanze verranno sostituite per prime.

Ribilanciamento della zona di disponibilità

Dimensionamento automatico Amazon EC2 bilancia la tua capacità in modo uniforme, nelle zone di disponibilità abilitate per il tuo gruppo con scalabilità automatica. Ciò permette di ridurre l'impatto di un'interruzione della zona di disponibilità. Se la distribuzione della capacità tra le zone di disponibilità diventa sbilanciata, Dimensionamento automatico Amazon EC2 ribilancia il gruppo con scalabilità automatica avviando istanze nelle zone di disponibilità abilitate con il minor numero di istanze e terminandole. La policy di terminazione controlla quali istanze vengono definite prioritarie per la terminazione.

Esistono diversi motivi per cui la distribuzione delle istanze tra le zone di disponibilità può risultare sbilanciata.

Rimozione delle istanze

Se distacchi le istanze dal gruppo con scalabilità automatica o termini esplicitamente le istanze e decrementi la capacità desiderata, impedendo così l'avvio delle istanze sostitutive, il gruppo potrebbe risultare sbilanciato. Se ciò accade, Dimensionamento automatico Amazon EC2 compensa ribilanciando le zone di disponibilità.

Utilizzo di zone di disponibilità diverse da quelle specificate in origine

Se espandi il gruppo con scalabilità automatica affinché includa ulteriori zone di disponibilità o modifichi le zone di disponibilità utilizzate, Dimensionamento automatico Amazon EC2 avvierà le istanze nelle nuove zone di disponibilità e le terminerà nelle altre, per garantire che si estendano uniformemente alle zone di disponibilità.

Interruzione di disponibilità

Le interruzioni di disponibilità sono rare. Tuttavia, se una zona diventa non disponibile e viene ripristinata in seguito, il gruppo con scalabilità automatica potrebbe risultare sbilanciato tra le zone di disponibilità. Dimensionamento automatico Amazon EC2 tenta di ribilanciare gradualmente il gruppo e il ribilanciamento potrebbe terminare le istanze in altre zone.

Per esempio, immaginate di avere un gruppo con scalabilità automatica con un tipo di istanza, due zone di disponibilità e una capacità desiderata di due istanze. In una situazione in cui in una zona di disponibilità si verifica un'interruzione, Dimensionamento automatico Amazon EC2 avvia automaticamente una nuova istanza nella zona di disponibilità integra per sostituire quella nella zona di disponibilità non integra. Quando la zona di disponibilità non integra torna a uno stato integro, Dimensionamento automatico Amazon EC2 avvierà automaticamente una nuova istanza in questa zona, che a sua volta terminerà un'istanza nella zona non interessata.

Note

In fase di ribilanciamento, Dimensionamento automatico Amazon EC2 avvia nuove istanze, prima di arrestare quelle vecchie, in modo che il ribilanciamento non comprometta le prestazioni o la disponibilità delle applicazioni.

Poiché Dimensionamento automatico Amazon EC2 tenta di avviare nuove istanze prima di terminare quelle vecchie, se si è raggiunta o si sta per raggiungere la capacità massima specificata, ciò potrebbe ostacolare o bloccare completamente le attività di ribilanciamento. Per evitare questo problema, durante un'attività di ribilanciamento il sistema è in grado di superare temporaneamente la capacità massima specificata di un gruppo con un margine del 10% (o con un margine di 1 istanza, a seconda del valore più alto). Il margine viene esteso

solo se il gruppo ha raggiunto o è vicino alla capacità massima e necessita di riequilibrio, sia a causa di ridefinizione delle aree richiesta dall'utente o per compensare problemi di zone di disponibilità. L'estensione dura solo il tempo necessario per ribilanciare il gruppo.

Configurazione delle politiche di terminazione per Amazon EC2 Auto Scaling

Una politica di terminazione fornisce i criteri seguiti da Amazon EC2 Auto Scaling per terminare le istanze in un ordine specifico.

Per impostazione predefinita, Amazon EC2 Auto Scaling utilizza una policy di terminazione progettata per terminare prima le istanze che utilizzano configurazioni obsolete. Puoi modificare la politica di terminazione per controllare quali istanze è più importante chiudere per prime.

Quando Amazon EC2 Auto Scaling termina le istanze, cerca di mantenere l'equilibrio tra le zone di disponibilità abilitate per il gruppo Auto Scaling. Il mantenimento dell'equilibrio zonale ha la precedenza sulla politica di cessazione. Se una zona di disponibilità ha più istanze di altre, Amazon EC2 Auto Scaling applica prima la politica di terminazione alla zona sbilanciata. Se le zone di disponibilità sono bilanciate, applica la politica di terminazione a tutte le zone.

Argomenti

- [Come funziona la politica di terminazione predefinita](#)
- [Policy di terminazione di default e gruppi di istanze miste](#)
- [Politiche di terminazione predefinite](#)
- [Modificare la politica di terminazione per un gruppo Auto Scaling](#)

Come funziona la politica di terminazione predefinita

Quando Amazon EC2 Auto Scaling deve terminare un'istanza, identifica innanzitutto quale zona (o zone) di disponibilità ha il maggior numero di istanze e almeno un'istanza non protetta dallo scalabilità in. Quindi, procede alla valutazione delle istanze non protette all'interno della zona di disponibilità identificata come segue:

Istanze che utilizzano configurazioni obsolete

- Per i gruppi che utilizzano un modello di avvio: determina se una delle istanze utilizza configurazioni obsolete, assegnando le priorità in questo ordine:

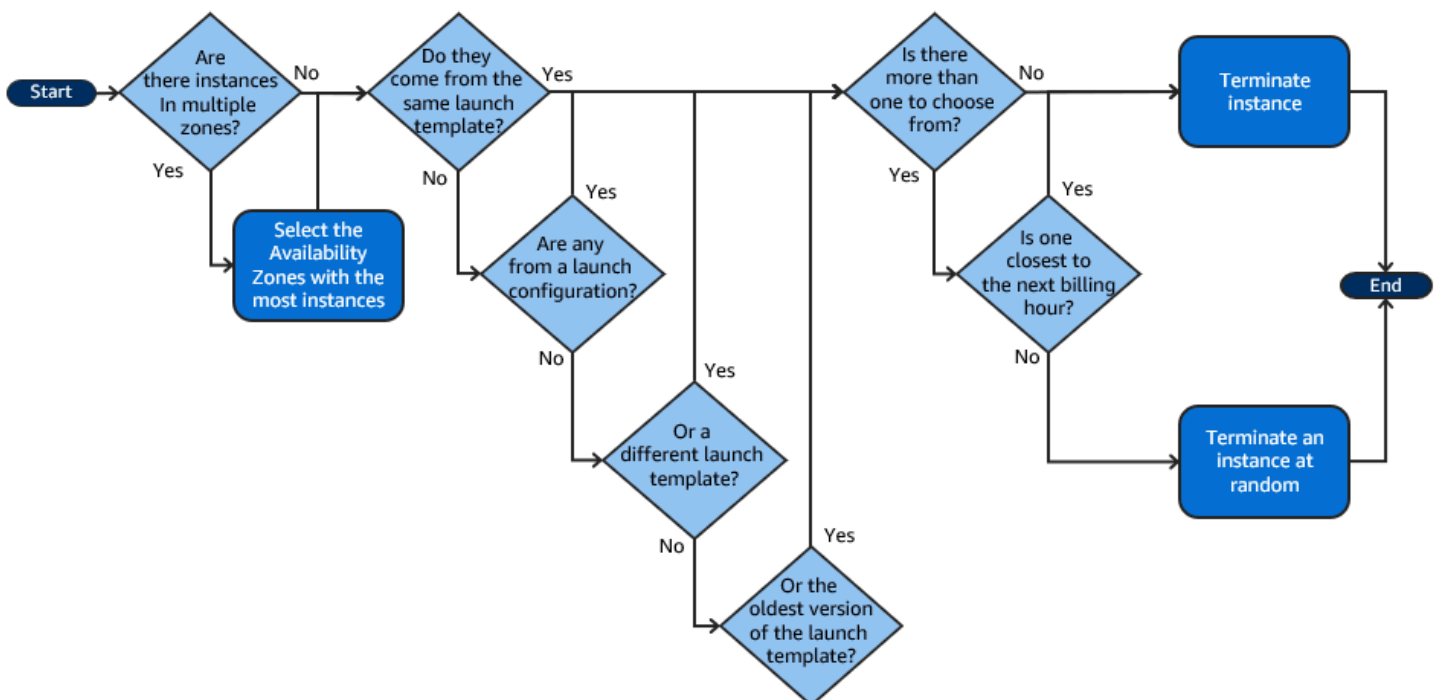
1. Innanzitutto, controlla le istanze avviate con una configurazione di avvio.
 2. Quindi, controlla le istanze avviate utilizzando un modello di avvio diverso anziché il modello di lancio corrente.
 3. Infine, verifica le istanze che utilizzano la versione più vecchia del modello di lancio corrente.
- Per i gruppi che utilizzano una configurazione di avvio: determina se una delle istanze utilizza la configurazione di avvio più vecchia.

Se non viene trovata alcuna istanza con configurazioni obsolete o se ci sono più istanze tra cui scegliere, Amazon EC2 Auto Scaling considera il criterio successivo delle istanze che si avvicinano all'ora di fatturazione successiva.

Istanze che si avvicinano all'ora di fatturazione successiva

Determina se alcune delle istanze che soddisfano i criteri precedenti sono più vicine all'ora di fatturazione successiva. Se più istanze sono ugualmente vicine, terminane una a caso. Questo ti aiuta a massimizzare l'uso delle istanze con fatturazione oraria. Tuttavia, la maggior parte dell'utilizzo di EC2 viene ora fatturato al secondo, quindi questa ottimizzazione offre meno vantaggi. Per ulteriori informazioni, consulta [Prezzi di Amazon EC2](#).

Il seguente diagramma di flusso illustra come funziona la politica di terminazione predefinita per i gruppi che utilizzano un modello di avvio.



Policy di terminazione di default e gruppi di istanze miste

[Amazon EC2 Auto Scaling applica criteri aggiuntivi quando termina le istanze in gruppi di istanze miste.](#)

Quando Amazon EC2 Auto Scaling deve terminare un'istanza, identifica innanzitutto quale opzione di acquisto (Spot o On-Demand) deve essere interrotta in base alle impostazioni del gruppo. Ciò garantisce che il gruppo tenda verso il rapporto specificato tra istanze Spot e On-Demand nel tempo.

Quindi applica la politica di risoluzione in modo indipendente all'interno di ciascuna zona di disponibilità. Determina quale istanza Spot o On-Demand in quale zona di disponibilità chiudere per mantenere le zone di disponibilità bilanciate. La stessa logica si applica a un gruppo di istanze miste con pesi definiti per i tipi di istanze.

All'interno di ciascuna zona, la politica di terminazione predefinita funziona come segue per determinare quale istanza non protetta all'interno dell'opzione di acquisto identificata può essere interrotta:

1. Determina se una delle istanze può essere terminata per migliorare l'allineamento con la strategia di [allocazione specificata per il gruppo Auto Scaling](#). Se non viene identificata alcuna istanza da ottimizzare o se ci sono più istanze tra cui scegliere, la valutazione continua.
2. Determina se una delle istanze utilizza configurazioni obsolete, assegnando le priorità in questo ordine:
 - a. Innanzitutto, controlla le istanze avviate con una configurazione di avvio.
 - b. Quindi, controlla le istanze avviate utilizzando un modello di avvio diverso anziché il modello di lancio corrente.
 - c. Infine, verifica le istanze che utilizzano la versione più vecchia del modello di lancio corrente.

Se non viene trovata alcuna istanza con configurazioni obsolete o se ci sono più istanze tra cui scegliere, la valutazione continua.

3. Determina se una delle istanze è la più vicina all'ora di fatturazione successiva. Se più istanze sono ugualmente vicine, sceglie una a caso.

Politiche di terminazione predefinite

È possibile scegliere tra le seguenti politiche di terminazione predefinite:

- **Default**— Termina le istanze in base alla politica di terminazione predefinita.

- **AllocationStrategy**— Termina le istanze nel gruppo Auto Scaling per allineare le istanze rimanenti alla strategia di allocazione per il tipo di istanza che sta terminando (istanza Spot o istanza On-Demand). Questa policy è utile quando i tipi di istanza preferiti sono cambiati. Se la strategia di allocazione Spot è `lowest-price`, puoi gradualmente ribilanciare la distribuzione di istanze Spot tra i pool con il prezzo N più basso. Se la strategia di allocazione Spot è `capacity-optimized`, puoi gradualmente ribilanciare la distribuzione di istanze Spot tra i pool dove la capacità Spot è più disponibile. Inoltre, puoi sostituire gradualmente le istanze on demand di un tipo con priorità inferiore con istanze on demand di un tipo con priorità superiore.
- **OldestLaunchTemplate**— Termina le istanze che hanno il modello di avvio più vecchio. Con questa policy, le istanze che usano il modello di avvio non corrente vengono terminate per prime, seguite da quelle che utilizzano la versione meno recente del modello di avvio corrente. Questa policy è utile quando si aggiorna un gruppo e si eliminano le istanze da una configurazione precedente.
- **OldestLaunchConfiguration**— Termina le istanze che hanno la configurazione di avvio più vecchia. Questa policy è utile quando si aggiorna un gruppo e si eliminano le istanze da una configurazione precedente. Con questa policy, le istanze che usano la configurazione di avvio non corrente vengono terminate per prime.
- **ClosestToNextInstanceHour**— Termina le istanze più vicine all'ora di fatturazione successiva. Questa policy consente di ottimizzare l'uso delle istanze che hanno una tariffa oraria.
- **NewestInstance**— Termina l'istanza più recente del gruppo. Questa policy è utile quando si esegue il test di una nuova configurazione di avvio, ma non si desidera mantenerla in produzione.
- **OldestInstance**— Termina l'istanza più vecchia del gruppo. Questa opzione è utile quando si aggiornano le istanze nel gruppo con scalabilità automatica a un nuovo tipo di istanza EC2. Puoi sostituire gradualmente le istanze del vecchio tipo con le istanze del nuovo tipo.

Note

Dimensionamento automatico Amazon EC2 bilancia sempre prima le istanze tra le zone di disponibilità, indipendentemente dalla policy di terminazione utilizzata. Di conseguenza, è possibile che si verifichino situazioni in cui alcune istanze più recenti vengono terminate prima di quelle meno recenti. Ad esempio, quando è stata aggiunta una zona di disponibilità più recente o quando una di esse ha più istanze rispetto alle altre zone di disponibilità utilizzate dal gruppo.

Modificare la politica di terminazione per un gruppo Auto Scaling

Per modificare la politica di terminazione per il gruppo Auto Scaling, utilizzate uno dei seguenti metodi.

Console

Non puoi modificare la politica di terminazione quando crei inizialmente un gruppo Auto Scaling nella console Amazon EC2 Auto Scaling. La policy di terminazione di default viene utilizzata automaticamente. Dopo aver creato il gruppo Auto Scaling, è possibile sostituire la politica predefinita con una politica di terminazione diversa o con più politiche di terminazione elencate nell'ordine in cui devono essere applicate.

Per modificare la politica di terminazione per un gruppo Auto Scaling

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Nella scheda Details (Dettagli) scegliere Advanced configurations (Configurazioni avanzate), Edit (Modifica).
4. Per Termination policies (Policy di terminazione), scegli una o più policy di terminazione. Se si scelgono più policy, elencarle nell'ordine in cui devono essere valutate.

È anche possibile scegliere Custom termination policy (Policy di terminazione personalizzata) e quindi scegliere una funzione Lambda che soddisfi le proprie esigenze. Se sono state create versioni e alias per la funzione Lambda, è possibile scegliere una versione o un alias dal menu a discesa Version/Alias (Versione/Alias). Per utilizzare la versione non pubblicata della funzione Lambda, mantieni Version/Alias (Versione/Alias) impostato sul valore di default. Per ulteriori informazioni, consulta [Creare una policy di terminazione personalizzata con Lambda..](#)

Note

Quando si utilizzano più policy, il loro ordine deve essere impostato correttamente:

- Se utilizzi la policy Default (Di default), deve essere l'ultima policy dell'elenco.

- Se utilizzi Custom termination policy (Policy di terminazione personalizzata), deve essere la prima policy dell'elenco.

5. Scegli Aggiorna.

AWS CLI

Viene utilizzata automaticamente la policy di terminazione di default, a meno che non ne venga specificata una diversa.

Per modificare la politica di terminazione per un gruppo Auto Scaling

Utilizzare uno dei seguenti comandi:

- [create-auto-scaling-group](#)
- [update-auto-scaling-group](#)

Puoi utilizzare queste policy singolarmente oppure combinarle in un elenco di policy. Ad esempio, esegui questo comando per aggiornare un gruppo con scalabilità automatica in modo da utilizzare prima la policy `OldestLaunchConfiguration` e poi la policy `ClosestToNextInstanceHour`.

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg --  
termination-policies "OldestLaunchConfiguration" "ClosestToNextInstanceHour"
```

Se utilizzi la policy di terminazione `Default`, collocala nell'ultima posizione dell'elenco delle policy di terminazione. Ad esempio, `--termination-policies "OldestLaunchConfiguration" "Default"`.

Per utilizzare una politica di terminazione personalizzata, è necessario prima creare la politica di risoluzione utilizzando `AWS Lambda`. Per specificare la funzione Lambda da utilizzare come policy di terminazione, impostala alla prima posizione dell'elenco delle policy di terminazione. Ad esempio, `--termination-policies "arn:aws:lambda:us-west-2:123456789012:function:HelloFunction:prod" "OldestLaunchConfiguration"`. Per ulteriori informazioni, consulta [Creare una policy di terminazione personalizzata con Lambda](#).

Creare una policy di terminazione personalizzata con Lambda.

Dimensionamento automatico Amazon EC2 utilizza le policy di terminazione per assegnare la priorità alle istanze da terminare per prime quando si riducono le dimensioni del gruppo con scalabilità automatica (indicato come scaling in (riduzione orizzontale)). Il gruppo con scalabilità automatica utilizza una policy di terminazione di default, ma è possibile scegliere o creare policy di terminazione personalizzate. Per ulteriori informazioni su come scegliere una policy di terminazione di default, consulta [Configurazione delle politiche di terminazione per Amazon EC2 Auto Scaling](#).

In questa sezione viene descritto come creare una policy di terminazione personalizzata utilizzando una funzione AWS Lambda , che Dimensionamento automatico Amazon EC2 richiama in risposta a determinati eventi. La funzione Lambda creata elabora le informazioni nei dati di input inviati da Dimensionamento automatico Amazon EC2 e restituisce un elenco di istanze pronte per la terminazione.

Una policy di terminazione personalizzata fornisce un controllo migliore su quali istanze vengono terminate e quando. Ad esempio, quando il gruppo con scalabilità automatica viene ridotto, Dimensionamento automatico Amazon EC2 non è in grado di determinare se siano in esecuzione carichi di lavoro che non devono essere interrotti. Con una funzione Lambda, puoi convalidare la richiesta di terminazione e attendere che il carico di lavoro venga completato, prima di restituire l'ID istanza a Dimensionamento automatico Amazon EC2 per la terminazione.

Indice

- [Dati di input](#)
- [Dati di risposta](#)
- [Considerazioni](#)
- [Creazione della funzione Lambda](#)
- [Limitazioni](#)

Dati di input

Dimensionamento automatico Amazon EC2 genera un payload JSON per gli eventi di riduzione orizzontale e lo fa anche quando le istanze stanno per essere terminate a causa della durata massima dell'istanza o delle funzionalità di aggiornamento dell'istanza. Inoltre, genera un payload JSON per gli eventi di dimensionamento che può avviare durante il ribilanciamento del gruppo tra le zone di disponibilità.

Questo payload contiene informazioni sulla capacità che Dimensionamento automatico Amazon EC2 deve terminare, un elenco di istanze suggerite per la terminazione e l'evento che l'ha avviata.

Di seguito è riportato un esempio di payload:

```
{
  "AutoScalingGroupARN": "arn:aws:autoscaling:us-east-1:<account-
id>:autoScalingGroup:d4738357-2d40-4038-ae7e-b00ae0227003:autoScalingGroupName/my-asg",
  "AutoScalingGroupName": "my-asg",
  "CapacityToTerminate": [
    {
      "AvailabilityZone": "us-east-1b",
      "Capacity": 2,
      "InstanceMarketOption": "on-demand"
    },
    {
      "AvailabilityZone": "us-east-1b",
      "Capacity": 1,
      "InstanceMarketOption": "spot"
    },
    {
      "AvailabilityZone": "us-east-1c",
      "Capacity": 3,
      "InstanceMarketOption": "on-demand"
    }
  ],
  "Instances": [
    {
      "AvailabilityZone": "us-east-1b",
      "InstanceId": "i-0056faf8da3e1f75d",
      "InstanceType": "t2.nano",
      "InstanceMarketOption": "on-demand"
    },
    {
      "AvailabilityZone": "us-east-1c",
      "InstanceId": "i-02e1c69383a3ed501",
      "InstanceType": "t2.nano",
      "InstanceMarketOption": "on-demand"
    },
    {
      "AvailabilityZone": "us-east-1c",
      "InstanceId": "i-036bc44b6092c01c7",
      "InstanceType": "t2.nano",
      "InstanceMarketOption": "on-demand"
    }
  ]
}
```

```
    },  
    ...  
  ],  
  "Cause": "SCALE_IN"  
}
```

Il payload include il nome del gruppo con scalabilità automatica, il relativo Amazon Resource Name (ARN) e i seguenti elementi:

- `CapacityToTerminate` descrive la quantità di capacità Spot o on demand impostata per la terminazione in una determinata zona di disponibilità.
- `Instances` rappresenta le istanze suggerite da Dimensionamento automatico Amazon EC2 per la terminazione in base alle informazioni contenute in `CapacityToTerminate`.
- `Cause` descrive l'evento che ha attivato la terminazione: `SCALE_IN`, `INSTANCE_REFRESH`, `MAX_INSTANCE_LIFETIME` oppure `REBALANCE`.

Le seguenti informazioni delineano i fattori più significativi nel modo in cui Dimensionamento automatico Amazon EC2 genera le `Instances` nei dati di input:

- Il mantenimento del bilanciamento tra le zone di disponibilità ha la precedenza quando un'istanza viene terminata a causa di eventi di dimensionamento e di terminazione basate sull'aggiornamento dell'istanza. Di conseguenza, se una zona di disponibilità ha più istanze delle altre zone di disponibilità usate dal gruppo, i dati di input contengono istanze che sono ammissibili per la terminazione solo dalla zona di disponibilità sbilanciata. Se le zone di disponibilità utilizzate dal gruppo sono bilanciate, i dati di input contengono istanze provenienti da tutte le zone di disponibilità del gruppo.
- Quando utilizzi una [mixed instances policy \(policy istanze miste\)](#), anche mantenere le capacità Spot e on demand bilanciate in base alle percentuali desiderate per ogni opzione di acquisto ha la precedenza. Innanzitutto identifichiamo quale dei due tipi (Spot o on demand) dovrebbe essere terminato. Inoltre, individua quali istanze (all'interno dell'opzione di acquisto identificata) terminare e in quali zone di disponibilità, in modo che risultino in un maggior bilanciamento delle zone di disponibilità.

Dati di risposta

I dati di input e i dati di risposta collaborano per restringere l'elenco delle istanze da terminare.

Con l'input fornito, la risposta dalla funzione Lambda dovrebbe essere simile al seguente esempio:

```
{
  "InstanceIDs": [
    "i-02e1c69383a3ed501",
    "i-036bc44b6092c01c7",
    ...
  ]
}
```

Gli InstanceIDs nella risposta rappresentano le istanze pronte per la terminazione.

In alternativa, è possibile restituire un set diverso di istanze pronte per la terminazione, che sovrascrive le istanze nei dati di input. Se quando viene richiamata la funzione Lambda nessuna istanza è pronta per la terminazione, è anche possibile scegliere di non restituire alcuna istanza.

Se nessuna istanza è pronta per la terminazione, la risposta dalla funzione Lambda dovrebbe essere simile alla seguente:

```
{
  "InstanceIDs": [ ]
}
```

Considerazioni

Quando usi una policy di terminazione personalizzata, tieni presente quanto segue:

- La restituzione di un'istanza per prima, nei dati di risposta, non ne garantisce la terminazione. Se quando viene richiamata la funzione Lambda viene restituita una quantità superiore al numero richiesto di istanze, Dimensionamento automatico Amazon EC2 valuta ogni istanza rispetto alle altre policy di terminazione specificate per il gruppo con scalabilità automatica. Quando sono presenti più policy di terminazione, tenta di applicare quella successiva nell'elenco e, se ci sono più istanze per la terminazione di quante ne siano necessarie, passa alla policy successiva, e così via. Se non vengono specificate altre policy di terminazione, per determinare quali istanze terminare, viene impiegata quella di default.
- Se non vengono restituite istanze o la funzione Lambda scade, Dimensionamento automatico Amazon EC2 attende per un breve periodo, prima di richiamare nuovamente la funzione. Per qualsiasi evento di riduzione orizzontale, continua a provare finché la capacità desiderata del gruppo sarà inferiore alla capacità corrente. Ad esempio, per terminazioni basate su

aggiornamento, continua a provare per un'ora. Successivamente, se continua a non riuscire a terminare alcuna istanza, l'operazione di aggiornamento non riesce. Con la durata massima dell'istanza, Dimensionamento automatico Amazon EC2 continua a tentare di terminare l'istanza identificata come eccedente la durata massima.

- Poiché la funzione viene riprovata ripetutamente, assicurati di testare e correggere eventuali errori permanenti nel codice, prima di utilizzare una funzione Lambda come policy di terminazione personalizzata.
- Se sostituisci i dati di input con il tuo elenco di istanze da terminare e la terminazione di queste istanze sbilancia le zone di disponibilità, Dimensionamento automatico Amazon EC2 ribilancia gradualmente la distribuzione della capacità tra le zone di disponibilità. Innanzitutto, richiama la funzione Lambda per vedere se ci siano istanze pronte per la terminazione, in modo da determinare se iniziare il ribilanciamento. Se ci sono istanze pronte per essere terminate, avvia prima quelle nuove. Al termine dell'avvio delle istanze, rileva che la capacità corrente del gruppo è superiore alla capacità desiderata e avvia un evento di riduzione orizzontale.
- Una policy di terminazione personalizzata non influisce sulla possibilità di utilizzare anche una protezione di dimensionamento per proteggere determinate istanze dall'interruzione. Per ulteriori informazioni, consulta [Utilizzare la protezione di riduzione orizzontale dell'istanza](#).

Creazione della funzione Lambda

Inizia creando la funzione Lambda, in modo da poter specificare il relativo Amazon Resource Name (ARN) nelle policy di terminazione per il tuo gruppo con scalabilità automatica.

Creazione di una funzione Lambda (console)

1. Aprire la [pagina Functions \(Funzioni\)](#) nella console Lambda.
2. Sulla barra di navigazione nella parte superiore della schermata, seleziona la stessa Regione utilizzata durante la creazione del gruppo con dimensionamento automatico.
3. Scegli Create function (Crea funzione) e Author from scratch (Crea da zero).
4. In Basic information (Informazioni di base), inserisci un nome per la funzione in Function name (Nome funzione).
5. Scegli Crea funzione. Verrai riportato alla configurazione e al codice della funzione.
6. Con la funzione aperta nella console, in Function code (Codice della funzione), incolla il codice nell'editor.
7. Seleziona Deploy (Implementa).

8. Facoltativamente, crea una versione pubblicata della funzione Lambda scegliendo la scheda Versions (Versioni) Scheda e quindi Publish new version (Pubblica nuova versione). Per ulteriori informazioni sul controllo delle versioni in Lambda, consulta [Versioni delle funzioni Lambda](#) nella Guida per gli sviluppatori di AWS Lambda .
9. Se hai scelto di pubblicare una versione e desideri associare un alias a questa versione della funzione Lambda, scegli la scheda Aliases (Alias). Per ulteriori informazioni sugli alias in Lambda, consulta [Alias delle funzioni Lambda](#) nella Guida per gli sviluppatori di AWS Lambda
10. Quindi, seleziona la scheda Configuration (Configurazione) e poi Permissions (Autorizzazioni).
11. Scorri fino a Resource-based policy (Policy basata su risorse), quindi, scegli Add permissions (Aggiungi autorizzazioni). Le policy basate su risorse vengono impiegate per concedere le autorizzazioni per richiamare la funzione al principale specificato nelle policy stesse. In questo caso, il principale sarà il [Dimensionamento automatico Amazon EC2 service-linked role \(Ruolo Dimensionamento automatico Amazon EC2 collegato ai servizi\)](#) associato al gruppo con scalabilità automatica.
12. Nella sezione Policy statement (Istruzione della policy), configura le autorizzazioni:
 - a. Scegli Account AWS.
 - b. In (Principal (Principale), inserisci l'ARN del ruolo collegato ai servizi chiamanti, ad esempio **arn:aws:iam::<aws-account-id>:role/aws-service-role/autoscaling.amazonaws.com/AWSServiceRoleForAutoScaling**.
 - c. Per Azione, scegli lambda: InvokeFunction
 - d. In Statement ID (ID istruzione), inserisci un ID di istruzione univoco, ad esempio **AllowInvokeByAutoScaling**.
 - e. Selezionare Salva.
13. Dopo aver seguito queste istruzioni, continua specificando l'ARN della tua funzione nelle policy di terminazione per il tuo gruppo con scalabilità automatica. Per ulteriori informazioni, consulta [Modificare la politica di terminazione per un gruppo Auto Scaling](#).

Note

Per esempi da utilizzare come riferimento per lo sviluppo della funzione Lambda, consulta il [GitHub repository](#) per Amazon EC2 Auto Scaling.

Limitazioni

- Nelle policy di terminazione per un gruppo con scalabilità automatica è possibile specificare una sola funzione Lambda. Se sono specificate più policy di terminazione, è necessario specificare prima la funzione Lambda.
- È possibile fare riferimento alla funzione Lambda utilizzando un ARN non qualificato (senza suffisso) o un ARN qualificato che abbia una versione o un alias come suffisso. Se utilizzi un ARN non qualificato (ad esempio, `function:my-function`), la policy basata su risorse deve essere creata sulla versione non pubblicata della funzione. Se utilizzi un ARN qualificato (ad esempio, `function:my-function:1` or `function:my-function:prod`), la policy basata su risorse deve essere creata su quella specifica versione pubblicata della funzione.
- Non è possibile utilizzare un ARN qualificato con il suffisso `$LATEST`. Se tenti di aggiungere una policy di terminazione personalizzata che si riferisce a un ARN qualificato con il suffisso `$LATEST`, ciò risulterà in un errore.
- Il numero di istanze fornite nei dati di input è limitato a 30.000 istanze. Se sono presenti più di 30.000 istanze che potrebbero essere terminate, i dati di input includono `"HasMoreInstances": true` per indicare che viene restituito il numero massimo di istanze.
- Il runtime massimo per la funzione Lambda è di due secondi (2000 millisecondi). La best practice è impostare il valore di timeout della funzione Lambda in base al runtime previsto. Le funzioni Lambda hanno un timeout di default di tre secondi, ma quest'ultimo può essere ridotto.
- Se il runtime supera il limite di 2 secondi, qualsiasi azione di scalabilità verrà sospesa fino a quando il runtime non scende al di sotto di questa soglia. Per le funzioni Lambda con tempi di esecuzione costantemente più lunghi, trova un modo per ridurre il tempo di esecuzione, ad esempio memorizzando nella cache i risultati in modo che possano essere recuperati durante le successive chiamate Lambda.

Utilizzare la protezione di riduzione orizzontale dell'istanza

La protezione scalabile in base alle istanze ti consente di controllare quali istanze Amazon EC2 Auto Scaling può terminare. Un caso d'uso comune di questa funzionalità è la scalabilità dei carichi di lavoro basati su container. Per ulteriori informazioni, consulta [Progetta le tue applicazioni su Dimensionamento automatico Amazon EC2 per gestire senza problemi la terminazione delle istanze](#).

Per impostazione predefinita, la protezione scale-in dell'istanza è disabilitata quando si crea un gruppo Auto Scaling. Ciò significa che Amazon EC2 Auto Scaling può terminare qualsiasi istanza del gruppo.

Puoi proteggere le istanze non appena vengono avviate abilitando l'impostazione di protezione per la riduzione delle istanze nel gruppo con dimensionamento automatico. La protezione per la riduzione dell'istanza inizia quando lo stato dell'istanza è `InService`. Quindi, per controllare quali istanze possono terminare, disabilita l'impostazione della protezione di dimensionamento sulle singole istanze all'interno del gruppo con dimensionamento automatico. In questo modo, puoi continuare a proteggere determinate istanze da terminazioni indesiderate.

Argomenti

- [Considerazioni](#)
- [Modificare la protezione scalabile per un gruppo Auto Scaling](#)
- [Modificare la protezione scalabile per un'istanza](#)

Considerazioni

Di seguito sono riportate le considerazioni relative all'utilizzo della protezione scalabile in base alle istanze:

- Se tutte le istanze in un gruppo con scalabilità automatica sono protette dalla riduzione orizzontale e si verifica un evento di riduzione orizzontale, la capacità desiderata diminuisce. Tuttavia, il gruppo con scalabilità automatica non può terminare il numero di istanze richiesto finché le loro impostazioni di protezione per la riduzione non vengono disabilitate. In AWS Management Console, la cronologia delle attività per il gruppo Auto Scaling include il seguente messaggio se tutte le istanze di un gruppo Auto Scaling sono protette dalla scalabilità in quando si verifica un evento di scalabilità in: `Could not scale to desired capacity because all remaining instances are protected from scale-in.`
- Se distacchi un'istanza protetta dalla riduzione, l'impostazione di protezione per la riduzione delle istanze viene persa. Quando si rialleggi l'istanza al gruppo, questa eredita da esso l'impostazione corrente di protezione per la riduzione delle istanze. Quando Dimensionamento automatico Amazon EC2 avvia una nuova istanza o sposta un'istanza da un warm pool nel gruppo con scalabilità automatica, questa eredita l'impostazione di protezione per la riduzione delle istanze del gruppo con scalabilità automatica.
- La protezione per la riduzione delle istanze non protegge le istanze Auto Scaling da quanto segue:
 - Sostituzione del controllo dell'integrità nel caso in cui l'istanza non superi i controlli dell'integrità. Per ulteriori informazioni, consulta [Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico](#).

- Interruzioni dell'istanza Spot. Un'istanza Spot viene terminata quando la capacità non è più disponibile o il prezzo spot supera il prezzo massimo.
- Una prenotazione Capacity Block termina. Amazon EC2 recupera le istanze Capacity Block anche se sono protette dallo scale-in.
- Terminazione manuale tramite il comando. `terminate-instance-in-auto-scaling-group` Per ulteriori informazioni, consulta [Terminare un'istanza nel gruppo con dimensionamento automatico \(AWS CLI\)](#).
- Terminazione manuale tramite la console Amazon EC2, i comandi CLI e le operazioni API. Per proteggere le istanze Auto Scaling dalla terminazione manuale, occorre abilitare la protezione dalla terminazione di Amazon EC2. (Ciò non impedisce ad Amazon EC2 Auto Scaling di terminare le istanze o la terminazione manuale tramite il comando.) `terminate-instance-in-auto-scaling-group` Per informazioni sull'attivazione della protezione dalle terminazioni di Amazon EC2 in un modello di avvio, consulta. [Crea un modello di avvio utilizzando le impostazioni avanzate](#)

Modificare la protezione scalabile per un gruppo Auto Scaling

È possibile attivare o disattivare l'impostazione di protezione di riduzione delle istanze per un gruppo con scalabilità automatica. Quando la abiliti, tutte le nuove istanze lanciate dal gruppo avranno la protezione scalabile in base alle istanze abilitata.

L'attivazione o la disabilitazione di questa impostazione per un gruppo Auto Scaling non influisce sulle istanze esistenti.

Console

Per abilitare la protezione scalabile in base a un nuovo gruppo Auto Scaling

Quando si crea il gruppo Auto Scaling, nella pagina Configura le dimensioni del gruppo e le politiche di ridimensionamento, in Protezione scale-in dell'istanza, selezionare la casella di controllo Abilita protezione scale-in dell'istanza.

Per abilitare o disabilitare la protezione scalabile per un gruppo esistente

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con scalabilità automatica.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Nella scheda Details (Dettagli) scegliere Advanced configurations (Configurazioni avanzate), Edit (Modifica).
4. Per la protezione scalabile in base alle istanze, seleziona o deseleziona la casella di controllo Abilita protezione scalabile in base alle esigenze per abilitare o disabilitare questa opzione.
5. Scegli Aggiorna.

AWS CLI

Per abilitare la protezione scalabile in base a un nuovo gruppo Auto Scaling

Utilizzate il [create-auto-scaling-group](#) comando seguente per abilitare la protezione scale-in dell'istanza.

```
aws autoscaling create-auto-scaling-group --auto-scaling-group-name my-asg --new-instances-protected-from-scale-in ...
```

Per abilitare la protezione scalabile per un gruppo esistente

Utilizzate il [update-auto-scaling-group](#) comando seguente per abilitare la protezione scale-in dell'istanza per il gruppo Auto Scaling specificato.

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg --new-instances-protected-from-scale-in
```

Per disabilitare la protezione scalabile in base a un gruppo esistente

Utilizza il comando seguente per disabilitare la protezione per la riduzione delle istanze per il gruppo specificato.

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg --no-new-instances-protected-from-scale-in
```

Modificare la protezione scalabile per un'istanza

Per impostazione predefinita, un'istanza riceve l'impostazione di protezione per la riduzione delle istanze dal suo gruppo con scalabilità automatica. Tuttavia, è possibile abilitare o disabilitare la protezione scalabile in base alla scalabilità delle istanze per le singole istanze dopo l'avvio.

Console

Per abilitare o disabilitare la protezione scalabile per un'istanza

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Nella scheda Instance management (Gestione istanze) in Instances (Istanze), selezionare un'istanza.
4. Per abilitare la protezione per la riduzione delle istanze, scegli Actions (Operazioni), Set scale-in protection (Imposta protezione per la riduzione). Quando richiesto, scegli Set scale-in protection (Imposta protezione per la riduzione).
5. Per disabilitare la protezione per la riduzione delle istanze, scegli Actions (Operazioni), Remove scale-in protection (Rimuovi protezione della riduzione). Quando richiesto, scegli Remove scale-in protection (Rimuovi protezione per la riduzione).

AWS CLI

Per abilitare la protezione scalabile per un'istanza

Utilizzate il [set-instance-protection](#) comando seguente per abilitare la protezione scalabile in base all'istanza specificata.

```
aws autoscaling set-instance-protection --instance-ids i-5f2e8a0d --auto-scaling-group-name my-asg --protected-from-scale-in
```

Per disabilitare la protezione scalabile in un'istanza

Per disabilitare la protezione per la riduzione delle istanze per l'istanza specificata, utilizza il comando seguente.

```
aws autoscaling set-instance-protection --instance-ids i-5f2e8a0d --auto-scaling-group-name my-asg --no-protected-from-scale-in
```


Note

Ricorda che la protezione scale-in delle istanze non garantisce che le istanze non vengano terminate in caso di errore umano, ad esempio se qualcuno interrompe manualmente un'istanza utilizzando la console Amazon EC2 o AWS CLI. Se desideri proteggere l'istanza da una terminazione accidentale, usa la protezione dalla terminazione di Amazon EC2. Tuttavia, anche se la protezione dalla terminazione e la protezione per la riduzione delle istanze sono abilitate, i dati salvati nell'archivio istanze possono andare persi qualora un controllo dell'integrità determini che un'istanza non è integra o il gruppo stesso venga eliminato accidentalmente. Come per qualsiasi ambiente, la best practice consiste nell'eseguire il backup dei dati frequentemente o in qualsiasi momento appropriato in base ai requisiti di business continuity.

Progetta le tue applicazioni su Dimensionamento automatico Amazon EC2 per gestire senza problemi la terminazione delle istanze

Questo argomento illustra i diversi approcci che puoi adottare se hai applicazioni in esecuzione su istanze che normalmente non dovrebbero terminare inaspettatamente quando Dimensionamento automatico Amazon EC2 risponde a un evento di dimensionamento.

Ad esempio, supponiamo di avere una coda Amazon SQS che raccoglie i messaggi in arrivo per lavori di lunga durata. Quando arriva un nuovo messaggio, un'istanza del gruppo con dimensionamento automatico recupera il messaggio e inizia a elaborarlo. L'elaborazione di ogni messaggio richiede 3 ore. All'aumentare del numero di messaggi, nuove istanze vengono aggiunte automaticamente al gruppo con dimensionamento automatico. Man mano che il numero di messaggi diminuisce, le istanze esistenti vengono automaticamente terminate. In questo caso, il Dimensionamento automatico Amazon EC2 deve decidere quale istanza terminare. Per impostazione predefinita, è possibile che Dimensionamento automatico Amazon EC2 interrompa un'istanza dopo 2,9 ore di elaborazione di un processo di 3 ore, anziché un'istanza attualmente inattiva. Per evitare problemi di terminazioni impreviste quando si utilizza Dimensionamento automatico Amazon EC2, è necessario progettare l'applicazione in modo da rispondere a questo scenario.

È possibile utilizzare le seguenti funzionalità per impedire al gruppo con dimensionamento automatico di terminare le istanze che non sono ancora pronte per la chiusura o di terminare le istanze troppo rapidamente perché possano completare i lavori assegnati. Tutte e tre queste funzionalità possono essere utilizzate in combinazione o separatamente.

Indice

- [Protezione del dimensionamento verticale dell'istanza](#)
- [Policy di terminazione personalizzata](#)
- [Hook del ciclo di vita di terminazione](#)

Important

Quando progetti le tue applicazioni su Dimensionamento automatico Amazon EC2 per gestire senza problemi la terminazione delle istanze, tieni presente i seguenti punti.

- Se un'istanza non è integra, Dimensionamento automatico Amazon EC2 la sostituirà indipendentemente dalla funzionalità utilizzata (a meno che non sospenda il processo `ReplaceUnhealthy`). Puoi utilizzare un hook del ciclo di vita per consentire all'applicazione di chiudersi correttamente o copiare tutti i dati che devi ripristinare prima che l'istanza venga terminata.
- Non è garantito che un hook del ciclo di vita di terminazione venga eseguito o terminato prima che un'istanza venga terminata. Se qualcosa non funziona, il Dimensionamento automatico Amazon EC2 termina comunque l'istanza.

Protezione del dimensionamento verticale dell'istanza

Puoi utilizzare la protezione di dimensionamento dell'istanza in molte situazioni in cui la chiusura delle istanze è un'azione critica che dovrebbe essere negata per impostazione predefinita e consentita esplicitamente solo per istanze specifiche. Ad esempio, quando esegui carichi di lavoro containerizzati, è comune voler proteggere tutte le istanze e rimuovere la protezione solo per le istanze senza attività correnti o pianificate. Servizi come Amazon ECS hanno sviluppato integrazioni con protezione di dimensionamento delle istanze nei loro prodotti.

È possibile abilitare la protezione di dimensionamento al gruppo con dimensionamento automatico per applicare la protezione scalabile alle istanze al momento della creazione e abilitarla per le istanze esistenti. Quando un'istanza non ha più attività da svolgere, può disattivare la protezione. L'istanza può continuare a cercare nuovi lavori e riattivare la protezione quando vengono assegnati nuovi lavori.

Le applicazioni possono impostare la protezione da un piano di controllo centralizzato che controlla se un'istanza è terminabile o meno, oppure dalle istanze stesse. Tuttavia, un'ampia flotta potrebbe

incorrere in problemi di limitazione se un gran numero di istanze cambia continuamente la protezione di dimensionamento.

Per ulteriori informazioni, consulta [Utilizzare la protezione di riduzione orizzontale dell'istanza.](#)

Policy di terminazione personalizzata

Come per la protezione del dimensionamento delle istanze, una policy di terminazione personalizzata consente di impedire l'interruzione di istanze specifiche da parte del gruppo con dimensionamento automatico.

Come impostazione predefinita, il gruppo con dimensionamento automatico utilizza una policy di terminazione predefinita per determinare quali istanze terminare per prime. Se desideri un maggiore controllo su quali istanze terminano per prime, puoi implementare una policy di terminazione personalizzata utilizzando una funzione Lambda. Dimensionamento automatico Amazon EC2 chiama la funzione ogni volta che deve decidere quale istanza terminare. Terminerà solo un'istanza restituita dalla funzione. Se la funzione presenta errori, scade o produce un elenco vuoto, Dimensionamento automatico Amazon EC2 non termina le istanze.

Una policy di terminazione personalizzata è utile se è noto quando un'istanza è sufficientemente ridondante o sottoutilizzata da poter essere terminata. A tale scopo, è necessario implementare l'applicazione con un piano di controllo che monitori il carico di lavoro in tutto il gruppo. In questo modo, se un'istanza sta ancora elaborando processi, la funzione Lambda sa di non includerla.

Per ulteriori informazioni, consulta [Creare una policy di terminazione personalizzata con Lambda.](#)

Hook del ciclo di vita di terminazione

Un hook del ciclo di vita della terminazione prolunga la durata di un'istanza già selezionata per la terminazione. Offre più tempo per completare tutti i messaggi o le richieste attualmente assegnati all'istanza o per salvare lo stato di avanzamento e trasferire il lavoro su un'altra istanza.

Per molti carichi di lavoro, un hook del ciclo di vita può essere sufficiente per chiudere correttamente un'applicazione su un'istanza selezionata per la chiusura. Si tratta di un approccio ottimale e non può essere utilizzato per impedire l'interruzione in caso di errore.

Per utilizzare un hook del ciclo di vita, è necessario sapere quando un'istanza è selezionata per essere terminata. Puoi saperlo in due modi:

Opzione	Descrizione	Ideale per	Collegamento alla documentazione
All'interno dell'istanza	Il servizio di metadati di istanza (IMDS) è un endpoint sicuro che consente di verificare lo stato di un'istanza direttamente dall'istanza. Se i metadati vengono restituiti con <code>Terminated</code> , viene pianificata la terminazione dell'istanza.	Applicazioni in cui è necessario eseguire un'azione sull'istanza prima che l'istanza venga terminata.	Recupero dello stato del ciclo di vita di destinazione
All'esterno dell'istanza	Quando un'istanza viene terminata, viene generata una notifica di evento. Puoi creare regole utilizzando Amazon EventBridge, Amazon SQS o Amazon SNS per acquisire questi eventi e richiamare una risposta, ad esempio con una funzione Lambda.	Applicazioni che devono agire al di fuori dell'istanza.	Creazione di una destinazione di notifica

Per utilizzare un hook del ciclo di vita, devi anche sapere quando un'istanza è pronta per essere terminata. Amazon EC2 Auto Scaling non dirà ad Amazon EC2 di terminare l'istanza finché non riceve [CompleteLifecycleAction](#) una chiamata o non scade il timeout, a seconda dell'evento che si verifica per primo.

Per impostazione predefinita, un'istanza può continuare a funzionare per un'ora (heartbeat timeout) causa di un hook del ciclo di vita della terminazione. È possibile configurare il timeout predefinito se un'ora non è sufficiente per completare l'azione del ciclo di vita. Quando un'azione del ciclo di vita è effettivamente in corso, puoi prolungare il timeout con chiamate API. [RecordLifecycleActionHeartbeat](#)

Per ulteriori informazioni, consulta [Hook del ciclo di vita di Amazon EC2 Auto Scaling](#).

Sospendi e riprendi i processi di Amazon EC2 Auto Scaling

Questo argomento descrive come sospendere e quindi riprendere uno o più processi per il gruppo Auto Scaling per disabilitare temporaneamente determinate operazioni.

La sospensione dei processi può essere utile quando è necessario indagare o risolvere un problema senza interferenze dovute al ridimensionamento delle politiche o alle azioni pianificate. Inoltre, impedisce ad Amazon EC2 Auto Scaling di contrassegnare le istanze come non integre e di sostituirle mentre apporti modifiche al gruppo Auto Scaling.

Argomenti

- [Tipi di processi](#)
- [Considerazioni](#)
- [Sospendere i processi](#)
- [Riprendere i processi](#)
- [In che modo i processi sospesi influiscono sugli altri processi](#)

Note

Oltre alle sospensioni avviate, Dimensionamento automatico Amazon EC2 può anche sospendere i processi per i gruppi Auto Scaling che falliscono ripetutamente l'avvio delle istanze. Ciò è noto anche come sospensione amministrativa. La sospensione amministrativa si applica generalmente ai gruppi Auto Scaling che hanno tentato di avviare istanze per più di 24 ore ma non sono mai riusciti a farlo. È possibile ripristinare i processi sospesi da Dimensionamento automatico Amazon EC2 per ragioni amministrative.

Tipi di processi

La funzione suspend-resume (sospensione e ripresa) supporta i seguenti processi:

- **Launch**— Aggiunge istanze al gruppo Auto Scaling quando il gruppo si ridimensiona orizzontalmente o quando Amazon EC2 Auto Scaling sceglie di avviare le istanze per altri motivi, ad esempio quando aggiunge istanze a un pool caldo.
- **Terminate**— Rimuove le istanze dal gruppo Auto Scaling quando il gruppo aumenta o quando Amazon EC2 Auto Scaling sceglie di terminare le istanze per altri motivi, ad esempio quando

un'istanza viene terminata per aver superato la sua durata di vita massima o non aver superato un controllo di integrità.

- `AddToLoadBalancer`— Aggiunge istanze al gruppo target del sistema di bilanciamento del carico collegato o al Classic Load Balancer quando vengono avviate. Per ulteriori informazioni, consulta [Utilizzare sistema di bilanciamento del carico elastico per distribuire il traffico tra le istanze nel gruppo con dimensionamento automatico.](#)
- `AlarmNotification`— Accetta notifiche dagli CloudWatch allarmi associati a politiche di scalabilità dinamica. Per ulteriori informazioni, consulta [Dimensionamento dinamico per Dimensionamento automatico Amazon EC2.](#)
- `AZRebalance`— Bilancia il numero di istanze EC2 nel gruppo in modo uniforme in tutte le zone di disponibilità specificate quando il gruppo diventa sbilanciato, ad esempio quando una zona di disponibilità precedentemente non disponibile torna a uno stato integro. Per ulteriori informazioni, consulta [Attività di ribilanciamento.](#)
- `HealthCheck`— Verifica lo stato delle istanze e contrassegna un'istanza come non integra se Amazon EC2 o Elastic Load Balancing comunica ad Amazon EC2 Auto Scaling che l'istanza non è integra. Questo processo può sostituire lo stato di integrità di un'istanza che si imposta manualmente. Per ulteriori informazioni, consulta [Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico.](#)
- `InstanceRefresh`— Termina e sostituisce le istanze utilizzando la funzionalità di aggiornamento dell'istanza. Per ulteriori informazioni, consulta [Usa un aggiornamento dell'istanza per aggiornare le istanze in un gruppo di Auto Scaling.](#)
- `ReplaceUnhealthy`— Termina le istanze contrassegnate come non integre e quindi crea nuove istanze per sostituirle. Per ulteriori informazioni, consulta [Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico.](#)
- `ScheduledActions`— Esegue le azioni di ridimensionamento pianificate create dall'utente o create automaticamente quando si crea un piano di scalabilità e si attiva la AWS Auto Scaling scalabilità predittiva. Per ulteriori informazioni, consulta [Dimensionamento programmato per Dimensionamento automatico Amazon EC2.](#)

Considerazioni

Considera quanto segue, prima di sospendere i processi:

- La sospensione `AlarmNotification` consente di interrompere temporaneamente le politiche di tracciamento degli obiettivi, dei passaggi e delle semplici politiche di scalabilità del gruppo

senza eliminare le politiche di ridimensionamento o gli allarmi associati. CloudWatch Per arrestare temporaneamente le policy di dimensionamento individuali, consulta [Disabilitare una policy di dimensionamento per un gruppo con scalabilità automatica](#).

- Puoi scegliere di sospendere `ReplaceUnhealthy` i processi `HealthCheck` e di riavvio delle istanze senza che Amazon EC2 Auto Scaling chiuda le istanze sulla base dei relativi controlli di integrità. Tuttavia, se hai bisogno di Amazon EC2 Auto Scaling per continuare a eseguire controlli di integrità sulle istanze rimanenti, utilizza invece la funzione di standby. Per ulteriori informazioni, consulta [Rimuovere temporaneamente le istanze dal gruppo con scalabilità automatica](#).
- Se sospendi i processi `Launch` e `Terminate` o `AZRebalance` e quindi apporti modifiche al gruppo con scalabilità automatica, ad esempio staccando istanze o modificando le zone di disponibilità specificate, il gruppo può diventare sbilanciato tra le zone di disponibilità. Se si verifica ciò, dopo aver ripreso i processi sospesi, il Dimensionamento automatico Amazon EC2 ridistribuisce gradualmente le istanze in modo uniforme tra le zone di disponibilità.
- Se sospendi il `Terminate` processo, puoi comunque forzare la chiusura delle istanze utilizzando il comando con l'[delete-auto-scaling-group](#) opzione `force delete`.
- La sospensione del `Terminate` processo si applica solo alle istanze che si trovano attualmente nello stato `InService`. Non impedisce la chiusura di istanze in altri stati, ad esempio `Pending`, o istanze che non riescono a riavviarsi correttamente dallo standby.
- Il `RemoveFromLoadBalancerLowPriority` processo può essere ignorato quando è presente nelle chiamate per descrivere i gruppi di Auto Scaling utilizzando AWS CLI o SDK. Questo processo è obsoleto e viene mantenuto solo per motivi di compatibilità con le versioni precedenti.

Sospendere i processi

Per sospendere un processo per un gruppo Auto Scaling, utilizzate uno dei seguenti metodi:

Console

Come sospendere un processo

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Nella scheda Details (Dettagli) scegliere Advanced configurations (Configurazioni avanzate), Edit (Modifica).
4. In Suspended processes (Processi sospesi) seleziona il processo da sospendere.
5. Scegli Aggiorna.

AWS CLI

Utilizza il comando [suspend-processes](#) per sospendere i singoli processi.

```
aws autoscaling suspend-processes --auto-scaling-group-name my-asg --scaling-processes HealthCheck ReplaceUnhealthy
```

Per sospendere tutti i processi, omettere l'opzione `--scaling-processes`, come segue.

```
aws autoscaling suspend-processes --auto-scaling-group-name my-asg
```

Riprendere i processi

Per riprendere un processo sospeso per un gruppo Auto Scaling, utilizzate uno dei seguenti metodi:

Console

Come riprendere un processo sospeso

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Nella scheda Details (Dettagli) scegliere Advanced configurations (Configurazioni avanzate), Edit (Modifica).
4. In Suspended processes (Processi sospesi) elimina il processo sospeso.
5. Scegli Aggiorna.

AWS CLI

[Per riprendere un processo sospeso, utilizzate il seguente comando resume-processes.](#)


```
aws autoscaling resume-processes --auto-scaling-group-name my-asg --scaling-  
processes HealthCheck
```

Per riprendere tutti i processi sospesi, omettere l'opzione `--scaling-processes`, come segue.

```
aws autoscaling resume-processes --auto-scaling-group-name my-asg
```

In che modo i processi sospesi influiscono sugli altri processi

Le sezioni seguenti descrivono cosa succede quando processi diversi vengono sospesi singolarmente.

Argomenti

- [Launchè sospeso](#)
- [Terminateè sospeso](#)
- [AddToLoadBalancerè sospeso](#)
- [AlarmNotificationè sospeso](#)
- [AZRebalanceè sospeso](#)
- [HealthCheckè sospeso](#)
- [InstanceRefreshè sospeso](#)
- [ReplaceUnhealthyè sospeso](#)
- [ScheduledActionsè sospeso](#)
- [Ulteriori considerazioni](#)

Launchè sospeso

- AlarmNotification è ancora attivo, ma il gruppo con scalabilità automatica non è in grado di avviare attività di aumento orizzontale per gli allarmi in violazione.
- ScheduledActions è attivo, ma il gruppo con scalabilità automatica non è in grado di avviare attività di aumento orizzontale per le azioni pianificate che si verificano.
- AZRebalance smette di ribilanciare il gruppo.

- `ReplaceUnhealthy` continua a terminare istanze non integre, ma non avvia sostituzioni. Quando riprendi il processo `Launch`, il Dimensionamento automatico Amazon EC2 sostituisce immediatamente tutte le istanze che ha terminato durante il periodo in cui `Launch` è stato sospeso.
- `InstanceRefresh` non sostituisce le istanze.

Terminate è sospeso

- `AlarmNotification` è ancora attivo, ma il gruppo con scalabilità automatica non è in grado di avviare attività di riduzione orizzontale in caso di allarmi che sono in violazione.
- `ScheduledActions` è attivo, ma il gruppo con scalabilità automatica non è in grado di avviare attività di riduzione orizzontale per le azioni pianificate che si verificano.
- `AZRebalance` è ancora attivo, ma non funziona correttamente. È possibile avviare nuove istanze senza terminare quelle precedenti. Questo potrebbe causare la crescita del gruppo con scalabilità automatica fino al 10% in più rispetto alle dimensioni massime, poiché durante le attività di ribilanciamento ciò è temporaneamente consentito. Il gruppo con scalabilità automatica potrebbe restare al di sopra della sua dimensione massima fino a riprendere il processo `Terminate`.
- `ReplaceUnhealthy` non è attivo ma `HealthCheck` lo è. Quando `Terminate` riprende, il processo `ReplaceUnhealthy` viene messo immediatamente in esecuzione. Se le istanze sono state contrassegnate come non integre durante la sospensione di `Terminate`, verranno immediatamente sostituite.
- `InstanceRefresh` non sostituisce le istanze.

AddToLoadBalancer è sospeso

- Dimensionamento automatico Amazon EC2 avvia le istanze, ma non le aggiunge al gruppo di destinazione del load balancer o al Classic Load Balancer. Se riprendi il processo `AddToLoadBalancer`, riprende l'aggiunta di istanze al load balancer al momento dell'avvio. Tuttavia, non aggiunge le istanze che sono state avviate mentre questo processo era sospeso. Dovrai registrare tali istanze manualmente.

AlarmNotification è sospeso

- Amazon EC2 Auto Scaling non richiama le politiche di scalabilità in caso di violazione di CloudWatch una soglia di allarme. Quando riprendi `AlarmNotification`, Dimensionamento

automatico Amazon EC2 valuta le policy per cui si è verificato un utilizzo fuori limite delle soglie di allarme.

AZRebalance è sospeso

- Il Dimensionamento automatico Amazon EC2 non tenta di ridistribuire le istanze dopo determinati eventi. Tuttavia, se si verifica un evento di aumento o riduzione orizzontale, il processo di dimensionamento cerca ancora di bilanciare le zone di disponibilità. Ad esempio, durante l'aumento orizzontale, viene avviata l'istanza nella zona di disponibilità con il minor numero di istanze. Se il gruppo diventa sbilanciato mentre **AZRebalance** è sospeso e viene ripreso, Dimensionamento automatico Amazon EC2 tenta di ribilanciarlo. Prima chiama `Launch` e poi `Terminate`.

HealthCheck è sospeso

- Dimensionamento automatico Amazon EC2 smette di contrassegnare le istanze come non integre in seguito ai controlli di integrità di EC2 e di Elastic Load Balancing. I controlli dell'integrità personalizzati continuano a funzionare correttamente. Dopo aver sospeso **HealthCheck**, se necessario, è possibile impostare manualmente lo stato di integrità delle istanze nel tuo gruppo e fare in modo che `ReplaceUnhealthy` le sostituisca.

InstanceRefresh è sospeso

- Il Dimensionamento automatico Amazon EC2 smette di sostituire le istanze in seguito all'aggiornamento di una di esse. Se è in corso un aggiornamento dell'istanza, questo interrompe l'operazione senza annullarla.

ReplaceUnhealthy è sospeso

- Dimensionamento automatico Amazon EC2 smette di sostituire le istanze contrassegnate come non integre. Le istanze che non superano i controlli dell'integrità di EC2 o di Elastic Load Balancing saranno comunque contrassegnate come non integre. Dopo aver ripreso il processo `ReplaceUnhealthy`, Dimensionamento automatico Amazon EC2 sostituisce le istanze che sono state contrassegnate come non integre durante la sospensione del processo. Il processo `ReplaceUnhealthy` chiama prima `Terminate` e poi `Launch`.

ScheduledActions è sospeso

- Durante il periodo di sospensione, il Dimensionamento automatico Amazon EC2 non esegue operazioni pianificate per l'esecuzione. Quando riprendi ScheduledActions, il Dimensionamento automatico Amazon EC2 considera solo le azioni pianificate i cui tempi pianificati non sono ancora scaduti.

Ulteriori considerazioni

Inoltre, quando Launch o Terminate sono sospesi, le seguenti caratteristiche potrebbero non funzionare correttamente:

- Durata massima dell'istanza: quando Launch o Terminate è sospesa, la funzionalità di durata massima dell'istanza non può sostituire alcuna istanza.
- Interruzioni delle istanze Spot: se Terminate è sospeso e il gruppo Auto Scaling dispone di istanze Spot, queste possono comunque terminare nel caso in cui la capacità Spot non sia più disponibile. Mentre Launch è sospeso, il Dimensionamento automatico Amazon EC2 non è in grado di avviare le istanze di sostituzione da un altro pool di istanze Spot o dallo stesso pool di quando è di nuovo disponibile.
- Ribilanciamento della capacità: se Terminate è sospeso e utilizzi il ribilanciamento della capacità per gestire le interruzioni delle istanze Spot, il servizio Spot Amazon EC2 può comunque terminare le istanze nel caso in cui la capacità Spot non sia più disponibile. Se Launch è sospeso, il Dimensionamento automatico Amazon EC2 non è in grado di avviare le istanze di sostituzione da un altro pool di istanze Spot o dallo stesso pool di quando è di nuovo disponibile.
- Collegamento e scollegamento di istanze: quando Launch le istanze Terminate sono sospese, puoi scollegare le istanze collegate al tuo gruppo Auto Scaling, ma mentre Launch sono sospese non puoi collegare nuove istanze al gruppo.
- Istanze in standby: quando Launch Terminate sono sospese, puoi mettere un'istanza Standby nello stato, ma mentre Launch è sospesa non puoi riportare in servizio un'istanza nello stato in cui si trova. Standby

Monitora i tuoi gruppi con dimensionamento automatico

Il monitoraggio è importante per mantenere l'affidabilità, la disponibilità e le prestazioni di Dimensionamento automatico Amazon EC2 e delle tue soluzioni Cloud AWS. AWS fornisce i seguenti strumenti di monitoraggio per controllare Dimensionamento automatico Amazon EC2, segnalare eventuali problemi ed eseguire operazioni automatiche quando appropriato:

Controlli dell'integrità

Amazon EC2 Auto Scaling esegue periodicamente controlli dell'integrità sulle istanze del gruppo con dimensionamento automatico. Se un'istanza non supera il controllo dell'integrità, viene contrassegnata come non integra e terminata, mentre Dimensionamento automatico Amazon EC2 avvia una nuova istanza in sostituzione. Per ulteriori informazioni, consulta [Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico](#).

AWS Health Dashboard

Il AWS Health Dashboard visualizza le informazioni e fornisce anche notifiche che vengono richiamate da modifiche nell'integrità delle risorse AWS. Le informazioni vengono presentate in due modi: su un pannello di controllo che mostra eventi recenti e prossimi organizzati per categoria e in un log completo che mostra tutti gli eventi degli ultimi 90 giorni. Per ulteriori informazioni, consulta [AWS Health Dashboard notifiche per Amazon EC2 Auto Scaling](#).

CloudTrail

Con AWS CloudTrail, puoi tenere traccia delle chiamate effettuate all'API Dimensionamento automatico Amazon EC2 da o per conto del tuo Account AWS. CloudTrail archivia le informazioni in file di log nel bucket Amazon S3 specificato. È possibile utilizzare questi file di log per monitorare l'attività dei gruppi Auto Scaling. I registri includono le richieste effettuate, gli indirizzi IP di origine da cui provengono le richieste, chi ha effettuato la richiesta, quando è stata effettuata la richiesta e così via. Per ulteriori informazioni, consulta [Registra le chiamate API di Amazon EC2 Auto Scaling con AWS CloudTrail](#).

Raccolta dei log sulle istanze Amazon EC2.

È possibile utilizzare CloudWatch per raccogliere log dai sistemi operativi per le istanze EC2. Per ulteriori informazioni, consulta [Raccolta di parametri e log da istanze Amazon EC2 e da server on-premise con l'agente di CloudWatch](#) e [Visualizzazione dati di log inviati a CloudWatch Logs](#) nella Guida per l'utente di Amazon CloudWatch.

Per informazioni su altri servizi AWS che possono aiutarti a registrare e raccogliere dati sui tuoi carichi di lavoro, consulta la [Guida alla registrazione e al monitoraggio per i proprietari delle applicazioni](#) nella Guida prescrittiva AWS.

Amazon CloudWatch

Amazon CloudWatch ti aiuta ad analizzare i log e, in tempo reale, a monitorare le metriche delle tue risorse AWS e applicazioni ospitate. Puoi raccogliere i parametri e tenerne traccia, creare pannelli di controllo personalizzati e impostare allarmi per inviare una notifica o intraprendere azioni quando un parametro specificato raggiunge una determinata soglia. Ad esempio, è possibile ricevere una notifica quando l'attività di rete è improvvisamente superiore o inferiore al valore previsto di un parametro. Per ulteriori informazioni su come utilizzare questo servizio per monitorare i parametri dei tuoi gruppi con dimensionamento automatico e istanze, consulta [Monitorare i parametri di CloudWatch per gruppi e istanze Auto Scaling](#).

CloudWatch monitora anche le metriche di utilizzo delle API AWS per Dimensionamento automatico Amazon EC2. Puoi utilizzare queste metriche per configurare allarmi che ti avvisano quando il volume delle chiamate API viola una soglia da te definita. Per ulteriori informazioni, consulta i [Parametri di utilizzo AWS](#) nella Guida per l'utente di Amazon CloudWatch.

AWS Compute Optimizer

L'Ottimizzatore di calcolo fornisce suggerimenti sulle istanze Amazon EC2 che possono aiutarti a decidere se passare a una nuova tipologia di istanza. Analizza se il tipo di istanza del gruppo con dimensionamento automatico è ottimale e genera suggerimenti per ridurre i costi e migliorare le prestazioni dei carichi di lavoro. Per ulteriori informazioni, consulta [AWS Compute Optimizer Utilizzatelo per ottenere consigli sul tipo di istanza per un gruppo di Auto Scaling](#).

Amazon EventBridge

Amazon EventBridge è un servizio bus di eventi serverless che semplifica la connessione delle applicazioni ai dati provenienti da un'ampia gamma di origini. EventBridge fornisce un flusso di dati in tempo reale dalle tue applicazioni, dalle applicazioni Software-as-a-Service (SaaS) e dai servizi AWS, quindi instrada tali dati a destinazioni come Lambda. Questo ti consente di monitorare gli eventi che si verificano nei servizi e creare architetture basate su eventi. Per ulteriori informazioni, consulta [Utilizzato EventBridge per gestire gli eventi di Auto Scaling](#).

AWS Security Hub

Utilizza [AWS Security Hub](#) per monitorare l'utilizzo di Dimensionamento automatico Amazon EC2, in quanto è correlato alle best practice in materia di sicurezza. Centrale di sicurezza utilizza controlli di sicurezza di investigazione per valutare le configurazioni delle risorse e gli standard di sicurezza per aiutarti a rispettare vari framework di conformità. Per ulteriori informazioni sull'utilizzo di Security Hub volto a valutare le risorse Dimensionamento automatico Amazon EC2, consulta [Controlli di Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di AWS Security Hub.

Amazon Simple Notification Service

È possibile configurare i gruppi Auto Scaling per l'invio di notifiche Amazon SNS quando Amazon EC2 Auto Scaling avvia o termina le istanze. Per ulteriori informazioni, consulta [Opzioni di notifica Amazon SNS per Amazon EC2 Auto Scaling](#).

Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico

Amazon EC2 Auto Scaling monitora continuamente lo stato di integrità delle istanze in un gruppo Auto Scaling per mantenere la capacità desiderata.

Tutte le istanze in un gruppo Auto Scaling iniziano con Healthy uno stato. Si presuppone che le istanze siano integre, a meno che Amazon EC2 Auto Scaling non riceva notifica della loro non integrità. Può ricevere notifiche da varie fonti quando un'istanza non è integra e deve essere sostituita. Queste fonti includono:

- Amazon EC2
- Sistema di bilanciamento del carico elastico
- VPC Lattice
- Controlli sanitari personalizzati definiti dall'utente

Quando Amazon EC2 Auto Scaling determina che InService un'istanza non è integra, la sostituisce con una nuova istanza per mantenere la capacità desiderata del gruppo. La nuova istanza viene avviata utilizzando le impostazioni correnti del gruppo con scalabilità automatica e il relativo modello di avvio o la configurazione di avvio associati.

Le istanze non integre possono verificarsi anche quando un'istanza si interrompe in modo imprevisto, ad esempio a causa dell'interruzione di un'istanza Spot o della chiusura manuale da parte di un utente. Anche in questo caso, Amazon EC2 Auto Scaling avvierà automaticamente un'istanza sostitutiva in questi casi per mantenere la capacità desiderata.

Indice

- [Informazioni sui controlli dell'integrità del gruppo con dimensionamento automatico](#)
- [Come visualizzare il motivo dei fallimenti dei controlli dell'integrità](#)
- [Impostazione del periodo di tolleranza del controllo dell'integrità per un gruppo con scalabilità automatica](#)

Informazioni sui controlli dell'integrità del gruppo con dimensionamento automatico

In questo argomento viene fornita una panoramica dei tipi di controlli dell'integrità predefiniti e disponibili e viene illustrato il loro funzionamento.

Indice

- [Tipi di controllo dell'integrità](#)
- [Controlli dell'integrità Amazon EC2](#)
- [Controlli dello stato di Elastic Load Balancing](#)
- [Controlli dell'integrità VPC Lattice](#)
- [Come Amazon EC2 Auto Scaling riduce al minimo i tempi di inattività](#)
- [Considerazioni sul controllo dell'integrità](#)
- [Controlli dell'integrità personalizzati](#)
- [Risorse correlate](#)

Tipi di controllo dell'integrità

Il Dimensionamento automatico Amazon EC2 può determinare lo stato di integrità di un'istanza tramite uno o più dei seguenti controlli dell'integrità:

Tipo di controllo dello stato	Che cosa controlla
Controlli dello stato Amazon EC2 ed eventi pianificati	<ul style="list-style-type: none"> • Verifica che l'istanza sia in esecuzione • Verifica la presenza di problemi hardware o software sottostanti che possano compromettere l'istanza <p>Questo è il tipo di controllo dell'integrità predefinito per un gruppo con scalabilità automatica.</p>
Controlli dello stato di Elastic Load Balancing	<ul style="list-style-type: none"> • Verifica se l'istanza viene segnalata come integra dal load balancer per stabilire se l'istanza sia disponibile per gestire le richieste o meno <p>Per eseguire questo tipo di controllo dell'integrità, è necessario abilitarlo per il gruppo con scalabilità automatica.</p>
Controlli dell'integrità VPC Lattice	<ul style="list-style-type: none"> • Verifica se l'istanza viene segnalata come integra dal VPC Lattice per stabilire se l'istanza sia disponibile per gestire le richieste o meno <p>Per eseguire questo tipo di controllo dell'integrità, è necessario abilitarlo per il gruppo con scalabilità automatica.</p>
Controlli dell'integrità personalizzati	<ul style="list-style-type: none"> • Controlla eventuali altri problemi che potrebbero indicare problemi di integrità dell'istanza, in base ai controlli dell'integrità personalizzati

Controlli dell'integrità Amazon EC2

Dopo l'avvio di un'istanza, viene allegato al gruppo con dimensionamento automatico ed entra nello stato InService. Per ulteriori informazioni sul ciclo di vita completo delle istanze in un gruppo con scalabilità automatica, consulta [Ciclo di vita delle istanze di Amazon EC2 Auto Scaling](#).

Il Dimensionamento automatico Amazon EC2 verifica che tutte le istanze all'interno del gruppo con scalabilità automatica siano in esecuzione e in buone condizioni controllando periodicamente lo stato di integrità delle istanze.

Verifiche di stato

I controlli dell'integrità di Amazon EC2 Auto Scaling utilizzano i risultati delle verifiche dello stato di Amazon EC2 per determinare l'integrità di un'istanza. Se lo stato dell'istanza Amazon EC2 è diverso da `running` oppure se lo stato del sistema è `impaired`, il Dimensionamento automatico Amazon EC2 considera l'istanza non integra e la sostituisce. Ciò si applica quando l'istanza è in uno dei seguenti stati:

- `stopping`
- `stopped`
- `shutting-down`
- `terminated`

I controlli dello stato Amazon EC2 non richiedono alcuna configurazione speciale e sono sempre abilitati. Per ulteriori informazioni, consulta [Tipi di controllo dell'integrità](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Important

Il Dimensionamento automatico Amazon EC2 consente a questi controlli dell'integrità di fallire occasionalmente senza intraprendere alcuna azione. Quando un controllo dello stato non riesce, Amazon EC2 Auto Scaling attende alcuni minuti AWS per risolvere il problema. Non contrassegna immediatamente un'istanza come `Unhealthy` quando lo stato dei controlli dello stato diventa `impaired`.

Se il Dimensionamento automatico Amazon EC2 rileva che lo stato di un'istanza non è più `running`, tuttavia, questa situazione è trattata come un fallimento immediato. In questo caso, contrassegna immediatamente l'istanza come `Unhealthy` e la sostituisce.

Eventi pianificati

Amazon EC2 può occasionalmente pianificare eventi sulle istanze da eseguire dopo un determinato timestamp. Per ulteriori informazioni, consulta [Eventi pianificati per le istanze](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Se una delle istanze è interessata da un evento pianificato, Amazon EC2 Auto Scaling la considera non integra e la sostituisce. L'istanza non inizia a terminare fino al raggiungimento della data e dell'ora specificate nel timestamp.

Controlli dello stato di Elastic Load Balancing

Quando si abilitano i controlli dell'integrità Elastic Load Balancing per il gruppo con scalabilità automatica, Dimensionamento automatico Amazon EC2 può utilizzare questi risultati dei controlli dell'integrità per determinare l'integrità di un'istanza.

Prima di poter abilitare i controlli dell'integrità Elastic Load Balancing per il gruppo con scalabilità automatica, è necessario effettuare le seguenti operazioni:

- Imposta un load balancer Elastic Load Balancer e configurare un controllo dell'integrità da utilizzare per determinare se le istanze siano integre.
- Collega un load balancer al gruppo con scalabilità automatica.

Dopo aver completato le azioni precedenti, si verifica quanto riportato di seguito:

- Il Dimensionamento automatico Amazon EC2 registra le istanze nel gruppo con scalabilità automatica con il load balancer.
- Dopo che un'istanza ha terminato la registrazione, passa allo stato `InService` e diventa disponibile per l'uso con il load balancer.

Per impostazione predefinita, Amazon EC2 Auto Scaling ignora i risultati dei controlli dell'integrità di Elastic Load Balancing. Tuttavia, è possibile abilitare questi controlli dell'integrità per il tuo gruppo con scalabilità automatica. Dopo aver eseguito questa operazione, quando Elastic Load Balancing segnala un'istanza registrata come `Unhealthy`, il Dimensionamento automatico Amazon EC2 contrassegna l'istanza come `Unhealthy` nel successivo controllo periodico dell'integrità e la sostituisce.

Se per il load balancer è abilitato lo svuotamento della connessione (ritardo di cancellazione), il Dimensionamento automatico Amazon EC2 attende il completamento delle richieste in transito o che il timeout massimo scada, prima di terminare le istanze non integre.

Per ulteriori informazioni su come abilitare i controlli dell'integrità Elastic Load Balancing per il gruppo con dimensionamento automatico, consulta [Collega un sistema di bilanciamento del carico Elastic Load Balancing al tuo gruppo Auto Scaling](#).

Note

Quando si abilitano i controlli dell'integrità di Elastic Load Balancing per un gruppo, il Dimensionamento automatico Amazon EC2 è in grado di sostituire le istanze segnalate come non integre da Elastic Load Balancing, ma solo dopo che il sistema di bilanciamento del carico sarà entrato nello stato `InService`. Per ulteriori informazioni, consulta [Verifica lo stato di collegamento del sistema di bilanciamento del carico](#).

Controlli dell'integrità VPC Lattice

Per impostazione predefinita, Dimensionamento automatico Amazon EC2 ignora i risultati dei controlli dell'integrità di VPC Lattice. In alternativa, puoi abilitare questi controlli dell'integrità per il tuo gruppo con dimensionamento automatico. Dopo aver eseguito questa operazione, quando VPC Lattice segnala un'istanza registrata come `Unhealthy`, il Dimensionamento automatico Amazon EC2 contrassegna l'istanza come `Unhealthy` nel successivo controllo periodico dell'integrità e la sostituisce. Il processo di registrazione delle istanze e quindi del relativo controllo dello stato è lo stesso utilizzato per i controlli dell'integrità di Elastic Load Balancing.

Per informazioni su come abilitare i controlli dell'integrità VPC Lattice per il gruppo con dimensionamento automatico, consulta [Allegare un gruppo di destinazione VPC Lattice al gruppo con dimensionamento automatico](#).

Note

Quando si abilitano i controlli dell'integrità di VPC Lattice per un gruppo, il Dimensionamento automatico Amazon EC2 è in grado di sostituire le istanze segnalate da VPC Lattice come non integre, ma solo dopo che il gruppo di destinazioni sarà entrato nello stato `InService`. Per ulteriori informazioni, consulta [Informazioni sullo stato di collegamento del gruppo di destinazione VPC Lattice](#).

Come Amazon EC2 Auto Scaling riduce al minimo i tempi di inattività

Come impostazione predefinita, le sostituzioni dei controlli dell'integrità richiedono prima la terminazione delle istanze, il che potrebbe impedire l'accettazione di nuove richieste fino all'avvio delle nuove istanze.

Se Amazon EC2 Auto Scaling determina che alcune istanze non sono più in esecuzione (o sono state Unhealthy contrassegnate con [set-instance-health](#) comando), le sostituisce immediatamente. Tuttavia, se altre istanze risultino non integre, il Dimensionamento automatico Amazon EC2 utilizza l'approccio riportato di seguito per riparare gli errori. Questo approccio riduce al minimo i tempi di inattività che potrebbero verificarsi a causa di problemi temporanei o controlli di integrità configurati in modo errato.

- Se è in corso un'attività di dimensionamento e la capacità del gruppo con scalabilità automatica è inferiore al 10% o più, il Dimensionamento automatico Amazon EC2 attende l'attività di dimensionamento in corso prima di sostituire le istanze non integre.
- Durante l'aumento orizzontale, il Dimensionamento automatico Amazon EC2 attende che le istanze superino un controllo dell'integrità iniziale. Attende inoltre il completamento del warmup dell'istanza predefinito per assicurarsi che le nuove istanze siano pronte.
- Al termine del warmup delle istanze e dell'aumento del gruppo a oltre il 90% della capacità desiderata, il Dimensionamento automatico Amazon EC2 sostituisce le istanze non integre come riportato di seguito:
 - Il Dimensionamento automatico Amazon EC2 sostituisce solo fino al 10% della capacità desiderata del gruppo alla volta. Lo fa fino a quando tutte le istanze non integre saranno state sostituite.
 - Quando si sostituiscono le istanze, attende che le istanze superino un controllo dell'integrità iniziale. Prima di continuare, attende anche che il warmup predefinito dell'istanza sia finito.

Note

Se le dimensioni di un gruppo con scalabilità automatica sono tanto piccole che il valore risultante del 10% è inferiore a uno, il Dimensionamento automatico Amazon EC2 sostituisce le istanze non integre una alla volta. Questo potrebbe comportare tempi di inattività del gruppo.

Inoltre, se tutte le istanze di un gruppo con dimensionamento automatico sono segnalate come non integre dai controlli dell'integrità di Elastic Load Balancing e il sistema di bilanciamento del carico è nello stato InService, il Dimensionamento automatico Amazon EC2 potrebbe contrassegnare un numero inferiore di istanze non integre alla volta. Ciò può comportare una sostituzione molto inferiore di istanze alla volta rispetto al 10% applicato in altri scenari. In questo modo è possibile risolvere il problema senza che il Dimensionamento automatico Amazon EC2 interrompa automaticamente l'intero gruppo.

Considerazioni sul controllo dell'integrità

Questa sezione contiene considerazioni per i controlli di integrità del Dimensionamento automatico Amazon EC2.

- In alternativa, se è necessario che qualcosa accada sull'istanza che sta terminando o sull'istanza che si sta avviando, puoi utilizzare gli hook del ciclo di vita. Questi hook ti consentono di eseguire un'operazione personalizzata quando Amazon EC2 Auto Scaling avvia o termina le istanze. Per ulteriori informazioni, consulta [Hook del ciclo di vita di Amazon EC2 Auto Scaling](#).
- Amazon EC2 Auto Scaling non offre un modo per rimuovere i controlli dello stato di Amazon EC2 e degli eventi programmati dai controlli dell'integrità. Se non vuoi sostituire le istanze, ti consigliamo di sospendere il processo `ReplaceUnhealthy` e `HealthCheck` per ogni singolo gruppo con scalabilità automatica. Per ulteriori informazioni, consulta [Sospendi e riprendi i processi di Amazon EC2 Auto Scaling](#).
- Per ripristinare manualmente lo stato di integrità di un'istanza non `integrHealthy`, puoi provare a utilizzare il comando `set-instance-health`. Se si riceverà un errore, ciò sarà probabilmente dovuto al fatto che l'istanza è già in corso di terminazione. In genere, reimpostare lo stato di integrità di un'istanza su `Healthy` con il `set-instance-health` comando è utile solo nei casi in cui il `ReplaceUnhealthy` processo o il `Terminate` processo sono sospesi.
- Amazon EC2 Auto Scaling non esegue controlli dell'integrità su istanze che non sono nello stato `Standby`. Per ulteriori informazioni, consulta [Rimuovere temporaneamente le istanze dal gruppo con scalabilità automatica](#).
- Quando l'istanza è terminata, i relativi indirizzi IP elastici sono disassociati e non vengono automaticamente associati alla nuova istanza. È necessario associare manualmente i indirizzi IP elastici alla nuova istanza o farlo automaticamente, con una soluzione basata su hook del ciclo di vita. Per ulteriori informazioni, consultare [Indirizzi IP elastici](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.
- Analogamente, quando l'istanza viene terminata, i relativi volumi EBS collegati vengono scollegati (o eliminati a seconda dell'attributo `DeleteOnTermination` del volume). È necessario collegare manualmente questi volumi EBS alla nuova istanza o farlo automaticamente con una soluzione basata su un hook del ciclo di vita. Per ulteriori informazioni, consulta [Collegare un volume Amazon EBS a un'istanza](#) nella Amazon EBS User Guide.

Controlli dell'integrità personalizzati

In alternativa, puoi eseguire attività di rilevamento dell'integrità personalizzate sulle istanze del gruppo con dimensionamento automatico e impostare lo stato di integrità di un'istanza come `Unhealthy` se l'operazione ha esito negativo. In questo modo, i controlli dell'integrità vengono estesi utilizzando una combinazione di controlli dell'integrità personalizzati, controlli di stato Amazon EC2 e controlli dell'integrità Elastic Load Balancing, se abilitati.

Puoi inviare le informazioni sull'integrità delle istanze direttamente dal tuo sistema al Dimensionamento automatico Amazon EC2 utilizzando la AWS CLI o l'SDK. Gli esempi seguenti mostrano come utilizzare AWS CLI per configurare lo stato di integrità di un'istanza e quindi verificare lo stato di integrità dell'istanza.

Utilizzate il [set-instance-health](#) comando seguente per impostare lo stato di integrità dell'istanza specificata su `Unhealthy`.

```
aws autoscaling set-instance-health --instance-id i-1234567890abcdef0 --health-status Unhealthy
```

Per impostazione predefinita, questo comando rispetta il periodo di tolleranza del controllo dell'integrità. Tuttavia, è possibile sovrascrivere questo comportamento e non rispettare il periodo di tolleranza includendo l'opzione `--no-should-respect-grace-period`.

Utilizzate il [describe-auto-scaling-groups](#) comando seguente per verificare che lo stato di integrità dell'istanza sia `Unhealthy`.

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-names my-asg
```

Di seguito è riportato un esempio di risposta che mostra che lo stato dell'istanza è `Unhealthy` e che questa sta per essere terminata.

```
{
  "AutoScalingGroups": [
    {
      ....
      "Instances": [
        {
          "ProtectedFromScaleIn": false,
          "AvailabilityZone": "us-west-2a",
          "LaunchTemplate": {
```

```
        "LaunchTemplateName": "my-launch-template",
        "Version": "1",
        "LaunchTemplateId": "lt-1234567890abcdef0"
    },
    "InstanceId": "i-1234567890abcdef0",
    "InstanceType": "t2.micro",
    "HealthStatus": "Unhealthy",
    "LifecycleState": "Terminating"
},
...
]
}
}
```

Risorse correlate

Per informazioni sulla risoluzione dei problemi dei controlli dell'integrità, consulta [Risoluzione dei problemi relativi ai controlli dell'integrità in Amazon EC2 Auto Scaling](#). Se i controlli di integrità non riescono, consulta questo argomento per il passaggio della procedura di risoluzione dei problemi. Nei seguenti argomenti viene descritto come individuare gli errori nel gruppo con scalabilità automatica e vengono offerti suggerimenti su come correggerli.

Amazon EC2 Auto Scaling monitora anche lo stato delle istanze avviate in un warm pool utilizzando Amazon EC2, Amazon EBS o controlli dell'integrità personalizzati. Per ulteriori informazioni, consulta [Visualizzare lo stato del controllo dell'integrità e il motivo degli errori relativi ad esso](#).

Come visualizzare il motivo dei fallimenti dei controlli dell'integrità

Utilizzando la procedura seguente, puoi visualizzare le informazioni su tutte le istanze sostituite a seguito di un controllo dell'integrità.

Come visualizzare il motivo dei fallimenti dei controlli dell'integrità (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina Gruppi Auto Scaling.

3. Nella scheda Activity (Attività) in Activity history (Cronologia attività), la colonna Status (Stato) indica se il gruppo con scalabilità automatica ha avviato o terminato correttamente le istanze.

Se ha terminato eventuali istanze non integre, il campo Cause (Causa) mostra la data e l'ora della terminazione e il motivo dell'errore di controllo dell'integrità. Ad esempio, `At 2022-05-14T20:11:53Z an instance was taken out of service in response to an ELB system health check failure.`

Per informazioni sui tipi di errori che è possibile riscontrare e su come gestirli, consulta [Risoluzione dei problemi relativi ai controlli dell'integrità in Amazon EC2 Auto Scaling](#).

Note

Dimensionamento automatico Amazon EC2 crea una nuova attività di dimensionamento per terminare l'istanza non integra, quindi la termina. Mentre l'istanza viene terminata, un'altra attività di dimensionamento avvia una nuova istanza.

Puoi modificare questo comportamento per avviare prima una nuova istanza utilizzando una policy di manutenzione dell'istanza. Con una policy di manutenzione delle istanze, puoi stabilire soglie per un gruppo con dimensionamento automatico in caso di eventi che portano alla sostituzione dell'istanza e il gruppo con dimensionamento automatico può sostituire solo le istanze all'interno di tale intervallo di soglia. Tuttavia, poiché Dimensionamento automatico Amazon EC2 interrompe immediatamente le istanze che non sono più in esecuzione, la soglia inferiore della policy di manutenzione delle istanze può essere superata se un'istanza si interrompe in modo imprevisto o se si arresta o si riavvia manualmente un'istanza. Per ulteriori informazioni, consulta [Policy di manutenzione delle istanze](#).

Impostazione del periodo di tolleranza del controllo dell'integrità per un gruppo con scalabilità automatica

Quando un controllo dell'integrità del Dimensionamento automatico Amazon EC2 determina che un'istanza `InService` non è integra, la sostituisce con una nuova. Il periodo di tolleranza per il controllo dell'integrità specifica il periodo di tempo minimo (in secondi) necessario per mantenere una nuova istanza in servizio prima di terminarla se risulta non integra.

Un esempio di caso d'uso potrebbe essere rappresentato da un requisito di Dimensionamento automatico Amazon EC2 per evitare di intervenire se i controlli dell'integrità di Elastic Load Balancing hanno esito negativo e se la causa è l'istanza ancora in fase di inizializzazione. I controlli dell'integrità di Elastic Load Balancing vengono eseguiti in parallelo, a partire dal momento in cui l'istanza

viene registrata con il sistema di bilanciamento del carico. Il periodo di tolleranza impedisce a Dimensionamento automatico Amazon EC2 di contrassegnare le istanze appena avviate come Unhealthy e di terminarle inutilmente se non superano immediatamente questi controlli dell'integrità dopo l'entrata nello stato InService.

Per impostazione predefinita, quando crei un gruppo con scalabilità automatica nella console il periodo di tolleranza del controllo dell'integrità è 300 secondi. Il suo valore predefinito è 0 secondi quando si crea un gruppo Auto Scaling utilizzando AWS CLI o un SDK. Il valore 0 disattiva il periodo di tolleranza per il controllo dell'integrità.

L'impostazione su un valore troppo alto riduce l'efficacia dei controlli dell'integrità di Dimensionamento automatico Amazon EC2. Se utilizzi un hook del ciclo di vita per l'avvio dell'istanza, puoi impostare il valore del periodo di tolleranza del controllo dell'integrità su 0. Grazie agli hook del ciclo di vita, Dimensionamento automatico Amazon EC2 offre un modo per assicurarsi che le istanze siano sempre inizializzate prima che entrino nello stato InService. Per ulteriori informazioni, consulta [Hook del ciclo di vita di Amazon EC2 Auto Scaling](#).

Il periodo di tolleranza si applica alle seguenti istanze:

- Istanze appena avviate
- Istanze che vengono rimesse in servizio dopo essere state messe in standby
- Istanze che vengono collegate manualmente al gruppo

Important

Durante il periodo di tolleranza del controllo dell'integrità, se Dimensionamento automatico Amazon EC2 rileva che un'istanza non è più nello stato `running` Amazon EC2, la contrassegna immediatamente come Unhealthy e la sostituisce. Ad esempio, se interrompi un'istanza in un gruppo con dimensionamento automatico, questa viene contrassegnata come Unhealthy e viene sostituita.

Impostazione del periodo di tolleranza del controllo dell'integrità per un gruppo

Il periodo di tolleranza del controllo dell'integrità può essere impostato per gruppi con scalabilità automatica nuovi ed esistenti.

Console

Modifica del periodo di tolleranza del controllo dell'integrità per un nuovo gruppo (console)

Quando crei il gruppo con scalabilità automatica, alla pagina Configure advanced options (Configura opzioni avanzate), per Health checks (Controlli dell'integrità), Health check grace period (Periodo di tolleranza del controllo dell'integrità), scegli la quantità di tempo (in secondi). Questo è il periodo di tempo che Dimensionamento automatico Amazon EC2 deve attendere prima di controllare lo stato di integrità di un'istanza dopo che questa è entrata nello stato InService.

AWS CLI

Modifica del periodo di tolleranza del controllo dell'integrità per un nuovo gruppo (AWS CLI)

Aggiungete l'`--health-check-grace-period` opzione al comando. [create-auto-scaling-group](#)
L'esempio seguente configura il periodo di tolleranza del controllo dell'integrità con un valore di **60** secondi per un nuovo gruppo con scalabilità automatica denominato *my-asg*.

```
aws autoscaling create-auto-scaling-group --auto-scaling-group-name my-asg --health-check-grace-period 60 ...
```

Console

Per modificare il periodo di tolleranza del controllo dell'integrità per un gruppo esistente (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Sulla barra di navigazione nella parte superiore della schermata, seleziona la Regione AWS in cui è stato creato il gruppo con dimensionamento automatico.
3. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

4. Nella scheda Dettagli scegli Controlli dell'integrità, Modifica.
5. In Health check grace period (Periodo di tolleranza del controllo dell'integrità), inserisci il tempo, in secondi. Questo è il periodo di tempo che Dimensionamento automatico Amazon EC2 deve attendere prima di controllare lo stato di integrità di un'istanza dopo che questa è entrata nello stato InService.

6. Scegli Aggiorna.

AWS CLI

Modifica del periodo di tolleranza del controllo dell'integrità per un gruppo esistente (AWS CLI)

Aggiunge l' `--health-check-grace-period` opzione al [update-auto-scaling-group](#) comando. L'esempio seguente configura il periodo di tolleranza del controllo dell'integrità con un valore di **120** secondi per un gruppo con scalabilità automatica esistente denominato *my-asg*.

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg --health-check-grace-period 120
```

Note

Consigliamo vivamente di impostare anche il tempo di riscaldamento predefinito dell'istanza per il gruppo Auto Scaling. Per ulteriori informazioni, consulta [Impostazione della preparazione di default dell'istanza per un gruppo con scalabilità automatica](#).

AWS Health Dashboard notifiche per Amazon EC2 Auto Scaling

AWS Health Dashboard Fornisce supporto per le notifiche provenienti da Amazon EC2 Auto Scaling. Queste notifiche offrono una panoramica dettagliata e soluzioni specifiche per problematiche a livello di prestazioni o di disponibilità delle risorse che potrebbero interessare le applicazioni. Sono attualmente disponibili solo gli eventi specifici dei gruppi di sicurezza e dei modelli di avvio mancanti.

Fa AWS Health Dashboard parte del servizio. AWS Health Questo servizio non richiede l'installazione e può essere visualizzato da qualsiasi utente autenticato nell'account. Per ulteriori informazioni, vedi [Guida introduttiva alla AWS Health dashboard](#).

Se visualizzi un messaggio simile a uno dei seguenti, questo deve essere gestito come un allarme su cui intervenire.

Esempio: il dimensionamento orizzontale del gruppo Auto Scaling non sta avendo luogo a causa di un gruppo di sicurezza mancante

```
Hello,
```

At 2020-01-11 04:00 UTC, we detected an issue with your Auto Scaling group [ARN] in Account AWS 123456789012.

A security group associated with this Auto Scaling group cannot be found. Each time a scale out operation is performed, it will be prevented until you make a change that fixes the issue.

We recommend that you review and update your Auto Scaling group configuration to change the launch template or launch configuration that depends on the unavailable security group.

Sincerely,
Amazon Web Services

Esempio: il dimensionamento orizzontale del gruppo Auto Scaling non sta avendo luogo a causa di un modello di avvio mancante

Hello,

At 2021-05-11 04:00 UTC, we detected an issue with your Auto Scaling group [ARN] in Account AWS 123456789012.

The launch template associated with this Auto Scaling group cannot be found. Each time a scale out operation is performed, it will be prevented until you make a change that fixes the issue.

We recommend that you review and update your Auto Scaling group configuration and specify an existing launch template to use.

Sincerely,
Amazon Web Services

Monitorare i parametri di CloudWatch per gruppi e istanze Auto Scaling

I parametri sono i concetti fondamentali in Amazon CloudWatch. Un parametro rappresenta un set in ordine cronologico di punti dati pubblicati in CloudWatch. Pensa a un parametro come a una variabile

da monitorare e ai punti di dati come i valori di questa variabile nel tempo. Puoi utilizzare questi parametri per verificare che le prestazioni del sistema siano quelle previste.

I parametri di Dimensionamento automatico Amazon EC2 per raccogliere informazioni sui gruppi con scalabilità automatica si trovano nel spazio dei nomi AWS/AutoScaling. I parametri delle istanze Amazon EC2 che raccolgono CPU e altri dati di utilizzo dalle istanze Auto Scaling si trovano nel spazio dei nomi AWS/EC2.

La console di Dimensionamento automatico Amazon EC2 visualizza una serie di grafici per i parametri del gruppo e i parametri aggregati delle istanze per il gruppo. In base alle tue esigenze, puoi decidere di accedere ai dati dei gruppi con scalabilità automatica da Amazon CloudWatch invece che dalla console di Dimensionamento automatico Amazon EC2.

Per ulteriori informazioni, consulta la [Guida per l'utente di Amazon CloudWatch](#).

Indice

- [Visualizzare i grafici di monitoraggio nella console Dimensionamento automatico Amazon EC2](#)
- [Parametri di Amazon CloudWatch per Dimensionamento automatico Amazon EC2](#)
- [Configurare il monitoraggio per le istanze Auto Scaling](#)

Visualizzare i grafici di monitoraggio nella console Dimensionamento automatico Amazon EC2

Nella sezione Dimensionamento automatico Amazon EC2 della console Amazon EC2, è possibile monitorare l'avanzamento minuto per minuto dei singoli gruppi Auto Scaling utilizzando i parametri di CloudWatch.

Puoi monitorare i seguenti tipi di parametri:

- Auto Scaling metrics (Parametri di Auto Scaling): i parametri di Auto Scaling sono attivati solo quando li si abilita. Per ulteriori informazioni, consulta [Abilitazione dei parametri del gruppo con scalabilità automatica \(console\)](#). Quando i parametri di Auto Scaling sono abilitati, i grafici di monitoraggio mostrano i dati pubblicati con granularità di un minuto per i parametri di Auto Scaling.
- Parametri EC2: i parametri delle istanze Amazon EC2 sono sempre abilitati. Se il monitoraggio dettagliato è abilitato, i grafici di monitoraggio mostrano i dati pubblicati con granularità di un minuto per i parametri dell'istanza. Per ulteriori informazioni, consulta [Configurare il monitoraggio per le istanze Auto Scaling](#).

Per visualizzare i grafici di monitoraggio nella console Dimensionamento automatico Amazon EC2

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con scalabilità automatica per il quale si desidera visualizzare i parametri.

Si aprirà un riquadro diviso nella parte inferiore della pagina Auto Scaling groups (Gruppi Auto Scaling).

3. Scegliere la scheda Monitoring (Monitoraggio).

Dimensionamento automatico Amazon EC2 visualizza grafici di monitoraggio per i parametri Auto Scaling.

4. Per visualizzare i grafici di monitoraggio dei parametri aggregati delle istanze per il gruppo, scegli EC2.

Operazioni del grafico

- Passa il mouse su un punto dati per visualizzare un pop-up di dati per un determinato periodo in UTC.
- Per ingrandire un grafico, scegli Enlarge (Ingrandisci) dallo strumento menu (i tre punti verticali) nella parte superiore destra del grafico. In alternativa, scegli l'icona di ingrandimento nella parte superiore del grafico.
- Regola il periodo di tempo per i dati visualizzati nel grafico selezionando uno dei valori predefiniti del periodo di tempo. Se il grafico viene ingrandito, è possibile scegliere Custom (Personalizzato) per definire il periodo di tempo.
- Scegli Update (Aggiorna) dallo strumento menu per aggiornare i dati in un grafico.
- Trascina il cursore sui dati del grafico per selezionare un intervallo specifico. È quindi possibile scegliere Apply time range (Applica intervallo di tempo) nello strumento menu.
- Scegli Visualizza log dallo strumento menu per visualizzare i flussi di registri associati (se presenti) nella console CloudWatch.
- Per visualizzare un grafico in CloudWatch, scegli View in metrics (Visualizza nei parametri) dallo strumento menu. In questo modo si accede alla pagina di CloudWatch per quel grafico. Qui è possibile visualizzare ulteriori informazioni o accedere a informazioni cronologiche per comprendere meglio come il gruppo con scalabilità automatica è cambiato per un periodo prolungato.

Parametri grafici per i gruppi Auto Scaling

Dopo aver creato un gruppo con scalabilità automatica, è possibile aprirle la console Dimensionamento automatico Amazon EC2 e visualizzare i grafici di monitoraggio del gruppo nella scheda Monitoring (Monitoraggio).

Nella sezione Auto Scaling, i parametri grafici includono i seguenti parametri. Questi parametri forniscono misurazioni che possono essere indicatori di un potenziale problema, come ad esempio il numero di istanze di terminazione o il numero di istanze in sospeso. Le definizioni di questi parametri sono riportate in [Parametri di Amazon CloudWatch per Dimensionamento automatico Amazon EC2](#).

Display name (Nome visualizzato)	Nome del parametro CloudWatch
Dimensione minima gruppo	GroupMinSize
Dimensione massima gruppo	GroupMaxSize
Capacità desiderata	GroupDesiredCapacity
Istanze in esecuzione	GroupInServiceInstances
Istanze in sospeso	GroupPendingInstances
Istanze in standby	GroupStandbyInstances
Istanze in terminazione	GroupTerminatingInstances
Istanze totali	GroupTotalInstances

Nella sezione EC2, puoi trovare i seguenti parametri grafici in base ai parametri chiave delle prestazioni per le tue istanze Amazon EC2. Questi parametri EC2 sono un aggregato di parametri per tutte le istanze del gruppo. Puoi trovare le definizioni per questi parametri in [Elenca i parametri di CloudWatch disponibili per le istanze](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Display name (Nome visualizzato)	Nome del parametro CloudWatch
Utilizzo CPU	CPUUtilization

Display name (Nome visualizzato)	Nome del parametro CloudWatch
Letture del disco	DiskReadBytes
Operazioni di lettura del disco	DiskReadOps
Scritture del disco	DiskWriteBytes
Operazioni di scrittura del disco	DiskWriteOps
Rete in ingresso	NetworkIn
Rete in uscita	NetworkOut
Status Check Failed (Any) (Verifica stato non riuscita (qualsiasi))	StatusCheckFailed
Status Check Failed (Instance) (Verifica stato non riuscita (istanza))	StatusCheckFailed_Instance
Status Check Failed (System) (Verifica stato non riuscita (sistema))	StatusCheckFailed_System

Inoltre, alcuni parametri sono disponibili per casi d'uso specifici nei parametri del grafico dimensionamento automatico.

I parametri seguenti sono utili per il gruppo che usa pesi che definiscono il contributo di ogni istanza alla capacità desiderata del gruppo in termini di numero di unità. Le definizioni di questi parametri sono riportate in [Parametri di Amazon CloudWatch per Dimensionamento automatico Amazon EC2](#).

Display name (Nome visualizzato)	Nome del parametro CloudWatch
Unità di capacità in servizio	GroupInServiceCapacity
Unità di capacità in sospenso	GroupPendingCapacity
Unità di capacità in standby	GroupStandbyCapacity
Unità di capacità in terminazione	GroupTerminatingCapacity
Unità di capacità totali	GroupTotalCapacity

I seguenti parametri sono utili se il gruppo utilizza la funzione [pool caldo](#). Le definizioni di questi parametri sono riportate in [Parametri di Amazon CloudWatch per Dimensionamento automatico Amazon EC2](#).

Display name (Nome visualizzato)	Nome del parametro CloudWatch
Dimensione minima del pool caldo	WarmPoolMinSize
Capacità desiderata del pool caldo	WarmPoolDesiredCapacity
Unità di capacità in attesa del pool caldo	WarmPoolPendingCapacity
Unità di capacità in terminazione del pool caldo	WarmPoolTerminatingCapacity
Unità di capacità preparate del pool caldo	WarmPoolWarmedCapacity
Unità di capacità totali del pool caldo avviate	WarmPoolTotalCapacity

Display name (Nome visualizzato)	Nome del parametro CloudWatch
Capacità desiderata del gruppo e del pool caldo	GroupAndWarmPoolDesiredCapacity
Unità di capacità totali del gruppo e del pool caldo avviate	GroupAndWarmPoolTotalCapacity

Risorse correlate

- Per visualizzare i parametri per le istanze, consulta [Graph metrics for your instances](#) (Crea grafici dei parametri per le tue istanze) nella Guida per l'utente di Amazon EC2 per le istanze Linux.
- I pannelli di controllo di CloudWatch sono home page personalizzabili nella console CloudWatch. Puoi utilizzare queste pagine per monitorare le risorse in un'unica visualizzazione, comprese quelle distribuite in diverse Regioni. Puoi utilizzare i pannelli di controllo di CloudWatch per creare viste personalizzate dei parametri e degli allarmi per le risorse AWS. Per ulteriori informazioni, consulta la [Guida per l'utente di Amazon CloudWatch](#).

Parametri di Amazon CloudWatch per Dimensionamento automatico Amazon EC2

Dimensionamento automatico Amazon EC2 pubblica i seguenti parametri nello spazio dei nomi AWS/AutoScaling. I parametri effettivi del gruppo con dimensionamento automatico rese disponibili dipenderanno dal fatto che i parametri di gruppo siano abilitati e dai parametri di gruppo abilitati. I parametri del gruppo sono disponibili con granularità di un minuto senza costi aggiuntivi, ma è necessario abilitarli.

Quando attivi i parametri del gruppo con dimensionamento automatico, Dimensionamento automatico Amazon EC2 invia dati campionati a CloudWatch ogni minuto sulla base del miglior tentativo. Nei rari casi in cui CloudWatch sta subendo un'interruzione del servizio, i dati non vengono compilati per colmare le lacune nella cronologia dei parametri di gruppo.

Indice

- [Parametri del gruppo con scalabilità automatica](#)

- [Dimensioni per parametri del gruppo con scalabilità automatica](#)
- [Parametri e dimensioni di dimensionamento predittivo](#)
- [Abilitazione dei parametri del gruppo con scalabilità automatica \(console\)](#)
- [Abilitazione dei parametri del gruppo con scalabilità automatica \(AWS CLI\)](#)

Parametri del gruppo con scalabilità automatica

Con questi parametri, si ottiene una visibilità quasi continua nella cronologia del gruppo con scalabilità automatica, ad esempio le modifiche delle dimensioni del gruppo nel tempo.

Parametro	Descrizione
GroupMinSize	Dimensione minima del gruppo con scalabilità automatica. Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.
GroupMaxSize	Dimensione massima del gruppo con scalabilità automatica. Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.
GroupDesiredCapacity	Numero di istanze che il gruppo con scalabilità automatica tenta di gestire. Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.
GroupInServiceInstances	Numero di istanze in esecuzione come parte del gruppo con scalabilità automatica. Questo parametro non include istanze in sospeso o in corso di terminazione. Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.
GroupPendingInstances	Numero di istanze in sospeso. Un'istanza in sospeso non è ancora operativa. Questo parametro non include istanze operative o in corso di terminazione.

Parametro	Descrizione
	<p>Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.</p>
GroupStandbyInstances	<p>Il numero di istanze con stato Standby. Le istanze in questo stato sono ancora in esecuzione ma non attivamente operative.</p> <p>Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.</p>
GroupTerminatingInstances	<p>Numero di istanze che sono in corso di terminazione. Questo parametro non include istanze operative o in sospeso.</p> <p>Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.</p>
GroupTotalInstances	<p>Numero totale di istanze nel gruppo con scalabilità automatica. Questo parametro identifica il numero di istanze operative, in sospeso e in corso di terminazione.</p> <p>Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.</p>

Quando configuri un gruppo di istanze miste per misurare la capacità desiderata in unità diverse, ad esempio assegnando pesi in base al numero di vCPU di ciascun tipo di istanza, i seguenti parametri contano il numero di unità utilizzate dal gruppo con dimensionamento automatico. Se non hai configurato un gruppo di istanze miste per misurare la capacità desiderata in unità diverse, i parametri seguenti sono uguali ai parametri corrispondenti definiti nella tabella precedente. Per ulteriori informazioni, consulta [Panoramica della configurazione](#).

Parametro	Descrizione
GroupInServiceCapacity	<p>Numero di unità di capacità in esecuzione come parte del gruppo con scalabilità automatica.</p> <p>Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.</p>

Parametro	Descrizione
GroupPendingCapacity	Numero di unità di capacità in sospeso. Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.
GroupStandbyCapacity	Numero di unità di capacità in stato Standby. Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.
GroupTerminatingCapacity	Il numero di unità di capacità in fase di terminazione. Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.
GroupTotalCapacity	Il numero totale di unità di capacità nel gruppo con scalabilità automatica. Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.

Dimensionamento automatico Amazon EC2 riporta i parametri descritti di seguito per gruppi con scalabilità automatica che dispongono di un warm pool. Per ulteriori informazioni, consulta [Warm pool per Dimensionamento automatico Amazon EC2](#).

Parametro	Descrizione
WarmPoolMinSize	Dimensione minima del warm pool. Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.
WarmPoolDesiredCapacity	La quantità di capacità che Dimensionamento automatico Amazon EC2 tenta di mantenere nel warm pool. Equivale alla dimensione massima del gruppo con scalabilità automatica meno la capacità desiderata o, se impostata

Parametro	Descrizione
	<p>, alla capacità massima preparata del gruppo con scalabilità automatica meno la capacità desiderata.</p> <p>Tuttavia, quando la dimensione minima del warm pool è uguale o superiore alla differenza tra le dimensioni massime (o, se impostata, la capacità massima preparata) e la capacità desiderata del gruppo con scalabilità automatica, la capacità desiderata del warm pool sarà equivalente a <code>WarmPoolMinSize</code>.</p> <p>Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.</p>
<p><code>WarmPoolPendingCapacity</code></p>	<p>La quantità di capacità nel warm pool in sospeso. Questo parametro non include istanze in funzione, arrestate o in corso di terminazione.</p> <p>Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.</p>
<p><code>WarmPoolTerminatingCapacity</code></p>	<p>La quantità di capacità nel warm pool in fase di terminazione. Questo parametro non include istanze in funzione, arrestate o in sospeso.</p> <p>Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.</p>
<p><code>WarmPoolWarmedCapacity</code></p>	<p>La quantità di capacità disponibile per accedere al gruppo con scalabilità automatica durante l'aumento orizzontale. Questo parametro non include istanze in sospeso o in corso di terminazione.</p> <p>Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.</p>

Parametro	Descrizione
WarmPoolTotalCapacity	<p>La capacità totale del warm pool, incluse le istanze in funzione, arrestate, in sospeso o in corso di terminazione.</p> <p>Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.</p>
GroupAndWarmPoolDesiredCapacity	<p>La capacità desiderata del gruppo con scalabilità automatica e del warm pool combinata.</p> <p>Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.</p>
GroupAndWarmPoolTotalCapacity	<p>La capacità totale del gruppo con scalabilità automatica e del warm pool combinata. Sono incluse le istanze in funzione, arrestate, in sospeso, in corso di terminazione o in servizio.</p> <p>Criteri di creazione di report: segnalato se è abilitata la raccolta dei parametri.</p>

Dimensioni per parametri del gruppo con scalabilità automatica

È possibile utilizzare le seguenti dimensioni per perfezionare i parametri elencati nelle tabelle precedenti.

Dimensione	Descrizione
AutoScalingGroupName	Filtri sul nome di un gruppo con scalabilità automatica.

Parametri e dimensioni di dimensionamento predittivo

Lo spazio dei nomi AWS/AutoScaling include i parametri descritti di seguito per il dimensionamento predittivo.

I parametri sono disponibili con una risoluzione di un'ora.

È possibile valutare l'accuratezza delle previsioni confrontando i valori previsti con i valori effettivi. Per ulteriori informazioni sulla valutazione dell'accuratezza delle previsioni tramite questi parametri, consulta [Monitora le metriche di scalabilità predittiva con CloudWatch](#).

Parametro	Descrizione	Dimensioni
PredictiveScalingLoadForecast	<p>La quantità di carico che si prevede venga generata dall'applicazione.</p> <p>Le statistiche Average, Minimum e Maximum sono utili, ma la statistica Sum non lo è.</p> <p>Criteri di reporting: riportato dopo la creazione della previsione iniziale.</p>	AutoScalingGroupName , PolicyName , PairIndex
PredictiveScalingCapacityForecast	<p>La quantità prevista di capacità necessaria per soddisfare la domanda delle applicazioni. Questo si basa sulla previsione del carico e sul livello di utilizzo target a cui si desidera mantenere le istanze Auto Scaling.</p> <p>Le statistiche Average, Minimum e Maximum sono utili, ma la statistica Sum non lo è.</p> <p>Criteri di reporting: riportato dopo la creazione della previsione iniziale.</p>	AutoScalingGroupName , PolicyName
PredictiveScalingMetricPairCorrelation	<p>La correlazione tra il parametro di dimensionamento e la media per istanza del parametro di carico. Il dimensionamento predittivo presuppone un'elevata correlazione. Pertanto, se si osserva un valore basso per questo parametro, è consigliabile non utilizzare una coppia di parametri.</p> <p>Le statistiche Average, Minimum e Maximum sono utili, ma la statistica Sum non lo è.</p>	AutoScalingGroupName , PolicyName , PairIndex

Parametro	Descrizione	Dimensioni
	Criteri di reporting: riportato dopo la creazione della previsione iniziale.	

Note

La dimensione `PairIndex` restituisce le informazioni associate all'indice della coppia di parametri in scala di carico assegnata da Dimensionamento automatico Amazon EC2. Attualmente, l'unico valore valido è `0`.

Abilitazione dei parametri del gruppo con scalabilità automatica (console)

Come abilitare i parametri del gruppo

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Nella scheda Monitoring (Monitoraggio), seleziona la casella di controllo Auto Scaling group metrics collection (Raccolta di parametri del gruppo con scalabilità automatica), Enable (Abilita), nella parte superiore della pagina, sotto Auto Scaling.

Come disabilitare i parametri del gruppo

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona il gruppo con scalabilità automatica.
3. Nella scheda Monitoring (Monitoraggio), deseleziona la casella di controllo Auto Scaling group metrics collection (Raccolta di parametri del gruppo con scalabilità automatica), Enable (Abilita).

Abilitazione dei parametri del gruppo con scalabilità automatica (AWS CLI)

Abilitazione dei parametri del gruppo con dimensionamento automatico

Abilitare uno o più parametri del gruppo utilizzando il comando [enable-metrics-collection](#). Ad esempio, il comando seguente abilita un singolo parametro per il gruppo con dimensionamento automatico specificato.

```
aws autoscaling enable-metrics-collection --auto-scaling-group-name my-asg \  
--metrics GroupDesiredCapacity --granularity "1Minute"
```

Se ometti l'opzione `--metrics`, tutti i parametri saranno abilitati.

```
aws autoscaling enable-metrics-collection --auto-scaling-group-name my-asg \  
--granularity "1Minute"
```

Come disabilitare i parametri del gruppo con dimensionamento automatico

Utilizza il comando [disable-metrics-collection](#) per disabilitare tutti i parametri di gruppo.

```
aws autoscaling disable-metrics-collection --auto-scaling-group-name my-asg
```

Configurare il monitoraggio per le istanze Auto Scaling

Amazon EC2 raccoglie ed elabora i dati non elaborati delle istanze e li elabora trasformandoli in tempo reale e leggibili per il gruppo con scalabilità automatica. È possibile configurare l'intervallo per il monitoraggio di questi parametri scegliendo la granularità di un minuto o cinque minuti.

Il monitoraggio dell'istanza viene attivato ogni volta che viene avviata un'istanza, utilizzando il monitoraggio di base (granularità di cinque minuti) o il monitoraggio dettagliato (granularità di un minuto). Per il monitoraggio dettagliato, sono previsti costi aggiuntivi. Per ulteriori informazioni, consulta [Prezzi di Amazon CloudWatch](#) e [Monitoraggio delle istanze tramite CloudWatch](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Prima di creare un gruppo con scalabilità automatica, è necessario creare un modello di avvio o una configurazione di avvio per consentire il tipo di monitoraggio appropriato per l'applicazione. Se si aggiunge una policy di dimensionamento al gruppo, si consiglia di utilizzare il monitoraggio dettagliato per ottenere i dati dei parametri per le istanze EC2 a una granularità di un minuto per garantire una risposta più rapida alle modifiche del carico.

Indice

- [Abilitazione del monitoraggio dettagliato \(console\)](#)
- [Abilitazione del monitoraggio dettagliato \(AWS CLI\)](#)

- [Passaggio tra il monitoraggio di base e quello dettagliato](#)
- [Raccolta di parametri aggiuntivi tramite l'agente CloudWatch](#)

Abilitazione del monitoraggio dettagliato (console)

Per impostazione predefinita, il monitoraggio base viene abilitato al momento della creazione di un modello di avvio o quando si utilizza la AWS Management Console per la creazione di una configurazione di avvio.

Come abilitare il monitoraggio dettagliato in un modello di avvio

Quando crei il modello di avvio utilizzando la AWS Management Console, nella sezione Advanced details (Dettagli avanzati), per Detailed CloudWatch monitoring (Monitoraggio CloudWatch dettagliato), scegli Enable (Abilita). In caso contrario, il monitoraggio di base è abilitato. Per ulteriori informazioni, consulta [Crea un modello di avvio utilizzando le impostazioni avanzate](#).

Come abilitare il monitoraggio dettagliato in una configurazione di avvio

Quando crei la configurazione di avvio tramite la AWS Management Console, nella pagina Additional configuration (Configurazione aggiuntiva) seleziona Enable EC2 instance detailed monitoring within CloudWatch (Abilita il monitoraggio dettagliato dell'istanza EC2 con CloudWatch). In caso contrario, il monitoraggio di base è abilitato. Per ulteriori informazioni, consulta [Creazione di una configurazione di avvio](#).

Abilitazione del monitoraggio dettagliato (AWS CLI)

Per impostazione predefinita, il monitoraggio base viene abilitato al momento della creazione di un modello di avvio utilizzando il parametro AWS CLI. Il monitoraggio dettagliato viene abilitato per impostazione predefinita al momento della creazione di una configurazione di avvio tramite la AWS CLI.

Come abilitare il monitoraggio dettagliato in un modello di avvio

Per i modelli di avvio, utilizza il comando [create-launch-template](#) e trasferisci un file JSON contenente le informazioni per la creazione del modello di avvio. Per abilitare rispettivamente il monitoraggio dettagliato e il monitoraggio base, imposta l'attributo di monitoraggio su "Monitoring": {"Enabled":true} o su "Monitoring":{"Enabled":false}.

Come abilitare il monitoraggio dettagliato in una configurazione di avvio

Utilizzare il comando [create-launch-configuration](#) con l'opzione `--instance-monitoring` per le configurazioni di avvio. Imposta questa opzione su `true` o `false` per abilitare rispettivamente il monitoraggio dettagliato e il monitoraggio base.

```
--instance-monitoring Enabled=true
```

Passaggio tra il monitoraggio di base e quello dettagliato

Per modificare il tipo di monitoraggio abilitato su nuove istanze EC2, aggiorna il modello di avvio o il gruppo con scalabilità automatica affinché utilizzi una nuova configurazione di avvio. Le istanze esistenti continuano a utilizzare il tipo di monitoraggio precedentemente abilitato. Per aggiornare tutte le istanze, terminale in modo che vengano sostituite dal gruppo con scalabilità automatica o aggiornale singolarmente utilizzando i comandi [monitor-instances](#) e [unmonitor-instances](#).

Note

Con le funzioni di aggiornamento dell'istanza e di durata massima dell'istanza, è inoltre possibile sostituire tutte le istanze nel gruppo con scalabilità automatica per avviare nuove istanze che utilizzano la nuova configurazione. Per ulteriori informazioni, consulta [Ricicla le istanze del gruppo con dimensionamento automatico](#).

Quando passi dal monitoraggio di base a quello dettagliato:

Se disponi di allarmi CloudWatch associati alle policy di dimensionamento in corso o alle policy di dimensionamento semplice per il tuo gruppo con dimensionamento automatico, utilizza il comando [put-metric-alarm](#) per aggiornare ogni allarme. Fai in modo che ogni periodo corrisponda al tipo di monitoraggio (300 secondi per il monitoraggio base e 60 secondi per il monitoraggio dettagliato). Se passi dal monitoraggio dettagliato al monitoraggio base, ma non aggiorni gli allarmi in modo che corrispondano al periodo di cinque minuti, essi continuano a controllare le statistiche ogni minuto. Potrebbero non trovare dati disponibili per quattro periodi ogni cinque.

Raccolta di parametri aggiuntivi tramite l'agente CloudWatch

Per raccogliere i parametri a livello di sistema operativo, come la memoria disponibile e utilizzata, è necessario installare l'agente CloudWatch. Potrebbe essere applicato un costo aggiuntivo. Puoi utilizzare l'agente CloudWatch per raccogliere parametri del sistema operativo e file di log da Amazon EC2. Per ulteriori informazioni, consulta [Parametri raccolti dall'agente CloudWatch](#) nella Guida per l'utente di Amazon CloudWatch.

Registra le chiamate API di Amazon EC2 Auto Scaling con AWS CloudTrail

Amazon EC2 Auto Scaling è integrato AWS CloudTrail con un servizio che fornisce un registro delle azioni intraprese da un utente, ruolo o servizio utilizzando Amazon EC2 Auto Scaling. CloudTrail acquisisce tutte le chiamate API per Amazon EC2 Auto Scaling come eventi. Le chiamate acquisite includono chiamate dalla console Amazon EC2 Auto Scaling e chiamate di codice alle API Amazon EC2 Auto Scaling.

Se crei un trail, puoi abilitare la distribuzione continua di CloudTrail eventi a un bucket Amazon S3, inclusi gli eventi per Amazon EC2 Auto Scaling. Se non configuri un percorso, puoi comunque visualizzare gli eventi più recenti nella CloudTrail console nella cronologia degli eventi. Utilizzando le informazioni raccolte da CloudTrail, è possibile determinare la richiesta effettuata ad Amazon EC2 Auto Scaling, l'indirizzo IP da cui è stata effettuata la richiesta, chi ha effettuato la richiesta, quando è stata effettuata e dettagli aggiuntivi.

Per ulteriori informazioni CloudTrail, consulta la Guida per l'[AWS CloudTrail utente](#).

Informazioni su Amazon EC2 Auto Scaling in CloudTrail

CloudTrail è abilitato sul tuo account Amazon Web Services al momento della creazione dell'account. Quando si verifica un'attività in Amazon EC2 Auto Scaling, tale attività viene registrata in CloudTrail un evento insieme ad altri eventi di Amazon Web Services nella cronologia degli eventi. Puoi visualizzare, cercare e scaricare gli eventi recenti nell'account Amazon Web Services. Per ulteriori informazioni, consulta [Visualizzazione degli eventi con CloudTrail cronologia degli eventi](#).

Per una registrazione continua degli eventi nell'account Amazon Web Services che includa gli eventi per Amazon EC2 Auto Scaling, crea un percorso. Un trail consente di CloudTrail inviare file di log a un bucket Amazon S3. Per impostazione predefinita, quando si crea un percorso nella console, questo si applicherà a tutte le Regioni. Il percorso registra gli eventi di tutte le Regioni nella partizione Amazon Web Services e distribuisce i file di log nel bucket Amazon S3 specificato. Inoltre, puoi configurare altri Amazon Web Services per analizzare ulteriormente e agire in base ai dati sugli eventi raccolti nei CloudTrail log. Per ulteriori informazioni, consulta gli argomenti seguenti:

- [Panoramica della creazione di un percorso](#)
- [CloudTrail servizi e integrazioni supportati](#)
- [Configurazione delle notifiche Amazon SNS per CloudTrail](#)

- [Ricezione di file di CloudTrail registro da più regioni](#) e [ricezione di file di CloudTrail registro da più account](#)

Tutte le azioni di Amazon EC2 Auto Scaling vengono registrate e CloudTrail documentate nell'[Amazon EC2 Auto Scaling API Reference](#). Ad esempio, le chiamate a `CreateLaunchConfigurationDescribeAutoScalingGroup`, e le `UpdateAutoScalingGroup`azioni generano voci nei file di registro. CloudTrail

Ogni evento o voce di log contiene informazioni sull'utente che ha generato la richiesta. Le informazioni di identità consentono di determinare quanto segue:

- Se la richiesta è stata effettuata con credenziali utente root o AWS Identity and Access Management (IAM).
- Se la richiesta è stata effettuata con le credenziali di sicurezza temporanee per un ruolo o un utente federato.
- Se la richiesta è stata effettuata da un altro servizio .

Per ulteriori informazioni, vedete l'[CloudTrailuserIdentityelemento](#).

Comprendere delle voci di file di log di Amazon EC2 Auto Scaling

Un trail è una configurazione che consente la distribuzione di eventi come file di log in un bucket Amazon S3 specificato dall'utente. CloudTrail i file di registro contengono una o più voci di registro. Un evento rappresenta una singola richiesta proveniente da qualsiasi fonte e include informazioni sull'azione richiesta, la data e l'ora dell'azione, i parametri della richiesta e così via. CloudTrail i file di registro non sono una traccia ordinata dello stack delle chiamate API pubbliche, quindi non vengono visualizzati in un ordine specifico.

L'esempio seguente mostra una voce di CloudTrail registro che illustra l'`CreateLaunchConfiguration`azione.

```
{
  "eventVersion": "1.05",
  "userIdentity": {
    "type": "Root",
    "principalId": "123456789012",
    "arn": "arn:aws:iam::123456789012:root",
    "accountId": "123456789012",
    "accessKeyId": "AKIAIOSFODNN7EXAMPLE",
```

```
    "sessionContext": {
      "attributes": {
        "mfaAuthenticated": "false",
        "creationDate": "2018-08-21T17:05:42Z"
      }
    }
  },
  "eventTime": "2018-08-21T17:07:49Z",
  "eventSource": "autoscaling.amazonaws.com",
  "eventName": "CreateLaunchConfiguration",
  "awsRegion": "us-west-2",
  "sourceIPAddress": "192.0.2.0",
  "userAgent": "Coral/Jakarta",
  "requestParameters": {
    "ebsOptimized": false,
    "instanceMonitoring": {
      "enabled": false
    },
    "instanceType": "t2.micro",
    "keyName": "EC2-key-pair-oregon",
    "blockDeviceMappings": [
      {
        "deviceName": "/dev/xvda",
        "ebs": {
          "deleteOnTermination": true,
          "volumeSize": 8,
          "snapshotId": "snap-01676e0a2c3c7de9e",
          "volumeType": "gp2"
        }
      }
    ],
    "launchConfigurationName": "launch_configuration_1",
    "imageId": "ami-6cd6f714d79675a5",
    "securityGroups": [
      "sg-00c429965fd921483"
    ]
  },
  "responseElements": null,
  "requestID": "0737e2ea-fb2d-11e3-bfd8-99133058e7bb",
  "eventID": "3fcfb182-98f8-4744-bd45-b38835ab61cb",
  "eventType": "AwsApiCall",
  "recipientAccountId": "123456789012"
}
```


Risorse correlate

Con CloudWatch Logs, è possibile monitorare e ricevere avvisi per eventi specifici acquisiti da CloudTrail. Gli eventi inviati a CloudWatch Logs sono quelli configurati per essere registrati dal tuo percorso, quindi assicurati di aver configurato il percorso o i percorsi per registrare i tipi di eventi che ti interessa monitorare. CloudWatch Logs possono monitorare le informazioni nei file di registro e avvisare l'utente quando vengono raggiunte determinate soglie. Puoi inoltre archiviare i dati del log in storage estremamente durevole. Per ulteriori informazioni, consulta la [Amazon CloudWatch Logs User Guide](#) e l'argomento [Monitoring CloudTrail log files with Amazon CloudWatch Logs](#) nella AWS CloudTrail User Guide.

Opzioni di notifica Amazon SNS per Amazon EC2 Auto Scaling

Puoi configurare il tuo gruppo Auto Scaling per avvisarti di eventi importanti che influiscono sulla tua applicazione. Con le notifiche, puoi anche eliminare il polling e non riscontrerai l'RequestLimitExceeded errore che a volte deriva dal polling.

Esistono due modi per ricevere notifiche su Amazon EC2 Auto Scaling:

- Amazon Simple Notification Service: Amazon SNS può avvisarti quando il tuo gruppo Auto Scaling avvia o chiude le istanze. Le notifiche Amazon SNS possono essere solo attivate o disattivate. Per ulteriori informazioni, consulta [Amazon SNS e Amazon EC2 Auto Scaling](#).
- Amazon EventBridge: EventBridge fornisce notifiche più avanzate, basate sugli eventi, corrispondenti a criteri specifici e inviate a una varietà di destinazioni, tra cui Amazon SNS. EventBridge può anche monitorare una gamma più ampia di eventi di Auto Scaling per un monitoraggio più preciso. Per ulteriori informazioni, consulta [Utilizzato EventBridge per gestire gli eventi di Auto Scaling](#).

Puoi anche eseguire un'azione personalizzata quando un'istanza entra in uno stato in sospeso durante l'avvio o la chiusura utilizzando gli hook e i servizi del ciclo di vita come Amazon EventBridge, SNS e Amazon SQS. I Lifecycle Hook possono inoltre fornire più tempo a una nuova istanza per completare uno script specificato nei dati utente prima che Amazon EC2 Auto Scaling aggiunga l'istanza al gruppo. Per ulteriori informazioni, consulta [Hook del ciclo di vita di Amazon EC2 Auto Scaling](#).

Amazon SNS e Amazon EC2 Auto Scaling

Questa sezione mostra come utilizzare Amazon SNS per monitorare quando il gruppo Auto Scaling avvia o chiude le istanze.

Ad esempio, se configuri il gruppo con scalabilità automatica affinché utilizzi il tipo di notifica `autoscaling: EC2_INSTANCE_TERMINATE` e il gruppo con scalabilità automatica termina un'istanza, viene inviata una notifica e-mail. Questa e-mail contiene i dettagli dell'istanza terminata, ad esempio il suo ID e il motivo della terminazione.

Tieni presente che quando Amazon EC2 Auto Scaling aggiunge o rimuove istanze dal gruppo, ti vengono inviate notifiche relative a queste modifiche, con una notifica inviata per istanza. Tuttavia, l'invio di queste notifiche avviene con la massima diligenza possibile e le istanze potrebbero comunque fallire dopo la notifica iniziale, ad esempio, se un successivo controllo dello stato di salute fallisce. Quindi, anche se Amazon EC2 Auto Scaling ti avvisa inizialmente, un'istanza potrebbe comunque fallire in seguito. Tieni presente che puoi configurare il tempo di attesa dopo il lancio di un'istanza Amazon EC2 Auto Scaling prima di eseguire il primo controllo di integrità. Per ulteriori informazioni, consulta [Impostazione del periodo di tolleranza del controllo dell'integrità per un gruppo con scalabilità automatica](#).

Per ulteriori informazioni su Amazon SNS in generale, consulta la [Amazon Simple Notification Service Developer Guide](#).

Indice

- [Notifiche SNS](#)
- [Configurare le notifiche Amazon SNS per Amazon EC2 Auto Scaling](#)
 - [Creare un argomento Amazon SNS](#)
 - [Sottoscrizione all'argomento Amazon SNS](#)
 - [Conferma la sottoscrizione ad Amazon SNS](#)
 - [Configurazione del gruppo con scalabilità automatica per l'invio di notifiche](#)
 - [Test della notifica](#)
 - [Eliminazione della configurazione delle notifiche](#)
- [Policy della chiave per un argomento Amazon SNS crittografato](#)

Notifiche SNS

Amazon EC2 Auto Scaling supporta l'invio di notifiche da Amazon SNS quando si verificano i seguenti eventi.

Evento	Descrizione
autoscaling:EC2_INSTANCE_LAUNCH	Avvio dell'istanza riuscito
autoscaling:EC2_INSTANCE_LAUNCH_ERROR	Avvio dell'istanza non riuscito
autoscaling:EC2_INSTANCE_TERMINATE	Terminazione dell'istanza riuscita
autoscaling:EC2_INSTANCE_TERMINATE_ERROR	Terminazione dell'istanza non riuscita

Il messaggio include le informazioni seguenti:

- **Event:** l'evento.
- **AccountId:** l'ID account Amazon Web Services.
- **AutoScalingGroupName:** il nome del gruppo con dimensionamento automatico.
- **AutoScalingGroupARN:** l'ARN del gruppo con scalabilità automatica.
- **EC2InstanceId:** l'ID dell'istanza EC2.

Ad esempio:

```
Service: AWS Auto Scaling
Time: 2016-09-30T19:00:36.414Z
RequestId: 4e6156f4-a9e2-4bda-a7fd-33f2ae528958
Event: autoscaling:EC2_INSTANCE_LAUNCH
AccountId: 123456789012
AutoScalingGroupName: my-asg
AutoScalingGroupARN: arn:aws:autoscaling:region:123456789012:autoScalingGroup...
ActivityId: 4e6156f4-a9e2-4bda-a7fd-33f2ae528958
Description: Launching a new EC2 instance: i-0598c7d356eba48d7
```

```
Cause: At 2016-09-30T18:59:38Z a user request update of AutoScalingGroup constraints
to ...
StartTime: 2016-09-30T19:00:04.445Z
EndTime: 2016-09-30T19:00:36.414Z
StatusCode: InProgress
StatusMessage:
Progress: 50
EC2InstanceId: i-0598c7d356eba48d7
Details: {"Subnet ID":"subnet-id","Availability Zone":"zone"}
Origin: AutoScalingGroup
Destination: EC2
```

Configurare le notifiche Amazon SNS per Amazon EC2 Auto Scaling

Per usare Amazon SNS per l'invio di notifiche e-mail, devi prima creare un argomento, quindi iscrivere i tuoi indirizzi e-mail all'argomento.

Creare un argomento Amazon SNS.

Un argomento SNS è un punto di accesso logico, un canale di comunicazione che il gruppo con scalabilità automatica utilizza per inviare le notifiche. Puoi creare un argomento specificando un nome.

Quando crei un nome per l'argomento SNS, il nome deve soddisfare i seguenti requisiti:

- Deve contenere da 1 a 256 caratteri.
- Deve contenere caratteri ASCII maiuscoli e minuscoli, numeri, trattini e caratteri di sottolineatura.

Per le istruzioni, consulta [Creazione di un argomento Amazon SNS](#) nella Guida per lo Sviluppatore di Amazon Simple Notification Service.

Sottoscrizione all'argomento Amazon SNS

Per ricevere le notifiche che il gruppo con scalabilità automatica invia all'argomento, è necessario sottoscrivere un endpoint all'argomento. In questa procedura, come Endpoint, specifica l'indirizzo e-mail a cui desideri ricevere le notifiche da Amazon EC2 Auto Scaling.

Per ulteriori informazioni, consulta [Sottoscrizione a un argomento di Amazon SNS](#) nella Guida per lo Sviluppatore di Amazon Simple Notification Service.

Conferma la sottoscrizione ad Amazon SNS

Amazon SNS invia un'e-mail di conferma all'indirizzo specificato nella fase precedente.

Prima di passare alla fase successiva, assicurati di aver aperto il messaggio e-mail da AWS Notifications (Notifiche AWS), poi scegli il link per confermare la sottoscrizione.

Riceverai un messaggio di conferma da. AWS Amazon SNS è ora configurato per ricevere notifiche e inviare la notifica come e-mail all'indirizzo specificato.

Configurazione del gruppo con scalabilità automatica per l'invio di notifiche

Puoi configurare il gruppo con scalabilità automatica per l'invio di notifiche ad Amazon SNS quando avviene un evento di dimensionamento, ad esempio l'avvio o la terminazione delle istanze. Amazon SNS invia una notifica con informazioni sulle istanze all'indirizzo e-mail specificato.

Configurazione delle notifiche Amazon SNS per il gruppo con scalabilità automatica (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si apre un riquadro diviso nella parte inferiore della pagina, con le informazioni sul gruppo selezionato.

3. Nella scheda Activity (Attività), Activity notifications (Attività notifiche), scegli Create notification (Crea notifiche).
4. Nel riquadro Create notifications (Crea notifiche), procedi come segue:
 - a. In SNS Topic (Argomento SNS), seleziona l'argomento SNS.
 - b. In Event types (Tipi di eventi), seleziona gli eventi per cui inviare le notifiche.
 - c. Scegli Crea.

Configurazione delle notifiche Amazon SNS per il gruppo con scalabilità automatica (AWS CLI)

Utilizza il seguente comando [put-notification-configuration](#).

```
aws autoscaling put-notification-configuration --auto-scaling-group-name my-  
asg --topic-arn arn --notification-types "autoscaling:EC2_INSTANCE_LAUNCH"  
"autoscaling:EC2_INSTANCE_TERMINATE"
```

Test della notifica

Per generare una notifica per un evento di avvio, aggiorna il gruppo con scalabilità automatica aumentandone di 1 la capacità desiderata. Riceverai una notifica entro qualche minuto dall'avvio dell'istanza.

Come modificare la capacità desiderata (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Viene visualizzato un riquadro diviso nella parte inferiore della pagina Auto Scaling Groups (Gruppi Auto Scaling) in cui vengono visualizzate le informazioni sul gruppo selezionato.

3. Nella scheda Details (Dettagli) scegliere Group details (Dettagli gruppo), Edit (Modifica).
4. Per Desired capacity (Capacità desiderata), aumentare il valore corrente di 1. Se questo valore supera la capacità massima, dovrai aumentare di un'unità anche il valore di Maximum capacity (Capacità massima).
5. Scegli Aggiorna.
6. Dopo qualche minuto, si riceve la notifica dell'evento. Se l'istanza aggiuntiva avviata per questo test non è necessaria, puoi ridurre di 1 il valore di Desired capacity (Capacità desiderata). Dopo qualche minuto, si riceve la notifica dell'evento.

Eliminazione della configurazione delle notifiche

Se non utilizzi più la configurazione delle notifiche di Amazon EC2 Auto Scaling, puoi eliminarla.

Come eliminare la configurazione delle notifiche Amazon EC2 Auto Scaling (console)

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona il gruppo con scalabilità automatica.
3. In Activity (Attività), seleziona la casella di controllo accanto alla notifica che desideri eliminare e quindi scegli Actions (Operazioni), Delete (Elimina).

Come eliminare la configurazione delle notifiche Amazon EC2 Auto Scaling (AWS CLI)

Utilizza il seguente comando delete-notification-configuration.

```
aws autoscaling delete-notification-configuration --auto-scaling-group-name my-asg --  
topic-arn arn
```

Per ulteriori informazioni sull'eliminazione dell'argomento Amazon SNS e di tutte le iscrizioni associate al gruppo con scalabilità automatica, consulta [Eliminazione di una sottoscrizione e di un argomento Amazon SNS](#) nella Guida per gli sviluppatori di Amazon Simple Notification Service.

Policy della chiave per un argomento Amazon SNS crittografato

L'argomento Amazon SNS specificato potrebbe essere crittografato con una chiave gestita dal cliente creata con AWS Key Management Service. Per autorizzare Dimensionamento automatico Amazon EC2 a pubblicare su argomenti crittografati, devi prima creare la tua chiave KMS e quindi aggiungere la seguente dichiarazione alla policy della chiave KMS. Sostituisci l'ARN di esempio con l'ARN del ruolo collegato ai servizi appropriato, cioè l'accesso permesso alla chiave. Per ulteriori informazioni, consulta [Configura le autorizzazioni AWS KMS](#) nella Guida per gli sviluppatori di Amazon Simple Notification Service.

In questo esempio, l'informativa sulla politica fornisce al ruolo collegato al servizio denominato le `AWSServiceRoleForAutoScaling` autorizzazioni per utilizzare la chiave gestita dal cliente. Per ulteriori informazioni sul ruolo collegato ai servizi di Dimensionamento automatico Amazon EC2, consulta [Ruoli collegati ai servizi per Dimensionamento automatico Amazon EC2](#).

```
{  
  "Sid": "Allow service-linked role use of the customer managed key",  
  "Effect": "Allow",  
  "Principal": {  
    "AWS": "arn:aws:iam::123456789012:role/aws-service-role/autoscaling.amazonaws.com/  
AWSServiceRoleForAutoScaling"  
  },  
  "Action": [  
    "kms:GenerateDataKey*",  
    "kms:Decrypt"  
  ],  
  "Resource": "*" }  
}
```

Le chiavi di condizione `aws:SourceArn` e `aws:SourceAccount` non sono supportate nelle policy chiave che consentono a Dimensionamento automatico Amazon EC2 di pubblicare su argomenti crittografati.

AWS servizi integrati con Amazon EC2 Auto Scaling

Amazon EC2 Auto Scaling può essere integrato con altri servizi. AWS Consulta le seguenti opzioni di integrazione per ulteriori informazioni sul funzionamento di ciascun servizio con Dimensionamento automatico Amazon EC2.

Argomenti

- [Utilizza il ribilanciamento della capacità per gestire le interruzioni spot di Amazon EC2](#)
- [Utilizza le prenotazioni della capacità on demand per prenotare la capacità in zone di disponibilità specifiche](#)
- [Crea gruppi di Auto Scaling dalla riga di comando usando AWS CloudShell](#)
- [Creazione di gruppi Auto Scaling con AWS CloudFormation](#)
- [AWS Compute Optimizer Utilizzatelo per ottenere consigli sul tipo di istanza per un gruppo di Auto Scaling](#)
- [Utilizzare sistema di bilanciamento del carico elastico per distribuire il traffico tra le istanze nel gruppo con dimensionamento automatico.](#)
- [Indirizza il traffico verso il tuo gruppo con dimensionamento automatico con un gruppo target VPC Lattice](#)
- [Utilizzato EventBridge per gestire gli eventi di Auto Scaling](#)
- [Fornisci connettività di rete per le istanze Auto Scaling utilizzando Amazon VPC](#)

Utilizza il ribilanciamento della capacità per gestire le interruzioni spot di Amazon EC2

Puoi configurare Amazon EC2 Auto Scaling per monitorare e rispondere automaticamente alle modifiche che influiscono sulla disponibilità delle istanze Spot. Il ribilanciamento della capacità consente di mantenere la disponibilità del carico di lavoro aumentando proattivamente il parco istanze con una nuova istanza spot prima che un'istanza in esecuzione venga interrotta da Amazon EC2.

L'obiettivo del ribilanciamento della capacità è di mantenere l'elaborazione del carico di lavoro senza interruzioni. Quando le istanze Spot sono ad alto rischio di interruzione, il servizio Spot Amazon EC2 notifica ad Amazon EC2 Auto Scaling un suggerimento di ribilanciamento dell'istanza EC2.

Quando abiliti il ribilanciamento della capacità per il gruppo con scalabilità automatica, Amazon EC2 Auto Scaling cerca di sostituire proattivamente le istanze Spot del gruppo per cui è suggerito il ribilanciamento. È possibile decidere di ribilanciare il carico di lavoro su Istanze Spot nuove o esistenti che non presentano un rischio elevato di interruzione. Ciò significa che il carico di lavoro può continuare a elaborare le operazioni, mentre Amazon EC2 Auto Scaling avvia una nuova Istanza Spot prima che un'istanza esistente venga interrotta.

Quando non utilizzi il ribilanciamento della capacità, Dimensionamento automatico Amazon EC2 non sostituisce le istanze Spot dopo fino a quando il servizio Amazon EC2 Spot non interrompe le istanze e il controllo dell'integrità fallisce. Prima di interrompere un'istanza, Amazon EC2 dà sempre sia una raccomandazione di ribilanciamento dell'istanza EC2 che un avviso di interruzione dell'istanza Spot di due minuti.

Indice

- [Panoramica](#)
- [Comportamento di ribilanciamento della capacità](#)
- [Considerazioni](#)
- [Abilitazione del ribilanciamento della capacità \(console\)](#)
- [Abilitare il ribilanciamento della capacità. \(AWS CLI\)](#)
- [Risorse correlate](#)
- [Limitazioni](#)

Panoramica

Per utilizzare il ribilanciamento della capacità con il gruppo con dimensionamento automatico, i passaggi di base sono:

1. Configurare l'utilizzo di più tipi di istanze e zone di disponibilità da parte del tuo gruppo con dimensionamento automatico. In questo modo, Dimensionamento automatico Amazon EC2 può esaminare la capacità disponibile per le istanze Spot in ogni zona di disponibilità. Per ulteriori informazioni, consulta [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#).
2. Se necessario, aggiungi gli hook del ciclo di vita per eseguire un arresto graduale dell'applicazione all'interno delle istanze che ricevono la notifica di ribilanciamento. Per ulteriori informazioni, consulta [Hook del ciclo di vita di Amazon EC2 Auto Scaling](#).

Di seguito sono riportati alcuni motivi per cui è possibile utilizzare un hook del ciclo di vita:

- Permettere l'arresto regolare dei dipendenti Amazon SQS
 - Completare la deregistrazione dal Domain Name System (DNS)
 - Eseguire il pull o i log dell'applicazione e caricarli su Amazon Simple Storage Service (Amazon S3)
3. Sviluppa un'azione personalizzata per l'hook del ciclo di vita. Per richiamare l'azione personalizzata il prima possibile, devi sapere quando un'istanza è pronta per essere terminata. Scoprilo rilevando lo stato del ciclo di vita dell'istanza.
- Per richiamare un'azione all'esterno dell'istanza, scrivi una EventBridge regola e automatizza l'azione da intraprendere quando un modello di evento corrisponde alla regola.
 - Per richiamare un'azione all'interno dell'istanza, configura l'istanza per eseguire uno script di spegnimento e recuperare lo stato del ciclo di vita tramite i metadati dell'istanza.

È fondamentale progettare l'azione personalizzata in modo che venga completata in meno di due minuti. In questo modo si garantisce che ci sia abbastanza tempo per completare le attività prima della chiusura dell'istanza.

Dopo aver completato questi passaggi, puoi iniziare a utilizzare il ribilanciamento della capacità.

Comportamento di ribilanciamento della capacità

Con il ribilanciamento della capacità, Dimensionamento automatico Amazon EC2 si comporta nel modo seguente quando un'istanza riceve una raccomandazione di ribilanciamento della capacità:

- Quando avvii la nuova istanza spot, Dimensionamento automatico Amazon EC2 attende che la nuova istanza superi il controllo dell'integrità, prima di terminare quella vecchia. Quando sostituisci più istanze, la terminazione di ogni istanza precedente inizia dopo l'avvio della nuova istanza e il controllo dell'integrità.
- Poiché Amazon EC2 Auto Scaling tenta di avviare nuove istanze prima di terminare quelle vecchie, se si è raggiunta o si sta per raggiungere la capacità massima specificata potrebbe ostacolare o bloccare completamente le attività di ribilanciamento. Per evitare questo problema, Dimensionamento automatico Amazon EC2 può temporaneamente superare la dimensione massima del gruppo fino al 10% della capacità desiderata.
- Se non hai aggiunto un hook del ciclo di vita, il tuo gruppo con dimensionamento automatico Amazon EC2 inizia a terminare le istanze precedenti non appena le nuove istanze superano il controllo dell'integrità.

- Aggiungendo un hook del ciclo di vita aumenti il tempo necessario prima di iniziare a terminare le istanze precedenti in base al valore di timeout specificato per l'hook del ciclo di vita.
- Se si utilizzano policy di dimensionamento o di dimensionamento pianificato, le attività di dimensionamento vengono eseguite in parallelo. Se è in corso un'attività di dimensionamento e il gruppo con scalabilità automatica è inferiore alla nuova capacità desiderata, le dimensioni di Amazon EC2 Auto Scaling vengono aumentate orizzontalmente prima che le istanze precedenti vengano terminate.

Se non è presente alcuna capacità per i tipi di istanza in una zona di disponibilità non riesce, Dimensionamento automatico Amazon EC2 continua a tentare di avviare istanze Spot nelle zone di disponibilità specificate fino a quando non riesce.

Nella peggiore delle ipotesi, se le nuove istanze non vengono avviate o vengono avviate ma il controllo dell'integrità non riesce, Dimensionamento automatico Amazon EC2 continua a tentare di riavviarle. Mentre cerca di avviare nuove istanze, quelle vecchie vengono infine interrotte e terminate forzatamente con un avviso di interruzione di due minuti.

Considerazioni

Quando utilizzi il ribilanciamento della capacità, considera quanto segue:

Progetta la tua applicazione in modo che sia tollerante alle interruzioni Spot

L'applicazione dovrebbe essere in grado di gestire modifiche dinamiche nel numero delle istanze e la possibilità che un'istanza spot venga interrotta in anticipo. Ad esempio, se il tuo gruppo con dimensionamento automatico si trova dietro un sistema di bilanciamento del carico di Elastic Load Balancing, Dimensionamento automatico Amazon EC2 attende che l'istanza deregistrata dal load balancer prima di chiamare l'hook del ciclo di vita. Se scaricare le connessioni e completare l'operazione del ciclo di vita richiede troppo tempo, l'istanza potrebbe essere interrotta mentre Dimensionamento automatico Amazon EC2 è in attesa della sua terminazione prima della terminazione dell'istanza.

Per Amazon EC2 non è sempre possibile inviare il segnale di raccomandazione per il ribilanciamento prima dell'avviso di interruzione dell'istanza spot di due minuti. Talvolta, il segnale di raccomandazione di ribilanciamento arriva insieme all'avviso di interruzione di due minuti. In questo caso, Dimensionamento automatico Amazon EC2 chiama l'hook del ciclo di vita e tenta di avviare immediatamente una nuova istanza Spot.

Evitare un rischio elevato di interruzione delle istanze spot sostitutive

Le Istanze spot sostitutive possono comportare un elevato rischio di interruzione se si utilizza la strategia di allocazione `lowest-price`. Questo perché avviamo le istanze nel pool con capacità disponibile al prezzo più basso in quel momento, anche se è probabile che le istanze Spot sostitutive vengano interrotte subito dopo l'avvio. Per evitare un rischio elevato di interruzioni, è vivamente consigliato di non utilizzare la strategia di allocazione `lowest-price`. Consigliamo invece la strategia di allocazione `price-capacity-optimized`. Questa strategia avvia istanze Spot sostitutive nei pool Spot che hanno il prezzo più basso possibile e meno probabilità di interruzioni. Pertanto, è meno probabile che vengano interrotti nelle prossime future.

Dimensionamento automatico Amazon EC2 avvierà una nuova istanza solo se la disponibilità è uguale o migliore

Uno degli obiettivi del ribilanciamento della capacità è migliorare la disponibilità di un'istanza spot. Se un'istanza spot esistente riceve una raccomandazione di ribilanciamento, Dimensionamento automatico Amazon EC2 avvierà una nuova istanza solo se la nuova istanza fornisce una disponibilità uguale o migliore rispetto all'istanza esistente. Se il rischio di interruzione è maggiore dell'istanza esistente, Dimensionamento automatico Amazon EC2 non avvierà una nuova istanza. Tuttavia, Dimensionamento automatico Amazon EC2 continuerà a valutare i pool di capacità Spot sulla base delle informazioni fornite dal servizio Amazon EC2 Spot e avvierà una nuova istanza se la disponibilità migliorerà.

È possibile che l'istanza esistente venga interrotta senza che Dimensionamento automatico Amazon EC2 avvii in modo proattivo una nuova istanza. In questo caso, Dimensionamento automatico Amazon EC2 tenterà di avviare una nuova istanza non appena riceve l'avviso di interruzione dell'istanza spot. Ciò accade indipendentemente dal fatto che la nuova istanza presenta un rischio elevato di interruzione.

Il ribilanciamento della capacità non aumenta il tasso di interruzione dell'istanza Spot

Quando si abilita il ribilanciamento della capacità, non aumenta il [tasso di interruzione dell'istanza spot](#) (il numero di istanze Spot che vengono recuperate quando Amazon EC2 ha bisogno di capacità). Tuttavia, se il ribilanciamento della capacità rileva che un'istanza è a rischio di interruzione, Dimensionamento automatico Amazon EC2 proverà immediatamente ad avviare una nuova istanza. Pertanto, potrebbero essere sostituite più istanze che se avessi aspettato che Dimensionamento automatico Amazon EC2 avesse avviato una nuova istanza dopo l'interruzione di quella a rischio.

Sebbene sia possibile sostituire più istanze mediante l'abilitazione del ribilanciamento della capacità, è meglio prendersi più tempo. In questo modo avrai più tempo per agire prima che le istanze vengano interrotte. Con un [Avviso di interruzione dell'istanza Spot](#), in genere hai solo fino a due minuti per interrompere l'istanza. Con il ribilanciamento della capacità che avvia una nuova istanza in anticipo, offri ai processi esistenti maggiori possibilità di completamento sull'istanza a rischio. È inoltre possibile avviare le procedure di chiusura dell'istanza, impedire la pianificazione di nuovi lavori sull'istanza a rischio e preparare l'istanza appena avviata per il controllo dell'applicazione. Con la sostituzione proattiva offerta dal ribilanciamento della capacità, puoi beneficiare di una continuità regolare.

Il seguente esempio dimostra i rischi e i benefici dell'utilizzo del ribilanciamento della capacità:

- 14:00: viene ricevuto un suggerimento di ribilanciamento per l'istanza A e Dimensionamento automatico Amazon EC2 inizia immediatamente a provare ad avviare un'istanza sostitutiva B, dandoti il tempo di iniziare le procedure di arresto.
- 14:30: viene ricevuto un suggerimento di ribilanciamento per l'istanza B, sostituita con l'istanza C, dandoti il tempo di iniziare le procedure di arresto.
- 14:32: se il ribilanciamento della capacità non fosse abilitato e se un avviso di interruzione dell'istanza Spot fosse stato ricevuto alle 14:32 per l'istanza A, avresti avuto solo fino a due minuti per agire. Tuttavia, l'istanza A avrebbe continuato a funzionare fino a quel momento.

Abilitazione del ribilanciamento della capacità (console)

È possibile abilitare o disabilitare il ribilanciamento della capacità durante la creazione o l'aggiornamento di un gruppo con scalabilità automatica.

Come abilitare il ribilanciamento della capacità per un nuovo gruppo con scalabilità automatica

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Selezionare Create Auto Scaling group (Crea un gruppo con scalabilità automatica).
3. Per il Passaggio 1: Scegli il modello o la configurazione di avvio, inserisci un nome per il gruppo con dimensionamento automatico, quindi scegli un modello di avvio e Avanti per andare al passaggio successivo.
4. Per il passaggio 2: scegli le opzioni di avvio dell'istanza, per i requisiti del tipo di istanza, scegli le impostazioni per creare un gruppo di istanze misto. Tra esse figurano i tipi di istanze che può avviare, le opzioni di acquisto delle istanze e le strategie di allocazione per le istanze Spot

e On-Demand. Queste impostazioni non sono configurate di default. Per configurarle, devi selezionare Override launch template (Sovrascrivi il modello di avvio). Per ulteriori informazioni sulla creazione di gruppi di istanze miste, consulta [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#).

5. In Rete, scegli le opzioni desiderate. Verifica che le sottoreti da utilizzare siano in diverse zone di disponibilità.
6. Nella sezione Strategie di allocazione, scegli una strategia di allocazione Spot. Per abilitare o disabilitare il ribilanciamento della capacità, seleziona o deseleziona la casella di controllo Ribilanciamento della capacità. Questa opzione viene visualizzata solo se si richiede una percentuale del gruppo con dimensionamento automatico da avviare come istanze Spot nella sezione Opzioni di acquisto delle istanze.
7. Crea il gruppo con scalabilità automatica.
8. (Facoltativo) Aggiungi gli hook del ciclo di vita in base alle esigenze. Per ulteriori informazioni, consulta [Aggiungere un hook del ciclo di vita](#).

Come abilitare o disabilitare il ribilanciamento della capacità per un gruppo con dimensionamento automatico esistente

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico. Si aprirà un riquadro diviso nella parte inferiore della pagina.
3. Nella tabella Details (Dettagli), scegli Allocation strategies (Strategie di allocazione), quindi Edit (Modifica).
4. Nella sezione Strategie di allocazione, abilita o disabilita il ribilanciamento della capacità selezionando o deselezionando la casella di controllo nel ribilanciamento della capacità.
5. Scegli Aggiorna.

Abilitare il ribilanciamento della capacità. (AWS CLI)

Gli esempi seguenti mostrano come utilizzare per abilitare e AWS CLI disabilitare il ribilanciamento della capacità.

Utilizzate il [update-auto-scaling-group](#) comando [create-auto-scaling-group](#) con il seguente parametro:

- `--capacity-rebalance/--no-capacity-rebalance`— Valore booleano che indica se il ribilanciamento della capacità è abilitato.

Prima di richiamare il [create-auto-scaling-group](#) comando, è necessario il nome di un modello di avvio configurato per l'uso con un gruppo Auto Scaling. Per ulteriori informazioni, consulta [Creare un modello di avvio per un gruppo con dimensionamento automatico](#).

Note

Le procedure seguenti illustrano come utilizzare un file di configurazione formattato in JSON o YAML. Se si utilizza la AWS CLI versione 1, è necessario specificare un file di configurazione in formato JSON. Se si utilizza la AWS CLI versione 2, è possibile specificare un file di configurazione formattato in YAML o JSON.

JSON

Creazione e configurazione di un nuovo gruppo con scalabilità automatica

- Utilizzate il seguente [create-auto-scaling-group](#) comando per creare un nuovo gruppo Auto Scaling e abilitare il ribilanciamento della capacità. Questo comando fa riferimento a un file JSON come unico parametro per il gruppo con dimensionamento automatico.

```
aws autoscaling create-auto-scaling-group --cli-input-json file://~/config.json
```

Se non disponi di un file di configurazione CLI che specifica una [policy di istanze miste](#), creane uno.

Aggiungi la seguente riga all'oggetto JSON di primo livello nel file di configurazione.

```
{  
  "CapacityRebalance": true  
}
```

Di seguito è riportato un esempio del file `config.json`.

```
{  
  "AutoScalingGroupName": "my-asg",  
  "DesiredCapacity": 12,  
}
```



```
"MinSize": 12,  
"MaxSize": 15,  
"CapacityRebalance": true,  
"MixedInstancesPolicy": {  
  "InstancesDistribution": {  
    "OnDemandBaseCapacity": 0,  
    "OnDemandPercentageAboveBaseCapacity": 25,  
    "SpotAllocationStrategy": "price-capacity-optimized"  
  },  
  "LaunchTemplate": {  
    "LaunchTemplateSpecification": {  
      "LaunchTemplateName": "my-launch-template",  
      "Version": "$Default"  
    },  
    "Overrides": [  
      {  
        "InstanceType": "c5.large"  
      },  
      {  
        "InstanceType": "c5a.large"  
      },  
      {  
        "InstanceType": "m5.large"  
      },  
      {  
        "InstanceType": "m5a.large"  
      },  
      {  
        "InstanceType": "c4.large"  
      },  
      {  
        "InstanceType": "m4.large"  
      },  
      {  
        "InstanceType": "c3.large"  
      },  
      {  
        "InstanceType": "m3.large"  
      }  
    ]  
  },  
  "TargetGroupARNs": "arn:aws:elasticloadbalancing:us-west-2:123456789012:targetgroup/my-alb-target-group/943f017f100becff",
```

```
"VPCZoneIdentifier": "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782"  
}
```

YAML

Creazione e configurazione di un nuovo gruppo con scalabilità automatica

- Utilizzate il seguente [create-auto-scaling-group](#) comando per creare un nuovo gruppo Auto Scaling e abilitare il ribilanciamento della capacità. Questo fa riferimento a un file YAML, non a un file YAML, come unico parametro per il gruppo con dimensionamento automatico.

```
aws autoscaling create-auto-scaling-group --cli-input-yaml file://~/config.yaml
```

Aggiungi la seguente riga al file di configurazione formattato in YAML.

```
CapacityRebalance: true
```

Di seguito è riportato un esempio del file config.yaml.

```
---  
AutoScalingGroupName: my-asg  
DesiredCapacity: 12  
MinSize: 12  
MaxSize: 15  
CapacityRebalance: true  
MixedInstancesPolicy:  
  InstancesDistribution:  
    OnDemandBaseCapacity: 0  
    OnDemandPercentageAboveBaseCapacity: 25  
    SpotAllocationStrategy: price-capacity-optimized  
  LaunchTemplate:  
    LaunchTemplateSpecification:  
      LaunchTemplateName: my-launch-template  
      Version: $Default  
    Overrides:  
      - InstanceType: c5.large  
      - InstanceType: c5a.large  
      - InstanceType: m5.large  
      - InstanceType: m5a.large  
      - InstanceType: c4.large
```

- InstanceType: *m4.large*
- InstanceType: *c3.large*
- InstanceType: *m3.large*

TargetGroupARNs:

- *arn:aws:elasticloadbalancing:us-west-2:123456789012:targetgroup/my-alb-target-group/943f017f100becff*

VPCZoneIdentifier: *subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782*

Come abilitare il ribilanciamento della capacità per un gruppo con scalabilità automatica esistente

- Utilizzare il [update-auto-scaling-group](#) comando seguente per abilitare il ribilanciamento della capacità.

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg \  
--capacity-rebalance
```

Come verificare che il ribilanciamento della capacità per un gruppo con scalabilità automatica sia abilitato

- Utilizzare il [describe-auto-scaling-groups](#) comando seguente per verificare che Capacity Rebalancing sia abilitato e per visualizzarne i dettagli.

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

Di seguito è riportata una risposta di esempio.

```
{  
  "AutoScalingGroups": [  
    {  
      "AutoScalingGroupName": "my-asg",  
      "AutoScalingGroupARN": "arn",  
      ...  
      "CapacityRebalance": true  
    }  
  ]  
}
```

Come disabilitare il ribilanciamento della capacità

Utilizzare il [update-auto-scaling-group](#) comando con l'`--no-capacity-rebalance` opzione per disabilitare Capacity Rebalancing.

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg \  
--no-capacity-rebalance
```

Risorse correlate

Per ulteriori informazioni sul Capacity Rebalancing, consulta [Gestisci in modo proattivo il ciclo di vita delle istanze Spot utilizzando la nuova funzionalità di ribilanciamento della capacità per EC2 Auto Scaling](#) sul blog di elaborazione. AWS

Per informazioni sugli eventi di suggerimento di ribilanciamento dell'istanza EC2, consulta [Suggerimenti per il ribilanciamento delle istanze EC2](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Per ulteriori informazioni sugli hook del ciclo di vita, consulta le seguenti risorse.

- [Tutorial : Configurazione di un hook del ciclo di vita che richiama una funzione Lambda EventBridge](#)(utilizzando)
- [Tutorial: configurazione dei dati utente per recuperare lo stato del ciclo di vita di destinazione tramite i metadati dell'istanza](#)

Limitazioni

- Dimensionamento automatico Amazon EC2 può sostituire l'istanza che riceve la notifica di ribilanciamento solo se l'istanza non è protetta dal dimensionamento verticale. Tuttavia, la protezione scalabile non impedisce l'interruzione dovuta a un'interruzione Spot. Per ulteriori informazioni, consulta [Utilizzare la protezione di riduzione orizzontale dell'istanza](#).
- Il supporto per il ribilanciamento della capacità è disponibile in tutte le Regioni AWS commerciali in cui è disponibile Dimensionamento automatico Amazon EC2, esclusa la regione Medio Oriente (EAU).

Utilizza le prenotazioni della capacità on demand per prenotare la capacità in zone di disponibilità specifiche

Le prenotazioni della capacità on demand Amazon EC2 consentono di prenotare la capacità di elaborazione in zone di disponibilità specifiche. Per iniziare a usare le Prenotazioni della capacità, crea una prenotazione della capacità in una zona di disponibilità specifica. Quindi, è possibile avviare istanze nella capacità riservata, visualizzarne la capacità di utilizzo in tempo reale e aumentarne o diminuirne la capacità in base alle esigenze.

Le prenotazioni delle capacità sono configurate come `open` o `targeted`. Se la Prenotazione della capacità è `open`, tutte le istanze nuove ed esistenti che hanno attributi corrispondenti vengono eseguiti automaticamente nella capacità della Prenotazione della capacità. Se la Prenotazione della capacità è `targeted`, le istanze devono specificamente puntarla per l'esecuzione nella capacità riservata.

Ad esempio, non è possibile creare un gruppo con dimensionamento automatico che specifichi più tipi di avvio `targeted`. In questo modo si ha un maggiore controllo su quando utilizzare specifiche prenotazioni della capacità.

I passaggi di base sono:

1. Crea prenotazioni della capacità in più zone di disponibilità con lo stesso tipo di istanza, piattaforma e numero di istanze.
2. Prenotazioni di capacità di gruppo utilizzando AWS Resource Groups.
3. Crea un gruppo con dimensionamento automatico con un modello di avvio destinato al gruppo di risorse, utilizzando le stesse zone di disponibilità delle prenotazioni della capacità.

Indice

- [Fase 1: creazione di prenotazioni della capacità](#)
- [Fase 2: creazione di un gruppo di prenotazione della capacità](#)
- [Fase 3: creazione di un modello di avvio](#)
- [Fase 4: creazione di un gruppo con dimensionamento automatico](#)
- [Risorse correlate](#)

Fase 1: creazione di prenotazioni della capacità

Il primo passaggio consiste nel creare una prenotazione della capacità in ogni zona di disponibilità in cui verrà distribuito il gruppo con dimensionamento automatico.

Note

È possibile creare prenotazioni `targeted` solo quando si creano per la prima volta le prenotazioni della capacità.

Console

Creazione di una prenotazione della capacità

1. Aprire la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Scegliere Prenotazioni della capacità, quindi Create Prenotazione di capacità (Crea Prenotazione di capacità).
3. Nella pagina Crea una Prenotazione della capacità, fare attenzione alle seguenti impostazioni nella sezione Dettagli istanza. Il tipo di istanza, piattaforma e la zona di disponibilità delle istanze lanciate devono corrispondere al tipo di istanza, piattaforma e zona di disponibilità qui specificata o la Prenotazione della capacità non viene applicata.
 - a. Tipo di istanza – Il tipo di istanza da avviare nella capacità riservata.
 - b. Per Platform, scegli il sistema operativo per le tue istanze.
 - c. Per Zona di disponibilità, scegli la prima zona di disponibilità in cui desideri prenotare la capacità.
 - d. Per la capacità totale, scegli il numero di istanze di cui hai bisogno. Calcola il numero totale di istanze necessarie per il tuo gruppo con dimensionamento automatico diviso per il numero di zone di disponibilità che intendi utilizzare.
4. In Dettagli prenotazione della capacità, per fine prenotazione della capacità scegli una delle seguenti opzioni:
 - In un momento specifico: annulla automaticamente la prenotazione della capacità alla data e all'ora specificate.
 - Manualmente: riserva la capacità fino a quando non la annulli esplicitamente.

5. Per idoneità delle istanze, scegli Mirato: solo istanze che hanno come target la prenotazione della capacità.
6. (Facoltativo) Per i tag, specifica i tag da associare alla prenotazione della capacità.
7. Scegli Crea.
8. Prendi nota dell'ID della prenotazione della capacità appena creata. È necessario per configurare il gruppo prenotazione della capacità.

Ripeti questa procedura per ogni zona di disponibilità che desideri abilitare per il tuo gruppo con dimensionamento automatico, modificando solo il valore dell'opzione Zona di disponibilità.

AWS CLI

Creazione di una prenotazione della capacità

Utilizzate il seguente [create-capacity-reservation](#) comando per creare le prenotazioni di capacità. Sostituisci i valori di esempio per `--availability-zone`, `--instance-type`, `--instance-platform` e `--instance-count`.

```
aws ec2 create-capacity-reservation \
  --availability-zone us-east-1a \
  --instance-type c5.xlarge \
  --instance-platform Linux/UNIX \
  --instance-count 3 \
  --instance-match-criteria targeted
```

Esempio di ID prenotazione della capacità risultante

```
{
  "CapacityReservation": {
    "CapacityReservationId": "cr-1234567890abcdef1",
    "OwnerId": "123456789012",
    "CapacityReservationArn": "arn:aws:ec2:us-east-1:123456789012:capacity-
reservation/cr-1234567890abcdef1",
    "InstanceType": "c5.xlarge",
    "InstancePlatform": "Linux/UNIX",
    "AvailabilityZone": "us-east-1a",
    "Tenancy": "default",
    "TotalInstanceCount": 3,
    "AvailableInstanceCount": 3,
    "EbsOptimized": false,
    "EphemeralStorage": false,
```

```
    "State": "active",
    "StartDate": "2023-07-26T21:36:14+00:00",
    "EndDateType": "unlimited",
    "InstanceMatchCriteria": "targeted",
    "CreateDate": "2023-07-26T21:36:14+00:00"
  }
}
```

Prendi nota dell'ID della prenotazione della capacità appena creata. È necessario per configurare il gruppo prenotazione della capacità.

Ripeti questo comando per ogni zona di disponibilità che desideri abilitare per il tuo gruppo con dimensionamento automatico, modificando solo il valore dell'opzione `--availability-zone`.

Fase 2: creazione di un gruppo di prenotazione della capacità

Al termine della creazione delle prenotazioni di capacità, è possibile raggrupparle utilizzando il servizio AWS Resource Groups. AWS Resource Groups supporta diversi tipi di gruppi per usi diversi. Amazon EC2 utilizza un gruppo per finalità speciali, noto come gruppo di risorse collegato al servizio, per indirizzare un gruppo di prenotazioni della capacità. Per interagire con questo gruppo di risorse collegato al servizio, puoi utilizzare AWS CLI o un SDK ma non la console. Per ulteriori informazioni sui gruppi di risorse collegati ai servizi, consultare [Configurazioni dei servizi per i gruppi di risorse](#) AWS Guida per l'utente dei gruppi di risorse.

Per creare un gruppo di prenotazione della capacità utilizzando il AWS CLI

Utilizza il comando [create-group](#) per creare un gruppo di risorse che può contenere solo prenotazioni della capacità. In questo esempio, il gruppo di risorse è denominato *my-cr-group*.

```
aws resource-groups create-group \
  --name my-cr-group \
  --configuration '{"Type":"AWS::EC2::CapacityReservationPool"}'
'{"Type":"AWS::ResourceGroups::Generic", "Parameters": [{"Name": "allowed-resource-
types", "Values": ["AWS::EC2::CapacityReservation"]}]}'
```

Di seguito è riportata una risposta di esempio.

```
{
  "Group": {
    "GroupArn": "arn:aws:resource-groups:us-east-1:123456789012:group/my-cr-group",
    "Name": "my-cr-group"
  }
}
```



```

    },
    "GroupConfiguration": {
      "Configuration": [
        {
          "Type": "AWS::EC2::CapacityReservationPool"
        },
        {
          "Type": "AWS::ResourceGroups::Generic",
          "Parameters": [
            {
              "Name": "allowed-resource-types",
              "Values": [
                "AWS::EC2::CapacityReservation"
              ]
            }
          ]
        }
      ]
    },
    "Status": "UPDATE_COMPLETE"
  }
}

```

Notare l'ARN del nuovo gruppo di risorse. Ad esempio, non è possibile creare un gruppo con dimensionamento automatico per creare un gruppo con dimensionamento automatico.

Per associare le tue prenotazioni della capacità al gruppo appena creato utilizzando AWS CLI

Utilizza il seguente comando [group-resources](#) per associare le prenotazioni della capacità al gruppo di prenotazione di capacità appena creato. Per l'opzione `--resource-arns`, specifica le prenotazioni della capacità utilizzando i relativi ARN. Gli ARN vanno creati utilizzando la regione pertinente, l'ID del tuo account e gli ID di prenotazione che hai annotato in precedenza. In questo esempio, le prenotazioni con ID `cr-1234567890abcdef1` e `cr-54321abcdef567890` verranno raggruppate nel gruppo denominato `my-cr-group`.

```

aws resource-groups group-resources \
  --group my-cr-group \
  --resource-arns \
    arn:aws:ec2:region:account-id:capacity-reservation/cr-1234567890abcdef1 \
    arn:aws:ec2:region:account-id:capacity-reservation/cr-54321abcdef567890

```

Di seguito è riportata una risposta di esempio.

```
{
  "Succeeded": [
    "arn:aws:ec2:us-east-1:123456789012:capacity-reservation/cr-1234567890abcdef1",
    "arn:aws:ec2:us-east-1:123456789012:capacity-reservation/cr-54321abcdef567890"
  ],
  "Failed": [],
  "Pending": []
}
```

Per informazioni sulla modifica o l'eliminazione del gruppo di risorse, consulta i [AWS documenti di riferimento all'API per gruppi di risorse](#).

Fase 3: creazione di un modello di avvio

Console

Per creare un modello di avvio

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/>.
2. Nel pannello di navigazione, in Istanze, scegli Modelli di avvio.
3. Scegli Crea modello di avvio. Immetti un nome e una descrizione per la versione iniziale del modello di avvio.
4. In Auto Scaling guidance (Guida Auto Scaling), seleziona la casella di controllo.
5. Creazione del modello di avvio. Scegli un AMI e un tipo di istanza che corrisponda alle prenotazioni della capacità che si intende utilizzare e, facoltativamente, una coppia di chiavi, uno o più gruppi di sicurezza ed eventuali volumi EBS o volumi di archivio dell'istanza aggiuntivi per le istanze.
6. Espandere Impostazioni avanzate ed eseguire le operazioni descritte di seguito:
 - a. Per Prenotazioni della capacità, scegli Destinazione per gruppo.
 - b. Per Prenotazione della capacità: destinazione per gruppo, scegli il gruppo di prenotazioni della capacità creato nella sezione precedente, quindi scegli Salva.
7. Scegli Crea modello di avvio.
8. Nella pagina di conferma, scegli Crea gruppo con dimensionamento automatico).

AWS CLI

Per creare un modello di avvio

Utilizzate il [create-launch-template](#) comando seguente per creare un modello di avvio che specifichi che la Capacity Reservation è destinata a un gruppo di risorse specifico. Sostituisci i valori di esempio per `--launch-template-name`. Sostituisci `c5.xlarge` con il tipo di istanza utilizzato nella prenotazione della capacità e `ami-0123456789EXAMPLE` con l'ID dell'AMI che desideri utilizzare. Sostituisci `arn:aws:resource-groups:region:account-id:group/my-cr-group` con l'ARN del gruppo risorsa creato all'inizio della sezione precedente.

```
aws ec2 create-launch-template \  
  --launch-template-name my-launch-template \  
  --launch-template-data \  
    '{"InstanceType": "c5.xlarge",  
     "ImageId": "ami-0123456789EXAMPLE",  
     "CapacityReservationSpecification":  
       {"CapacityReservationTarget":  
         { "CapacityReservationResourceGroupArn": "arn:aws:resource-  
groups:region:account-id:group/my-cr-group" }  
       }  
    }'
```

Di seguito è riportata una risposta di esempio.

```
{  
  "LaunchTemplate": {  
    "LaunchTemplateId": "lt-0dd77bd41dEXAMPLE",  
    "LaunchTemplateName": "my-launch-template",  
    "CreateTime": "2023-07-26T21:42:48+00:00",  
    "CreatedBy": "arn:aws:iam::123456789012:user/Bob",  
    "DefaultVersionNumber": 1,  
    "LatestVersionNumber": 1  
  }  
}
```

Fase 4: creazione di un gruppo con dimensionamento automatico

Console

Crea il tuo gruppo con dimensionamento automatico come al solito, ma quando scegli le sottoreti VPC, scegli una sottorete da ogni zona di disponibilità che corrisponda alle `targeted` prenotazioni della capacità che hai creato. Quindi, quando il gruppo con dimensionamento automatico avvia un'istanza on demand in una di queste zone di disponibilità, l'istanza verrà eseguita nella capacità riservata per quella zona di disponibilità. Se il gruppo di risorse esaurisce le prenotazioni della capacità prima che la capacità desiderata sia soddisfatta, lanciamo qualsiasi altra capacità oltre la capacità riservata come normale capacità on demand.

Creazione di un semplice gruppo con dimensionamento automatico

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Nella barra di navigazione nella parte superiore dello schermo, scegli lo stesso Regione AWS che hai usato per creare il modello di lancio.
3. Selezionare Crea un gruppo con dimensionamento automatico.
4. Nella pagina Scegli il modello o la configurazione di avvio, in Nome gruppo con dimensionamento automatico, inserisci un nome per il gruppo con dimensionamento automatico.
5. Come Launch template (Modello di avvio) scegli un modello di avvio esistente.
6. Per la Versione del modello di avvio, scegli se durante il dimensionamento orizzontale il gruppo con dimensionamento automatico debba usare la versione di default, quella più recente o una versione specifica del modello di avvio.
7. Nella pagina Scegli le opzioni di avvio dell'istanza, salta la sezione Requisiti del tipo di istanza per utilizzare il tipo di istanza EC2 specificato nel modello di avvio.
8. In Rete, per VPC, scegli un VPC. Il gruppo con dimensionamento automatico deve essere creato nello stesso VPC del gruppo di sicurezza specificato nel modello di avvio. Se non hai specificato un gruppo di sicurezza nel modello di avvio, puoi scegliere qualsiasi VPC con sottoreti nelle stesse zone di disponibilità delle prenotazioni della capacità.
9. Per le zone di disponibilità delle e le sottoreti, scegli le sottoreti da ciascuna zona di disponibilità che desideri includere, in base alle zone di disponibilità in cui si trovano le tue prenotazioni della capacità.
10. Scegliere Next (Successivo) due volte.

11. Nella pagina Configura le dimensioni del gruppo e le policy di dimensionamento, per Capacità desiderata, inserisci il numero iniziale di istanze da avviare. Quando si modifica questo numero in un valore non compreso nei limiti di capacità minima o massima, è necessario aggiornare i valori di Minimum capacity (Capacità minima) o Maximum capacity (Capacità massima). Per ulteriori informazioni, consulta [Impostare i limiti di dimensionamento per il gruppo con dimensionamento automatico](#).
12. Selezionare Passa a revisione.
13. Nella pagina Verifica, scegliere Crea gruppo con dimensionamento automatico.

AWS CLI

Creazione di un semplice gruppo con dimensionamento automatico

Usa il [create-auto-scaling-group](#) comando seguente e specifica il nome e la versione del modello di lancio come valore per l'`--launch-template` opzione. Sostituisci i valori di esempio per `--auto-scaling-group-name`, `--min-size`, `--max-size` e `--vpc-zone-identifier`.

Per l'opzione `--availability-zones`, specifica le zone di disponibilità per le quali hai creato le prenotazioni della capacità. Ad esempio, se le prenotazioni della capacità specificano le zone di disponibilità `us-east-1a` e `us-east-1b`, è necessario creare il gruppo con dimensionamento automatico nelle stesse zone. Quindi, quando il gruppo con dimensionamento automatico avvia un'istanza on demand in una di queste zone di disponibilità, l'istanza verrà eseguita nella capacità riservata per quella zona di disponibilità. Se il gruppo di risorse esaurisce le prenotazioni della capacità prima che la capacità desiderata sia soddisfatta, lanciamo qualsiasi altra capacità oltre la capacità riservata come normale capacità on demand.

```
aws autoscaling create-auto-scaling-group \  
  --auto-scaling-group-name my-asg \  
  --launch-template LaunchTemplateName=my-launch-template,Version='1' \  
  --min-size 6 \  
  --max-size 6 \  
  --vpc-zone-identifier "subnet-5f46ec3b,subnet-0ecac448" \  
  --availability-zones us-east-1a us-east-1b
```

Risorse correlate

Per un esempio di implementazione, consultate il AWS CloudFormation modello nel seguente GitHub repository di AWS esempi: <https://github.com/aws-samples/aws-auto-scaling-backed-by-on-demand-capacity-reservations/>.

I seguenti argomenti correlati possono essere utili per imparare a conoscere prenotazioni della capacità.

- Prenotazione della capacità on demand
 - [Creazione di una prenotazione della capacità](#) nella Guida per l'utente di Amazon EC2 per istanze Linux
 - [Creazione di una prenotazione della capacità on demand](#) nella Guida per l'utente di Amazon EC2 per istanze Linux
 - [Rivolgiti a un gruppo di prenotazioni di capacità on demand di Amazon EC2 sul blog AWS Cloud Operations & Migrations](#)
- Blocchi di capacità (prenotazioni della capacità con una durata definita)
 - [Blocchi di capacità per il machine learning](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux
 - [Usa Capacity Blocks per carichi di lavoro di machine learning](#)

Crea gruppi di Auto Scaling dalla riga di comando usando AWS CloudShell

Nella [versione supportata Regioni AWS](#), puoi eseguire AWS CLI comandi utilizzando AWS CloudShell una shell preautenticata basata su browser che si avvia direttamente da AWS Management Console. È possibile eseguire AWS CLI comandi sui servizi utilizzando la shell preferita (Bash o Z shell). PowerShell

È possibile eseguire l' AWS Management Console avvio AWS CloudShell da uno dei due metodi seguenti:

- Scegli l' AWS CloudShell icona nella barra di navigazione della console. Si trova alla destra della casella di ricerca.
- Utilizza la casella di ricerca sulla barra di navigazione della console per CloudShell cercare, quindi scegli l'CloudShell opzione.

Quando viene AWS CloudShell avviato per la prima volta in una nuova finestra del browser, un pannello di benvenuto mostra ed elenca le funzionalità principali. Dopo aver chiuso questo pannello, vengono forniti aggiornamenti dello stato, mentre la shell configura e inoltra le credenziali della console. Quando viene visualizzato il prompt dei comandi, la shell è pronta per l'interazione.

Per ulteriori informazioni su questo servizio, consulta la [Guida per l'utente di AWS CloudShell](#).

Creazione di gruppi Auto Scaling con AWS CloudFormation

Amazon EC2 Auto Scaling è integrato AWS CloudFormation con un servizio che ti aiuta a modellare e configurare le AWS tue risorse in modo da poter dedicare meno tempo alla creazione e alla gestione delle risorse e dell'infrastruttura. Crei un modello che descrive tutte le AWS risorse che desideri (come i gruppi di Auto Scaling) e predispone e AWS CloudFormation configura tali risorse per te.

Quando lo utilizzi AWS CloudFormation, puoi riutilizzare il modello per configurare le risorse di Amazon EC2 Auto Scaling in modo coerente e ripetuto. Descrivi le tue risorse una sola volta, quindi fornisci le stesse risorse più e più volte in più Account AWS regioni.

Auto Scaling e modelli di Amazon EC2 AWS CloudFormation

Per eseguire il provisioning e configurare le risorse per Amazon EC2 Auto Scaling e i servizi correlati, devi conoscere i [modelli AWS CloudFormation](#). I modelli sono file di testo formattati in JSON o YAML. Questi modelli descrivono le risorse che desideri fornire nei tuoi AWS CloudFormation stack. Se non conosci JSON o YAML, puoi usare AWS CloudFormation Designer per iniziare a usare i modelli. AWS CloudFormation [Per ulteriori informazioni, consulta Cos'è Designer? AWS CloudFormation](#) nella Guida AWS CloudFormation per l'utente.

Per iniziare a creare modelli di stack personalizzati per Dimensionamento automatico Amazon EC2, completa le seguenti attività:

- Crea un modello di lancio utilizzando [AWS::EC2::LaunchTemplate](#).
-

Per una procedura dettagliata che mostra come implementare un gruppo con dimensionamento automatico dietro un Application Load Balancer, consulta [Spiegazione passo per passo: creazione di un'applicazione con dimensionamento e sistema di bilanciamento del carico](#) nella Guida per l'utente di AWS CloudFormation .

Puoi trovare altri esempi utili di frammenti di modello che creano gruppi di Auto Scaling e risorse correlate nelle seguenti sezioni della AWS CloudFormation Guida per l'utente:

- Riferimento al tipo di risorsa [Amazon EC2 Auto Scaling](#) [Riferimento al tipo di risorsa Amazon Scaling](#)
- [Configura le risorse di Amazon EC2 Auto Scaling con AWS CloudFormation](#)

Scopri di più su AWS CloudFormation

Per ulteriori informazioni AWS CloudFormation, consulta le seguenti risorse:

- [AWS CloudFormation](#)
- [AWS CloudFormation Guida per l'utente](#)
- [AWS CloudFormation Documentazione di riferimento delle API](#)
- [AWS CloudFormation Guida per l'utente dell'interfaccia a riga di comando](#)

AWS Compute Optimizer Utilizzatelo per ottenere consigli sul tipo di istanza per un gruppo di Auto Scaling

AWS fornisce consigli sulle istanze di Amazon EC2 per aiutarti a migliorare le prestazioni, risparmiare denaro o entrambi, utilizzando funzionalità fornite da AWS Compute Optimizer. Puoi utilizzare questi suggerimenti per decidere se passare o meno a un nuovo tipo di istanza.

Per fornire i suggerimenti, Compute Optimizer analizza le specifiche esistenti dell'istanza e la cronologia dei parametri recenti. I dati compilati vengono quindi utilizzati per suggerire quali tipi di istanza Amazon EC2 sono ottimizzati per gestire il carico di lavoro delle prestazioni. I suggerimenti vengono restituiti insieme ai prezzi orari delle istanze.

Note

Per ricevere suggerimenti da Compute Optimizer, devi prima scegliere Compute Optimizer. Per ulteriori informazioni, consulta [Nozioni di base su AWS Compute Optimizer](#) nella Guida per l'utente di AWS Compute Optimizer .

Indice

- [Limitazioni](#)
- [Risultati](#)
- [Visualizzare le raccomandazioni](#)
- [Considerazioni sulla valutazione dei suggerimenti](#)

Limitazioni

Compute Optimizer genera suggerimenti per le istanze nei gruppi Auto Scaling che sono configurati per avviare ed eseguire i tipi di istanza M, C, R, T e X. Tuttavia, non genera raccomandazioni per i tipi di istanze -g basati su processori AWS Graviton2 (ad esempio, C6g) e per i tipi di istanze -n che hanno prestazioni di larghezza di banda di rete più elevate (ad esempio, M5n).

I gruppi Auto Scaling devono inoltre essere configurati per eseguire un singolo tipo di istanza (cioè, nessun tipo di istanza misto), non devono avere una policy di dimensionamento allegata e avere gli stessi valori per la capacità desiderata, minima e massima (ad esempio, un gruppo Auto Scaling con un numero fisso di istanze). Compute Optimizer genera suggerimenti per le istanze nei gruppi Auto Scaling che soddisfano tutti questi requisiti di configurazione.

Risultati

Compute Optimizer classifica i risultati per i gruppi Auto Scaling come segue:

- **Non ottimizzato:** un gruppo Auto Scaling viene considerato non ottimizzato quando Compute Optimizer identifica un suggerimento che può fornire prestazioni migliori per il carico di lavoro.
- **Ottimizzato:** un gruppo Auto Scaling viene considerato ottimizzato quando Compute Optimizer rileva che il gruppo con provisioning può eseguire correttamente il carico di lavoro, in base al tipo di istanza scelto. Per le risorse ottimizzate, Compute Optimizer può talvolta suggerire un tipo di istanza di nuova generazione.
- **Nessuno:** non ci sono suggerimenti per questo gruppo Auto Scaling. Ciò potrebbe verificarsi se Computer Optimizer è stato attivato da meno di 12 ore o quando il gruppo Auto Scaling è in esecuzione da meno di 30 ore o quando il gruppo Auto Scaling o il tipo di istanza non è supportato da Compute Optimizer. Per ulteriori informazioni, consulta la sezione [Limitazioni](#).

Visualizzare le raccomandazioni

Dopo aver scelto Compute Optimizer, è possibile visualizzare i suggerimenti e i risultati generati per i gruppi Auto Scaling. Se hai scelto Compute Optimizer, i suggerimenti potrebbero non essere disponibili per un periodo di tempo fino a un massimo di 12 ore.

Come visualizzare i suggerimenti generati per un gruppo Auto Scaling

1. Apri la console di Compute Optimizer all'indirizzo <https://console.aws.amazon.com/compute-optimizer/>.

Visualizzerai la pagina Dashboard (Pannello di controllo).

2. Scegli Visualizza suggerimenti per tutti i gruppi Auto Scaling.
3. Seleziona il gruppo Auto Scaling.
4. Selezionare View Details (Visualizza dettagli).

La visualizzazione cambia per fornire fino a tre diversi suggerimenti dell'istanza in una visualizzazione preconfigurata, in base alle impostazioni di default della tabella. Fornisce inoltre dati CloudWatch metrici recenti (utilizzo medio della CPU, rete media in entrata e rete media in uscita) per il gruppo Auto Scaling.

Valuta se vuoi utilizzare una delle raccomandazioni. Decidi se ottimizzare per migliorare le prestazioni, per ridurre i costi o per entrambe le motivazioni.

Per modificare il tipo di istanza nel gruppo Auto Scaling, aggiorna il modello di avvio o il gruppo Auto Scaling per utilizzare una nuova configurazione di avvio. Le istanze esistenti continuano a utilizzare la configurazione precedente. Per aggiornare le istanze esistenti, terminarle in modo che vengano sostituite dal gruppo Auto Scaling o permettere la scalabilità automatica per sostituire gradualmente istanze meno recenti con istanze più recenti, in base alle [policy di terminazione](#).

Note

Con la durata massima e le funzioni di aggiornamento dell'istanza, è inoltre possibile sostituire quelle esistenti nel gruppo Auto Scaling per avviare nuove istanze che utilizzano la nuova configurazione di avvio o il nuovo modello di avvio. Per ulteriori informazioni, consulta [Sostituire le istanze Auto Scaling basate sulla durata massima dell'istanza](#) e [Usa un aggiornamento dell'istanza per aggiornare le istanze in un gruppo di Auto Scaling](#).

Considerazioni sulla valutazione dei suggerimenti

Prima di passare a un nuovo tipo di istanza, considera quanto segue:

- I suggerimenti non prevedono l'utilizzo che ne farai. I suggerimenti si basano sullo storico del tuo utilizzo negli ultimi 14 giorni. Assicurati di scegliere un tipo di istanza che soddisfi le tue esigenze di utilizzo future.
- Concentrati sui parametri dei grafici per determinare se l'utilizzo effettivo sia o meno inferiore alla capacità dell'istanza. Puoi anche visualizzare i dati metrici (media, picco, percentile) CloudWatch per valutare ulteriormente i consigli sulle istanze EC2. Ad esempio, verifica se i parametri in percentuale della CPU durante il giorno cambiano e se si verificano picchi che devono essere gestiti. Per ulteriori informazioni, consulta [Visualizzazione delle metriche disponibili](#) nella Amazon CloudWatch User Guide.
- Compute Optimizer può fornire suggerimenti per le istanze a prestazioni espandibili, ossia le istanze T3, T3a e T2. Se periodicamente ti espandi oltre il tuo riferimento, assicurati di poter continuare a farlo, in base alle vCPU del nuovo tipo di istanza. Per ulteriori informazioni, consulta [Crediti CPU e prestazioni di base per istanze a prestazioni espandibili](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.
- Se hai acquistato un'Istanza riservata, è possibile che l'Istanza on demand venga fatturata come Istanza riservata. Prima di modificare il tipo di istanza corrente, valuta innanzitutto l'impatto sull'utilizzo e sulla copertura dell'Istanza riservata.
- Laddove possibile, valuta il passaggio a istanze di ultima generazione.
- Quando esegui la migrazione a una famiglia di istanze diversa, assicurati che il tipo di istanza corrente e quello nuovo siano compatibili in termini, ad esempio, di virtualizzazione, architettura o tipo di rete. Per ulteriori informazioni, consulta [Compatibilità per il ridimensionamento delle istanze](#) nella Guida dell'utente di Amazon EC2 per le istanze Linux.
- Infine, prendi in considerazione la valutazione del rischio delle prestazioni fornita per ogni suggerimento. Il rischio delle prestazioni indica l'impegno che potrebbe essere richiesto per stabilire se il tipo di istanza suggerito soddisfi i requisiti di prestazioni del carico di lavoro. Ti suggeriamo, inoltre, di eseguire test rigorosi per il carico e le prestazioni prima e dopo aver apportato eventuali modifiche.

Risorse aggiuntive

Oltre agli argomenti in questa pagina, consulta le pagine seguenti:

- [Tipi di istanza di Amazon EC2](#)
- [AWS Compute Optimizer Guida per l'utente](#)

Utilizzare sistema di bilanciamento del carico elastico per distribuire il traffico tra le istanze nel gruppo con dimensionamento automatico.

Il sistema di bilanciamento del carico elastico distribuisce e bilancia automaticamente il traffico di applicazioni in ingresso tra tutte le istanze EC2 server in esecuzione. Il sistema di bilanciamento del carico elastico aiuta a gestire le richieste in ingresso eseguendo il routing del traffico in modo ottimale in modo che nessuna istanza venga sovraccaricata.

Per utilizzare il sistema di bilanciamento del carico elastico con il gruppo con dimensionamento automatico, [collego il sistema di bilanciamento del carico al gruppo con dimensionamento automatico](#). In questo modo il gruppo viene registrato con il sistema di bilanciamento del carico elastico, che agisce come un singolo punto di contatto per tutto il traffico Web in entrata nel gruppo con dimensionamento automatico.

Quando utilizzi il sistema di bilanciamento del carico elastico con un gruppo con dimensionamento automatico, non è necessario che registri le singole istanze EC2 con il sistema di bilanciamento del carico. Le istanze avviate dal gruppo con dimensionamento automatico vengono registrate automaticamente con il sistema di bilanciamento del carico. Allo stesso modo, le istanze terminate dal gruppo con dimensionamento automatico vengono deregistrate automaticamente dal sistema di bilanciamento del carico.

Dopo aver collegato un sistema di bilanciamento del carico al gruppo con dimensionamento automatico, potrai configurare il gruppo con dimensionamento automatico in modo che utilizzi parametri di bilanciamento del carico elastico (ad esempio il conteggio richieste Application Load Balancer per destinazione) per dimensionare il numero di istanze nel gruppo man mano che cambia la domanda.

Facoltativamente, puoi aggiungere controlli dell'integrità del bilanciamento del carico elastico al tuo gruppo con dimensionamento automatico in modo che Dimensionamento automatico Amazon EC2 possa identificare e sostituire le istanze non integre in base a questi controlli aggiuntivi. Altrimenti, puoi creare un CloudWatch allarme che ti avvisi se il numero di host sani del gruppo target è inferiore a quello consentito.

Indice

- [Tipi di sistema di bilanciamento del carico elastico](#)
- [Preparati a collegare un sistema di bilanciamento del carico Elastic Load Balancing al tuo gruppo Auto Scaling](#)
- [Collega un sistema di bilanciamento del carico Elastic Load Balancing al tuo gruppo Auto Scaling](#)
- [Configurare un Application Load Balancer o Network Load Balancer dalla console di Dimensionamento automatico Amazon EC2.](#)
- [Verifica lo stato di collegamento del sistema di bilanciamento del carico](#)
- [Aggiungere o rimuovere zone di disponibilità](#)
- [Esempi di utilizzo di Elastic Load Balancing con \(\) AWS Command Line InterfaceAWS CLI](#)

Tipi di sistema di bilanciamento del carico elastico

Il sistema di bilanciamento del carico elastico fornisce quattro tipi di sistema di bilanciamento del carico utilizzabili con il gruppo con dimensionamento automatico: Application Load Balancer, Network Load Balancer, Gateway Load Balancer e Classic Load Balancer.

Esiste una differenza fondamentale nel modo in cui i tipi di sistema di bilanciamento del carico vengono configurati. Con Application Load Balancer, Network Load Balancer e Gateway Load Balancer le istanze vengono registrate come destinatarie, con un gruppo destinatario verso il quale viene instradato il traffico. Con Classic Load Balancer le istanze vengono registrate direttamente con il sistema di bilanciamento del carico.

Application Load Balancer

Instrada e fornisce bilanciamento del carico a livello di applicazione (HTTP/HTTPS) e supporta il routing in base al routing. Un Application Load Balancer può instradare le richieste alle porte su una o più destinazioni registrate, ad esempio le istanze EC2, nel cloud privato virtuale (VPC).

Network Load Balancer

Instrada e fornisce bilanciamento del carico a livello di trasporto (TCP/UDP Layer-4) in base alle informazioni relative all'indirizzo estratte dall'intestazione del Layer-4. Il Network Load Balancer è in grado di gestire i picchi di traffico, mantenere l'IP sorgente del client e utilizzare un IP fisso per l'intera durata del sistema di bilanciamento del carico.

Gateway Load Balancer

Distribuisce il traffico a un parco istanze di appliance. Fornisce scalabilità, disponibilità e semplicità per le appliance virtuali di terze parti, come firewall, sistemi di rilevamento e prevenzione delle intrusioni e altre. I Gateway Load Balancer funzionano con appliance virtuali che supportano il protocollo GENEVE. Richiedono un'ulteriore integrazione tecnica quindi, prima di scegliere un Gateway Load Balancer, consulta la guida per l'utente.

Classic Load Balancer

Instrada e fornisce bilanciamento del carico sia a livello di trasporto (TCP/SSL) sia a livello di applicazione (HTTP/HTTPS).

Per una comprensione più approfondita dei diversi tipi di sistemi di bilanciamento del carico disponibili, consulta le seguenti risorse:

- [Cos'è l'Elastic Load Balancing?](#)
- [Cos'è un Application Load Balancer?](#)
- [Cos'è un Network Load Balancer?](#)
- [Cos'è un Gateway Load Balancer?](#)
- [Cos'è un Classic Load Balancer?](#)

Preparati a collegare un sistema di bilanciamento del carico Elastic Load Balancing al tuo gruppo Auto Scaling

Prima di collegare un sistema di bilanciamento del carico Elastic Load Balancing al gruppo Auto Scaling, è necessario completare i seguenti prerequisiti:

- È necessario aver già creato il sistema di bilanciamento del carico e il gruppo target utilizzati per indirizzare il traffico verso il gruppo Auto Scaling.

Esistono due modi per creare il load balancer e il gruppo target:

- Utilizzo di Elastic Load Balancing: segui le procedure nella documentazione di Elastic Load Balancing per creare e configurare il load balancer e il gruppo target prima di creare il gruppo Auto Scaling. Puoi saltare il passaggio per registrare le istanze Amazon EC2. Amazon EC2 Auto Scaling si occupa automaticamente della registrazione (e dell'annullamento della registrazione) delle istanze quando colleghi un gruppo target al gruppo Auto Scaling. Per ulteriori informazioni,

consulta [Nozioni di base dell'Elastic Load Balancing](#) nella Guida per l'utente di Elastic Load Balancing.

- Utilizzo di Amazon EC2 Auto Scaling: crea, configura e collega il sistema di bilanciamento del carico e il gruppo target con una configurazione di base dalla console Amazon EC2 Auto Scaling. Per ulteriori informazioni, consulta [Configurare un Application Load Balancer o Network Load Balancer dalla console di Dimensionamento automatico Amazon EC2](#).
- Prima di creare un sistema di bilanciamento del carico, conosci il tipo di bilanciamento del carico di cui hai bisogno. Per ulteriori informazioni, consulta [Tipi di sistema di bilanciamento del carico elastico](#).
- Il sistema di bilanciamento del carico e il relativo gruppo target devono trovarsi nello stesso Account AWS VPC e nella stessa regione del gruppo Auto Scaling.
- I gruppi di destinazione devono avere il tipo di destinazione `instance`. Quando si utilizza un gruppo con dimensionamento automatico, non è possibile specificare un tipo di destinazione `ip`.
- Se il modello di avvio per il gruppo Auto Scaling non contiene il gruppo di sicurezza corretto per consentire il traffico in entrata necessario dal sistema di bilanciamento del carico, è necessario aggiornare il modello di avvio. Le regole suggerite variano a seconda del tipo di sistema di bilanciamento del carico e dei tipi di back-end che utilizza. Ad esempio, per instradare il traffico ai server Web, permetti l'accesso HTTP in ingresso sulla porta 80 dal sistema di bilanciamento del carico. Le istanze esistenti non vengono aggiornate con le nuove impostazioni quando il modello di avvio viene modificato. Per aggiornare le istanze esistenti, puoi avviare un aggiornamento dell'istanza per sostituire le istanze. Per ulteriori informazioni, consulta [Usa un aggiornamento dell'istanza per aggiornare le istanze in un gruppo di Auto Scaling](#).
- I gruppi di sicurezza nel modello di avvio devono inoltre consentire l'accesso dal load balancer sulla porta corretta affinché Elastic Load Balancing possa eseguire i controlli di integrità.
- Quando si distribuiscono appliance virtuali dietro un Gateway Load Balancer, l'Amazon Machine Image (AMI) nel modello di avvio deve specificare l'ID di un'AMI che supporta il protocollo GENEVE per consentire al gruppo Auto Scaling di scambiare traffico con un Gateway Load Balancer. Inoltre, i gruppi di sicurezza nel modello di lancio devono consentire il traffico UDP sulla porta 6081.

Tip

Se disponi di script di bootstrap che richiedono del tempo per essere completati, è possibile aggiungere un hook del ciclo di vita di avvio al gruppo con dimensionamento automatico in modo da ritardare la registrazione delle istanze dietro il sistema di bilanciamento del carico prima che gli script di bootstrap siano stati completati correttamente e che le applicazioni

nelle istanze siano pronte per accettare il traffico. Non puoi aggiungere un hook del ciclo di vita quando si crea inizialmente un gruppo con dimensionamento automatico nella console Dimensionamento automatico Amazon EC2. Tuttavia, è possibile aggiungere un lifecycle hook dopo la creazione del gruppo. Per ulteriori informazioni, consulta [Hook del ciclo di vita di Amazon EC2 Auto Scaling](#).

Configura i controlli sanitari per gli obiettivi

Puoi configurare i controlli di integrità per i tuoi obiettivi registrati con un sistema di bilanciamento del carico Elastic Load Balancing per assicurarti che siano in grado di gestire correttamente il traffico. I passaggi specifici variano in base al tipo di bilanciamento del carico in uso. Per ulteriori informazioni, consulta le seguenti risorse:

- Application Load Balancer: consulta i [controlli Health per i tuoi gruppi target](#) nella User Guide for Application Load Balancers.
- Network Load Balancer: consulta i [controlli Health per i tuoi gruppi target](#) nella Guida per l'utente di Network Load Balancer.
- Gateway Load Balancer: consulta i [controlli Health per i gruppi target](#) nella User Guide for Gateway Load Balancer.
- Classic Load Balancer: consulta [Configurare i controlli di integrità per il tuo Classic Load Balancer](#) nella Guida per l'utente di Classic Load Balancer.

Per impostazione predefinita, Amazon EC2 Auto Scaling non considera un'istanza non integra e la sostituisce se non supera i controlli di integrità di Elastic Load Balancing. I controlli dell'integrità di default per un gruppo con dimensionamento automatico sono solo i controlli dell'integrità EC2. Per ulteriori informazioni, consulta [Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico](#).

Per consentire ad Amazon EC2 Auto Scaling di sostituire le istanze segnalate come non integre da Elastic Load Balancing, puoi configurare il tuo gruppo Auto Scaling per utilizzare i controlli di integrità di Elastic Load Balancing. In questo modo, Amazon EC2 Auto Scaling considera l'istanza non integra se non supera i controlli di integrità di EC2 o i controlli di integrità di Elastic Load Balancing. Se si allegano più gruppi di destinazione del sistema di bilanciamento del carico o Classic Load Balancer a un gruppo, tutti i sistemi devono segnalare che l'istanza è integra, per considerarla tale. Se uno di essi segnala un'istanza come non integra, il gruppo con dimensionamento automatico sostituisce l'istanza, anche se altri la segnalano come integra.

Per informazioni su come abilitare questi controlli di integrità per il gruppo Auto Scaling, vedere. [Collega un sistema di bilanciamento del carico Elastic Load Balancing al tuo gruppo Auto Scaling](#)

Note

Per assicurarti che questi controlli di integrità inizino il prima possibile, assicurati che il periodo di tolleranza per il controllo dello stato di salute del tuo gruppo non sia troppo alto, ma sufficientemente alto da consentire ai controlli di integrità di Elastic Load Balancing di determinare se un obiettivo è disponibile per gestire le richieste. Per ulteriori informazioni, consulta [Impostazione del periodo di tolleranza del controllo dell'integrità per un gruppo con scalabilità automatica](#).

Collega un sistema di bilanciamento del carico Elastic Load Balancing al tuo gruppo Auto Scaling

Questo argomento descrive come collegare un sistema di bilanciamento del carico Elastic Load Balancing a un gruppo Auto Scaling. Descrive inoltre come attivare i controlli di integrità di Elastic Load Balancing per consentire ad Amazon EC2 Auto Scaling di sostituire le istanze che Elastic Load Balancing segnala come non integre.

Per impostazione predefinita, Dimensionamento automatico Amazon EC2 sostituisce solo le istanze non integre o irraggiungibili in base ai controlli dell'integrità di Amazon EC2. Se attivi i controlli di integrità di Elastic Load Balancing, Amazon EC2 Auto Scaling può sostituire un'istanza in esecuzione se uno dei sistemi di bilanciamento del carico Elastic Load Balancing collegati al gruppo Auto Scaling la segnala come non integra.

Per un tutorial su come collegare un Application Load Balancer al gruppo Auto Scaling, consulta. [Tutorial: Configurazione di un'applicazione con dimensionamento e bilanciamento del carico](#)

Important

Prima di continuare, completa tutti i [prerequisiti indicati](#) nella sezione precedente.

Indice

- [Collega un gruppo target o un Classic Load Balancer](#)
- [Scollegare un gruppo target o un Classic Load Balancer](#)

Collega un gruppo target o un Classic Load Balancer

Quando crei o aggiorni un gruppo Auto Scaling, puoi collegare uno o più gruppi target o Classic Load Balancer. Quando si collega un Application Load Balancer, Network Load Balancer o Gateway Load Balancer, si collega un gruppo target anziché il load balancer stesso.

Segui i passaggi indicati in questa sezione per utilizzare la console per:

- Collega un gruppo target o un Classic Load Balancer a un gruppo Auto Scaling
- Attiva i controlli di integrità per Elastic Load Balancing

Come collegare un sistema di bilanciamento del carico esistente durante la creazione di un nuovo gruppo con dimensionamento automatico

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Nella barra di navigazione nella parte superiore dello schermo, scegli il dispositivo in Regione AWS cui hai creato il sistema di bilanciamento del carico.
3. Selezionare Crea un gruppo con dimensionamento automatico).
4. Nei passaggi 1 e 2, scegli le opzioni desiderate e procedi al Passaggio 3: configura le opzioni avanzate.
5. In Sistema di bilanciamento del carico, scegli Collega a un sistema di bilanciamento del carico esistente.
6. In Collega a un sistema di bilanciamento del carico esistente scegli una delle seguenti operazioni:
 - a. Per Application Load Balancer, Network Load Balancer e Gateway Load Balancer:

Scegli Scegli tra i gruppi di destinazione del sistema di bilanciamento del carico, quindi scegli un gruppo di destinazione nel campo Gruppi di destinazione del sistema di bilanciamento del carico esistenti.
 - b. Per Classic Load Balancer:

Scegli Scegli tra i Classic Load Balancer e quindi scegli il sistema di bilanciamento del carico nel campo Classic Load Balancer.
7. (Facoltativo) Per i controlli dell'integrità, tipi di controlli dell'integrità aggiuntivi, seleziona Attiva i controlli dell'integrità di Elastic Load Balancing.

8. (Facoltativo) In Periodo di tolleranza del controllo dell'integrità, inserisci il tempo in secondi. Questo è il periodo di tempo che Dimensionamento automatico Amazon EC2 deve attendere prima di controllare lo stato di integrità di un'istanza dopo che questa è entrata nello stato InService. Per ulteriori informazioni, consulta [Impostazione del periodo di tolleranza del controllo dell'integrità per un gruppo con scalabilità automatica](#).
9. Procedi e crea il gruppo con dimensionamento automatico. Le istanze verranno registrate automaticamente nel sistema di bilanciamento del carico dopo la creazione del gruppo con dimensionamento automatico.

Come collegare un sistema di bilanciamento del carico al tuo gruppo con dimensionamento automatico dopo la sua creazione

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina Gruppi con dimensionamento automatico.

3. Nella scheda Dettagli scegliere Sistema di bilanciamento del carico, Modifica.
4. In Sistema di bilanciamento del carico, eseguire una delle operazioni seguenti:
 - a. In Gruppi di destinazione Application, Network o Gateway Load Balancer, seleziona la relativa casella di controllo e scegli un gruppo di destinazione.
 - b. In Classic Load Balancer, seleziona la relativa casella di controllo e quindi il sistema di bilanciamento del carico.
5. Scegli Aggiorna.

Dopo aver collegato il sistema di bilanciamento del carico, puoi facoltativamente attivare i controlli di integrità che lo utilizzano.

Per attivare i controlli di integrità di Elastic Load Balancing

1. Nella scheda Dettagli scegli Controlli dell'integrità, Modifica.
2. Per i controlli dell'integrità, tipi di controlli dell'integrità aggiuntivi, seleziona Attiva i controlli dell'integrità di Elastic Load Balancing.

3. In Periodo di tolleranza del controllo dell'integrità, inserisci il tempo in secondi. Questo è il periodo di tempo che Dimensionamento automatico Amazon EC2 deve attendere prima di controllare lo stato di integrità di un'istanza dopo che questa è entrata nello stato InService. Per ulteriori informazioni, consulta [Impostazione del periodo di tolleranza del controllo dell'integrità per un gruppo con scalabilità automatica](#).
4. Scegli Aggiorna.

Note

Puoi monitorare lo stato del sistema di bilanciamento del carico mentre è collegato utilizzando la AWS CLI. Quando Dimensionamento automatico Amazon EC2 ha registrato correttamente le istanze e almeno una di esse supera i controlli dell'integrità, viene visualizzato lo stato InService. Per ulteriori informazioni, consulta [Verifica lo stato di collegamento del sistema di bilanciamento del carico](#).

Scollegare un gruppo target o un Classic Load Balancer

Quando il sistema di bilanciamento del carico non ti serve più, utilizza la procedura seguente per scollegarlo dal gruppo con dimensionamento automatico.

Come scollegare un sistema di bilanciamento del carico da un gruppo

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto a un gruppo esistente.

Si aprirà un riquadro diviso nella parte inferiore della pagina Gruppi con dimensionamento automatico.

3. Nella scheda Dettagli scegliere Sistema di bilanciamento del carico, Modifica.
4. In Sistema di bilanciamento del carico, eseguire una delle operazioni seguenti:
 - a. In Gruppi di destinazione Application, Network o Gateway , scegli l'icona di eliminazione (X) accanto al gruppo di destinazione.
 - b. In Classic , scegli l'icona di eliminazione (X) accanto al sistema di bilanciamento del carico.
5. Scegli Aggiorna.

Al termine del distacco del gruppo target, puoi disattivare i controlli di integrità di Elastic Load Balancing.

Per disattivare i controlli di integrità di Elastic Load Balancing

1. Nella scheda Dettagli scegli Controlli dell'integrità, Modifica.
2. Per i controlli di integrità, Tipi di controlli sanitari aggiuntivi, deseleziona Attiva i controlli di integrità di Elastic Load Balancing.
3. Scegli Aggiorna.

Configurare un Application Load Balancer o Network Load Balancer dalla console di Dimensionamento automatico Amazon EC2.

Per creare e collegare un Application Load Balancer o un Network Load Balancer durante la creazione del gruppo con dimensionamento automatico, utilizza la procedura seguente.

Come creare e collegare un nuovo sistema di bilanciamento del carico durante la creazione di un nuovo gruppo con dimensionamento automatico

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Selezionare Crea un gruppo con dimensionamento automatico).
3. Nei passaggi 1 e 2, scegli le opzioni desiderate e procedi al Passaggio 3: configura le opzioni avanzate.
4. In Sistema di bilanciamento del carico, scegli Collega a un nuovo sistema di bilanciamento del carico).
 - a. In Collega a un nuovo sistema di bilanciamento del carico, per tipo di sistema di bilanciamento del carico, scegli se creare un Application Load Balancer o un Network Load Balancer.
 - b. In nome del sistema di bilanciamento del carico name , inserisci un nome per il sistema di bilanciamento del carico o mantieni quello predefinito.
 - c. In Schema del sistema di bilanciamento del carico, scegli se creare un sistema di bilanciamento del carico pubblico connesso a Internet o mantenere l'impostazione di default, che prevede un sistema di bilanciamento del carico interno.

- d. In Zone di disponibilità e sottoreti, seleziona la sottorete pubblica per ciascuna zona di disponibilità in cui hai scelto di avviare le istanze EC2. (La prepopolazione avverrà nel passaggio 2).
 - e. In Listener e instradamento, aggiorna il numero di porta per il listener (se necessario) e in Instradamento di default, scegli Crea gruppo di destinazione. In alternativa, puoi scegliere un gruppo di destinazione esistente dall'elenco a discesa.
 - f. Se nell'ultimo passaggio hai scelto Crea un gruppo di destinazione, in Nuovo nome del gruppo di destinazione, inserisci un nome per il gruppo di destinazione o mantieni quello di default.
 - g. Per aggiungere tag al proprio sistema di bilanciamento del carico, scegli Aggiungi tag e fornisci una chiave e un valore di tag per ogni tag.
5. (Facoltativo) Per i controlli dell'integrità, tipi di controlli dell'integrità aggiuntivi, seleziona Attiva i controlli dell'integrità di Elastic Load Balancing.
 6. (Facoltativo) In Periodo di tolleranza del controllo dell'integrità, inserisci il tempo in secondi. Questo è il periodo di tempo che Dimensionamento automatico Amazon EC2 deve attendere prima di controllare lo stato di integrità di un'istanza dopo che questa è entrata nello stato InService. Per ulteriori informazioni, consulta [Impostazione del periodo di tolleranza del controllo dell'integrità per un gruppo con scalabilità automatica](#).
 7. Procedi e crea il gruppo con dimensionamento automatico. Le istanze verranno registrate automaticamente nel sistema di bilanciamento del carico dopo la creazione del gruppo con dimensionamento automatico.

Note

Dopo aver creato il gruppo con dimensionamento automatico, è possibile utilizzare la console Elastic Load Balancing per creare altri listener. Ciò è utile se hai bisogno di creare un listener con un protocollo sicuro, ad esempio HTTPS, o un listener UDP. È possibile aggiungere altri listener al sistema di bilanciamento del carico esistenti, purché si utilizzino porte distinte.

Verifica lo stato di collegamento del sistema di bilanciamento del carico

Dopo aver collegato un sistema di bilanciamento del carico, questo entra in stato Adding durante la registrazione delle istanze nel gruppo. In seguito alla registrazione di tutte le istanze del gruppo nel sistema di bilanciamento del carico, questo entra in stato Added. Dopo che almeno un'istanza

sarà passata i controlli dell'integrità, entrerà nello stato `InService`. Una volta che il sistema di bilanciamento del carico sarà entrato nello stato `InService`, Dimensionamento automatico Amazon EC2 sarà in grado di terminare e sostituire tutte le istanze indicate come non integre. Se nessuna istanza registrata supera i controlli dell'integrità (ad esempio, a causa di un controllo configurato erroneamente), il sistema di bilanciamento del carico non entra nello stato `InService`. Dimensionamento automatico Amazon EC2 non termina e sostituisce le istanze.

Quando distacchi un sistema di bilanciamento del carico, questo entra in stato `Removing` durante l'annullamento della registrazione delle istanze nel gruppo. Dopo l'annullamento della registrazione, le istanze rimangono in esecuzione. Per impostazione predefinita il sistema imposta il sistema di svuotamento della (ritardo di deregistrazione) connessione per Application Load Balancer, Network Load Balancer e Gateway Load Balancer. Se lo svuotamento della connessione è abilitato, Elastic Load Balancing attende il completamento delle richieste in transito o la scadenza del timeout massimo (a seconda di quale situazione si verifica prima), per deregistrare le istanze.

Puoi verificare lo stato degli allegati utilizzando AWS Command Line Interface (AWS CLI) o AWS gli SDK. Non è possibile verificare lo stato del collegamento dalla console.

Da utilizzare AWS CLI per verificare lo stato dell'allegato

Il [describe-traffic-sources](#) comando seguente restituisce lo stato degli allegati di tutte le sorgenti di traffico per il gruppo Auto Scaling specificato.

```
aws autoscaling describe-traffic-sources --auto-scaling-group-name my-asg
```

L'esempio restituisce l'ARN del gruppo target Elastic Load Balancing collegato al gruppo con dimensionamento automatico, insieme allo stato di collegamento del gruppo target nell'elemento.

State

```
{
  "TrafficSources": [
    {
      "Identifier": "arn:aws:elasticloadbalancing:region:account-id:targetgroup/my-targets/1234567890123456",
      "State": "InService",
      "Type": "elbv2"
    }
  ]
}
```

Aggiungere o rimuovere zone di disponibilità

È possibile sfruttare i vantaggi della sicurezza e dell'affidabilità della ridondanza geografica includendo il gruppo con dimensionamento automatico su più zone di disponibilità di una regione che stai utilizzando e quindi allegando un sistema di bilanciamento del carico per distribuire il traffico in ingresso tra le zone.

Quando una zona di disponibilità perde di integrità o non è disponibile, Dimensionamento automatico Amazon EC2 avvia nuove istanze in una zona di disponibilità integra. Quando la zona di disponibilità non integra torna allo stato integro, Dimensionamento automatico Amazon EC2 ridistribuisce automaticamente le istanze dell'applicazione in modo uniforme su tutte le zone di disponibilità designate per il gruppo con dimensionamento automatico. Dimensionamento automatico Amazon EC2 esegue questa operazione tentando di avviare nuove istanze nella zona di disponibilità con il minor numero di istanze. Se il tentativo non riesce, Dimensionamento automatico Amazon EC2 tenta di avviare le istanze in altre zone di disponibilità, finché l'operazione non va a buon fine.

Elastic Load Balancing crea un nodo del sistema di bilanciamento del carico per ciascuna zona di disponibilità che abiliti per quest'ultimo. Se attivi il sistema di bilanciamento del carico su più zone, ogni nodo del sistema di bilanciamento del carico distribuisce le richieste nelle destinazioni registrate in tutte le zone di disponibilità attivate. Se il sistema di bilanciamento del carico su più zone non è attivo, ogni nodo del sistema di bilanciamento del carico distribuisce le richieste in modo uniforme sulle istanze registrate solo nella relativa zona di disponibilità.

Quando crei il gruppo con dimensionamento automatico, è necessario che specifichi almeno una zona di disponibilità. In seguito, è possibile espandere la disponibilità dell'applicazione aggiungendo una zona di disponibilità al gruppo con dimensionamento automatico e abilitando tale zona di disponibilità per il sistema di bilanciamento del carico (se tale funzionalità è supportata).

Indice

- [Aggiunta di una zona di disponibilità](#)
- [Rimozione di una zona di disponibilità](#)
- [Risorse correlate](#)
- [Limitazioni](#)

Aggiunta di una zona di disponibilità

Utilizza la procedura seguente per espandere il gruppo con dimensionamento automatico e il sistema di bilanciamento del carico su una sottorete in una zona di disponibilità aggiuntiva.

Per aggiungere una zona di disponibilità

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto a un gruppo esistente.

Si aprirà un riquadro diviso nella parte inferiore della pagina Gruppi con dimensionamento automatico.

3. Nella scheda Dettagli scegliere Rete, Modifica.
4. In Sottoreti, scegli la sottorete corrispondente alla zona di disponibilità che desideri aggiungere al gruppo con dimensionamento automatico.
5. Scegli Aggiorna.
6. Per aggiornare le zone di disponibilità per il sistema di bilanciamento del carico in modo che condivida le stesse zone del gruppo con dimensionamento automatico, completa la procedura seguente:
 - a. Nel pannello di navigazione, sotto Bilanciamento del carico, scegli Sistemi di bilanciamento del carico.
 - b. Scegli il sistema di bilanciamento del carico.
 - c. Esegui una di queste operazioni:
 - Per Application Load Balancer e Network Load Balancer:
 1. Nella scheda Descrizione, per Zone di disponibilità, scegli Modifica.
 2. In Modifica sottoreti, per Zone di disponibilità, seleziona la casella di controllo corrispondente alla zona di disponibilità da aggiungere. Se è presente una sola sottorete per la zona di disponibilità, è già selezionata. Se è presente più di una sottorete per la zona di disponibilità, dovrai selezionarne una.
 - Per Classic Load Balancer in un VPC:
 1. Nella scheda Istanze, scegli Modifica zone di disponibilità.
 2. Nella pagina Aggiungi e rimuovi sottoreti per Sottoreti disponibili seleziona la sottorete da aggiungere facendo clic sull'icona (+). La sottorete viene spostata nella sezione Sottoreti selezionate.
 - d. Selezionare Salva.

Rimozione di una zona di disponibilità

Utilizza la procedura seguente per rimuovere una zona di disponibilità dal gruppo con dimensionamento automatico e dal sistema di bilanciamento del carico.

Per rimuovere una zona di disponibilità

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto a un gruppo esistente.

Si aprirà un riquadro diviso nella parte inferiore della pagina Gruppi con dimensionamento automatico.

3. Nella scheda Dettagli scegliere Rete, Modifica.
4. In Subnets, scegli l'icona di eliminazione (X) per la sottorete corrispondente alla zona di disponibilità che desideri rimuovere dal gruppo con dimensionamento automatico. Se è presente più di una sottorete per quella zona, scegli l'icona di eliminazione (X) per ciascuna zona.
5. Scegli Aggiorna.
6. Per aggiornare le zone di disponibilità per il sistema di bilanciamento del carico in modo che condivida le stesse zone del gruppo con dimensionamento automatico, completa la procedura seguente:
 - a. Nel pannello di navigazione, sotto Bilanciamento del carico, scegli Sistemi di bilanciamento del carico.
 - b. Scegli il sistema di bilanciamento del carico.
 - c. Esegui una di queste operazioni:
 - Per Application Load Balancer e Network Load Balancer:
 1. Nella scheda Descrizione, per Zone di disponibilità, scegli Modifica.
 2. In Modifica sottoreti, per Zone di disponibilità, seleziona la casella di controllo per rimuovere la sottorete per zona di disponibilità.
 - Per Classic Load Balancer in un VPC:
 1. Nella scheda Istanze, scegli Modifica zone di disponibilità.
 2. Nella pagina Aggiungi e rimuovi sottoreti per Sottoreti disponibili rimuovi la sottorete facendo clic sull'icona di eliminazione (-). La sottorete viene spostata in Sottoreti disponibili.

- d. Selezionare Salva.

Risorse correlate

Quando cambi le zone di disponibilità, il dimensionamento automatico Amazon EC2 ribilancia il tuo gruppo. Ciò significa sostituire e ridistribuire alcune istanze. Per ulteriori informazioni, consulta [Esempio: distribuzione di istanze tra le zone di disponibilità](#).

Se nelle zone di disponibilità sono state registrate destinazioni che non sono abilitate per il sistema di bilanciamento del carico, il sistema di bilanciamento del carico non indirizza il traffico verso di esse. Per ulteriori informazioni consultare la guida [Come funziona Elastic Load Balancing](#) all'interno della Guida per l'utente di Elastic Load Balancing.

Limitazioni

Nell'aggiornare le zone di disponibilità abilitate per il sistema di bilanciamento del carico, devi essere consapevole delle seguenti limitazioni:

- Quando attivi una zona di disponibilità per il tuo sistema di bilanciamento del carico, devi specificare una sottorete per la zona di disponibilità. Ricorda che puoi abilitare una sola sottorete per zona di disponibilità, per il tuo sistema di bilanciamento del carico.
- Per i sistema di bilanciamento del carico connessi a Internet, le sottoreti specificate devono avere a disposizione almeno otto indirizzi IP.
- Per gli Application Load Balancer, devi abilitare almeno due zone di disponibilità.
- Per i Network Load Balancer non è possibile disabilitare le zone di disponibilità abilitate, ma è possibile abilitarne altre.
- Per Gateway Load Balancer, non è possibile disabilitare le zone di disponibilità abilitate, ma è possibile abilitarne di aggiuntive.

Esempi di utilizzo di Elastic Load Balancing con () AWS Command Line InterfaceAWS CLI

Utilizza il AWS CLI per collegare, scollegare e descrivere i sistemi di bilanciamento del carico e i gruppi target, aggiungere e rimuovere i controlli di integrità Elastic Load Balancing e modificare le zone di disponibilità abilitate.

Questo argomento mostra esempi di AWS CLI comandi che eseguono attività comuni per Amazon EC2 Auto Scaling.

⚠ Important

Per ulteriori esempi di comandi, vedere [aws elbv2](#) e [aws elb](#) nei AWS CLI Riferimenti ai comandi.

Indice

- [Collega il tuo gruppo target o Classic Load Balancer](#)
- [Descrivi i tuoi gruppi target o i Classic Load Balancer](#)
- [Come aggiungere controlli dell'integrità di Elastic Load Balancing](#)
- [Modifica le zone di disponibilità](#)
- [Scollega il tuo gruppo target o Classic Load Balancer](#)
- [Rimuovi Elastic Load Balancing dai controlli dell'integrità](#)
- [Comandi legacy](#)

Collega il tuo gruppo target o Classic Load Balancer

Usa il seguente [create-auto-scaling-group](#) comando per creare un gruppo Auto Scaling e collegare contemporaneamente un gruppo target specificandone l'Amazon Resource Name (ARN). Il gruppo di destinazione può essere associato a un Application Load Balancer, un Network Load Balancer o un Gateway Load Balancer.

Sostituisci i valori di esempio per `--auto-scaling-group-name`, `--vpc-zone-identifier`, `--min-size` e `--max-size`. Per l'opzione `--launch-template`, sostituisci *my-launch-template* e *1* con il nome e la versione di un modello di avvio per il tuo gruppo con dimensionamento automatico. Per l'opzione `--traffic-sources`, sostituisci l'ARN di esempio con l'ARN di un gruppo di destinazione per un Application Load Balancer, un Network Load Balancer o un Gateway Load Balancer.

```
aws autoscaling create-auto-scaling-group --auto-scaling-group-name my-asg \  
  --launch-template LaunchTemplateName=my-launch-template,Version='1' \  
  --vpc-zone-identifier "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782" \  
  --min-size 1 --max-size 5 \  
  --traffic-sources "arn:aws:elasticloadbalancing:us-east-1:123456789012:loadbalancer/targetgroup/tg-12345678"
```

```
--traffic-sources "Identifier=arn:aws:elasticloadbalancing:region:account-id:targetgroup/my-targets/12345678EXAMPLE1"
```

Utilizzate il [attach-traffic-sources](#) comando per collegare gruppi target aggiuntivi al gruppo Auto Scaling dopo la creazione.

Il comando seguente aggiunge un altro gruppo target allo stesso gruppo.

```
aws autoscaling attach-traffic-sources --auto-scaling-group-name my-asg \  
  --traffic-sources "Identifier=arn:aws:elasticloadbalancing:region:account-id:targetgroup/my-targets/12345678EXAMPLE2"
```

In alternativa, per collegare un Classic Load Balancer al gruppo, specifica le opzioni `--traffic-sources` e `--type` quando utilizzi `create-auto-scaling-group` o `attach-traffic-sources`, come nell'esempio seguente. Sostituisci *my-classic-load-balancer* con il nome di un Classic Load Balancer. Per l'opzione `--type`, specifica il valore di **elb**.

```
--traffic-sources "Identifier=my-classic-load-balancer" --type elb
```

Descrivi i tuoi gruppi target o i Classic Load Balancer

Per descrivere i sistemi di bilanciamento del carico o i gruppi target collegati al gruppo Auto Scaling, utilizzate il seguente comando. [describe-traffic-sources](#) Sostituisci *my-asg* con il nome del tuo gruppo.

```
aws autoscaling describe-traffic-sources --auto-scaling-group-name my-asg
```

L'esempio restituisce l'ARN dei gruppi di destinazione del bilanciamento del carico allegati al gruppo con dimensionamento automatico.

```
{  
  "TrafficSources": [  
    {  
      "Identifier": "arn:aws:elasticloadbalancing:region:account-id:targetgroup/my-targets/12345678EXAMPLE1",  
      "State": "InService",  
      "Type": "elbv2"  
    },  
    {
```

```
        "Identifier": "arn:aws:elasticloadbalancing:region:account-  
id:targetgroup/my-targets/12345678EXAMPLE",  
        "State": "InService",  
        "Type": "elbv2"  
    }  
]  
}
```

Per una spiegazione del campo State nell'output, consulta [Verifica lo stato di collegamento del sistema di bilanciamento del carico](#).

Come aggiungere controlli dell'integrità di Elastic Load Balancing

Per aggiungere i controlli di integrità di Elastic Load Balancing ai controlli di integrità eseguiti dal gruppo Auto Scaling sulle istanze, utilizzate il comando [update-auto-scaling-group](#) seguente e **ELB** specificate come valore per l'opzione. `--health-check-type` Sostituisci *my-asg* con il nome del tuo gruppo.

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg \  
--health-check-type "ELB"
```

Le nuove istanze spesso richiedono del tempo per un breve riscaldamento prima di poter superare un controllo di integrità. Se il periodo di prova non garantisce un tempo di riscaldamento sufficiente, le istanze potrebbero non sembrare pronte a servire il traffico. Dimensionamento automatico Amazon EC2 potrebbe ritenere queste istanze non integre e sostituirle.

Per aggiornare il periodo di tolleranza del controllo dell'integrità, utilizza l'opzione `--health-check-grace-period` quando utilizzi `update-auto-scaling-group`, come nell'esempio seguente. Sostituisci *300* con il numero di secondi per mantenere le nuove istanze in servizio prima di interromperle se risultano non integre.

```
--health-check-grace-period 300
```

Per ulteriori informazioni, consulta [Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico](#).

Modifica le zone di disponibilità

La modifica delle zone di disponibilità presenta alcune limitazioni da tenere presente. Per ulteriori informazioni, consulta [Limitazioni](#).

Per modificare le zone di disponibilità per un Application Load Balancer o un Network Load Balancer

1. Prima di modificare le zone di disponibilità del sistema di bilanciamento del carico, è consigliabile aggiornare innanzitutto le zone di disponibilità del gruppo con dimensionamento automatico per verificare la disponibilità dei tipi di istanza nelle zone specificate.

Per aggiornare le zone di disponibilità per il gruppo Auto Scaling, utilizzare il comando seguente [update-auto-scaling-group](#). Sostituisci gli ID di sottorete di esempio con gli ID delle sottoreti nelle zone di disponibilità per abilitarli. Le sottoreti specificate sostituiscono le sottoreti abilitate in precedenza. Sostituisci *my-asg* con il nome del tuo gruppo.

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg \  
--vpc-zone-identifier "subnet-41767929, subnet-cb663da2, subnet-8360a9e7"
```

2. Utilizzate il [describe-auto-scaling-groups](#) comando seguente per verificare che le istanze nelle nuove sottoreti siano state avviate. Se le istanze sono state avviate, viene visualizzato un elenco delle istanze e dei relativi stati. Sostituisci *my-asg* con il nome del tuo gruppo.

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

3. Usa il seguente comando [set-subnets](#) per specificare le sottoreti per il tuo sistema di bilanciamento del carico. Sostituisci gli ID di sottorete di esempio con gli ID delle sottoreti nelle zone di disponibilità per abilitarli. Puoi specificare una sola sottorete per ogni zona di disponibilità. Le sottoreti specificate sostituiscono le sottoreti abilitate in precedenza. Sostituisci *my-lb-arn* con l'ARN del tuo sistema di bilanciamento del carico.

```
aws elbv2 set-subnets --load-balancer-arn my-lb-arn \  
--subnets subnet-41767929 subnet-cb663da2 subnet-8360a9e7
```

Per modificare le zone di disponibilità per un Classic Load Balancer

1. Prima di modificare le zone di disponibilità del sistema di bilanciamento del carico, è consigliabile aggiornare innanzitutto le zone di disponibilità del gruppo con dimensionamento automatico per verificare la disponibilità dei tipi di istanza nelle zone specificate.

Per aggiornare le zone di disponibilità per il gruppo Auto Scaling, utilizzare il comando seguente [update-auto-scaling-group](#). Sostituisci gli ID di sottorete di esempio con gli ID delle sottoreti nelle

zone di disponibilità per abilitarli. Le sottoreti specificate sostituiscono le sottoreti abilitate in precedenza. Sostituisci *my-asg* con il nome del tuo gruppo.

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg \  
--vpc-zone-identifier "subnet-41767929,subnet-cb663da2"
```

- Utilizzate il [describe-auto-scaling-groups](#) comando seguente per verificare che le istanze nelle nuove sottoreti siano state avviate. Se le istanze sono state avviate, viene visualizzato un elenco delle istanze e dei relativi stati. Sostituisci *my-asg* con il nome del tuo gruppo.

```
aws autoscaling describe-auto-scaling-groups --auto-scaling-group-name my-asg
```

- Usa il seguente comando [attach-load-balancer-to-subnets](#) per abilitare una nuova zona di disponibilità per il tuo Classic Load Balancer. Sostituisci gli ID di sottorete di esempio con gli ID delle sottoreti nella zona di disponibilità per abilitarli. Sostituisci *my-lb* con il nome del tuo sistema di bilanciamento del carico.

```
aws elb attach-load-balancer-to-subnets --load-balancer-name my-lb \  
--subnets subnet-cb663da2
```

[Per disabilitare una zona di disponibilità, utilizzare il seguente comando -subnets.detach-load-balancer-from](#) Sostituisci gli ID di sottorete di esempio con gli ID delle sottoreti nella zona di disponibilità per disabilitarli. Sostituisci *my-lb* con il nome del tuo sistema di bilanciamento del carico.

```
aws elb detach-load-balancer-from-subnets --load-balancer-name my-lb \  
--subnets subnet-8360a9e7
```

Scollega il tuo gruppo target o Classic Load Balancer

Il [detach-traffic-sources](#) comando seguente separa un gruppo target dal gruppo Auto Scaling quando non è più necessario.

Per l'opzione `--auto-scaling-group-name`, sostituisci *my-asg* con il nome del tuo gruppo. Per l'opzione `--traffic-sources`, sostituisci l'ARN di esempio con l'ARN di un gruppo di destinazione per un Application Load Balancer, un Network Load Balancer o un Gateway Load Balancer.

```
aws autoscaling detach-traffic-sources --auto-scaling-group-name my-asg \  
--traffic-sources arn:aws:elasticloadbalancing:us-east-1:123456789012:loadbalancer/target
```



```
--traffic-sources "Identifier=arn:aws:elasticloadbalancing:region:account-id:targetgroup/my-targets/1234567890123456"
```

Per scollegare un Classic Load Balancer dal gruppo, specifica le opzioni `--traffic-sources` e `--type`, come nell'esempio seguente. Sostituisci `my-classic-load-balancer` con il nome di un Classic Load Balancer. Per l'opzione `--type`, specifica il valore di **e1b**.

```
--traffic-sources "Identifier=my-classic-load-balancer" --type elb
```

Rimuovi Elastic Load Balancing dai controlli dell'integrità

Per rimuovere i controlli di integrità di Elastic Load Balancing dal gruppo Auto Scaling, utilizzate il comando [update-auto-scaling-group](#) seguente e **EC2** specificatelo come valore per l'opzione. `--health-check-type` Sostituisci `my-asg` con il nome del tuo gruppo.

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg \  
--health-check-type "EC2"
```

Per ulteriori informazioni, consulta [Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico](#).

Comandi legacy

Gli esempi seguenti mostrano come utilizzare i comandi CLI legacy per collegare, scollegare e descrivere sistemi di bilanciamento del carico e gruppi target. Rimangono in questo documento come riferimento per tutti i clienti che desiderano utilizzarli. Continuiamo a supportare i comandi CLI legacy, ma ti consigliamo di utilizzare i nuovi comandi CLI "sorgenti di traffico", che possono collegare e scollegare più tipi di sorgenti di traffico. È possibile utilizzare sia i comandi CLI legacy che i comandi CLI "sorgenti di traffico" nello stesso gruppo con dimensionamento automatico.

Collega il tuo gruppo target o Classic Load Balancer (legacy)

Come collegare il tuo gruppo target

Il [create-auto-scaling-group](#) comando seguente crea un gruppo Auto Scaling con un gruppo target collegato. Specifica il nome della risorsa Amazon Resource Name (ARN) di un gruppo di destinazione per un Application Load Balancer, un Network Load Balancer o un Gateway Load Balancer.

```
aws autoscaling create-auto-scaling-group --auto-scaling-group-name my-asg \  
--target-group-arn arn:aws:elasticloadbalancing:region:account-id:targetgroup/my-targets/1234567890123456
```

```
--launch-template LaunchTemplateName=my-launch-template,Version='1' \  
--vpc-zone-identifier "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782" \  
--target-group-arns "arn:aws:elasticloadbalancing:region:account-id:targetgroup/my-  
targets/1234567890123456" \  
--min-size 1 --max-size 5
```

Il seguente comando [attach-load-balancer-target-groups](#) collega un gruppo target a un gruppo Auto Scaling esistente.

```
aws autoscaling attach-load-balancer-target-groups --auto-scaling-group-name my-asg \  
--target-group-arns "arn:aws:elasticloadbalancing:region:account-id:targetgroup/my-  
targets/1234567890123456"
```

Come collegare il proprio Classic Load Balancer

Il [create-auto-scaling-group](#) comando seguente crea un gruppo Auto Scaling con un Classic Load Balancer collegato.

```
aws autoscaling create-auto-scaling-group --auto-scaling-group-name my-asg \  
--launch-configuration-name my-launch-config \  
--vpc-zone-identifier "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782" \  
--load-balancer-names "my-load-balancer" \  
--min-size 1 --max-size 5
```

Il [attach-load-balancers](#) comando seguente collega il Classic Load Balancer specificato a un gruppo Auto Scaling esistente.

```
aws autoscaling attach-load-balancers --auto-scaling-group-name my-asg \  
--load-balancer-names my-lb
```

Come descrivere il proprio gruppo target o Classic Load Balancer (legacy)

Come descrivere gruppi target

Per descrivere i gruppi target associati a un gruppo di Auto Scaling, utilizzate il comando [describe-load-balancer-target-groups](#). L'esempio seguente elenca i gruppi di destinazione per *my-asg*.

```
aws autoscaling describe-load-balancer-target-groups --auto-scaling-group-name my-asg
```

Per descrivere un Classic Load Balancer

Per descrivere i Classic Load Balancer associati a un gruppo Auto Scaling, utilizzate il comando [describe-load-balancers](#). Nell'esempio seguente vengono elencati i Classic Load Balancer per *my-asg*.

```
aws autoscaling describe-load-balancers --auto-scaling-group-name my-asg
```

Scollega il tuo gruppo target o Classic Load Balancer (legacy)

Per scollegare un gruppo target

Il seguente comando [detach-load-balancer-target-groups](#) separa un gruppo target dal gruppo Auto Scaling quando non è più necessario.

```
aws autoscaling detach-load-balancer-target-groups --auto-scaling-group-name my-asg \  
--target-group-arns "arn:aws:elasticloadbalancing:region:account-id:targetgroup/my-  
targets/1234567890123456"
```

Come collegare un Classic Load Balancer

Il [detach-load-balancers](#) comando seguente scollega un Classic Load Balancer dal gruppo Auto Scaling quando non è più necessario.

```
aws autoscaling detach-load-balancers --auto-scaling-group-name my-asg \  
--load-balancer-names my-lb
```

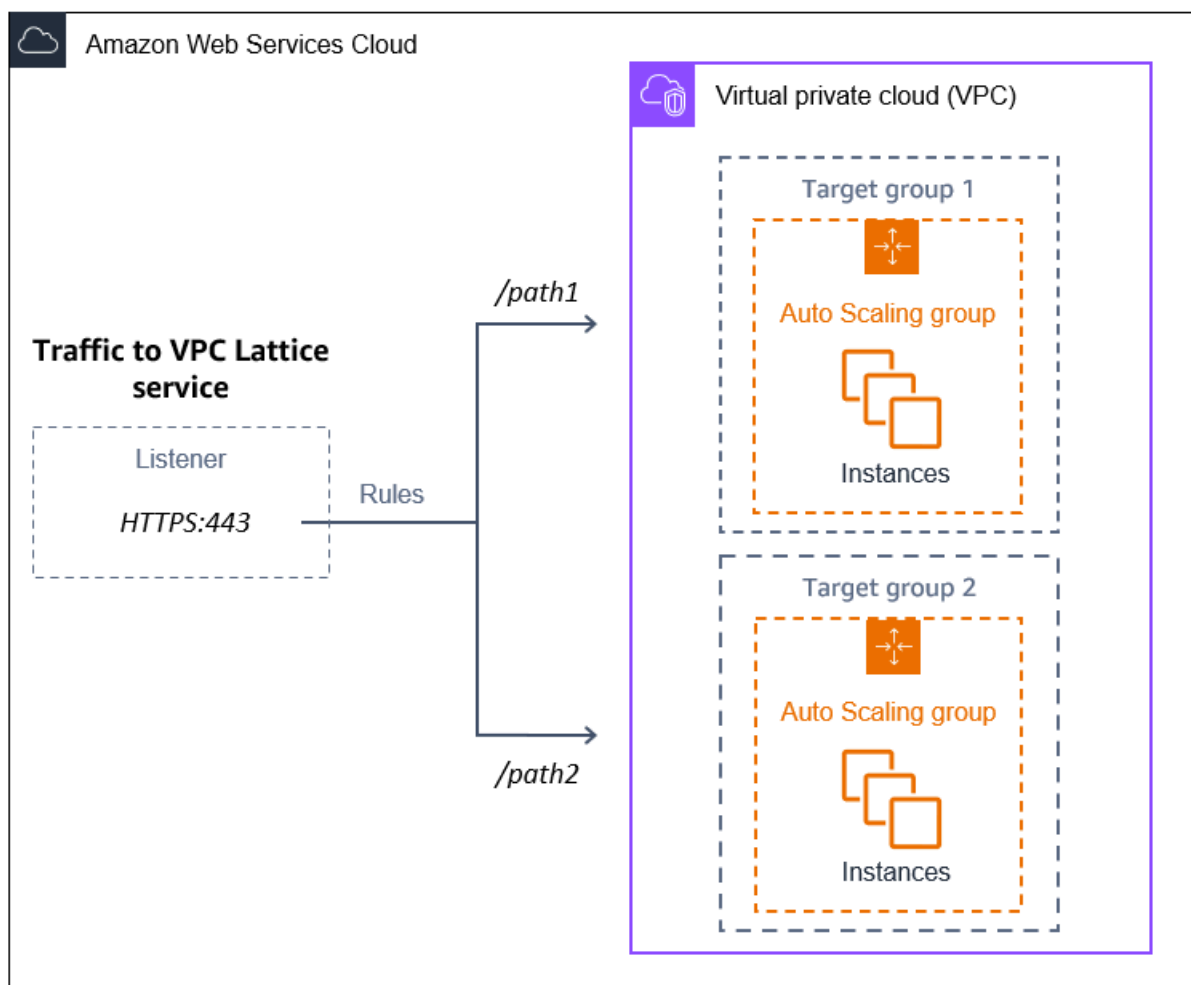
Indirizza il traffico verso il tuo gruppo con dimensionamento automatico con un gruppo target VPC Lattice

Puoi usare Amazon VPC Lattice per gestire il flusso di traffico e le chiamate API tra le tue applicazioni e i tuoi servizi eseguiti su risorse separate, come i gruppi con dimensionamento automatico o le funzioni Lambda. VPC Lattice è un servizio di rete di applicazioni che ti consente di connettere, proteggere e monitorare tutti i tuoi servizi su più account e cloud privati virtuali (VPC). Per saperne di più su VPC Lattice, consulta [Cos'è VPC Lattice?](#)

Per iniziare a usare VPC Lattice, crea innanzitutto le risorse VPC Lattice necessarie che consentano alle risorse in un VPC associato a una rete di servizi di connettersi tra loro. Queste risorse includono i servizi, gli ascoltatori, le regole dell'ascoltatore e i gruppi di destinazione.

Per associare un gruppo con dimensionamento automatico a un servizio VPC Lattice, crea un gruppo di destinazioni per il servizio che instradi le richieste alle istanze registrate per ID istanza e aggiungi un ascoltatore al servizio che invia le richieste al gruppo di destinazioni. Quindi, collega il gruppo target al tuo gruppo con dimensionamento automatico. Dimensionamento automatico Amazon EC2 registra automaticamente le istanze EC2 come destinazioni con il gruppo di destinazioni. Successivamente, quando Dimensionamento automatico Amazon EC2 deve terminare un'istanza, annulla automaticamente la registrazione dell'istanza dal gruppo di destinazione prima della chiusura.

Dopo aver collegato il gruppo target, questo diventa il punto di ingresso per tutte le richieste in arrivo al gruppo con dimensionamento automatico. Come mostra l'esempio nel diagramma seguente, le richieste in entrata possono quindi essere indirizzate al gruppo di destinazione appropriato utilizzando le regole dell'ascoltatore specificate per un servizio VPC Lattice.



Quando il traffico viene indirizzato tramite VPC Lattice al gruppo con dimensionamento automatico, VPC Lattice bilancia le richieste tra le istanze del gruppo utilizzando il sistema di bilanciamento del

carico round robin. VPC Lattice è inoltre in grado di monitorare lo stato delle istanze registrate e instradare il traffico solo alle istanze integre.

Per mantenere le istanze disponibili per le richieste in arrivo, puoi aggiungere opzionalmente i controlli dell'integrità di VPC Lattice al tuo gruppo con dimensionamento automatico. In questo modo, in caso di errore di una delle istanze EC2, il tuo gruppo con dimensionamento automatico avvia automaticamente una nuova istanza in sostituzione. Il comportamento dei controlli dell'integrità VPC Lattice è simile al comportamento dei controlli dell'integrità Elastic Load Balancing. I controlli dell'integrità di default per un gruppo con dimensionamento automatico sono solo i controlli dell'integrità EC2.

Per ulteriori informazioni su VPC Lattice, consulta [Semplifica la connettività, la sicurezza e il monitoraggio da servizio a servizio con Amazon VPC Lattice](#), ora disponibile a tutti sul blog. AWS

Indice

- [Preparazione al gruppo di destinazione VPC Lattice al gruppo con dimensionamento automatico](#)
- [Allegare un gruppo di destinazione VPC Lattice al gruppo con dimensionamento automatico](#)
- [Informazioni sullo stato di collegamento del gruppo di destinazione VPC Lattice](#)

Preparazione al gruppo di destinazione VPC Lattice al gruppo con dimensionamento automatico

Prima di collegare un gruppo di destinazione VPC Lattice al gruppo con dimensionamento automatico, devi completare i seguenti prerequisiti:

- È necessario aver già creato una rete di servizi VPC Lattice, un servizio, un ascoltatore e un gruppo target. Per ulteriori informazioni, consulta gli argomenti seguenti nella Guida per l'utente di VPC Lattice:
 - [Reti di servizio](#)
 - [Servizi](#)
 - [Listener](#)
 - [Gruppi target](#)
- Il gruppo target deve appartenere allo stesso Account AWS VPC e alla stessa regione del gruppo Auto Scaling.
- I gruppi di destinazione devono avere il tipo di destinazione `instance`. Quando si utilizza un gruppo con dimensionamento automatico, non è possibile specificare un tipo di destinazione `ip`.

- È necessario disporre di autorizzazioni IAM sufficienti per allegare il gruppo di destinazione al gruppo con dimensionamento automatico. La seguente policy di esempio mostra le autorizzazioni minime richieste necessarie per collegare e scollegare i gruppi target.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "autoscaling:AttachTrafficSources",
        "autoscaling:DetachTrafficSources",
        "autoscaling:DescribeTrafficSources",
        "vpc-lattice:RegisterTargets",
        "vpc-lattice:DeregisterTargets"
      ],
      "Resource": "*"
    }
  ]
}
```

- Se il modello di avvio per il tuo gruppo con dimensionamento automatico non contiene le impostazioni corrette per VPC Lattice, ad esempio un gruppo di sicurezza compatibile, devi aggiornare il modello di avvio. Le istanze esistenti non vengono aggiornate con le nuove impostazioni quando il modello di avvio viene modificato. Per aggiornare le istanze esistenti, puoi avviare un aggiornamento dell'istanza per sostituire le istanze. Per ulteriori informazioni, consulta [Usa un aggiornamento dell'istanza per aggiornare le istanze in un gruppo di Auto Scaling](#).
- Prima di abilitare i controlli dell'integrità di VPC Lattice sul tuo gruppo con dimensionamento automatico, puoi configurare un controllo dello stato basato sull'applicazione per verificare che l'applicazione risponda come previsto. Per ulteriori informazioni, consulta [Controlli dell'integrità dei propri gruppi target](#) nella Guida per l'utente di VPC Lattice.

Gruppi di sicurezza: regole in entrata e in uscita

I gruppi di sicurezza fungono da firewall per le istanze EC2 associate, controllando sia il traffico in entrata che in uscita a livello di istanza.

Note

La configurazione di rete è sufficientemente complessa da consentire di creare un nuovo gruppo di sicurezza per l'utilizzo con VPC Lattice. Inoltre, semplifica l'assistenza nel caso AWS Support in cui sia necessario contattarle. Le seguenti sezioni si basano sul presupposto che l'utente segua questa raccomandazione.

Per saperne di più sulla creazione di gruppi di sicurezza per VPC Lattice che puoi utilizzare con il tuo gruppo con dimensionamento automatico, consulta [Controllare il traffico utilizzando i gruppi di sicurezza nella Guida per l'utente di VPC Lattice](#). Per risolvere i problemi relativi al flusso di traffico, consulta la Guida per l'utente VPC Lattice.

Per informazioni su come creare un gruppo di sicurezza, consulta la sezione [Creazione di un gruppo di sicurezza](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux e utilizza la tabella seguente per determinare quali opzioni selezionare.

Opzione	Valore
Nome	Un nome facile da ricordare.
Descrizione	Una descrizione che aiuti nell'identificazione del gruppo di sicurezza.
VPC	Lo stesso VPC del gruppo con dimensionamento automatico.

Regole in entrata

Al momento della sua creazione, un gruppo di sicurezza è privo di regole in entrata. Non è consentito alcun traffico in entrata da client all'interno di una rete di servizi VPC Lattice verso l'istanza fino a quando al gruppo di sicurezza non vengono aggiunte regole in entrata.

Per consentire ai client all'interno di una rete di servizi VPC Lattice di connettersi alle istanze del gruppo con d, il gruppo di sicurezza per il gruppo con dimensionamento automatico deve essere configurato correttamente. In questo caso, assegnagli una regola in entrata per consentire il traffico dal nome dell'elenco di prefissi AWS gestiti per VPC Lattice, anziché da un indirizzo IP specifico.

L'elenco dei prefissi VPC Lattice è un intervallo di indirizzi IP utilizzati da VPC Lattice nella notazione CIDR. Per ulteriori informazioni, consulta [Work with AWS-managed prefix lists](#) nella Amazon VPC User Guide.

Per informazioni su come aggiungere regole a un gruppo di sicurezza, consulta [Aggiungi regole al tuo gruppo di sicurezza](#) nella Guida per l'utente di Amazon VPC e utilizza la tabella seguente per determinare quali opzioni selezionare.

Opzione	Valore
Regola HTTP	Tipo: HTTP Origine: com.amazonaws. <i>region</i> .vpc-lattice
Regola HTTPS	Tipo: HTTPS Origine: com.amazonaws. <i>region</i> .vpc-lattice

Il gruppo di sicurezza è stateful: consente il traffico dai client all'interno della rete di servizi VPC Lattice alle istanze nel gruppo con dimensionamento automatico e rispedisce le risposte al client che aveva lasciato in precedenza.

Regole in uscita

Per impostazione predefinita, un gruppo di sicurezza include una regola in uscita che autorizza tutto il traffico in uscita. Facoltativamente, puoi rimuovere questa regola predefinita e aggiungere una regola in uscita per soddisfare esigenze di sicurezza specifiche.

Limitazioni

- [I gruppi di istanze miste](#) non sono supportati. Se provi a collegare un gruppo target VPC Lattice a un gruppo con dimensionamento automatico con una policy di istanze miste, ricevi il messaggio di errore Attualmente, i gruppi con dimensionamento automatico con istanze miste non possono essere integrati con un servizio VPC Lattice. Questo perché l'algoritmo di bilanciamento del carico distribuisce uniformemente il carico su tutte le risorse disponibili e presuppone che le istanze siano sufficientemente simili da gestire carichi uguali.

Allegare un gruppo di destinazione VPC Lattice al gruppo con dimensionamento automatico

In questo argomento viene descritto come allegare un gruppo di destinazione VPC Lattice a un gruppo con dimensionamento automatico. Descrive inoltre come attivare i controlli dell'integrità di VPC Lattice per consentire a Dimensionamento automatico Amazon EC2 di sostituire le istanze che VPC Lattice segnala come non integre.

Per impostazione predefinita, Dimensionamento automatico Amazon EC2 sostituisce solo le istanze non integre o irraggiungibili in base ai controlli dell'integrità di Amazon EC2. Se attivi i controlli dell'integrità di VPC Lattice, Dimensionamento automatico Amazon EC2 può sostituire un'istanza in esecuzione se uno dei gruppi target VPC Lattice che colleghi al gruppo con dimensionamento automatico la segnala come non integra. Per ulteriori informazioni, consulta [Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico](#).

Important

Prima di continuare, completa tutti i [prerequisiti indicati](#) nella sezione precedente.

Collega un gruppo target VPC Lattice

È possibile associare uno o più gruppi target a un gruppo Auto Scaling quando si crea o si aggiorna il gruppo.

Console

Segui i passaggi indicati in questa sezione per utilizzare la console per:

- Collegare un gruppo di destinazione VPC Lattice a un gruppo con dimensionamento automatico
- Attivare i controlli dell'integrità per VPC Lattice

Come collegare un gruppo di destinazione VPC Lattice a un nuovo gruppo con dimensionamento automatico

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.

2. Sulla barra di navigazione nella parte superiore della schermata, seleziona la Regione AWS in cui è stato creato il gruppo di destinazioni.
3. Seleziona Crea un gruppo con dimensionamento automatico.
4. Nei passaggi 1 e 2, scegli le opzioni desiderate e procedi al Passaggio 3: configura le opzioni avanzate.
5. Per le opzioni di integrazione VPC Lattice, scegli il servizio Collega a un servizio VPC Lattice.
6. In Scegli il gruppo target VPC Lattice, scegli il tuo gruppo target.
7. (Facoltativo) Per i controlli dell'integrità, tipi di controlli dell'integrità aggiuntivi, seleziona Attiva i controlli dell'integrità di VPC Lattice.
8. (Facoltativo) In Periodo di tolleranza del controllo dell'integrità, inserisci il tempo in secondi. Questo è il periodo di tempo che Dimensionamento automatico Amazon EC2 deve attendere prima di controllare lo stato di integrità di un'istanza dopo che questa è entrata nello stato InService. Per ulteriori informazioni, consulta [Impostazione del periodo di tolleranza del controllo dell'integrità per un gruppo con scalabilità automatica](#).
9. Procedi e crea il gruppo con dimensionamento automatico. Le istanze verranno registrate automaticamente nel gruppo di destinazione di VPC Lattice carico dopo la creazione del gruppo con dimensionamento automatico.

Per collegare un gruppo target VPC Lattice a un gruppo con dimensionamento automatico esistente

Per collegare un gruppo target per un servizio a un gruppo con dimensionamento automatico, utilizza la procedura seguente.

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.
2. Seleziona la casella di controllo accanto al gruppo con dimensionamento automatico.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Nella scheda Dettagli, scegli le opzioni di integrazione VPC Lattice, Modifica.
4. Per le opzioni di integrazione VPC Lattice, scegli il servizio Collega a un servizio VPC Lattice.
5. In Scegli il gruppo target VPC Lattice, scegli il tuo gruppo target.
6. Scegli Aggiorna.

Quando hai finito di collegare il gruppo target, puoi facoltativamente attivare i controlli dell'integrità che lo utilizzano.

Attivare i controlli dell'integrità per VPC Lattice

1. Nella scheda Dettagli scegli Controlli dell'integrità, Modifica.
2. Per i controlli dell'integrità, tipi di controlli dell'integrità aggiuntivi, seleziona Attiva i controlli dell'integrità di VPC Lattice.
3. In Periodo di tolleranza del controllo dell'integrità, inserisci il tempo in secondi. Questo è il periodo di tempo che Dimensionamento automatico Amazon EC2 deve attendere prima di controllare lo stato di integrità di un'istanza dopo che questa è entrata nello stato InService. Per ulteriori informazioni, consulta [Impostazione del periodo di tolleranza del controllo dell'integrità per un gruppo con scalabilità automatica](#).
4. Scegli Aggiorna.

AWS CLI

Segui la procedura descritta in questa sezione per utilizzare: AWS CLI

- Collegare un gruppo di destinazione VPC Lattice a un gruppo con dimensionamento automatico
- Attivare i controlli dell'integrità per VPC Lattice

Come collegare un gruppo di destinazione VPC Lattice a un gruppo con dimensionamento automatico

Utilizza il seguente [create-auto-scaling-group](#) comando per creare un gruppo Auto Scaling e collegare contemporaneamente un gruppo target VPC Lattice specificando il relativo Amazon Resource Name (ARN).

Sostituisci i valori di esempio per `--auto-scaling-group-name`, `--vpc-zone-identifier`, `--min-size` e `--max-size`. Per l'opzione `--launch-template`, sostituisci *my-launch-template* e *1* con il nome e la versione del modello di avvio che hai creato per le istanze registrate in un gruppo target VPC Lattice. Come `--traffic-sources` opzione, sostituisci l'ARN di esempio con l'ARN del gruppo di destinazione VPC Lattice.

```
aws autoscaling create-auto-scaling-group --auto-scaling-group-name my-asg \  
--launch-template LaunchTemplateName=my-launch-template,Version='1' \  
--vpc-zone-identifier "subnet-5ea0c127,subnet-6194ea3b,subnet-c934b782" \  

```

```
--min-size 1 --max-size 5 \  
--traffic-sources "Identifier=arn:aws:vpc-lattice:region:account-id:targetgroup/  
tg-0e2f2665eEXAMPLE"
```

Usa il seguente [attach-traffic-sources](#) comando per collegare un gruppo target VPC Lattice a un gruppo Auto Scaling dopo che è già stato creato.

```
aws autoscaling attach-traffic-sources --auto-scaling-group-name my-asg \  
--traffic-sources "Identifier=arn:aws:vpc-lattice:region:account-id:targetgroup/  
tg-0e2f2665eEXAMPLE"
```

Come attivare i controlli dell'integrità per VPC Lattice

Se hai configurato un controllo dell'integrità basato sulle applicazioni per il tuo gruppo target VPC Lattice, puoi attivare questi controlli dell'integrità. Usa il [update-auto-scaling-group](#) comando [create-auto-scaling-group](#) o con l' `--health-check-type` opzione e il valore di **VPC_LATTICE**. Per specificare il periodo di tolleranza per i controlli dell'integrità eseguiti dal gruppo con dimensionamento automatico, includi l'opzione `--health-check-grace-period` e fornisci il suo valore in secondi.

```
--health-check-type "VPC_LATTICE" --health-check-grace-period 60
```

Scollegare un gruppo target VPC Lattice

Quando il sistema VPC Lattice non ti serve più, utilizza la procedura seguente per scollegarlo dal gruppo con dimensionamento automatico.

Console

Segui i passaggi indicati in questa sezione per utilizzare la console per:

- Come scollegare un gruppo target VPC Lattice da un gruppo con dimensionamento automatico
- Come disattivare i controlli dell'integrità per VPC Lattice

Come scollegare un gruppo target VPC Lattice da un gruppo con dimensionamento automatico

1. Apri la console Amazon EC2 all'indirizzo <https://console.aws.amazon.com/ec2/> e scegli Gruppi con dimensionamento automatico dal pannello di navigazione.

2. Seleziona la casella di controllo accanto a un gruppo esistente.

Si aprirà un riquadro diviso nella parte inferiore della pagina.

3. Nella scheda Dettagli, scegli le opzioni di integrazione VPC Lattice, Modifica.
4. In Opzioni di integrazione VPC Lattice, scegli l'icona di eliminazione (X) accanto al gruppo target.
5. Scegli Aggiorna.

Quando hai finito di scollegare il gruppo target, puoi disattivare i controlli dell'integrità di VPC Lattice.

Come disattivare i controlli dell'integrità per VPC Lattice

1. Nella scheda Dettagli scegli Controlli dell'integrità, Modifica.
2. Per i controlli dell'integrità, tipi di controlli dell'integrità aggiuntivi, deseleziona Attiva i controlli dell'integrità di VPC Lattice.
3. Scegli Aggiorna.

AWS CLI

Segui la procedura descritta in questa sezione per AWS CLI utilizzare:

- Come scollegare un gruppo target VPC Lattice da un gruppo con dimensionamento automatico
- Come disattivare i controlli dell'integrità per VPC Lattice

Usa il [detach-traffic-sources](#) comando per staccare un gruppo target dal tuo gruppo Auto Scaling quando non ne hai più bisogno.

```
aws autoscaling detach-traffic-sources --auto-scaling-group-name my-asg \  
  --traffic-sources "Identifier=arn:aws:vpc-lattice:region:account-id:targetgroup/  
tg-0e2f2665eEXAMPLE"
```

Per aggiornare i controlli di integrità su un gruppo Auto Scaling in modo che non utilizzi più i controlli di integrità VPC Lattice, usa il comando. [update-auto-scaling-group](#) Includi l'opzione `--health-check-type` e un valore pari a **EC2**.

```
aws autoscaling update-auto-scaling-group --auto-scaling-group-name my-asg \  
  --health-check-type EC2
```

```
--health-check-type "EC2"
```

Informazioni sullo stato di collegamento del gruppo di destinazione VPC Lattice

Dopo aver collegato un gruppo target VPC Lattice a un gruppo con dimensionamento automatico, entra nello stato `Adding` durante la registrazione delle istanze nel gruppo. In seguito alla registrazione di tutte le istanze del gruppo nel sistema di bilanciamento del carico, questo entra in stato `Added`. Dopo che almeno un'istanza sarà passata i controlli dell'integrità, entrerà nello stato `InService`. Una volta che il gruppo target sarà entrato nello stato `InService`, Dimensionamento automatico Amazon EC2 sarà in grado di terminare e sostituire tutte le istanze indicate come non integre. Se nessuna istanza registrata supera i controlli dell'integrità (ad esempio, a causa di un controllo configurato erroneamente), il gruppo target non entra nello stato `InService`. Dimensionamento automatico Amazon EC2 non termina e sostituisce le istanze.

Quando si scollega un gruppo target per un servizio, questo entra nello stato di `Removing` delle istanze del gruppo. Dopo l'annullamento della registrazione, le istanze rimangono in esecuzione. Per impostazione predefinita viene abilitato lo svuotamento della connessione (ritardo annullamento della registrazione). Se lo svuotamento della connessione è abilitato, VPC Lattice attende il completamento delle richieste in transito o la scadenza del timeout massimo (a seconda di quale situazione si verifica prima), per annullare la registrazione delle istanze.

Puoi verificare lo stato degli allegati utilizzando AWS Command Line Interface (AWS CLI) o gli SDK. AWS Non è possibile verificare lo stato del collegamento dalla console.

Da utilizzare AWS CLI per verificare lo stato dell'allegato

Il [describe-traffic-sources](#) comando seguente restituisce lo stato degli allegati di tutte le sorgenti di traffico per il gruppo Auto Scaling specificato.

```
aws autoscaling describe-traffic-sources --auto-scaling-group-name my-asg
```

L'esempio restituisce l'ARN del gruppo target Elastic Load Balancing collegato al gruppo con dimensionamento automatico, insieme allo stato di collegamento del gruppo target nell'elemento `State`.

```
{  
  "TrafficSources": [  

```

```
{
  "Identifier": "arn:aws:vpc-lattice:region:account-
id:targetgroup/tg-0e2f2665eEXAMPLE",
  "State": "InService",
  "Type": "vpc-lattice"
}
]
```

Utilizzato EventBridge per gestire gli eventi di Auto Scaling

Amazon EventBridge, precedentemente chiamato CloudWatch Events, ti aiuta a configurare regole basate sugli eventi che monitorano le risorse e avviano azioni mirate che utilizzano altri servizi. AWS

Gli eventi di Amazon EC2 Auto Scaling vengono forniti quasi in tempo EventBridge reale. È possibile stabilire EventBridge regole che richiamano azioni e notifiche programmatiche in risposta a una varietà di questi eventi. Ad esempio, mentre le istanze sono in fase di avvio o chiusura, è possibile richiamare una AWS Lambda funzione per eseguire un'attività preconfigurata.

Gli obiettivi delle EventBridge regole possono includere AWS Lambda funzioni, argomenti di Amazon SNS, destinazioni API, bus di eventi e molto altro Account AWS ancora. Per informazioni sugli obiettivi supportati, consulta [Amazon EventBridge targets](#) nella Amazon EventBridge User Guide.

Inizia creando EventBridge regole con un esempio che utilizza un argomento e una EventBridge regola di Amazon SNS. Quindi, quando un utente avvia l'aggiornamento di un'istanza, Amazon SNS ti invia una notifica via e-mail ogni volta che viene raggiunto un checkpoint. Per ulteriori informazioni, consulta [Crea EventBridge regole, ad esempio eventi di aggiornamento.](#)

Indice

- [Riferimento eventi Dimensionamento automatico Amazon EC2](#)
- [Esempi di eventi e modelli relativi al pool caldo](#)
- [Crea EventBridge regole](#)

Riferimento eventi Dimensionamento automatico Amazon EC2

Con Amazon EventBridge, puoi creare regole che corrispondono agli eventi in arrivo e indirizzarli verso le destinazioni per l'elaborazione.

Indice

- [Eventi dell'operazione del ciclo di vita](#)
- [Eventi di dimensionamento riusciti](#)
- [Eventi di dimensionamento non riusciti](#)
- [Eventi di aggiornamento dell'istanza](#)

Eventi dell'operazione del ciclo di vita

Quando aggiungi lifecycle hook al tuo gruppo Auto Scaling, Amazon EC2 Auto Scaling invia eventi a quando un'istanza passa EventBridge allo stato di attesa. Gli eventi vengono prodotti nel miglior modo possibile.

Event types (Tipi di evento)

- [Azione di dimensionamento orizzontale del ciclo di vita](#)
- [Operazione del ciclo di vita del dimensionamento verticale](#)

Azione di dimensionamento orizzontale del ciclo di vita

Il seguente evento di esempio mostra che il Dimensionamento automatico Amazon EC2 ha impostato lo stato Pending:Wait per un'istanza a causa di un hook del ciclo di vita dell'avvio.

```
{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Instance-launch Lifecycle Action",
  "source": "aws.autoscaling",
  "account": "123456789012",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-west-2",
  "resources": [
    "auto-scaling-group-arn"
  ],
  "detail": {
    "LifecycleActionToken": "87654321-4321-4321-4321-210987654321",
    "AutoScalingGroupName": "my-asg",
    "LifecycleHookName": "my-lifecycle-hook",
    "EC2InstanceId": "i-1234567890abcdef0",
    "LifecycleTransition": "autoscaling:EC2_INSTANCE_LAUNCHING",
    "NotificationMetadata": "additional-info",
    "Origin": "EC2",
```



```

    "Destination": "AutoScalingGroup"
  }
}

```

Operazione del ciclo di vita del dimensionamento verticale

Il seguente evento di esempio mostra che il Dimensionamento automatico Amazon EC2 ha impostato lo stato `Terminating:Wait` per un'istanza a causa di un hook del ciclo di vita della terminazione.

Important

Quando un gruppo con dimensionamento automatico restituisce istanze a un pool caldo con dimensionamento verticale, anche la restituzione delle istanze al pool caldo può generare eventi `EC2 Instance-terminate Lifecycle Action`. Gli eventi che vengono consegnati quando un'istanza passa allo stato di attesa su dimensionamento verticale hanno `WarmPool` come valore per `Destination`. Per ulteriori informazioni, consulta [Instance reuse policy](#).

```

{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Instance-terminate Lifecycle Action",
  "source": "aws.autoscaling",
  "account": "123456789012",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-west-2",
  "resources": [
    "auto-scaling-group-arn"
  ],
  "detail": {
    "LifecycleActionToken": "87654321-4321-4321-4321-210987654321",
    "AutoScalingGroupName": "my-asg",
    "LifecycleHookName": "my-lifecycle-hook",
    "EC2InstanceId": "i-1234567890abcdef0",
    "LifecycleTransition": "autoscaling:EC2_INSTANCE_TERMINATING",
    "NotificationMetadata": "additional-info",
    "Origin": "AutoScalingGroup",
    "Destination": "EC2"
  }
}

```

Eventi di dimensionamento riusciti

Gli esempi seguenti mostrano i tipi di eventi per eventi di dimensionamento di successo. Gli eventi vengono prodotti nel miglior modo possibile.

Event types (Tipi di evento)

- [Evento di dimensionamento orizzontale riuscito](#)
- [Evento di dimensionamento verticale riuscito](#)

Evento di dimensionamento orizzontale riuscito

Il seguente evento di esempio mostra che il Dimensionamento automatico Amazon EC2 è riuscito ad avviare un'istanza.

```
{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Instance Launch Successful",
  "source": "aws.autoscaling",
  "account": "123456789012",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-west-2",
  "resources": [
    "auto-scaling-group-arn",
    "instance-arn"
  ],
  "detail": {
    "StatusCode": "InProgress",
    "Description": "Launching a new EC2 instance: i-12345678",
    "AutoScalingGroupName": "my-asg",
    "ActivityId": "87654321-4321-4321-4321-210987654321",
    "Details": {
      "Availability Zone": "us-west-2b",
      "Subnet ID": "subnet-12345678"
    },
    "RequestId": "12345678-1234-1234-1234-123456789012",
    "StatusMessage": "",
    "EndTime": "yyyy-mm-ddThh:mm:ssZ",
    "EC2InstanceId": "i-1234567890abcdef0",
    "StartTime": "yyyy-mm-ddThh:mm:ssZ",
    "Cause": "description-text",
    "Origin": "EC2",
  }
}
```

```
"Destination": "AutoScalingGroup"  
}  
}
```

Evento di dimensionamento verticale riuscito

Il seguente evento di esempio mostra che il Dimensionamento automatico Amazon EC2 è riuscito a terminare un'istanza.

Important

Quando un gruppo con dimensionamento automatico restituisce istanze a un pool caldo con dimensionamento verticale, anche la restituzione delle istanze al pool caldo può generare eventi EC2 Instance Terminate Successful. Gli eventi che vengono consegnati quando un'istanza ritorna correttamente nel pool caldo hanno WarmPool1 come valore per Destination. Per ulteriori informazioni, consulta [Instance reuse policy](#).

```
{  
  "version": "0",  
  "id": "12345678-1234-1234-1234-123456789012",  
  "detail-type": "EC2 Instance Terminate Successful",  
  "source": "aws.autoscaling",  
  "account": "123456789012",  
  "time": "yyyy-mm-ddThh:mm:ssZ",  
  "region": "us-west-2",  
  "resources": [  
    "auto-scaling-group-arn",  
    "instance-arn"  
  ],  
  "detail": {  
    "StatusCode": "InProgress",  
    "Description": "Terminating EC2 instance: i-12345678",  
    "AutoScalingGroupName": "my-asg",  
    "ActivityId": "87654321-4321-4321-4321-210987654321",  
    "Details": {  
      "Availability Zone": "us-west-2b",  
      "Subnet ID": "subnet-12345678"  
    },  
    "RequestId": "12345678-1234-1234-1234-123456789012",  
    "StatusMessage": "",  
    "EndTime": "yyyy-mm-ddThh:mm:ssZ",  
  }  
}
```

```
"EC2InstanceId": "i-1234567890abcdef0",
"StartTime": "yyyy-mm-ddThh:mm:ssZ",
"Cause": "description-text",
"Origin": "AutoScalingGroup",
"Destination": "EC2"
}
}
```

Eventi di dimensionamento non riusciti

Gli esempi seguenti mostrano i tipi di eventi per eventi di dimensionamento non riusciti. Gli eventi vengono prodotti nel miglior modo possibile.

Event types (Tipi di evento)

- [Evento di dimensionamento orizzontale non riuscito](#)
- [Evento di dimensionamento verticale non riuscito](#)

Evento di dimensionamento orizzontale non riuscito

Il seguente evento di esempio mostra che il Dimensionamento automatico Amazon EC2 non è riuscito ad avviare un'istanza.

```
{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Instance Launch Unsuccessful",
  "source": "aws.autoscaling",
  "account": "123456789012",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-west-2",
  "resources": [
    "auto-scaling-group-arn",
    "instance-arn"
  ],
  "detail": {
    "StatusCode": "Failed",
    "AutoScalingGroupName": "my-asg",
    "ActivityId": "87654321-4321-4321-4321-210987654321",
    "Details": {
      "Availability Zone": "us-west-2b",
      "Subnet ID": "subnet-12345678"
    }
  }
}
```

```

    },
    "RequestId": "12345678-1234-1234-1234-123456789012",
    "StatusMessage": "message-text",
    "EndTime": "yyyy-mm-ddThh:mm:ssZ",
    "EC2InstanceId": "i-1234567890abcdef0",
    "StartTime": "yyyy-mm-ddThh:mm:ssZ",
    "Cause": "description-text",
    "Origin": "EC2",
    "Destination": "AutoScalingGroup"
  }
}

```

Evento di dimensionamento verticale non riuscito

Il seguente evento di esempio mostra che il Dimensionamento automatico Amazon EC2 non è riuscito a terminare un'istanza.

Important

Quando un gruppo con dimensionamento automatico restituisce istanze a un pool caldo con dimensionamento verticale, anche la mancata restituzione delle istanze al pool caldo può generare eventi EC2 Instance Terminate Unsuccessful. Gli eventi che vengono consegnati quando un'istanza non ritorna correttamente nel pool caldo hanno WarmPool come valore per Destination. Per ulteriori informazioni, consulta [Instance reuse policy](#).

```

{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Instance Terminate Unsuccessful",
  "source": "aws.autoscaling",
  "account": "123456789012",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-west-2",
  "resources": [
    "auto-scaling-group-arn",
    "instance-arn"
  ],
  "detail": {
    "StatusCode": "Failed",
    "AutoScalingGroupName": "my-asg",
    "ActivityId": "87654321-4321-4321-4321-210987654321",

```

```

"Details": {
  "Availability Zone": "us-west-2b",
  "Subnet ID": "subnet-12345678"
},
"RequestId": "12345678-1234-1234-1234-123456789012",
"StatusMessage": "message-text",
"EndTime": "yyyy-mm-ddThh:mm:ssZ",
"EC2InstanceId": "i-1234567890abcdef0",
"StartTime": "yyyy-mm-ddThh:mm:ssZ",
"Cause": "description-text",
"Origin": "AutoScalingGroup",
"Destination": "EC2"
}
}

```

Eventi di aggiornamento dell'istanza

Di seguito sono riportati eventi di esempio della funzione di aggiornamento dell'istanza. Gli eventi vengono prodotti nel miglior modo possibile.

Event types (Tipi di evento)

- [Checkpoint raggiunto](#)
- [Aggiornamento dell'istanza iniziato](#)
- [Aggiornamento dell'istanza riuscito](#)
- [Aggiornamento dell'istanza non riuscito](#)
- [Aggiornamento dell'istanza annullato](#)

Checkpoint raggiunto

Quando il numero di istanze sostituite raggiunge la soglia percentuale definita per il punto di controllo, Dimensionamento automatico Amazon EC2 invia il seguente evento.

```

{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Auto Scaling Instance Refresh Checkpoint Reached",
  "source": "aws.autoscaling",
  "account": "123456789012",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-west-2",

```

```

"resources": [
  "auto-scaling-group-arn"
],
"detail": {
  "InstanceRefreshId": "ab00cf8f-9126-4f3c-8010-dbb8cad6fb86",
  "AutoScalingGroupName": "my-asg",
  "CheckpointPercentage": "50",
  "CheckpointDelay": "300"
}
}

```

Aggiornamento dell'istanza iniziato

Dimensionamento automatico Amazon EC2 emette il seguente evento quando lo stato di un aggiornamento di un'istanza cambia in InProgress.

```

{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Auto Scaling Instance Refresh Started",
  "source": "aws.autoscaling",
  "account": "123456789012",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-west-2",
  "resources": [
    "auto-scaling-group-arn"
  ],
  "detail": {
    "InstanceRefreshId": "c613620e-07e2-4ed2-a9e2-ef8258911ade",
    "AutoScalingGroupName": "my-asg"
  }
}

```

Aggiornamento dell'istanza riuscito

Dimensionamento automatico Amazon EC2 emette il seguente evento quando lo stato di un aggiornamento di un'istanza cambia in Succeeded.

```

{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Auto Scaling Instance Refresh Succeeded",
  "source": "aws.autoscaling",

```

```

"account": "123456789012",
"time": "yyyy-mm-ddThh:mm:ssZ",
"region": "us-west-2",
"resources": [
  "auto-scaling-group-arn"
],
"detail": {
  "InstanceRefreshId": "c613620e-07e2-4ed2-a9e2-ef8258911ade",
  "AutoScalingGroupName": "my-asg"
}
}

```

Aggiornamento dell'istanza non riuscito

Dimensionamento automatico Amazon EC2 emette il seguente evento quando lo stato di un aggiornamento di un'istanza cambia in Failed.

```

{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Auto Scaling Instance Refresh Failed",
  "source": "aws.autoscaling",
  "account": "123456789012",
  "time": "yyyy-mm-ddThh:mm:ssZ",
  "region": "us-west-2",
  "resources": [
    "auto-scaling-group-arn"
  ],
  "detail": {
    "InstanceRefreshId": "c613620e-07e2-4ed2-a9e2-ef8258911ade",
    "AutoScalingGroupName": "my-asg"
  }
}

```

Aggiornamento dell'istanza annullato

Dimensionamento automatico Amazon EC2 emette il seguente evento quando lo stato di un aggiornamento di un'istanza cambia in Cancelled.

```

{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Auto Scaling Instance Refresh Cancelled",

```



```
"source": "aws.autoscaling",
"account": "123456789012",
"time": "yyyy-mm-ddThh:mm:ssZ",
"region": "us-west-2",
"resources": [
  "auto-scaling-group-arn"
],
"detail": {
  "InstanceRefreshId": "c613620e-07e2-4ed2-a9e2-ef8258911ade",
  "AutoScalingGroupName": "my-asg"
}
}
```

Esempi di eventi e modelli relativi al pool caldo

Amazon EC2 Auto Scaling supporta diversi modelli predefiniti in Amazon EventBridge. Ciò semplifica il modo in cui viene creato un modello di eventi. Seleziona i valori dei campi su un modulo e EventBridge genera il modello automaticamente. Al momento, il Dimensionamento automatico Amazon EC2 non supporta modelli predefiniti per eventi emessi da un gruppo con scalabilità automatica con un warm pool. È necessario immettere il modello come oggetto JSON. Questa sezione e l'argomento [Crea EventBridge regole per eventi in piscina riscaldata](#) mostrano come utilizzare un modello di evento per selezionare gli eventi e inviarli alle destinazioni.

Per creare EventBridge regole che filtrino gli eventi relativi ai warm pool a cui EventBridge invia Amazon EC2 Auto Scaling, Origin includi i campi e detail dalla Destination sezione dell'evento.

I valori di Origin e Destination possono essere i seguenti:

EC2 | AutoScalingGroup | WarmPool

Indice

- [Eventi di esempio](#)
- [Esempio di modelli di eventi](#)

Eventi di esempio

Quando aggiungi lifecycle hook al tuo gruppo Auto Scaling, Amazon EC2 Auto Scaling invia eventi a quando un'istanza passa EventBridge allo stato di attesa. Per ulteriori informazioni, consulta [Utilizzare un hook del ciclo di vita con un warm pool](#).

Questa sezione include esempi di questi eventi quando il gruppo con dimensionamento automatico ha un pool caldo. Gli eventi vengono emessi sulla base del best effort.

Note

Per gli eventi a cui Amazon EC2 Auto Scaling invia quando il ridimensionamento ha EventBridge esito positivo, consulta [Eventi di dimensionamento riusciti](#). Per gli eventi in cui il dimensionamento non riesce, consulta [Eventi di dimensionamento non riusciti](#).

Esempi di eventi

- [Azione di dimensionamento orizzontale del ciclo di vita](#)
- [Operazione del ciclo di vita del dimensionamento verticale](#)

Azione di dimensionamento orizzontale del ciclo di vita

Gli eventi che vengono consegnati quando un'istanza passa allo stato di attesa su dimensionamento orizzontale hanno `EC2 Instance-launch Lifecycle Action` come valore per `detail-type`. Nell'oggetto `detail`, i valori degli attributi `Origin` e `Destination` mostrano la provenienza e la direzione dell'istanza.

In questo esempio di evento di dimensionamento orizzontale, viene avviata una nuova istanza e il suo stato cambia in `Warm:Pending:Wait` perché viene aggiunta al pool caldo. Per ulteriori informazioni, consulta [Transizioni dello stato del ciclo di vita delle istanze in un warm pool](#).

```
{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Instance-launch Lifecycle Action",
  "source": "aws.autoscaling",
  "account": "123456789012",
  "time": "2021-01-13T00:12:37.214Z",
  "region": "us-west-2",
  "resources": [
    "auto-scaling-group-arn"
  ],
  "detail": {
    "LifecycleActionToken": "71514b9d-6a40-4b26-8523-05e7eEXAMPLE",
    "AutoScalingGroupName": "my-asg",
    "LifecycleHookName": "my-launch-lifecycle-hook",
```

```

    "EC2InstanceId": "i-1234567890abcdef0",
    "LifecycleTransition": "autoscaling:EC2_INSTANCE_LAUNCHING",
    "NotificationMetadata": "additional-info",
    "Origin": "EC2",
    "Destination": "WarmPool"
  }
}

```

In questo caso evento di dimensionamento orizzontale di esempio, lo stato di un'istanza cambia in Pending:Wait perché viene aggiunto al gruppo con dimensionamento automatico dal pool caldo. Per ulteriori informazioni, consulta [Transizioni dello stato del ciclo di vita delle istanze in un warm pool](#).

```

{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Instance-launch Lifecycle Action",
  "source": "aws.autoscaling",
  "account": "123456789012",
  "time": "2021-01-19T00:35:52.359Z",
  "region": "us-west-2",
  "resources": [
    "auto-scaling-group-arn"
  ],
  "detail": {
    "LifecycleActionToken": "19cc4d4a-e450-4d1c-b448-0de67EXAMPLE",
    "AutoScalingGroupName": "my-asg",
    "LifecycleHookName": "my-launch-lifecycle-hook",
    "EC2InstanceId": "i-1234567890abcdef0",
    "LifecycleTransition": "autoscaling:EC2_INSTANCE_LAUNCHING",
    "NotificationMetadata": "additional-info",
    "Origin": "WarmPool",
    "Destination": "AutoScalingGroup"
  }
}

```

Operazione del ciclo di vita del dimensionamento verticale

Gli eventi che vengono consegnati quando un'istanza passa allo stato di attesa su dimensionamento verticale hanno EC2 Instance-terminate Lifecycle Action come valore per detail-type. Nell'oggetto detail, i valori degli attributi Origin e Destination mostrano la provenienza e la direzione dell'istanza.

In questo caso di esempio di dimensionamento verticale, lo stato di un'istanza cambia in `Warmed:Pending:Wait` perché viene restituito al pool caldo. Per ulteriori informazioni, consulta [Transizioni dello stato del ciclo di vita delle istanze in un warm pool](#).

```
{
  "version": "0",
  "id": "12345678-1234-1234-1234-123456789012",
  "detail-type": "EC2 Instance-terminate Lifecycle Action",
  "source": "aws.autoscaling",
  "account": "123456789012",
  "time": "2022-03-28T00:12:37.214Z",
  "region": "us-west-2",
  "resources": [
    "auto-scaling-group-arn"
  ],
  "detail": {
    "LifecycleActionToken": "42694b3d-4b70-6a62-8523-09a1eEXAMPLE",
    "AutoScalingGroupName": "my-asg",
    "LifecycleHookName": "my-termination-lifecycle-hook",
    "EC2InstanceId": "i-1234567890abcdef0",
    "LifecycleTransition": "autoscaling:EC2_INSTANCE_TERMINATING",
    "NotificationMetadata": "additional-info",
    "Origin": "AutoScalingGroup",
    "Destination": "WarmPool"
  }
}
```

Esempio di modelli di eventi

Nella sezione precedente vengono forniti eventi di esempio emessi da Amazon EC2 Auto Scaling.

EventBridge i modelli di eventi hanno la stessa struttura degli eventi a cui corrispondono. Il modello cita i campi che desideri abbinare e fornisce i valori che stai cercando.

I campi riportati di seguito nell'evento costituiscono il modello di evento definito nella regola e che richiamano un'operazione:

`"source": "aws.autoscaling"`

Identifica che l'evento proviene da Amazon EC2 Auto Scaling.

`"detail-type": "EC2 Instance-launch Lifecycle Action"`

Identifica il tipo di evento.

"Origin": "EC2"

Identifica la provenienza dell'istanza.

"Destination": "WarmPool"

Identifica la destinazione dell'istanza.

Usa il modello di evento di esempio riportato di seguito per acquisire tutti gli eventi EC2 Instance-launch Lifecycle Action associati alle istanze che entrano nel pool caldo.

```
{
  "source": [ "aws.autoscaling" ],
  "detail-type": [ "EC2 Instance-launch Lifecycle Action" ],
  "detail": {
    "Origin": [ "EC2" ],
    "Destination": [ "WarmPool" ]
  }
}
```

Usa il modello di evento di esempio riportato di seguito per acquisire tutti gli eventi EC2 Instance-launch Lifecycle Action associati alle istanze che lasciano il pool caldo a causa di un aumento orizzontale.

```
{
  "source": [ "aws.autoscaling" ],
  "detail-type": [ "EC2 Instance-launch Lifecycle Action" ],
  "detail": {
    "Origin": [ "WarmPool" ],
    "Destination": [ "AutoScalingGroup" ]
  }
}
```

Usa il modello di evento di esempio riportato di seguito per acquisire tutti gli eventi EC2 Instance-launch Lifecycle Action associati alle istanze avviate direttamente nel gruppo con dimensionamento automatico.

```
{
  "source": [ "aws.autoscaling" ],
  "detail-type": [ "EC2 Instance-launch Lifecycle Action" ],
  "detail": {
```

```
"Origin": [ "EC2" ],
"Destination": [ "AutoScalingGroup" ]
}
}
```

Usa il modello di evento di esempio riportato di seguito per acquisire tutti gli eventi EC2 Instance-terminate Lifecycle Action associati alle istanze che lasciano il pool caldo a causa di una riduzione orizzontale.

```
{
  "source": [ "aws.autoscaling" ],
  "detail-type": [ "EC2 Instance-terminate Lifecycle Action" ],
  "detail": {
    "Origin": [ "AutoScalingGroup" ],
    "Destination": [ "WarmPool" ]
  }
}
```

Usa il seguente modello o eventi di esempio per acquisire tutti gli eventi associati a EC2 Instance-launch Lifecycle Action, indipendentemente dall'origine o dalla destinazione.

```
{
  "source": [ "aws.autoscaling" ],
  "detail-type": [ "EC2 Instance-launch Lifecycle Action" ]
}
```

Crea EventBridge regole

Quando un evento viene emesso da Amazon EC2 Auto Scaling, una notifica dell'evento viene inviata ad EventBridge Amazon come file JSON. Puoi scrivere una EventBridge regola per automatizzare le azioni da intraprendere quando un modello di evento corrisponde alla regola. Se EventBridge rileva un pattern di eventi che corrisponde a uno schema definito in una regola, EventBridge richiama l'obiettivo (o i target) specificati nella regola.

È possibile utilizzare le procedure di esempio di questa sezione come punto di partenza.

Può essere utile anche la seguente documentazione:

- Per eseguire azioni personalizzate sulle istanze durante l'avvio o prima che vengano terminate utilizzando una funzione Lambda, consulta [Tutorial : Configurazione di un hook del ciclo di vita che richiama una funzione Lambda](#).

- Per richiamare una funzione Lambda sulle chiamate API registrate CloudTrail con, [consulta Tutorial: AWS Log API call using EventBridge in Amazon EventBridge User Guide](#).
- Per ulteriori informazioni su come creare regole per gli eventi, consulta la sezione [Creazione di EventBridge regole Amazon che reagiscono agli eventi](#) nella Amazon EventBridge User Guide.

Argomenti

- [Crea EventBridge regole, ad esempio eventi di aggiornamento.](#)
- [Crea EventBridge regole per eventi in piscina riscaldata](#)

Crea EventBridge regole, ad esempio eventi di aggiornamento.

L'esempio seguente crea una EventBridge regola per inviare una notifica via e-mail. Lo fa ogni volta che il gruppo con scalabilità automatica emette un evento quando viene raggiunto un punto di controllo durante un aggiornamento delle istanze. La procedura per impostare le notifiche via e-mail utilizzando Amazon SNS è inclusa. Per usare Amazon SNS per l'invio di notifiche e-mail, devi prima creare un argomento, quindi iscrivere i tuoi indirizzi e-mail all'argomento.

Per ulteriori informazioni sulla caratteristica di aggiornamento di un'istanza, consulta [Usa un aggiornamento dell'istanza per aggiornare le istanze in un gruppo di Auto Scaling](#).

Creare un argomento Amazon SNS.

Un argomento SNS è un punto di accesso logico, un canale di comunicazione che il gruppo Auto Scaling utilizza per inviare le notifiche. Puoi creare un argomento specificando un nome.

I nomi dell'argomento devono soddisfare i seguenti requisiti:

- Hanno un numero di caratteri compreso tra 1 e 256.
- Deve contenere caratteri ASCII maiuscoli e minuscoli, numeri, trattini e caratteri di sottolineatura.

Per le istruzioni, consulta [Creazione di un argomento Amazon SNS](#) nella Guida per lo Sviluppatore di Amazon Simple Notification Service.

Sottoscrizione all'argomento Amazon SNS

Per ricevere le notifiche che il gruppo con scalabilità automatica invia all'argomento, è necessario sottoscrivere un endpoint all'argomento. In questa procedura, come Endpoint, specifica l'indirizzo e-mail a cui desideri ricevere le notifiche da Amazon EC2 Auto Scaling.

Per ulteriori informazioni, consulta [Sottoscrizione a un argomento di Amazon SNS](#) nella Guida per lo Sviluppatore di Amazon Simple Notification Service.

Conferma la sottoscrizione ad Amazon SNS

Amazon SNS invia un'e-mail di conferma all'indirizzo specificato nella fase precedente.

Assicurati di aprire l'e-mail da AWS Notifiche e di scegliere il link per confermare l'iscrizione prima di continuare con il passaggio successivo.

Riceverai un messaggio di conferma da AWS Amazon SNS è ora configurato per ricevere notifiche e inviare la notifica come e-mail all'indirizzo specificato.

Instradamento degli eventi al tuo argomento Amazon SNS

Crea una regola che corrisponda agli eventi selezionati e li instradi al tuo argomento Amazon SNS per dare notifica agli indirizzi e-mail sottoscritti.

Per creare una regola che invia notifiche al tuo argomento Amazon SNS

1. Apri la EventBridge console Amazon all'[indirizzo https://console.aws.amazon.com/events/](https://console.aws.amazon.com/events/).
2. Nel pannello di navigazione, scegli Regole.
3. Scegli Crea regola.
4. Per Define rule detail (Definisci dettagli della regola), effettua le seguenti operazioni:

- a. Immettere un Name (Nome) per la regola e, facoltativamente, una descrizione.

Una regola non può avere lo stesso nome di un'altra regola nella stessa regione e sullo stesso bus di eventi.

- b. Per Event bus (Bus di eventi), scegli default. Quando un AWS servizio del tuo account genera un evento, passa sempre al bus eventi predefinito del tuo account.
 - c. Per Rule type (Tipo di regola), scegli Rule with an event pattern (Regola con un modello di eventi).
 - d. Seleziona Successivo.
5. Per Build event pattern (Crea modello di eventi), procedi come segue:
 - a. Per Event source, scegli AWS eventi o eventi EventBridge partner.
 - b. Per Event pattern (Modello eventi), procedi come segue:

- i. In Event source (Origine eventi), selezionare Servizi AWS.
 - ii. Per Servizio AWS, scegli Auto Scaling (Scalabilità automatica).
 - iii. In Event type (Tipo di evento), scegli Instance Refresh (Aggiornamento delle istanze).
 - iv. Per impostazione predefinita, la regola corrisponde a qualsiasi evento di aggiornamento delle istanze. Per creare una regola che notifica all'utente quando viene raggiunto un punto di controllo durante l'aggiornamento delle istanze, scegli Specific instance event(s) (Evento/i di istanza specifico) e seleziona EC2 Auto Scaling Instance Refresh Checkpoint Reached (Punto di controllo di aggiornamento istanze EC2 Auto Scaling raggiunto).
 - v. Per impostazione predefinita, la regola corrisponde a qualsiasi gruppo con scalabilità automatica nella Regione. Affinché la regola corrisponda a un gruppo con scalabilità automatica specifico, seleziona Specific group name(s) (Nomi gruppi specifici), quindi seleziona uno o più gruppi con scalabilità automatica.
 - vi. Seleziona Successivo.
6. Per Select target(s) (Seleziona destinazione/i), esegui queste operazioni:
- a. Per Target types (Tipi di target), scegli Servizio AWS.
 - b. Per Select a target (Seleziona un target), scegli SNS topic (Argomento SNS).
 - c. Per Topic (Argomento), scegli il tuo argomento Amazon SNS.
 - d. (Facoltativo) In Additional settings (Impostazioni aggiuntive), facoltativamente puoi configurare impostazioni aggiuntive. Per ulteriori informazioni, consulta la sezione [Creazione di EventBridge regole Amazon che reagiscono agli eventi](#) nella Amazon EventBridge User Guide.
 - e. Seleziona Successivo.
7. (Facoltativo) Per Tags (Tag), se desideri puoi assegnare uno o più tag alla regola, quindi scegli Next (Successivo).
8. Per Review and create (Verifica e crea), verifica i dettagli della regola e se necessario, modificali. Quindi scegli Create role (Crea ruolo).

Crea EventBridge regole per eventi in piscina riscaldata

L'esempio seguente crea una EventBridge regola per richiamare azioni programmatiche. Lo fa ogni volta che il gruppo con scalabilità automatica emette un evento quando una nuova istanza viene aggiunta al warm pool.

Prima di creare la regola, create la AWS Lambda funzione che desiderate che la regola utilizzi come destinazione. È necessario specificare questa funzione come destinazione della regola. La procedura seguente fornisce solo i passaggi per creare la EventBridge regola che agisce quando nuove istanze entrano nel pool caldo. Per un'esercitazione introduttiva che mostra come creare una semplice funzione Lambda da richiamare quando un evento in arrivo corrisponde a una regola, consulta [Tutorial : Configurazione di un hook del ciclo di vita che richiama una funzione Lambda](#).

Per ulteriori informazioni sulla creazione e l'utilizzo dei warm pool, consulta [Warm pool per Dimensionamento automatico Amazon EC2](#).

Creazione di una regola di evento per richiamare una funzione Lambda

1. Apri la EventBridge console Amazon all'[indirizzo https://console.aws.amazon.com/events/](https://console.aws.amazon.com/events/).
2. Nel pannello di navigazione, scegli Regole.
3. Scegli Crea regola.
4. Per Define rule detail (Definisci dettagli della regola), effettua le seguenti operazioni:
 - a. Immettere un Name (Nome) per la regola e, facoltativamente, una descrizione.

Una regola non può avere lo stesso nome di un'altra regola nella stessa regione e sullo stesso bus di eventi.
 - b. Per Event bus (Bus di eventi), scegli default. Quando un Servizio AWS utente genera un evento, questo passa sempre al bus eventi predefinito del tuo account.
 - c. Per Rule type (Tipo di regola), scegli Rule with an event pattern (Regola con un modello di eventi).
 - d. Seleziona Successivo.
5. Per Build event pattern (Crea modello di eventi), procedi come segue:
 - a. Per Event source, scegli AWS eventi o eventi EventBridge partner.
 - b. Per Event pattern (Modello di eventi), scegli Custom pattern (JSON editor) (Modello personalizzato (editor JSON)) e incolla il seguente modello nella casella Event pattern (Modello di eventi), sostituendo il testo in *corsivo* con il nome del gruppo con scalabilità automatica.

```
{
  "source": [ "aws.autoscaling" ],
  "detail-type": [ "EC2 Instance-launch Lifecycle Action" ],
  "detail": {
```

```
"AutoScalingGroupName": [ "my-asg" ],  
"Origin": [ "EC2" ],  
"Destination": [ "WarmPool" ]  
}  
}
```

Per creare una regola che corrisponda ad altri eventi, modifica il modello di evento. Per ulteriori informazioni, consulta la pagina [Esempio di modelli di eventi](#).

- c. Seleziona Next (Successivo).
6. Per Select target(s) (Seleziona destinazione/i), esegui queste operazioni:
 - a. Per Target types (Tipi di target), scegli Servizio AWS.
 - b. Per Select a target (Seleziona destinazione), scegli Lambda function (Funzione Lambda).
 - c. In Function (Funzione), scegli la funzione a cui inviare gli eventi.
 - d. (Facoltativo) Per Configure version/alias (Configura versione/alias), inserisci le impostazioni di versione e alias per la funzione Lambda di destinazione.
 - e. (Facoltativo) In Additional settings (Impostazioni aggiuntive), inserisci eventuali impostazioni aggiuntive, a seconda di quanto appropriato per l'applicazione. Per ulteriori informazioni, consulta la sezione [Creazione di EventBridge regole Amazon che reagiscono agli eventi](#) nella Amazon EventBridge User Guide.
 - f. Seleziona Successivo.
7. (Facoltativo) Per Tags (Tag), se desideri puoi assegnare uno o più tag alla regola, quindi scegli Next (Successivo).
8. Per Review and create (Verifica e crea), verifica i dettagli della regola e se necessario, modificali. Quindi scegli Create role (Crea ruolo).

Fornisci connettività di rete per le istanze Auto Scaling utilizzando Amazon VPC

Amazon Virtual Private Cloud (Amazon VPC) è un servizio che ti consente di avviare AWS risorse come gruppi di Auto Scaling in una rete virtuale logicamente isolata da te definita.

Una sottorete in Amazon VPC è una suddivisione all'interno di una zona di disponibilità definita da un segmento dell'intervallo dell'indirizzo IP del VPC. Tramite le sottoreti, è possibile raggruppare le istanze in base alle esigenze di sicurezza e operative. La sottorete risiede interamente all'interno

della zona di disponibilità in cui è stata creata. Le istanze di Auto Scaling vengono avviate all'interno delle sottoreti.

Per abilitare la comunicazione tra Internet e le istanze nelle sottoreti, è necessario creare un Gateway Internet e allegarlo al VPC. Il Gateway Internet consente alle risorse all'interno delle sottoreti di connettersi a Internet attraverso l'edge di rete Amazon EC2. Se il traffico di una sottorete viene instradato a un Gateway Internet, la sottorete è nota come sottorete pubblica. Se il traffico di una sottorete non viene instradato verso un Gateway Internet, la sottorete è nota come sottorete privata. Utilizza una sottorete pubblica per le risorse che devono essere connesse a Internet e una sottorete privata per le risorse che non devono essere connesse a Internet. Per ulteriori informazioni sull'accesso Internet alle istanze in un VPC, consulta [Accesso a Internet](#) nella Guida per l'utente ad Amazon VPC.

Indice

- [VPC di default](#)
- [VPC non di default](#)
- [Considerazioni sulla scelta di sottoreti VPC](#)
- [Indirizzamento IP in un VPC](#)
- [Interfacce di rete in un VPC](#)
- [Tenancy di posizionamento delle istanze](#)
- [AWS Outposts](#)
- [Più risorse per l'approfondimento della conoscenza di VPC](#)

VPC di default

Se hai creato il tuo Account AWS dopo il 4 dicembre 2013 o stai creando il tuo gruppo Auto Scaling in un nuovo gruppo Regione AWS, creeremo un VPC predefinito per te. Il VPC di default viene fornito con una sottorete di default in ciascuna zona di disponibilità. Se disponi di un VPC di default, il gruppo Auto Scaling viene creato in questo VPC per impostazione predefinita.

Puoi visualizzare i VPC nella pagina [Your VPCs](#) (I tuoi VPC) della console Amazon VPC.

Per ulteriori informazioni sui VPC predefiniti, consulta [VPC predefiniti](#) in Guida per l'utente di Amazon VPC.

VPC non di default

Puoi scegliere di creare altri VPC accedendo alla pagina [VPC Dashboard](#) nella AWS Management Console e selezionando Create VPC (Crea VPC).

Per ulteriori informazioni, consulta la [Guida utente Amazon VPC](#).

Note

Un VPC si estende su tutte le zone di disponibilità nella sua Regione AWS. Quando aggiungi sottoreti al tuo VPC, scegli più zone di disponibilità per assicurarti che le applicazioni ospitate in tali sottoreti siano altamente disponibili. Una zona di disponibilità consiste in uno o più data center separati con alimentazione, rete e connettività ridondanti in una Regione AWS. Le zone di disponibilità consentono di rendere le applicazioni di produzione altamente disponibili, tolleranti ai guasti e scalabili.

Considerazioni sulla scelta di sottoreti VPC

Nello scegliere le sottoreti VPC per il gruppo con dimensionamento automatico, tieni in considerazione quanto segue:

- Se si allega un sistema di bilanciamento del carico Elastic Load Balancing al gruppo con dimensionamento automatico, le istanze possono essere avviate in sottoreti pubbliche o private. È tuttavia necessario che il sistema di bilanciamento del carico venga creato in sottoreti pubbliche per supportare la risoluzione DNS.
- Se si accede alle istanze di Auto Scaling direttamente tramite SSH, le istanze possono essere avviate solo nelle sottoreti pubbliche.
- Se accedi a istanze di Auto Scaling senza ingresso AWS Systems Manager utilizzando Session Manager, le istanze possono essere avviate in sottoreti pubbliche o private.
- Se si utilizzano sottoreti private, è possibile permettere alle istanze di Auto Scaling di accedere a Internet utilizzando un Gateway NAT pubblico.
- Per impostazione predefinita, le sottoreti di default in un VPC predefinito sono pubbliche.

Indirizzamento IP in un VPC

Quando si avviano delle istanze Auto Scaling in un VPC, alle istanze viene assegnato automaticamente un indirizzo IP privato dall'intervallo CIDR della sottorete in cui viene avviata l'istanza. In questo modo le istanze possono comunicare con altre istanze nel VPC.

È possibile configurare il modello di avvio o configurazione di avvio per assegnare indirizzi IPv4 pubblici alle proprie istanze. L'assegnazione di indirizzi IP pubblici alle istanze consente loro di comunicare con Internet o altri servizi. AWS

Quando si avviano istanze in una sottorete configurata per assegnare automaticamente gli indirizzi IPv6, queste ricevono sia indirizzi IPv4 che IPv6. In caso contrario, ricevono solo indirizzi IPv4. Per ulteriori informazioni, consulta [IPv6 addresses Indirizzi IP pubblici](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Per ulteriori informazioni sulla specifica degli intervalli CIDR per il VPC o sottorete, consulta la [Guida per l'utente di Amazon VPC](#).

Dimensionamento automatico Amazon EC2 può assegnare automaticamente ulteriori indirizzi IP privati all'avvio dell'istanza quando si utilizza un modello di avvio che specifica interfacce di rete aggiuntive. Ad ogni interfaccia di rete viene assegnato un unico indirizzo IP privato dall'intervallo CIDR della sottorete in cui viene avviata l'istanza. In questo caso, il sistema non è più in grado di assegnare automaticamente un indirizzo IPv4 pubblico all'interfaccia di rete primaria. Non potrai connetterti alle istanze su un indirizzo IPv4 pubblico a meno che non associ gli indirizzi IP elastici disponibili alle istanze Auto Scaling.

Interfacce di rete in un VPC

Ogni istanza in un VPC dispone di un'interfaccia di rete di default (interfaccia di rete primaria). Non è possibile scollegare un'interfaccia di rete primaria da un'istanza. Puoi creare e allegare un'ulteriore interfaccia di rete a qualsiasi istanza nel VPC. Il numero di interfacce di rete che puoi allegare varia a seconda del tipo di istanza.

Quando avvii un'istanza utilizzando un modello di avvio, puoi specificare interfacce di rete aggiuntive. Tuttavia, l'avvio di un'istanza Auto Scaling con più interfacce di rete crea automaticamente ogni interfaccia nella stessa sottorete dell'istanza. Questo perché Dimensionamento automatico Amazon EC2 ignora le sottoreti definite nel modello di avvio a favore di quanto specificato nel gruppo Auto Scaling. Per ulteriori informazioni, consulta [Creazione di un modello di avvio per un gruppo Auto Scaling](#).

Se crei o alleghi due o più interfacce di rete della stessa sottorete a un'istanza, potresti riscontrare errori a livello di rete, ad esempio un routing asimmetrico, soprattutto su istanze che utilizzano una variante Linux non Amazon. Se hai bisogno di questo tipo di configurazione, dovrai anche configurare l'interfaccia di rete secondaria all'interno del sistema operativo. Per un esempio, vedi [Come posso far funzionare la mia interfaccia di rete secondaria nella mia istanza Ubuntu EC2?](#) nel AWS Knowledge Center.

Tenancy di posizionamento delle istanze

Per impostazione predefinita, tutte le istanze nel VPC vengono eseguite come istanze a tenancy condivisa. Dimensionamento automatico Amazon EC2 supporta anche istanze e host dedicati. Per ulteriori informazioni, consulta [Crea un modello di avvio utilizzando le impostazioni avanzate](#).

AWS Outposts

AWS Outposts estende un Amazon VPC da una AWS regione a un avamposto con i componenti VPC accessibili nella regione, inclusi gateway Internet, gateway privati virtuali, gateway di transito Amazon VPC ed endpoint VPC. Un Outpost è ospitato in una zona di disponibilità nella regione ed è un'estensione della zona di disponibilità che è possibile utilizzare per la resilienza.

Per ulteriori informazioni, consulta la [AWS Outposts Guida per l'utente](#).

Per un esempio di come distribuire un gruppo con dimensionamento automatico che serve il traffico da un Application Load Balancer all'interno di un Outpost, vedere il seguente post sul blog: [Configurazione di un sistema di Application Load Balancer su AWS Outposts](#).

Più risorse per l'approfondimento della conoscenza di VPC

Utilizza gli argomenti seguenti per avere ulteriori informazioni sui VPC e sulle sottoreti.

- Sottoreti private in un VPC
 - [Esempio: VPC con server in sottoreti private e NAT](#)
 - [Gateway NAT](#)
- Sottoreti pubbliche in un VPC
 - [Esempio: VPC per un ambiente di test](#)
 - [Esempio: VPC per server Web e di database](#)
- Sottoreti per l'Application Load Balancer
 - [Sottoreti per il sistema di bilanciamento del carico](#)

- **Informazioni generali sul VPC**
 - [Guida per l'utente di Amazon VPC](#)
 - [Connettere i VPC utilizzando il peering VPC](#)
 - [Interfacce di rete elastiche](#)
 - [Utilizzare gli endpoint VPC per la connettività privata](#)

Sicurezza in Amazon EC2 Auto Scaling

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 ulteriori informazioni sui programmi di conformità che si applicano ad Amazon EC2 Auto Scaling, [AWS consulta i servizi nell'ambito del programma di conformità \(servizi nell'ambito per di conformità\)](#).
- Sicurezza nel cloud: la tua responsabilità è determinata dal AWS servizio che utilizzi. Inoltre, sei responsabile anche per altri fattori, tra cui la riservatezza dei dati, i requisiti dell'azienda e le leggi e le normative applicabili.

Questa documentazione permette di comprendere come applicare il modello di responsabilità condivisa quando si usa Amazon EC2 Auto Scaling. Gli argomenti seguenti illustrano come configurare Amazon EC2 Auto Scaling in modo da soddisfare gli obiettivi di sicurezza e conformità. Scopri anche come utilizzare altri AWS servizi che ti aiutano a monitorare e proteggere le tue risorse di Amazon EC2 Auto Scaling.

Argomenti

- [Sicurezza dell'infrastruttura in Amazon EC2 Auto Scaling](#)
- [Resilienza in Amazon EC2 Auto Scaling](#)
- [Protezione dei dati in Amazon EC2 Auto Scaling](#)
- [Identity and Access Management per Dimensionamento automatico Amazon EC2](#)
- [Convalida della conformità per Dimensionamento automatico Amazon EC2](#)
- [Dimensionamento automatico Amazon EC2 e interfaccia VPC endpoint](#)

Sicurezza dell'infrastruttura in Amazon EC2 Auto Scaling

In quanto servizio gestito, Amazon EC2 Auto Scaling è protetto AWS dalla sicurezza di rete globale. Per informazioni sui servizi di AWS 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.

Utilizza chiamate API AWS pubblicate per accedere ad Amazon EC2 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. O puoi utilizzare [AWS Security Token Service](#) (AWS STS) per generare credenziali di sicurezza temporanee per sottoscrivere le richieste.

Puoi anche usare un endpoint cloud privato virtuale (VPC) per Dimensionamento automatico Amazon EC2. Gli endpoint VPC di interfaccia consentono alle risorse Amazon VPC di utilizzare i propri indirizzi IP privati per accedere a Dimensionamento automatico Amazon EC2, senza alcuna esposizione a Internet. Per ulteriori informazioni, consultare [Dimensionamento automatico Amazon EC2 e interfaccia VPC endpoint](#)

Risorse correlate

Per informazioni sulle funzionalità di isolamento del traffico di servizio fornito da Amazon EC2, [consulta la sezione Sicurezza dell'infrastruttura in Amazon EC2 nella Guida per l'utente di Amazon EC2](#) per le istanze Linux.

Resilienza in Amazon EC2 Auto Scaling

L'infrastruttura AWS globale è costruita attorno a zone di disponibilità. Regioni AWS Regioni AWS forniscono più zone di disponibilità fisicamente separate e isolate, collegate con reti a bassa latenza, ad alto throughput e altamente ridondanti. Con le zone di disponibilità, puoi progettare e gestire applicazioni e database che eseguono automaticamente il failover tra zone di disponibilità senza interruzioni. Le zone di disponibilità sono più disponibili, tolleranti ai guasti e scalabili rispetto alle infrastrutture a data center singolo o multiplo tradizionali.

[Per ulteriori informazioni sulle zone di disponibilità, vedere Global Regioni AWS Infrastructure.AWS](#)

Per sfruttare la ridondanza geografica della progettazione della zona di disponibilità, esegui le operazioni seguenti:

- Distribuisci il gruppo con scalabilità automatica su più zone di disponibilità.
- Mantieni almeno un'istanza in ogni zona di disponibilità.
- Collega un sistema di bilanciamento del carico per distribuire il traffico in entrata nelle stesse zone di disponibilità. Se utilizzi un Application Load Balancer, assicurati che ogni istanza EC2 riceva una quantità di traffico simile mantenendo abilitato il sistema di bilanciamento del carico tra zone. Ciò consente di limitare l'impatto dovuto all'aumento del carico sulle istanze esistenti durante un evento di failover e di ottenere una maggiore resilienza rispetto all'assenza del sistema di bilanciamento del carico tra zone.
- Assicurati che i controlli di integrità di Elastic Load Balancing siano configurati correttamente e che siano abilitati nel gruppo con scalabilità automatica. Quindi, se un'istanza non supera il controllo dell'integrità, Elastic Load Balancing interrompe l'invio di traffico e reindirizza il traffico verso istanze integre, mentre Dimensionamento automatico Amazon EC2 sostituisce l'istanza non integra.

Amazon EC2 Auto Scaling aiuta a supportare le esigenze di resilienza delle applicazioni nei seguenti modi:

- Controlla le istanze per individuare eventuali problemi di integrità e raggiungibilità. Quando un'istanza diventa non integra, la termina automaticamente e ne avvia una nuova.
- Se sono attive policy di dimensionamento dinamico, dimensiona automaticamente la capacità in base al traffico in ingresso.
- Rileva problemi di affidabilità dei CloudWatch parametri di Amazon che supportano le politiche di scalabilità e sospende le attività di scalabilità quando non sono disponibili parametri affidabili, ad esempio quando mancano punti dati.
- Cerca di mantenere un numero equivalente di istanze in ogni zona di disponibilità abilitata durante il dimensionamento del gruppo.
- Sfrutta le zone di disponibilità al fine di mantenere un'elevata disponibilità. Quando una zona di disponibilità perde di integrità, Dimensionamento automatico Amazon EC2 esegue le seguenti operazioni:
 - Avvia nuove istanze in un'altra zona di disponibilità abilitata per il gruppo con scalabilità automatica.
 - Ridistribuisce le istanze in tutte le zone di disponibilità abilitate quando la zona di disponibilità non integra torna allo stato integro.

- Cerca di avviare le istanze in altre zone di disponibilità abilitate se l'avvio di un'istanza in una determinata zona di disponibilità ha esito negativo.
- Esegue e annulla automaticamente la registrazione delle istanze che presentano i sistemi di bilanciamento del carico associati al gruppo con scalabilità automatica. In questo modo, non è necessario registrare e annullare la registrazione delle istanze separatamente.

Risorse correlate

Per informazioni sulle funzionalità che aiutano a supportare le tue esigenze di resilienza dei dati fornite da Amazon EBS, consulta [Resilience in Amazon Elastic Block Store nella Amazon EBS User Guide](#).

Protezione dei dati in Amazon EC2 Auto Scaling

Il modello di [responsabilità AWS condivisa modello](#) si applica alla protezione dei dati in Amazon EC2 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 con Amazon EC2 Auto Scaling o Servizi AWS altro utilizzando la console, l'API AWS CLI o gli SDK. 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.

Quando avvii un'istanza Amazon EC2, hai la possibilità di passare i dati utente all'istanza per eseguire una configurazione aggiuntiva all'avvio dell'istanza. Ti consigliamo inoltre di non inserire mai informazioni riservate o sensibili nei dati utente che verranno passati a un'istanza.

Utilizzalo AWS KMS keys per crittografare i volumi Amazon EBS

Puoi configurare il tuo gruppo Auto Scaling per crittografare i dati del volume Amazon EBS archiviati nel cloud con AWS KMS keys. Amazon EC2 Auto Scaling AWS supporta chiavi gestite e gestite dal cliente per crittografare i dati. Ricorda che l'opzione `KmsKeyId` per specificare una chiave gestita dal cliente non è disponibile, quando utilizzi una configurazione di avvio. Per specificare la chiave gestita dal cliente, utilizza piuttosto un modello di avvio. Per ulteriori informazioni, consulta [Creare un modello di avvio per un gruppo con dimensionamento automatico](#). [Per informazioni su come creare, archiviare e gestire le chiavi di AWS KMS crittografia, consulta la Guida per gli AWS Key Management Service sviluppatori](#).

Puoi anche configurare una chiave gestita dal cliente nell'AMI supportata da EBS prima di impostare il modello o la configurazione di avvio oppure utilizzare la crittografia di default per applicarla ai nuovi volumi e alle copie di snapshot EBS creati. Per ulteriori informazioni, consulta [Usare la crittografia con AMI supportate da EBS](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux e Encryption [by default nella](#) Amazon EBS User Guide.

Note

Per informazioni su come impostare le policy delle chiavi necessarie per avviare istanze di Auto Scaling quando si utilizza una chiave gestita dal cliente per la crittografia, vedi [Policy AWS KMS chiave richiesta per l'uso con volumi crittografati](#).

Risorse correlate

Per le linee guida sulla protezione dei dati fornite da Amazon EBS, consulta la sezione [Protezione dei dati in Amazon Elastic Block Store](#) nella Amazon EBS User Guide.

Policy AWS KMS chiave richiesta per l'uso con volumi crittografati

Amazon EC2 Auto Scaling [utilizza ruoli collegati ai servizi per delegare le autorizzazioni](#) ad altri. Servizi AWS I ruoli collegati ai servizi di Amazon EC2 Auto Scaling sono predefiniti e includono le autorizzazioni richieste da Amazon EC2 Auto Scaling per chiamare altri utenti per tuo conto. Servizi AWS Le autorizzazioni predefinite includono anche l'accesso ai tuoi. Chiavi gestite da AWS Tuttavia, non includono l'accesso alle chiavi gestite dal cliente, permettendoti di mantenere il pieno controllo su queste chiavi.

In questo argomento viene descritto come impostare le policy chiave necessarie per avviare istanze di Auto Scaling quando specifichi una chiave gestita dal cliente per la crittografia Amazon EBS.

Note

Dimensionamento automatico Amazon EC2 non ha bisogno di ulteriori autorizzazioni per l'utilizzo della Chiave gestita da AWS di default, per proteggere i volumi crittografati nel tuo account.

Indice

- [Panoramica](#)
- [Configurare le policy chiave](#)
- [Esempio 1: sezioni delle policy delle chiavi che permettono l'accesso alla chiave gestita dal cliente](#)
- [Esempio 2: sezioni delle policy delle chiavi che permettono l'accesso multiaccount alla chiave gestita dal cliente](#)

- [Modificare le policy delle chiavi nella console AWS KMS](#)

Panoramica

Quanto segue AWS KMS keys può essere usato per la crittografia Amazon EBS quando Amazon EC2 Auto Scaling avvia le istanze:

- [Chiave gestita da AWS](#)— Una chiave di crittografia nel tuo account che Amazon EBS crea, possiede e gestisce. Questa è la chiave di crittografia di default per un nuovo account. Chiave gestita da AWS Viene utilizzato per la crittografia a meno che non si specifichi una chiave gestita dal cliente.
- [Chiave gestita dal cliente](#): una chiave di crittografia personalizzata che puoi creare, possedere e gestire. Per ulteriori informazioni, consulta [Creazione di chiavi](#) nella Guida per gli sviluppatori di AWS Key Management Service .

Nota: la chiave deve essere simmetrica. Amazon EBS non supporta le chiavi gestite dal cliente asimmetriche.

Le chiavi gestite dal cliente vengono configurate quando crei snapshot crittografati o un modello di avvio che specifichi i volumi crittografati, oppure abiliti la crittografia di default.

Configurare le policy chiave

Le chiavi KMS devono avere una policy che permetta a Dimensionamento automatico Amazon EC2 di avviare istanze con volumi Amazon EBS crittografati con una chiave gestita dal cliente.

Utilizza gli esempi in questa pagina per configurare una policy chiave per fornire l'accesso Dimensionamento automatico Amazon EC2 alla tua chiave gestita dal cliente. Puoi modificare la policy della chiave gestita dal cliente sia al momento della creazione della chiave che in seguito.

Dovrai aggiungere almeno due istruzioni alla tua policy della chiave, affinché questa funzioni con Dimensionamento automatico Amazon EC2.

- La prima istruzione permette all'identità IAM specificata nell'elemento `Principal` di utilizzare la chiave gestita dal cliente direttamente. Include le autorizzazioni per eseguire `AWS KMS EncryptDecrypt`, `ReEncrypt*GenerateDataKey*`, e `DescribeKey` operazioni sulla chiave.
- La seconda istruzione consente all'identità IAM specificata nell'elemento `Principal` di utilizzare l'`CreateGrant` operazione per generare concessioni che delegano un sottoinsieme delle proprie

autorizzazioni a Servizi AWS quelle integrate con o con un altro principale. AWS KMS Questo permette di utilizzare la chiave per creare le risorse crittografate per te.

Quando aggiungi le nuove istruzioni alla policy della chiave, non modificare quelle già esistenti nella policy.

Per ciascuno degli esempi seguenti, gli argomenti che devono essere sostituiti, ad esempio un ID chiave o il nome di un ruolo collegato al servizio, vengono visualizzati come testo segnaposto utente. Nella maggior parte dei casi, è possibile sostituire il nome del ruolo collegato ai servizi con il nome di un ruolo collegato ai servizi Dimensionamento automatico Amazon EC2.

Per ulteriori informazioni, consulta le seguenti risorse:

- [Per creare una chiave con, vedi create-key. AWS CLI](#)
- Per aggiornare una politica chiave con il AWS CLI, vedi. [put-key-policy](#)
- Per trovare un ID e un Amazon Resource Name (ARN), della chiave, vedi [Come trovare l'ID e l'ARN della chiave](#) nella Guida per gli sviluppatori di AWS Key Management Service .
- Per ulteriori informazioni sull'utilizzo dei ruoli collegati al servizio per Dimensionamento automatico Amazon EC2, consulta [Ruoli collegati ai servizi per Dimensionamento automatico Amazon EC2](#).
- [Per informazioni sulla crittografia Amazon EBS e KMS in generale, consulta la crittografia Amazon EBS nella Amazon EBSUser Guide e nella Developer Guide.AWS Key Management Service](#)

Esempio 1: sezioni delle policy delle chiavi che permettono l'accesso alla chiave gestita dal cliente

Aggiungi le seguenti due istruzioni alla policy della chiave gestita dal cliente, sostituendo l'ARN di esempio con l'ARN del ruolo collegato ai servizi appropriato, cioè l'accesso permesso alla chiave. In questo esempio, le due seguenti sezioni della policy concedono al ruolo collegato ai servizi denominato AWSServiceRoleForAutoScaling le autorizzazioni per utilizzare la chiave gestita dal cliente.

```
{
  "Sid": "Allow service-linked role use of the customer managed key",
  "Effect": "Allow",
  "Principal": {
    "AWS": [
```



```

        "arn:aws:iam::account-id:role/aws-service-role/
autoscaling.amazonaws.com/AWSServiceRoleForAutoScaling"
    ]
},
"Action": [
    "kms:Encrypt",
    "kms:Decrypt",
    "kms:ReEncrypt*",
    "kms:GenerateDataKey*",
    "kms:DescribeKey"
],
"Resource": "*"
}

```

```

{
  "Sid": "Allow attachment of persistent resources",
  "Effect": "Allow",
  "Principal": {
    "AWS": [
      "arn:aws:iam::account-id:role/aws-service-role/
autoscaling.amazonaws.com/AWSServiceRoleForAutoScaling"
    ]
  },
  "Action": [
    "kms:CreateGrant"
  ],
  "Resource": "*",
  "Condition": {
    "Bool": {
      "kms:GrantIsForAWSResource": true
    }
  }
}

```

Esempio 2: sezioni delle policy delle chiavi che permettono l'accesso multiaccount alla chiave gestita dal cliente

Se crei una chiave gestita dal cliente in un account diverso da quello del gruppo con dimensionamento automatico, è necessario utilizzare una concessione in combinazione con la policy della chiave per consentire l'accesso multi-account alla chiave stessa.

Esistono due passaggi, che devono essere completati nel seguente ordine:

1. Innanzitutto, aggiungi le seguenti due dichiarazioni di policy alla policy della chiave della chiave gestita dal cliente. Sostituisci l'ARN di esempio con l'ARN dell'altro account, assicurandoti di sostituire **111122223333** con l'ID account effettivo del gruppo Auto Scaling in Account AWS cui desideri creare il gruppo Auto Scaling. Ciò permette di dare a un utente o ruolo IAM dell'account specificato l'autorizzazione a creare una concessione per la chiave utilizzando il comando CLI che segue. Tuttavia, di per sé non fornisce l'accesso alla chiave agli utenti.

```
{
  "Sid": "Allow external account 111122223333 use of the customer managed key",
  "Effect": "Allow",
  "Principal": {
    "AWS": [
      "arn:aws:iam::111122223333:root"
    ]
  },
  "Action": [
    "kms:Encrypt",
    "kms:Decrypt",
    "kms:ReEncrypt*",
    "kms:GenerateDataKey*",
    "kms:DescribeKey"
  ],
  "Resource": "*"
}
```

```
{
  "Sid": "Allow attachment of persistent resources in external
account 111122223333",
  "Effect": "Allow",
  "Principal": {
    "AWS": [
      "arn:aws:iam::111122223333:root"
    ]
  },
  "Action": [
    "kms:CreateGrant"
  ],
  "Resource": "*"
}
```

2. Quindi, dall'account in cui desideri creare il gruppo con dimensionamento automatico, crea una concessione che deleghi le autorizzazioni pertinenti al ruolo collegato al servizio appropriato. L'elemento `Grantee Principal` della concessione è l'ARN del ruolo collegato ai servizi appropriato. La `key-id` è l'ARN della chiave.

Il seguente è un comando CLI, [create-grant](#), di esempio che fornisce al ruolo collegato ai servizi denominato `AWSServiceRoleForAutoScaling` nell'account `111122223333` le autorizzazioni per utilizzare la chiave gestita dal cliente nell'account `444455556666`.

```
aws kms create-grant \
  --region us-west-2 \
  --key-id arn:aws:kms:us-
west-2:444455556666:key/1a2b3c4d-5e6f-1a2b-3c4d-5e6f1a2b3c4d \
  --grantee-principal arn:aws:iam::111122223333:role/aws-service-role/
autoscaling.amazonaws.com/AWSServiceRoleForAutoScaling \
  --operations "Encrypt" "Decrypt" "ReEncryptFrom" "ReEncryptTo" "GenerateDataKey"
"GenerateDataKeyWithoutPlaintext" "DescribeKey" "CreateGrant"
```

Affinché questo comando possa completare la richiesta, l'utente che effettua la richiesta deve avere le autorizzazioni per l'operazione `CreateGrant`.

La policy IAM seguente permette a un'identità IAM (utente o ruolo) nell'account `111122223333` di creare una concessione per la chiave gestita dal cliente nell'account `444455556666`.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowCreationOfGrantForTheKMSKeyInExternalAccount444455556666",
      "Effect": "Allow",
      "Action": "kms:CreateGrant",
      "Resource": "arn:aws:kms:us-
west-2:444455556666:key/1a2b3c4d-5e6f-1a2b-3c4d-5e6f1a2b3c4d"
    }
  ]
}
```

Per ulteriori informazioni sulla creazione di una concessione per una chiave KMS in un diverso Account AWS, consulta [Concessioni in AWS KMS](#) nella Guida per gli sviluppatori AWS Key Management Service .

⚠ Important

Il nome del ruolo collegato al servizio specificato come principale assegnatario deve essere il nome di un ruolo esistente. Dopo aver creato la concessione, per assicurarti che la concessione consenta a Dimensionamento automatico Amazon EC2 di utilizzare la chiave KMS specificata, non eliminare e ricreare il ruolo collegato al servizio.

Modificare le policy delle chiavi nella console AWS KMS

Gli esempi nelle seguenti sezioni mostrano solo come aggiungere le istruzioni alla policy di una chiave, che è solo uno dei modi per modificare questo tipo di policy. Il modo più semplice per modificare una policy chiave consiste nell'utilizzare la visualizzazione predefinita della AWS KMS console per le policy chiave e rendere un'identità IAM (utente o ruolo) uno degli utenti chiave per la policy chiave appropriata. Per ulteriori informazioni, consulta [Using the AWS Management Console default view](#) nella AWS Key Management Service Developer Guide.

⚠ Important

Fai attenzione. Le dichiarazioni relative alla politica di visualizzazione predefinita della console includono le autorizzazioni per eseguire AWS KMS Revoke operazioni sulla chiave gestita dal cliente. Se concedi Account AWS l'accesso a una chiave gestita dal cliente nel tuo account e revochi accidentalmente la concessione che ha concesso loro tale autorizzazione, gli utenti esterni non possono più accedere ai loro dati crittografati o alla chiave utilizzata per crittografare i loro dati.

Identity and Access Management per Dimensionamento automatico Amazon EC2

AWS Identity and Access Management (IAM) è un programma Servizio AWS che aiuta un amministratore a controllare in modo sicuro l'accesso alle AWS risorse. Gli amministratori IAM controllano chi è autenticato (con accesso effettuato) e autorizzato (che dispone di autorizzazioni) a utilizzare risorse Dimensionamento automatico Amazon EC2. IAM è un software Servizio AWS che puoi utilizzare senza costi aggiuntivi.

Per utilizzare Amazon EC2 Auto Scaling, hai bisogno di Account AWS un account e delle tue credenziali di sicurezza per accedere al tuo account. Per ulteriori informazioni, consulta le [credenziali AWS di sicurezza](#) nella Guida per l'utente IAM.

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

Controllo accessi

Per autenticare le richieste, puoi disporre di credenziali valide, ma a meno che non tu abbia le autorizzazioni, non puoi creare o accedere a risorse Dimensionamento automatico Amazon EC2. Ad esempio, è necessario disporre delle autorizzazioni per creare gruppi con dimensionamento automatico, istanze di avvio con modelli di avvio e così via.

Le sezioni seguenti forniscono dettagli su come un amministratore IAM può utilizzare IAM per proteggere le risorse Dimensionamento automatico Amazon EC2, controllando chi può eseguire le operazioni Dimensionamento automatico Amazon EC2.

Consigliamo di leggere prima gli argomenti relativi a Amazon EC2. Consulta [Identity and Access Management per Amazon EC2](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux. Dopo aver letto gli argomenti in questa sezione, avrai un quadro più preciso delle autorizzazioni per il controllo degli accessi offerte da Amazon EC2 e di come possono essere adattate alle autorizzazioni a livello di risorsa Dimensionamento automatico Amazon EC2.

Argomenti

- [Funzionamento di Dimensionamento automatico Amazon EC2 con IAM](#)
- [Autorizzazioni API per Dimensionamento automatico Amazon EC2](#)
- [AWS politiche gestite per Amazon EC2 Auto Scaling](#)
- [Ruoli collegati ai servizi per Dimensionamento automatico Amazon EC2](#)
- [Esempio di policy basate su identità di Dimensionamento automatico Amazon EC2](#)
- [Prevenzione del problema "confused deputy" tra servizi](#)
- [Supporto modello di avvio](#)
- [Ruoli IAM per le applicazioni in esecuzione sulle istanze Amazon EC2](#)

Funzionamento di Dimensionamento automatico Amazon EC2 con IAM

Prima di utilizzare IAM per gestire l'accesso a Dimensionamento automatico Amazon EC2, è necessario comprendere quali funzioni IAM sono disponibili per l'uso con Dimensionamento automatico Amazon EC2.

Funzionalità IAM che puoi utilizzare con il Dimensionamento automatico Amazon EC2

Funzionalità IAM	Supporto di Dimensionamento automatico Amazon EC2
Policy basate su identità	Sì
Policy basate su risorse	No
Azioni di policy	Sì
Risorse relative alle policy	Sì
Chiavi di condizione della policy (specifica del servizio)	Sì
Liste di controllo degli accessi (ACL)	No
ABAC (tag nelle policy)	Parziale
Credenziali temporanee	Sì
Ruoli di servizio	Sì
Ruoli collegati al servizio	Sì

Per avere una visione di alto livello del funzionamento di Amazon EC2 Auto Scaling e di Servizi AWS altro tipo con la maggior parte delle funzionalità IAM, [Servizi AWS consulta la sezione relativa alla compatibilità con IAM nella IAM User Guide](#).

Policy basate su identità Dimensionamento automatico Amazon EC2

Supporta le policy basate su identità	Sì
---------------------------------------	----

Le policy basate su identità sono documenti di policy di autorizzazione JSON che è possibile allegare a un'identità (utente, gruppo di utenti o ruolo IAM). Tali policy definiscono le azioni che utenti e ruoli possono eseguire, su quali risorse e in quali condizioni. Per informazioni su come creare una policy basata su identità, consulta [Creazione di policy IAM](#) nella Guida per l'utente di IAM.

Con le policy basate su identità di IAM, è possibile specificare quali operazioni e risorse sono consentite o respinte, nonché le condizioni in base alle quali le operazioni sono consentite o respinte. Non è possibile specificare l'entità principale in una policy basata sull'identità perché si applica all'utente o al ruolo a cui è associato. Per informazioni su tutti gli elementi utilizzabili in una policy JSON, consulta [Guida di riferimento agli elementi delle policy JSON IAM](#) nella Guida per l'utente di IAM.

Policy basate su risorse di Dimensionamento automatico Amazon EC2

Supporta le policy basate su risorse	No
--------------------------------------	----

Le policy basate su risorse sono documenti di policy JSON che è possibile allegare a una risorsa. Gli esempi più comuni di policy basate su risorse sono le policy di attendibilità dei ruoli IAM e le policy dei bucket Amazon S3. Nei servizi che supportano policy basate sulle risorse, gli amministratori dei servizi possono utilizzarle per controllare l'accesso a una risorsa specifica. Quando è allegata a una risorsa, una policy definisce le azioni che un principale può eseguire su tale risorsa e a quali condizioni. È necessario [specificare un principale](#) in una policy basata sulle risorse. I principali possono includere account, utenti, ruoli, utenti federati o Servizi AWS

Per consentire l'accesso multi-account, puoi specificare un intero account o entità IAM in un altro account come principale in una policy basata sulle risorse. L'aggiunta di un principale multi-account a una policy basata sulle risorse rappresenta solo una parte della relazione di trust. Quando il principale e la risorsa sono diversi Account AWS, un amministratore IAM dell'account affidabile deve inoltre concedere all'entità principale (utente o ruolo) l'autorizzazione ad accedere alla risorsa. L'autorizzazione viene concessa collegando all'entità una policy basata sull'identità. Tuttavia, se una policy basata su risorse concede l'accesso a un principale nello stesso account, non sono richieste ulteriori policy basate su identità. Per ulteriori informazioni, consulta [Differenza tra i ruoli IAM e le policy basate su risorse](#) nella Guida per l'utente di IAM.

Operazioni di policy per Dimensionamento automatico Amazon EC2

Supporta le azioni di policy	Sì
------------------------------	----

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.

Per visualizzare un elenco delle operazioni Dimensionamento automatico Amazon EC2, consulta [Operazioni definite da Dimensionamento automatico Amazon EC2](#) nella Documentazione di riferimento alla autorizzazione di servizio.

Le operazioni delle policy in Dimensionamento automatico Amazon EC2 utilizzano il seguente prefisso prima dell'operazione:

```
autoscaling
```

Per specificare più operazioni in una sola istruzione, occorre separarle con la virgola.

```
"Action": [  
  "autoscaling:action1",  
  "autoscaling:action2"  
]
```

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

```
"Action": "autoscaling:Describe*"
```

Risorse per le policy per Dimensionamento automatico Amazon EC2

Supporta le risorse di policy

Sì

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

L'elemento JSON `Resource` della policy specifica l'oggetto o gli oggetti ai quali si applica l'azione. Le istruzioni devono includere un elemento `Resource` o un elemento `NotResource`. Come best practice, specifica una risorsa utilizzando il suo [nome della risorsa Amazon \(ARN\)](#). Puoi eseguire questa operazione per azioni che supportano un tipo di risorsa specifico, note come autorizzazioni a livello di risorsa.

Per le azioni che non supportano le autorizzazioni a livello di risorsa, ad esempio le operazioni di elenco, utilizza un carattere jolly (*) per indicare che l'istruzione si applica a tutte le risorse.

```
"Resource": "*" 
```

È possibile utilizzare gli ARN per identificare i gruppi con dimensionamento automatico e le configurazioni di avvio a cui si applica la policy IAM.

Un gruppo con dimensionamento automatico ha il seguente ARN.

```
"Resource": "arn:aws:autoscaling:region:account-id:autoScalingGroup:uuid:autoScalingGroupName/asg-name"
```

Una configurazione di avvio ha il seguente ARN.

```
"Resource": "arn:aws:autoscaling:region:account-id:launchConfiguration:uuid:launchConfigurationName/lc-name"
```

Per specificare un gruppo con dimensionamento automatico con l'operazione `CreateAutoScalingGroup`, devi sostituire l'UUID con un carattere jolly (*), come mostrato di seguito.

```
"Resource": "arn:aws:autoscaling:region:account-id:autoScalingGroup:*:autoScalingGroupName/asg-name"
```

Per specificare una configurazione di avvio con l'operazione `CreateLaunchConfiguration`, devi sostituire l'UUID con un carattere jolly (*), come mostrato di seguito.

```
"Resource": "arn:aws:autoscaling:region:account-id:launchConfiguration:*:launchConfigurationName/lc-name"
```

Per ulteriori informazioni sui tipi di risorsa di Dimensionamento automatico Amazon EC2 e sui relativi ARN, consulta [Risorse definite da Dimensionamento automatico Amazon EC2](#) nella Documentazione

di riferimento alla autorizzazione di servizio. Per informazioni sulle operazioni con cui è possibile specificare l'ARN di ogni risorsa, consulta [Operazioni definite da Dimensionamento automatico Amazon EC2](#).

Note

Per un esempio di policy IAM che utilizza gli ARN per controllare l'accesso ai gruppi con dimensionamento automatico, consulta [Verifica quale gruppo con dimensionamento automatico è possibile eliminare](#).

Non tutte le operazioni Dimensionamento automatico Amazon EC2 supportano le autorizzazioni a livello di risorsa. Per le operazioni che non supportano le autorizzazioni a livello di risorsa, è necessario utilizzare un carattere jolly (*) come risorsa.

Le seguenti Dimensionamento automatico Amazon EC2 non supportano le autorizzazioni a livello di risorsa.

- DescribeAccountLimits
- DescribeAdjustmentTypes
- DescribeAutoScalingGroups
- DescribeAutoScalingInstances
- DescribeAutoScalingNotificationTypes
- DescribeInstanceRefreshes
- DescribeLaunchConfigurations
- DescribeLifecycleHooks
- DescribeLifecycleHookTypes
- DescribeLoadBalancers
- DescribeLoadBalancerTargetGroups
- DescribeMetricCollectionTypes
- DescribeNotificationConfigurations
- DescribePolicies
- DescribeScalingActivities
- DescribeScalingProcessTypes

- DescribeScheduledActions
- DescribeTags
- DescribeTerminationPolicyTypes
- DescribeWarmPool

Chiavi di condizione della policy per Dimensionamento automatico Amazon EC2

Supporta le chiavi di condizione delle policy specifiche del servizio	Si
---	----

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 `Condition` (o blocco `Condition`) consente di specificare le condizioni in cui un'istruzione è in vigore. L'elemento `Condition` è facoltativo. Puoi compilare espressioni condizionali che utilizzano [operatori di condizione](#), ad esempio uguale a o minore di, per soddisfare la condizione nella policy con i valori nella richiesta.

Se specifichi più elementi `Condition` in un'istruzione o più chiavi in un singolo elemento `Condition`, questi vengono valutati da AWS utilizzando un'operazione AND logica. Se si specificano più valori per una singola chiave di condizione, AWS valuta la condizione utilizzando un'operazione logica. OR Tutte le condizioni devono essere soddisfatte prima che le autorizzazioni dell'istruzione vengano concesse.

Puoi anche utilizzare variabili segnaposto quando specifichi le condizioni. Ad esempio, puoi autorizzare un utente IAM ad accedere a una risorsa solo se è stata taggata con il relativo nome utente IAM. Per ulteriori informazioni, consulta [Elementi delle policy IAM: variabili e tag](#) nella Guida per l'utente di IAM.

AWS supporta chiavi di condizione globali e chiavi di condizione specifiche del servizio. Per visualizzare tutte le chiavi di condizione AWS globali, consulta le chiavi di [contesto delle condizioni AWS globali nella Guida](#) per l'utente IAM.

Dimensionamento automatico Amazon EC2 supporta le seguenti chiavi di condizione che possono essere utilizzate per controllare l'accesso alle operazioni supportate e sulla configurazione dei gruppi con dimensionamento automatico:

- `autoscaling:InstanceTypes`
- `autoscaling:LaunchConfigurationName`
- `autoscaling:LaunchTemplateVersionSpecified`
- `autoscaling:LoadBalancerNames`
- `autoscaling:MaxSize`
- `autoscaling:MinSize`
- `autoscaling:ResourceTag/key-name: tag-value`
- `autoscaling:TargetGroupARNs`
- `autoscaling:VPCZoneIdentifiers`

Le seguenti chiavi di condizione sono specifiche per creare richieste di configurazione di avvio:

- `autoscaling:ImageId`
- `autoscaling:InstanceType`
- `autoscaling:MetadataHttpEndpoint`
- `autoscaling:MetadataHttpPutResponseHopLimit`
- `autoscaling:MetadataHttpTokens`
- `autoscaling:SpotPrice`

Dimensionamento automatico Amazon EC2 supporta anche le seguenti chiavi di condizione globali che puoi utilizzare per definire le autorizzazioni in base ai tag nella richiesta o presenti nel gruppo con dimensionamento automatico. Per ulteriori informazioni, consulta [Tag di gruppi e istanze Auto Scaling](#).

- `aws:RequestTag/key-name: tag-value`
- `aws:ResourceTag/key-name: tag-value`
- `aws:TagKeys: [tag-key, ...]`

Per scoprire con quali operazioni API di Dimensionamento automatico Amazon EC2 puoi utilizzare una chiave di condizione, consulta [Operazioni definite da Dimensionamento automatico Amazon EC2](#) nella Documentazione di riferimento alla autorizzazione di servizio. Per ulteriori informazioni sull'utilizzo delle chiavi di condizione di Dimensionamento automatico Amazon EC2, consulta [Chiavi di condizione per Dimensionamento automatico Amazon EC2](#).

Note

Per esempi di policy IAM che utilizzano le chiavi di condizione per controllare l'accesso alle operazioni supportate e sull'applicazione della configurazione dei gruppi con dimensionamento automatico, consulta le seguenti risorse:

- [Richiesta di un modello di avvio e di un numero di versione](#)— Questo esempio impone che un modello di avvio e il numero di versione del modello di avvio debbano essere specificati durante la creazione o l'aggiornamento dei gruppi di Auto Scaling.
- [Controlla le dimensioni dei gruppi con dimensionamento automatico che possono essere creati](#)— Questo esempio impone vincoli sui possibili valori per le MaxSize proprietà MinSize and durante la creazione o l'aggiornamento di gruppi Auto Scaling con un tag specifico.
- [Verifica di quali policy di dimensionamento sia possibile eliminare](#)— Questo esempio impone che l'eliminazione delle policy di scaling sia consentita solo per i gruppi Auto Scaling senza un tag specifico.

ACL in Dimensionamento automatico Amazon EC2

Supporta le ACL

No

Le liste di controllo degli accessi (ACL) controllano quali principali (membri, utenti o ruoli dell'account) hanno le autorizzazioni ad accedere a una risorsa. Le ACL sono simili alle policy basate su risorse, sebbene non utilizzino il formato del documento di policy JSON.

ABAC con Dimensionamento automatico Amazon EC2

Supporta ABAC (tag nelle policy)

Parziale

Il controllo dell'accesso basato su attributi (ABAC) è una strategia di autorizzazione che definisce le autorizzazioni in base agli attributi. In AWS, questi attributi sono chiamati tag. Puoi allegare tag a entità IAM (utenti o ruoli) e a molte AWS risorse. L'assegnazione di tag alle entità e alle risorse è il primo passaggio di ABAC. In seguito, vengono progettate policy ABAC per consentire operazioni quando il tag dell'entità principale corrisponde al tag sulla risorsa a cui si sta provando ad accedere.

La strategia ABAC è utile in ambienti soggetti a una rapida crescita e aiuta in situazioni in cui la gestione delle policy diventa impegnativa.

Per controllare l'accesso basato su tag, fornisci informazioni sui tag nell'[elemento condizione](#) di una policy utilizzando le chiavi di condizione `aws:ResourceTag/key-name`, `aws:RequestTag/key-name` o `aws:TagKeys`.

Se un servizio supporta tutte e tre le chiavi di condizione per ogni tipo di risorsa, il valore per il servizio è Yes (Sì). Se un servizio supporta tutte e tre le chiavi di condizione solo per alcuni tipi di risorsa, allora il valore sarà Parziale.

Per ulteriori informazioni su ABAC, consulta [Che cos'è ABAC?](#) nella Guida per l'utente di IAM. Per visualizzare un tutorial con i passaggi per l'impostazione di ABAC, consulta [Utilizzo del controllo degli accessi basato su attributi \(ABAC\)](#) nella Guida per l'utente di IAM.

ABAC è possibile per le risorse che supportano i tag, ma non tutto supporta i tag. Le configurazioni di avvio e le policy di dimensionamento non supportano i tag, ma i gruppi con dimensionamento automatico supportano i tag.

Per ulteriori informazioni, consulta [Tag di gruppi e istanze Auto Scaling](#).

Utilizzo di credenziali temporanee con Dimensionamento automatico Amazon EC2

Supporta le credenziali temporanee	Sì
------------------------------------	----

Alcuni Servizi AWS non funzionano quando accedi utilizzando credenziali temporanee. Per ulteriori informazioni, incluse quelle che Servizi AWS funzionano con credenziali temporanee, consulta la sezione relativa alla [Servizi AWS compatibilità con IAM nella IAM User Guide](#).

Stai utilizzando credenziali temporanee se accedi AWS Management Console utilizzando qualsiasi metodo tranne nome utente e password. Ad esempio, quando accedi AWS utilizzando il link Single Sign-On (SSO) della tua azienda, tale processo crea automaticamente credenziali temporanee. Le credenziali temporanee vengono create in automatico anche quando accedi alla console come utente e poi cambi ruolo. Per ulteriori informazioni sullo scambio dei ruoli, consulta [Cambio di un ruolo \(console\)](#) nella Guida per l'utente di IAM.

È possibile creare manualmente credenziali temporanee utilizzando l'API o AWS CLI. AWS consiglia di generare quindi possibile utilizzare tali credenziali temporanee per accedere. AWS consiglia di generare

dinamicamente credenziali temporanee anziché utilizzare chiavi di accesso a lungo termine. Per ulteriori informazioni, consulta [Credenziali di sicurezza provvisorie in IAM](#).

Ruoli di servizio per Dimensionamento automatico Amazon EC2

Supporta i ruoli di servizio

Sì

Un ruolo di servizio è un [ruolo IAM](#) che un servizio assume per eseguire operazioni per tuo conto. Un amministratore IAM può creare, modificare ed eliminare un ruolo di servizio dall'interno di IAM. Per ulteriori informazioni, consulta la sezione [Creazione di un ruolo per delegare le autorizzazioni a un Servizio AWS](#) nella Guida per l'utente di IAM.

Quando si crea un hook del ciclo di vita che notifica un argomento Amazon SNS o una coda di Amazon SQS, è necessario specificare un ruolo per consentire a Dimensionamento automatico Amazon EC2 di accedere ad Amazon SNS o Amazon SQS per conto dell'utente. Utilizzare la console IAM per impostare il ruolo di servizio per il hook del ciclo di vita. La console consente di creare un ruolo con un set sufficiente di autorizzazioni utilizzando una policy gestita. Per ulteriori informazioni, consultare [Ricezione di notifiche tramite Amazon SNS](#) e [Ricezione di notifiche tramite Amazon SQS](#).

Quando crei un gruppo Auto Scaling, puoi facoltativamente assegnare un ruolo di servizio per consentire alle istanze Amazon EC2 di accedere ad altre istanze per tuo conto. Servizi AWS Il ruolo di servizio per le istanze Amazon EC2 (chiamate anche profilo dell'istanza Amazon EC2 per un modello di avvio o configurazione di avvio) è un tipo speciale di ruolo di servizio che viene assegnato a ogni istanza EC2 in un gruppo con dimensionamento automatico quando l'istanza viene avviata. Puoi utilizzare la console IAM e creare o modificare AWS CLI questo ruolo di servizio. Per ulteriori informazioni, consulta [Ruoli IAM per le applicazioni in esecuzione sulle istanze Amazon EC2](#).

Warning

La modifica delle autorizzazioni per un ruolo di servizio potrebbe compromettere la funzionalità di Dimensionamento automatico Amazon EC2. Modifica i ruoli del servizio solo quando Dimensionamento automatico Amazon EC2 fornisce le indicazioni per farlo.

Ruoli collegati ai servizi per Dimensionamento automatico Amazon EC2

Supporta i ruoli collegati ai servizi

Sì

Un ruolo collegato al servizio è un tipo di ruolo di servizio collegato a un servizio AWS. Il servizio può assumere il ruolo per eseguire un'azione per tuo conto. I ruoli collegati al servizio vengono visualizzati nel tuo account Account AWS e sono di proprietà del servizio. Un amministratore IAM può visualizzare le autorizzazioni per i ruoli collegati ai servizi, ma non modificarle.

Per dettagli su come creare e gestire i ruoli collegati ai servizi Amazon EC2 Auto Scaling, consulta [Ruoli collegati ai servizi per Dimensionamento automatico Amazon EC2](#).

Autorizzazioni API per Dimensionamento automatico Amazon EC2

Devi concedere agli utenti l'autorizzazione a chiamare le operazioni API di dimensionamento automatico Amazon EC2 di cui hanno bisogno, come descritto in [Operazioni di policy per Dimensionamento automatico Amazon EC2](#). Inoltre, per alcune azioni di Amazon EC2 Auto Scaling, devi concedere agli utenti l'autorizzazione a richiamare azioni specifiche da altre API. AWS

Autorizzazioni richieste da altre API AWS

Oltre alle autorizzazioni dell'API Amazon EC2 Auto Scaling, gli utenti devono disporre delle seguenti autorizzazioni da AWS altre API per eseguire correttamente l'azione associata.

Creazione di un gruppo con dimensionamento automatico
(`autoscaling:CreateAutoScalingGroup`)

- `iam:CreateServiceLinkedRole`— Per creare il ruolo predefinito collegato al servizio, se tale ruolo non esiste ancora.
- `iam:PassRole`— Per passare un ruolo IAM al servizio o alle istanze EC2 all'avvio. Necessario quando viene fornito un ruolo collegato a un servizio non predefinito, un ruolo IAM per un hook del ciclo di vita o un modello di avvio che specifica un profilo dell'istanza (un container per un ruolo IAM).
- `ec2:RunInstance`— Per avviare istanze quando viene fornito un modello di avvio.
- `ec2:CreateTags`— Per etichettare istanze e volumi all'avvio quando viene fornito un modello di lancio con una specifica di tag.

Creazione di un hook del ciclo di vita (`autoscaling:PutLifecycleHook`)

- `iam:PassRole`— Per passare un ruolo IAM al servizio. Necessario quando viene fornito un ruolo IAM.

Collega un gruppo target VPC Lattice (`autoscaling:AttachTrafficSources`)

- `vpc-lattice:RegisterTargets`— Per registrare automaticamente le istanze con il gruppo target.

Scollegare un gruppo target VPC Lattice (autoscaling:DetachTrafficSources)

- `vpc-lattice:DeregisterTargets`— Per annullare automaticamente la registrazione delle istanze con il gruppo target.

Crea una configurazione di avvio (autoscaling:CreateLaunchConfiguration).

- `ec2:DescribeImages`
- `ec2:DescribeInstances`
- `ec2:DescribeInstanceAttribute`
- `ec2:DescribeKeyPairs`
- `ec2:DescribeSecurityGroups`
- `ec2:DescribeSpotInstanceRequests`
- `ec2:DescribeVpcClassicLink`
- `iam:PassRole`— Per passare un ruolo IAM alle istanze EC2 al momento del lancio. Necessario quando una configurazione di avvio specifica un profilo dell'istanza (un container per un ruolo IAM).

AWS politiche gestite per Amazon EC2 Auto Scaling

Una politica AWS gestita è una politica autonoma creata e amministrata da AWS. Le politiche gestite sono progettate per fornire autorizzazioni per molti casi d'uso comuni, in modo da poter iniziare ad assegnare autorizzazioni a utenti, gruppi e ruoli.

Tieni presente che le policy AWS gestite potrebbero non concedere le autorizzazioni con il privilegio minimo per i tuoi casi d'uso specifici, poiché sono disponibili per tutti i clienti. AWS Consigliamo pertanto di ridurre ulteriormente le autorizzazioni definendo [policy gestite dal cliente](#) specifiche per i tuoi casi d'uso.

Non è possibile modificare le autorizzazioni definite nelle politiche gestite. Se AWS aggiorna le autorizzazioni definite in una politica AWS gestita, l'aggiornamento ha effetto su tutte le identità principali (utenti, gruppi e ruoli) a cui è associata la politica. AWS è più probabile che aggiorni una policy AWS gestita quando nel Servizio AWS viene lanciata una nuova o quando diventano disponibili nuove operazioni API per i servizi esistenti.

Per ulteriori informazioni, consultare [Policy gestite da AWS](#) nella Guida per l'utente di IAM.

Policy gestite da Dimensionamento automatico Amazon EC2

Puoi allegare le seguenti politiche gestite alle tue identità AWS Identity and Access Management (utenti o ruoli) (IAM). Ciascuna policy consente di accedere ad alcune o a tutte le operazioni API per Dimensionamento automatico Amazon EC2.

- [AutoScalingFullAccess](#)— Garantisce l'accesso completo alle identità di Amazon EC2 Auto Scaling for IAM che richiedono l'accesso completo ad Amazon EC2 Auto Scaling dagli o SDK, ma non l'accesso. AWS CLI AWS Management Console
- [AutoScalingReadOnlyAccess](#)— Concede l'accesso in sola lettura alle identità di Amazon EC2 Auto Scaling for IAM che effettuano chiamate solo agli o SDK. AWS CLI
- [AutoScalingConsoleFullAccess](#)— Garantisce l'accesso completo ad Amazon EC2 Auto Scaling utilizzando. AWS Management Console Questa policy funziona quando si utilizzano configurazioni di avvio, ma non quando si utilizzano modelli di avvio.
- [AutoScalingConsoleReadOnlyAccess](#)— Concede l'accesso in sola lettura ad Amazon EC2 Auto Scaling utilizzando. AWS Management Console Questa policy funziona quando si utilizzano configurazioni di avvio, ma non quando si utilizzano modelli di avvio.

Quando utilizzi modelli di avvio dalla console, devi concedere autorizzazioni specifiche aggiuntive per avviare i modelli, come illustrato in [Supporto modello di avvio](#). La console Dimensionamento automatico Amazon EC2 richiede autorizzazioni per le operazioni ec2, in modo da poter visualizzare informazioni sui modelli e sulle istanze di avvio tramite i modelli di avvio.

Policy gestita di AutoScalingServiceRolePolicy AWS

Non puoi [AutoScalingServiceRolePolicy](#) collegarti alle tue identità IAM. Questa policy è allegata a un ruolo collegato al servizio che consente a Dimensionamento automatico Amazon EC2 di avviare e terminare le istanze. Per ulteriori informazioni, consulta [Ruoli collegati ai servizi per Dimensionamento automatico Amazon EC2](#).

Amazon EC2 Auto Scaling si aggiorna alle politiche gestite AWS

Visualizza i dettagli sugli aggiornamenti delle politiche AWS gestite per Amazon EC2 Auto Scaling da quando questo servizio ha iniziato a tracciare queste modifiche. Per gli avvisi automatici sulle modifiche apportate a questa pagina, sottoscrivi il feed RSS nella pagina della cronologia dei documenti di Dimensionamento automatico Amazon EC2.

Modifica	Descrizione	Data
<p>Dimensionamento automatico Amazon EC2 aggiunge le autorizzazioni al suo ruolo collegato al servizio</p>	<p>La <code>AutoScalingServiceRolePolicy</code> policy ora concede le autorizzazioni per richiamare l'azione dell'API Amazon GetSecurityGroupsForVpc EC2 per ottenere tutti i gruppi di sicurezza di un VPC per migliorare la convalida e l'azione dell'API Amazon EC2 GetInstanceTypesFromInstanceRequirements per ottenere informazioni sui tipi di istanza che soddisfano un determinato set di requisiti di istanza. Per ulteriori informazioni, consulta Ruoli collegati ai servizi per Dimensionamento automatico Amazon EC2.</p>	<p>29 febbraio 2024</p>
<p>Dimensionamento automatico Amazon EC2 aggiunge le autorizzazioni al suo ruolo collegato al servizio</p>	<p>La policy <code>AutoScalingServiceRolePolicy</code> ora concede al servizio le autorizzazioni per accedere alle operazioni API necessari e per un'integrazione con VPC Lattice.</p> <ul style="list-style-type: none"> • Operazioni <code>GetTargetGroup</code> e <code>ListTargetGroup</code>. Necessarie per recuperare le informazioni relative ai gruppi di destinazione di VPC Lattice. • Operazioni <code>RegisterTargets</code> e <code>DeregisterTargets</code>. 	<p>6 dicembre 2022</p>

Modifica	Descrizione	Data
	<p><code>Targets</code> . Necessarie per registrare e annullare la registrazione delle istanze dai gruppi di destinazione di VPC Lattice.</p> <ul style="list-style-type: none"> • <code>ListTargets</code> . Consente a Dimensionamento automatico Amazon EC2 di recuperare informazioni sulle istanze registrate nei gruppi di destinazione di VPC Lattice. <p>Per ulteriori informazioni, consulta Ruoli collegati ai servizi per Dimensionamento automatico Amazon EC2.</p>	
<p>Dimensionamento automatico Amazon EC2 aggiunge le autorizzazioni al suo ruolo collegato al servizio</p>	<p>Per supportare l'utilizzo di un AWS Systems Manager parametro come alias per un ID AMI durante la creazione di un modello di avvio, la <code>AutoScalingServiceRolePolicy</code> policy ora concede l'autorizzazione a chiamare l'azione AWS Systems Manager GetParametersAPI. Per ulteriori informazioni, consulta Ruoli collegati ai servizi per Dimensionamento automatico Amazon EC2.</p>	<p>28 marzo 2022</p>

Modifica	Descrizione	Data
Dimensionamento automatico Amazon EC2 aggiunge le autorizzazioni al suo ruolo collegato al servizio	Per supportare la scalabilità predittiva, la <code>AutoScalingServiceRolePolicy</code> policy ora include l'autorizzazione a richiamare l'azione dell'API. <code>CloudWatchGetMetricData</code> Per ulteriori informazioni, consulta Ruoli collegati ai servizi per Dimensionamento automatico Amazon EC2 .	19 maggio 2021
Dimensionamento automatico Amazon EC2 ha iniziato a monitorare le modifiche	Amazon EC2 Auto Scaling ha iniziato a tracciare le modifiche per AWS le sue politiche gestite.	19 maggio 2021

Ruoli collegati ai servizi per Dimensionamento automatico Amazon EC2

Dimensionamento automatico Amazon EC2 utilizza ruoli collegati ai servizi per le autorizzazioni di cui ha bisogno per eseguire chiamate ad altri Servizi AWS per tuo conto. Un ruolo collegato ai servizi è un tipo unico di ruolo IAM collegato direttamente a un Servizio AWS

I ruoli collegati ai servizi offrono un modo sicuro per concedere autorizzazioni ad altri Servizi AWS , in quanto solo il servizio associato può assumere un ruolo collegato al servizio. Per ulteriori informazioni, consulta [Utilizzo di ruoli collegati a servizi](#) nella Guida per l'utente di IAM . I ruoli collegati ai servizi consentono inoltre di rendere visibili tutte le chiamate API. AWS CloudTrail In questo modo, il monitoraggio e i requisiti di controllo sono più semplici perché è possibile tenere traccia di tutte le operazioni che Dimensionamento automatico Amazon EC2 esegue per te. Per ulteriori informazioni, consulta [Registra le chiamate API di Amazon EC2 Auto Scaling con AWS CloudTrail](#).

Le seguenti sezioni descrivono come creare e gestire i ruoli legati al servizio Dimensionamento automatico Amazon EC2. Per iniziare, configura le autorizzazioni per consentire a un'identità IAM

(ad esempio un utente o un ruolo) di creare, modificare o eliminare un ruolo collegato ai servizi. Per ulteriori informazioni, consulta [Utilizzo di ruoli collegati a servizi](#) nella Guida per l'utente di IAM .

Indice

- [Panoramica](#)
- [Autorizzazioni concesse dal ruolo collegato ai servizi](#)
- [Creazione di un ruolo collegato ai servizi \(automatico\)](#)
- [Creazione di un ruolo collegato ai servizi \(manuale\)](#)
- [Modifica del ruolo collegato ai servizi](#)
- [Eliminazione del ruolo collegato ai servizi](#)
- [Regioni supportate per i ruoli collegati ai servizi di Dimensionamento automatico Amazon EC2](#)

Panoramica

Esistono due tipi di ruoli collegati ai servizi di Dimensionamento automatico Amazon EC2 :

- Il ruolo predefinito collegato al servizio per il tuo account, denominato. `AWSServiceRoleForAutoScaling` Questo ruolo viene automaticamente assegnato ai tuoi gruppi Auto Scaling, a meno che non specifichi un diverso ruolo collegato ai servizi.
- ***Un ruolo collegato al servizio con un suffisso personalizzato specificato al momento della creazione del ruolo, ad esempio `_mysuffix.AWSServiceRoleForAutoScaling`***

Le autorizzazioni di un ruolo collegato al servizio con un suffisso personalizzato sono identiche a quelle del ruolo collegato al servizio predefinito. In entrambi i casi, non è possibile modificare i ruoli o eliminarli se sono ancora in uso da parte di un gruppo con dimensionamento automatico. L'unica differenza è il suffisso del nome del ruolo.

Puoi specificare entrambi i ruoli quando modifichi le politiche AWS Key Management Service chiave per consentire la crittografia delle istanze lanciate da Amazon EC2 Auto Scaling con la tua chiave gestita dal cliente. Tuttavia, se prevedi di fornire l'accesso granulare a una determinata chiave gestita dal cliente, ti consigliamo di utilizzare un ruolo collegato a un servizio con suffisso personalizzato. L'utilizzo di un ruolo collegato a un servizio con suffisso personalizzato offre:

- Un maggiore controllo sulla chiave gestita dal cliente

- La capacità di tracciare quale gruppo di Auto Scaling ha effettuato una chiamata API nei tuoi log CloudTrail

Se crei chiavi gestite dai clienti a cui non tutti gli utenti devono avere accesso, segui questi passaggi per permettere l'uso di un ruolo collegato al servizio con suffisso personalizzato:

1. Creazione di un ruolo collegato al servizio con un suffisso personalizzato. Per ulteriori informazioni, consulta [Creazione di un ruolo collegato ai servizi \(manuale\)](#).
2. Assegnazione al ruolo collegato al servizio l'accesso a una chiave gestita dal cliente. Per ulteriori informazioni sulla policy chiave che permette alla chiave di essere utilizzata da un ruolo collegato al servizio, vedi [Policy AWS KMS chiave richiesta per l'uso con volumi crittografati](#).
3. Concedere agli utenti l'accesso al ruolo collegato al servizio creato. Per ulteriori informazioni sulla creazione di policy IAM, consulta [Controlla quale ruolo collegato al servizio può essere passato \(utilizzando\) PassRole](#). Se gli utenti provano a specificare un ruolo collegato al servizio senza l'autorizzazione necessaria per passare tale ruolo al servizio, riceveranno un messaggio d'errore.

Autorizzazioni concesse dal ruolo collegato ai servizi

Amazon EC2 Auto Scaling utilizza il ruolo collegato al servizio `AWSServiceRoleForAutoScaling` denominato o il tuo ruolo collegato al servizio con suffisso personalizzato.

Ai fini dell'assunzione del ruolo, il ruolo collegato ai servizi considera attendibile il seguente servizio:

- `autoscaling.amazonaws.com`

Il ruolo utilizza la policy [AutoScalingServiceRolePolicy](#), che include le seguenti autorizzazioni:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "EC2InstanceManagement",
      "Effect": "Allow",
      "Action": [
        "ec2:AttachClassicLinkVpc",
        "ec2:CancelSpotInstanceRequests",
        "ec2:CreateFleet",
```

```

    "ec2:CreateTags",
    "ec2>DeleteTags",
    "ec2:Describe*",
    "ec2:DetachClassicLinkVpc",
    "ec2:GetInstanceTypesFromInstanceRequirements",
    "ec2:GetSecurityGroupsForVpc",
    "ec2:ModifyInstanceAttribute",
    "ec2:RequestSpotInstances",
    "ec2:RunInstances",
    "ec2:StartInstances",
    "ec2:StopInstances",
    "ec2:TerminateInstances"
  ],
  "Resource": "*"
},
{
  "Sid": "EC2InstanceProfileManagement",
  "Effect": "Allow",
  "Action": [
    "iam:PassRole"
  ],
  "Resource": "*",
  "Condition": {
    "StringLike": {
      "iam:PassedToService": "ec2.amazonaws.com*"
    }
  }
},
{
  "Sid": "EC2SpotManagement",
  "Effect": "Allow",
  "Action": [
    "iam:CreateServiceLinkedRole"
  ],
  "Resource": "*",
  "Condition": {
    "StringEquals": {
      "iam:AWSServiceName": "spot.amazonaws.com"
    }
  }
},
{
  "Sid": "ELBManagement",
  "Effect": "Allow",

```



```

    "Action":[
      "elasticloadbalancing:Register*",
      "elasticloadbalancing:Deregister*",
      "elasticloadbalancing:Describe*"
    ],
    "Resource": "*"
  },
  {
    "Sid": "CWManagement",
    "Effect": "Allow",
    "Action": [
      "cloudwatch:DeleteAlarms",
      "cloudwatch:DescribeAlarms",
      "cloudwatch:GetMetricData",
      "cloudwatch:PutMetricAlarm"
    ],
    "Resource": "*"
  },
  {
    "Sid": "SNSManagement",
    "Effect": "Allow",
    "Action": [
      "sns:Publish"
    ],
    "Resource": "*"
  },
  {
    "Sid": "EventBridgeRuleManagement",
    "Effect": "Allow",
    "Action": [
      "events:PutRule",
      "events:PutTargets",
      "events:RemoveTargets",
      "events>DeleteRule",
      "events:DescribeRule"
    ],
    "Resource": "*",
    "Condition": {
      "StringEquals": {
        "events:ManagedBy": "autoscaling.amazonaws.com"
      }
    }
  },
  {

```

```
    "Sid": "SystemsManagerParameterManagement",
    "Effect": "Allow",
    "Action": [
        "ssm:GetParameters"
    ],
    "Resource": "*"
},
{
    "Sid": "VpcLatticeManagement",
    "Effect": "Allow",
    "Action": [
        "vpc-lattice:DeregisterTargets",
        "vpc-lattice:GetTargetGroup",
        "vpc-lattice:ListTargets",
        "vpc-lattice:ListTargetGroups",
        "vpc-lattice:RegisterTargets"
    ],
    "Resource": "*"
}
]
```

Il ruolo dispone delle autorizzazioni necessarie per eseguire tutte le operazioni seguenti:

- `ec2`— Crea, descrivi, modifica, avvia/interrompi e termina le istanze EC2.
- `iam`— [Passa i ruoli IAM](#) alle istanze EC2 in modo che le applicazioni in esecuzione sulle istanze possano accedere alle credenziali temporanee per il ruolo.
- `iam`— Crea il ruolo `AWSServiceRoleForEC2Spot` collegato ai servizi per consentire ad Amazon EC2 Auto Scaling di avviare istanze Spot per tuo conto.
- `elasticloadbalancing`— Registra e annulla la registrazione delle istanze con Elastic Load Balancing e verifica lo stato dei target registrati.
- `cloudwatch`— Crea, descrivi, modifica ed elimina CloudWatch allarmi per le politiche di scalabilità e recupera le metriche utilizzate per la scalabilità predittiva.
- `sns`— Pubblica notifiche su Amazon SNS all'avvio o alla chiusura delle istanze.
- `events`— Crea, descrivi, aggiorna ed elimina EventBridge regole per tuo conto.
- `ssm`— Leggi i parametri da Parameter Store quando usi un parametro Systems Manager come alias per un ID AMI in un modello di avvio.
- `vpc-lattice`— Registra e annulla la registrazione delle istanze con VPC Lattice e controlla lo stato dei target registrati.

Creazione di un ruolo collegato ai servizi (automatico)

Amazon EC2 Auto Scaling crea `AWSServiceRoleForAutoScaling` il ruolo collegato al servizio per te la prima volta che crei un gruppo Auto Scaling, a meno che non crei manualmente un ruolo collegato al servizio con suffisso personalizzato e lo specifichi durante la creazione del gruppo.

Important

Per creare il ruolo collegato ai servizi, devi disporre delle autorizzazioni IAM. In caso contrario, la creazione automatica non va a buon fine. Per ulteriori informazioni, consulta [Autorizzazioni del ruolo collegato ai servizi](#) nella Guida per l'utente di IAM o le informazioni sulle [Creazione di un ruolo collegato ai servizi](#) in questa guida.

Dimensionamento automatico Amazon EC2 ha iniziato a supportare i ruoli collegati ai servizi a marzo 2018. Se hai già creato un gruppo Auto Scaling in precedenza, Amazon EC2 Auto Scaling `AWSServiceRoleForAutoScaling` ha creato il ruolo nel tuo account. Per ulteriori informazioni, consulta [Visualizzazione di un nuovo ruolo nell'account Account AWS](#) nella Guida per l'utente di IAM.

Creazione di un ruolo collegato ai servizi (manuale)

Come creare un ruolo collegato ai servizi (console)

1. Aprire la console IAM all'indirizzo <https://console.aws.amazon.com/iam/>.
2. Nel pannello di navigazione seleziona Ruoli, quindi Crea ruolo.
3. In Seleziona tipo di entità attendibile, scegli Servizio AWS .
4. Per Choose the service that will use this role (Scegli il servizio che utilizzerà questo ruolo), scegli il caso d'uso EC2 Auto Scaling ed EC2 Auto Scaling .
5. Scegli Next: Permissions (Successivo: autorizzazioni), Next: Tags (Successivo: tag), quindi Next: Review (Successivo: verifica). Nota: non è possibile allegare tag ai ruoli collegati ai servizi durante la creazione.
6. ***Nella pagina di revisione, lascia vuoto il campo Nome ruolo per creare un ruolo collegato al servizio con il nome `AWSServiceRoleForAutoScaling` oppure inserisci un suffisso per creare un ruolo collegato al servizio con il suffisso `name _`. `AWSServiceRoleForAutoScaling`***

7. (Facoltativo) In Role description (Descrizione ruolo) modifica la descrizione per il ruolo collegato ai servizi.
8. Scegli Create role (Crea ruolo).

Come creare un ruolo collegato ai servizi (AWS CLI)

Utilizza il seguente comando [create-service-linked-role](#) CLI per creare un ruolo collegato al servizio per Amazon EC2 Auto Scaling con il suffisso name `_`. `AWSServiceRoleForAutoScaling`

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

L'output di questo comando include l'ARN del ruolo collegato ai servizi, che puoi utilizzare per concedere a tale ruolo l'accesso alla chiave gestita dal cliente.

```
{  
  "Role": {  
    "RoleId": "ABCDEF0123456789ABCDEF",  
    "CreateDate": "2018-08-30T21:59:18Z",  
    "RoleName": "AWSServiceRoleForAutoScaling_suffix",  
    "Arn": "arn:aws:iam::123456789012:role/aws-service-role/  
autoscaling.amazonaws.com/AWSServiceRoleForAutoScaling_suffix",  
    "Path": "/aws-service-role/autoscaling.amazonaws.com/",  
    "AssumeRolePolicyDocument": {  
      "Version": "2012-10-17",  
      "Statement": [  
        {  
          "Action": [  
            "sts:AssumeRole"  
          ],  
          "Principal": {  
            "Service": [  
              "autoscaling.amazonaws.com"  
            ]  
          },  
          "Effect": "Allow"  
        }  
      ]  
    }  
  }  
}
```

```
}
```

Per ulteriori informazioni, consulta [Creazione di un ruolo collegato ai servizi](#) nella Guida per l'utente di IAM.

Modifica del ruolo collegato ai servizi

Non è possibile modificare i ruoli collegati ai servizi che vengono creati per Dimensionamento automatico Amazon EC2. Dopo aver creato un ruolo collegato ai servizi, non è possibile modificare il nome del ruolo o delle sue autorizzazioni. Tuttavia, puoi modificare la descrizione del ruolo. 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 stai utilizzando un gruppo con dimensionamento automatico, ti suggeriamo di eliminare il ruolo collegato al servizio. L'eliminazione del ruolo impedisce di avere un'entità che non viene utilizzata o monitorata e gestita attivamente.

È possibile eliminare un ruolo collegato ai servizi solo dopo aver eliminato le risorse correlate. Questo ti protegge dalla possibilità di revocare inavvertitamente le autorizzazioni Dimensionamento automatico Amazon EC2 per le risorse. Se un ruolo collegato ai servizi viene utilizzato con più gruppi Auto Scaling, è necessario eliminare tutti quelli che utilizzano il ruolo collegato ai servizi, prima di poterlo eliminare. Per ulteriori informazioni, consulta [Eliminazione dell'infrastruttura Auto Scaling](#).

Per eliminare il ruolo collegato ai servizi, puoi utilizzare l'IAM. Per ulteriori informazioni, consulta [Eliminazione del ruolo collegato ai servizi](#) nella Guida per l'utente di IAM.

Se elimini il ruolo `AWSServiceRoleForAutoScaling` collegato al servizio, Amazon EC2 Auto Scaling crea nuovamente il ruolo quando crei un gruppo Auto Scaling e non specifichi un ruolo collegato al servizio diverso.

Regioni supportate per i ruoli collegati ai servizi di Dimensionamento automatico Amazon EC2

Amazon EC2 Auto Scaling supporta l'utilizzo di ruoli collegati al servizio in tutti i luoghi in cui Regioni AWS il servizio è disponibile.

Esempio di policy basate su identità di Dimensionamento automatico Amazon EC2

Per impostazione predefinita, un nuovo utente non Account AWS ha le autorizzazioni per fare nulla. Un amministratore IAM deve creare e assegnare policy IAM che concedano a un'identità IAM (ad esempio un utente o un ruolo) l'autorizzazione per eseguire operazioni API di Dimensionamento automatico Amazon EC2.

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.

Di seguito viene illustrato un esempio di policy di autorizzazione.

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": [
      "autoscaling:CreateAutoScalingGroup",
      "autoscaling:UpdateAutoScalingGroup",
      "autoscaling>DeleteAutoScalingGroup"
    ],
    "Resource": "*",
    "Condition": {
      "StringEquals": { "autoscaling:ResourceTag/purpose": "testing" }
    }
  },
  {
    "Effect": "Allow",
    "Action": "autoscaling:Describe*",
    "Resource": "*"
  }
]
```

Questa policy di esempio concede agli utenti l'autorizzazione per creare, modificare ed eliminare gruppi con dimensionamento automatico, ma solo se il gruppo utilizza il tag **purpose=testing**. Poiché le operazioni Describe non supportano le autorizzazioni a livello di risorsa, devi specificarle in un'istruzione separata senza condizioni. Per avviare istanze con un modello di avvio, l'utente deve inoltre disporre dell'autorizzazione ec2:RunInstances. Per ulteriori informazioni, consulta [Supporto modello di avvio](#).

Note

Puoi creare policy IAM personalizzate per permettere o negare a identità IAM (utenti o ruoli) le autorizzazioni per eseguire operazioni di Dimensionamento automatico Amazon EC2. Puoi allegare queste policy personalizzate alle identità IAM che hanno bisogno delle autorizzazioni specificate. I seguenti esempi mostrano le autorizzazioni per alcuni casi d'uso comuni. Alcune operazioni API di Dimensionamento automatico Amazon EC2 ti permettono di includere gruppi Auto Scaling specifici nella policy, che possono essere creati o modificati dall'operazione. È possibile limitare le risorse di destinazione per queste operazioni specificando gli ARN di singoli gruppi Auto Scaling. Come best practice, tuttavia, si suggerisce di utilizzare policy basate su tag che permettono (o negano) azioni su gruppi Auto Scaling con un tag specifico.

Argomenti

- [Controlla le dimensioni dei gruppi con dimensionamento automatico che possono essere creati](#)
- [Controllo di quali tag chiavi e valori tag possono essere utilizzati](#)
- [Verifica quale gruppo con dimensionamento automatico è possibile eliminare](#)
- [Verifica di quali policy di dimensionamento sia possibile eliminare](#)
- [Controlla l'accesso alle azioni di aggiornamento delle istanze](#)
- [Creazione di un ruolo collegato ai servizi](#)
- [Controlla quale ruolo collegato al servizio può essere passato \(utilizzando\) PassRole](#)

Controlla le dimensioni dei gruppi con dimensionamento automatico che possono essere creati

La policy seguente concede le autorizzazioni per creare e aggiornare tutti i gruppi con dimensionamento automatico con il tag **environment=development**, purché il richiedente non specifichi una dimensione minima inferiore a **1** o una dimensione massima maggiore di **10**. Quando possibile, è consigliabile utilizzare i tag per controllare l'accesso ai gruppi con dimensionamento automatico nel tuo account.

```
{
  "Version": "2012-10-17",
  "Statement": [{
```

```

    "Effect": "Allow",
    "Action": [
        "autoscaling:CreateAutoScalingGroup",
        "autoscaling:UpdateAutoScalingGroup"
    ],
    "Resource": "*",
    "Condition": {
        "StringEquals": { "autoscaling:ResourceTag/environment": "development" },
        "NumericGreaterThanEqualsIfExists": { "autoscaling:MinSize": 1 },
        "NumericLessThanEqualsIfExists": { "autoscaling:MaxSize": 10 }
    }
}

```

In alternativa, se utilizzi tag per controllare l'accesso ai gruppi con scalabilità automatica, puoi utilizzare gli ARN per identificare i gruppi con dimensionamento automatico a cui si applica la policy IAM.

Un gruppo con dimensionamento automatico ha il seguente ARN.

```

"Resource": "arn:aws:autoscaling:region:account-id:autoScalingGroup:*:autoScalingGroupName/my-asg"

```

È inoltre possibile specificare più ARN racchiudendoli in un elenco. Per ulteriori informazioni su come specificare gli ARN delle risorse Dimensionamento automatico Amazon EC2 nell'elemento Resource, consulta [Risorse per le policy per Dimensionamento automatico Amazon EC2](#).

Controllo di quali tag chiavi e valori tag possono essere utilizzati

Puoi inoltre usare le condizioni nelle tue policy IAM per controllare le chiavi e i valori dei tag che possono essere applicati ai gruppi con dimensionamento automatico. Per concedere le autorizzazioni a creare o taggare un gruppo con dimensionamento automatico solo se il richiedente specifica determinati tag, utilizza la chiave di condizione `aws:RequestTag`. Per permettere solo chiavi di tag specifiche, utilizza la chiave di condizioni `aws:TagKeys` con il modificatore `ForAllValues`.

La policy seguente richiede al richiedente di specificare un tag con la chiave **environment** nella richiesta. Il valore `"?*"` impone che vi sia un valore per la chiave tag. Quando si include un carattere jolly, è necessario utilizzare l'operatore condizionale `StringLike`.

```

{

```



```

"Version": "2012-10-17",
"Statement": [{
  "Effect": "Allow",
  "Action": [
    "autoscaling:CreateAutoScalingGroup",
    "autoscaling:CreateOrUpdateTags"
  ],
  "Resource": "*",
  "Condition": {
    "StringLike": { "aws:RequestTag/environment": "?*" }
  }
}]
}

```

La policy seguente specifica che il richiedente può solo applicare tag ai gruppi con scalabilità automatica con i tag **purpose=webserver** e **cost-center=cc123**, e permette solo i tag **purpose** e **cost-center** (non è possibile specificare altri tag).

```

{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": [
      "autoscaling:CreateAutoScalingGroup",
      "autoscaling:CreateOrUpdateTags"
    ],
    "Resource": "*",
    "Condition": {
      "StringEquals": {
        "aws:RequestTag/purpose": "webserver",
        "aws:RequestTag/cost-center": "cc123"
      },
      "ForAllValues:StringEquals": { "aws:TagKeys": [purpose, cost-center] }
    }
  }]
}

```

La policy seguente richiede al richiedente di specificare almeno un tag nella richiesta e consente solo le chiavi **cost-center** e **owner**.

```

{
  "Version": "2012-10-17",

```

```

"Statement": [{
  "Effect": "Allow",
  "Action": [
    "autoscaling:CreateAutoScalingGroup",
    "autoscaling:CreateOrUpdateTags"
  ],
  "Resource": "*",
  "Condition": {
    "ForAnyValue:StringEquals": { "aws:TagKeys": ["cost-center", "owner] }
  }
}]
}

```

Note

Per le condizioni, la chiave di condizione non fa distinzione tra maiuscole e minuscole, mentre il valore della condizione la fa. Pertanto, per applicare la distinzione tra maiuscole e minuscole per una chiave di tag, utilizza la chiave di condizione `aws:TagKeys`, specificando la chiave di tag come valore nella condizione.

Verifica quale gruppo con dimensionamento automatico è possibile eliminare

La seguente policy consente l'eliminazione di un gruppo con dimensionamento automatico solo se il gruppo ha il tag **environment=development**.

```

{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": "autoscaling:DeleteAutoScalingGroup",
    "Resource": "*",
    "Condition": {
      "StringEquals": { "aws:ResourceTag/environment": "development" }
    }
  ]
}

```

In alternativa, se non si utilizzano chiavi di condizione per controllare l'accesso ai gruppi con dimensionamento automatico, è possibile specificare gli ARN delle risorse nell'elemento `Resource` per controllare l'accesso.

La seguente policy fornisce agli utenti le autorizzazioni per utilizzare l'azione `DeleteAutoScalingGroup` API, ma solo per i gruppi con dimensionamento automatico il cui nome inizia con **devteam-**.

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": "autoscaling:DeleteAutoScalingGroup",
    "Resource": "arn:aws:autoscaling:region:account-
id:autoScalingGroup:*:autoScalingGroupName/devteam-*"
  }]
}
```

È inoltre possibile specificare più ARN racchiudendoli in un elenco. L'inclusione dell'UUID garantisce l'accesso al gruppo con dimensionamento automatico specifico. L'UUID per un nuovo gruppo è diverso dall'UUID per un gruppo eliminato con lo stesso nome.

```
"Resource": [
  "arn:aws:autoscaling:region:account-
id:autoScalingGroup:uuid:autoScalingGroupName/devteam-1",
  "arn:aws:autoscaling:region:account-
id:autoScalingGroup:uuid:autoScalingGroupName/devteam-2",
  "arn:aws:autoscaling:region:account-
id:autoScalingGroup:uuid:autoScalingGroupName/devteam-3"
]
```

Verifica di quali policy di dimensionamento sia possibile eliminare

La policy seguente concede le autorizzazioni per utilizzare l'operazione `DeletePolicy` al fine di eliminare una policy di dimensionamento. Tuttavia, nega anche l'operazione, se il gruppo con dimensionamento automatico su cui stai agendo ha il tag **environment=production**. Quando possibile, è consigliabile utilizzare i tag per controllare l'accesso ai gruppi con dimensionamento automatico nel tuo account.

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": "autoscaling:DeletePolicy",
```

```

    "Resource": "*"
  },
  {
    "Effect": "Deny",
    "Action": "autoscaling:DeletePolicy",
    "Resource": "*",
    "Condition": {
      "StringEquals": { "autoscaling:ResourceTag/environment": "production" }
    }
  }
]
}

```

Controlla l'accesso alle azioni di aggiornamento delle istanze

La seguente policy concede l'autorizzazione ad avviare, ripristinare e annullare un aggiornamento dell'istanza solo se il gruppo con dimensionamento automatico su cui si sta agendo ha il tag **environment=testing**. Poiché le operazioni Describe non supportano le autorizzazioni a livello di risorsa, devi specificarle in un'istruzione separata senza condizioni.

```

{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": [
      "autoscaling:StartInstanceRefresh",
      "autoscaling:CancelInstanceRefresh",
      "autoscaling:RollbackInstanceRefresh"
    ],
    "Resource": "*",
    "Condition": {
      "StringEquals": { "autoscaling:ResourceTag/environment": "testing" }
    }
  },
  {
    "Effect": "Allow",
    "Action": "autoscaling:DescribeInstanceRefreshes",
    "Resource": "*"
  }
]
}

```

Per specificare la configurazione desiderata nella chiamata StartInstanceRefresh, gli utenti potrebbero richiedere alcune autorizzazioni correlate, come ad esempio:

- `ec2:RunInstances` — Per avviare istanze EC2 utilizzando un modello di avvio, l'utente deve disporre dell'`ec2:RunInstances` autorizzazione in una policy IAM. Per ulteriori informazioni, consulta [Supporto modello di avvio](#).
- `ec2:CreateTags` — Per avviare le istanze EC2 da un modello di avvio che aggiunge tag alle istanze e ai volumi al momento della creazione, l'utente deve disporre dell'autorizzazione in una policy IAM. `ec2:CreateTags` Per ulteriori informazioni, consulta [Autorizzazioni richieste per assegnare tag a istanze e volumi](#).
- `iam:PassRole` — Per avviare le istanze EC2 da un modello di avvio che contiene un profilo di istanza (un contenitore per un ruolo IAM), l'utente deve inoltre disporre dell'`iam:PassRole` autorizzazione in una policy IAM. Per ulteriori informazioni e per un esempio di policy IAM, consulta [Ruoli IAM per le applicazioni in esecuzione sulle istanze Amazon EC2](#).
- `ssm:GetParameters` — Per avviare istanze EC2 da un modello di avvio che utilizza un AWS Systems Manager parametro, l'utente deve inoltre disporre dell'`ssm:GetParameters` autorizzazione in una policy IAM. Per ulteriori informazioni, consulta [Usa AWS Systems Manager i parametri anziché gli ID AMI nei modelli di avvio](#).

Creazione di un ruolo collegato ai servizi

Amazon EC2 Auto Scaling richiede le autorizzazioni per creare un ruolo collegato al servizio la prima volta che un utente del tuo account richiama le azioni dell'API Account AWS Amazon EC2 Auto Scaling. Se il ruolo collegato al servizio non esiste già, Dimensionamento automatico Amazon EC2 lo crea nel tuo account. Il ruolo collegato al servizio fornisce le autorizzazioni ad Amazon EC2 Auto Scaling in modo che possa chiamare altri utenti per tuo conto. Servizi AWS

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 di autorizzazione che permette a un utente di creare un ruolo collegato al servizio Dimensionamento automatico Amazon EC2 per Dimensionamento automatico Amazon EC2.

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
```

```

    "Action": "iam:CreateServiceLinkedRole",
    "Resource": "arn:aws:iam::*:role/aws-service-role/
autoscaling.amazonaws.com/AWSServiceRoleForAutoScaling",
    "Condition": {
        "StringLike": { "iam:AWSServiceName": "autoscaling.amazonaws.com" }
    }
  ]
}

```

Controlla quale ruolo collegato al servizio può essere passato (utilizzando) PassRole

Gli utenti che creano o aggiornano i gruppi con scalabilità automatica e specificano un suffisso personalizzato nel ruolo collegato ai servizi richiedono l'autorizzazione `iam:PassRole`.

Puoi utilizzare l'`iam:PassRole` autorizzazione per proteggere la sicurezza delle chiavi gestite dai AWS KMS clienti se consenti a diversi ruoli collegati al servizio l'accesso a chiavi diverse. In base alle esigenze dell'organizzazione, puoi avere una chiave per il team di sviluppo, un'altra per il team di QA e un'altra per il team finanziario. Innanzitutto, crea un ruolo collegato al servizio che abbia accesso alla chiave richiesta, ad esempio un ruolo collegato al servizio denominato `AWSServiceRoleForAutoScaling_devteamkeyaccess`. Quindi, allega la policy a un'identità IAM, come un utente o ruolo.

La policy seguente concede le autorizzazioni necessarie per passare il ruolo

`AWSServiceRoleForAutoScaling_devteamkeyaccess` a qualsiasi gruppo con dimensionamento automatico il cui nome inizi con **`devteam-`**. Se l'identità IAM che crea il gruppo con dimensionamento automatico tenta di specificare un diverso ruolo collegato ai servizi, viene restituito un messaggio errore. Se scelgono di non specificare un ruolo collegato al servizio, viene invece utilizzato il ruolo predefinito `AWSServiceRoleForAutoScaling`

```

{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Allow",
    "Action": "iam:PassRole",
    "Resource": "arn:aws:iam::account-id:role/aws-service-role/
autoscaling.amazonaws.com/AWSServiceRoleForAutoScaling_devteamkeyaccess",
    "Condition": {
      "StringEquals": { "iam:PassedToService": [ "autoscaling.amazonaws.com" ] },
      "StringLike": { "iam:AssociatedResourceARN":
[ "arn:aws:autoscaling:region:account-
id:autoScalingGroup:*:autoScalingGroupName/devteam-" ] }
    }
  ]
}

```

```
    }  
  }]  
}
```

Per ulteriori informazioni sui ruoli dei servizi personalizzati, consulta [Ruoli collegati ai servizi per Dimensionamento automatico Amazon EC2](#).

Prevenzione del problema "confused deputy" tra servizi

Con "confused deputy" si intende un problema di sicurezza in cui un'entità che non dispone dell'autorizzazione per eseguire una certa operazione può costringere un'entità con più privilegi a eseguire tale operazione.

Nel AWS, l'impersonificazione tra servizi può causare il confuso problema del vice. La rappresentazione tra servizi può verificarsi quando un servizio (il servizio chiamante) effettua una chiamata a un altro servizio (il servizio chiamato). Il servizio chiamante può essere manipolato per utilizzare le proprie autorizzazioni e agire sulle risorse di un altro cliente, a cui normalmente non avrebbe accesso.

Per evitare che ciò accada, AWS mette a disposizione strumenti che consentono di proteggere i dati relativi a tutti i servizi con responsabili del servizio a cui è stato concesso l'accesso alle risorse del vostro account. Si consiglia di utilizzare il [aws:SourceArn](#) e le chiavi di contesto delle condizioni globali di [aws:SourceAccount](#) nelle policy di trust per i ruoli del servizio di Dimensionamento automatico Amazon EC2. Queste chiavi limitano le autorizzazioni che Dimensionamento automatico Amazon EC2 fornisce a un altro servizio alla risorsa.

I valori per i SourceAccount campi SourceArn e vengono impostati quando Amazon EC2 Auto Scaling AWS Security Token Service utilizza AWS STS() per assumere un ruolo per tuo conto.

Per utilizzare le chiavi di condizione globali `aws:SourceArn` o `aws:SourceAccount`, impostare il valore sul nome della risorsa Amazon (ARN) o sull'account della risorsa archiviata da Dimensionamento automatico Amazon EC2. Quando possibile, utilizzare `aws:SourceArn`, che è più specifico. Impostare il valore sull'ARN o su un modello dell'ARN con caratteri jolly (*) per le parti sconosciute dell'ARN. Se non si conosce l'ARN della risorsa, utilizzare `aws:SourceAccount`.

L'esempio seguente mostra il modo in cui puoi utilizzare le chiavi di contesto delle condizioni globali `aws:SourceArn` e `aws:SourceAccount` in Dimensionamento automatico Amazon EC2 per prevenire il problema "confused deputy".

Esempio: utilizzo delle chiavi di condizione `aws:SourceArn` e `aws:SourceAccount`

Un ruolo che un servizio assume per eseguire operazioni a tuo nome viene chiamato [ruolo del servizio](#). Nei casi in cui desideri creare lifecycle hook che inviino notifiche a un luogo diverso da Amazon EventBridge, devi creare un ruolo di servizio per consentire ad Amazon EC2 Auto Scaling di inviare notifiche a un argomento Amazon SNS o a una coda Amazon SQS per tuo conto. Se desideri consentire l'associazione di un solo gruppo con dimensionamento automatico all'accesso tra servizi, è possibile specificare la policy di trust del ruolo del servizio come segue.

Questo esempio di policy di trust utilizza istruzioni della condizione per limitare la capacità di AssumeRole sul ruolo di servizio di eseguire solo le operazioni che influiscono sul gruppo con dimensionamento automatico specificato nell'account specificato. Le condizioni `aws:SourceArn` e `aws:SourceAccount` sono valutate in modo indipendente. Qualsiasi richiesta di utilizzare il ruolo di servizio deve soddisfare entrambe le condizioni.

Prima di utilizzare questa policy, è necessario sostituire l'ID dell'account, la Regione, l'UUID e il nome del gruppo con valori validi riferiti al proprio account.

```
{
  "Version": "2012-10-17",
  "Statement": {
    "Sid": "ConfusedDeputyPreventionExamplePolicy",
    "Effect": "Allow",
    "Principal": {
      "Service": "autoscaling.amazonaws.com"
    },
    "Action": "sts:AssumeRole",
    "Condition": {
      "ArnLike": {
        "aws:SourceArn":
"arn:aws:autoscaling:region:account_id:autoScalingGroup:uuid:autoScalingGroupName/my-
asg"
      },
      "StringEquals": {
        "aws:SourceAccount": "account_id"
      }
    }
  }
}
```

Nell'esempio precedente:

- L'elemento `Principal` specifica il servizio principale del servizio (`autoscaling.amazonaws.com`).
- L'elemento `Action` specifica l'azione `sts:AssumeRole`.
- L'elemento `Condition` specifica le chiavi di condizione globali `aws:SourceArn` e `aws:SourceAccount`. L'ARN dell'origine include l'ID account, quindi non è necessario utilizzare `aws:SourceAccount` con `aws:SourceArn`.

Informazioni aggiuntive

Per ulteriori informazioni, consulta [Chiavi di contesto delle condizioni globali di AWS](#), [Problema "confused deputy"](#) e [Modifica di una policy di trust del ruolo \(console\)](#) nella Guida per l'utente di IAM.

Supporto modello di avvio

Dimensionamento automatico Amazon EC2 supporta l'utilizzo di modelli di avvio Amazon EC2 con i tuoi gruppi Auto Scaling. Sugeriamo di permettere agli utenti di creare gruppi Auto Scaling da modelli di avvio, in quanto ciò permette loro di utilizzare le funzionalità più recenti di Dimensionamento automatico Amazon EC2 e Amazon EC2. Ad esempio, per utilizzare una [policy di istanze miste](#) gli utenti devono specificare un modello di avvio.

Per permettere agli utenti l'accesso completo per lavorare con le risorse Dimensionamento automatico Amazon EC2, i modelli di avvio e altre risorse EC2 nel proprio account, puoi utilizzare la policy `AmazonEC2FullAccess`. In alternativa, puoi creare policy IAM personalizzate per concedere agli utenti le autorizzazioni a granulari affinché possano lavorare con modelli di avvio, come descritto in questo argomento.

Una policy di esempio che puoi personalizzare per le tue esigenze

Di seguito viene illustrato un esempio di policy di autorizzazione di base che è possibile personalizzare per uso personale. La policy concede le autorizzazioni per creare, modificare ed eliminare tutti i gruppi con dimensionamento automatico, ma solo se il gruppo utilizza il tag **`purpose=testing`**. Quindi concede l'autorizzazione per tutte le operazioni `Describe`. Poiché le operazioni `Describe` non supportano le autorizzazioni a livello di risorsa, devi specificarle in un'istruzione separata senza condizioni.

Le identità IAM (utenti o ruoli) che presentano questa policy, dispongono dell'autorizzazione per creare o aggiornare un gruppo con dimensionamento automatico utilizzando un modello di avvio, in quanto dispongono anche dell'autorizzazione per utilizzare l'operazione `ec2:RunInstances`.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "autoscaling:CreateAutoScalingGroup",
        "autoscaling:UpdateAutoScalingGroup",
        "autoscaling>DeleteAutoScalingGroup"
      ],
      "Resource": "*",
      "Condition": {
        "StringEquals": { "autoscaling:ResourceTag/purpose": "testing" }
      }
    },
    {
      "Effect": "Allow",
      "Action": [
        "autoscaling:Describe*",
        "ec2:RunInstances"
      ],
      "Resource": "*"
    }
  ]
}
```

Gli utenti che creano o aggiornano i gruppi con dimensionamento automatico potrebbero richiedere alcune autorizzazioni correlate, come ad esempio:

- `ec2: CreateTags` — Per aggiungere tag alle istanze e ai volumi al momento della creazione, l'utente deve disporre dell'autorizzazione in una policy IAM. `ec2:CreateTags` Per ulteriori informazioni, consulta [Autorizzazioni richieste per assegnare tag a istanze e volumi](#).
- `iam: PassRole` — Per avviare istanze EC2 da un modello di avvio che contiene un profilo di istanza (un contenitore per un ruolo IAM), l'utente deve inoltre disporre dell'`iam:PassRole` autorizzazione in una policy IAM. Per ulteriori informazioni e per un esempio di policy IAM, consulta [Ruoli IAM per le applicazioni in esecuzione sulle istanze Amazon EC2](#).
- `ssm: GetParameters` — Per avviare istanze EC2 da un modello di avvio che utilizza un AWS Systems Manager parametro, l'utente deve inoltre disporre dell'`ssm:GetParameters` autorizzazione in una policy IAM. Per ulteriori informazioni, consulta [Usa AWS Systems Manager i parametri anziché gli ID AMI nei modelli di avvio](#).

Queste autorizzazioni per le azioni da completare all'avvio delle istanze vengono verificate quando l'utente interagisce con un gruppo con dimensionamento automatico. Per ulteriori informazioni, consulta [Convalida delle autorizzazioni per ec2:RunInstances e iam:PassRole](#).

I seguenti esempi mostrano le istruzioni di policy che è possibile utilizzare per controllare gli accessi degli utenti IAM per i modelli di avvio.

Argomenti

- [Richiesta di modelli di avvio che dispongono di un tag specifico](#)
- [Richiesta di un modello di avvio e di un numero di versione](#)
- [Richiesta dell'utilizzo del servizio metadati dell'istanza versione 2 \(IMDSv2\)](#)
- [Come limitare l'accesso alle risorse Amazon EC2](#)
- [Autorizzazioni richieste per assegnare tag a istanze e volumi](#)
- [Autorizzazioni aggiuntive per i modelli di avvio](#)
- [Convalida delle autorizzazioni per ec2:RunInstances e iam:PassRole](#)
- [Risorse correlate](#)

Richiesta di modelli di avvio che dispongono di un tag specifico

Quando si concedono le autorizzazioni `ec2:RunInstances`, è possibile specificare che gli utenti possono utilizzare solo i modelli di avvio con tag o ID specifici per limitare le autorizzazioni all'avvio delle istanze con un modello di avvio. Inoltre, è possibile controllare l'AMI e altre risorse alle quali chiunque utilizzi i modelli di avvio può fare riferimento e che può utilizzare per avviare le istanze specificando le autorizzazioni aggiuntive a livello di risorsa per la chiamata `RunInstances`.

L'esempio seguente limita le autorizzazioni dell'operazione `ec2:RunInstances` con modelli di avvio che si trovano nella Regione specificata e che dispongono del tag **purpose=testing**. Consente inoltre agli utenti di accedere alle risorse specificate in un modello di avvio: AMI, tipi di istanze, volumi, coppie di chiavi, interfacce di rete e gruppi di sicurezza.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "ec2:RunInstances",
```

```

    "Resource": "arn:aws:ec2:region:account-id:launch-template/*",
    "Condition": {
      "StringEquals": { "aws:ResourceTag/purpose": "testing" }
    }
  },
  {
    "Effect": "Allow",
    "Action": "ec2:RunInstances",
    "Resource": [
      "arn:aws:ec2:region::image/ami-*",
      "arn:aws:ec2:region:account-id:instance/*",
      "arn:aws:ec2:region:account-id:subnet/*",
      "arn:aws:ec2:region:account-id:volume/*",
      "arn:aws:ec2:region:account-id:key-pair/*",
      "arn:aws:ec2:region:account-id:network-interface/*",
      "arn:aws:ec2:region:account-id:security-group*"
    ]
  }
]
}

```

Per ulteriori informazioni sull'utilizzo di policy basate su tag con i modelli di avvio, consulta [Controllare l'accesso ai modelli di avvio con autorizzazioni IAM](#) nella Guida per l'utente Amazon EC2 per istanze Linux.

Richiesta di un modello di avvio e di un numero di versione

È possibile anche utilizzare le autorizzazioni IAM per imporre che un modello di avvio e il numero di versione del modello di avvio debbano essere specificati durante la creazione o l'aggiornamento dei gruppi con dimensionamento automatico.

L'esempio seguente consente agli utenti di creare e aggiornare i gruppi con dimensionamento automatico solo se sono specificati un modello di avvio e il numero di versione del modello di avvio. Se gli utenti con questa policy omettono il numero di versione per specificare `$Latest` o la versione del modello di avvio `$Default`, o tentano di utilizzare una configurazione di avvio, l'operazione non riesce.

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",

```

```

    "Action": [
      "autoscaling:CreateAutoScalingGroup",
      "autoscaling:UpdateAutoScalingGroup"
    ],
    "Resource": "*",
    "Condition": {
      "Bool": { "autoscaling:LaunchTemplateVersionSpecified": "true" }
    }
  },
  {
    "Effect": "Deny",
    "Action": [
      "autoscaling:CreateAutoScalingGroup",
      "autoscaling:UpdateAutoScalingGroup"
    ],
    "Resource": "*",
    "Condition": {
      "Null": { "autoscaling:LaunchConfigurationName": "false" }
    }
  }
]
}

```

Richiesta dell'utilizzo del servizio metadati dell'istanza versione 2 (IMDSv2)

Per una maggiore sicurezza, hai l'opzione di impostare le autorizzazioni degli utenti per richiedere l'utilizzo di un modello di avvio che necessita di IMDSv2. Per ulteriori informazioni, consulta [Configurazione del servizio di metadati dell'istanza](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

La policy seguente specifica che gli utenti non possono chiamare l'operazione `ec2:RunInstances` a meno che anche l'istanza non sia impostata per richiedere l'uso di IMDSv2 (indicato da `"ec2:MetadataHttpTokens":"required"`).

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "RequireImdsV2",
      "Effect": "Deny",
      "Action": "ec2:RunInstances",
      "Resource": "arn:aws:ec2:*:*:instance/*",

```

```

        "Condition": {
            "StringNotEquals": { "ec2:MetadataHttpTokens": "required" }
        }
    ]
}

```

Tip

Per forzare l'avvio di istanze sostitutive di dimensionamento automatico che utilizzano un nuovo modello di avvio o una nuova versione di un modello di avvio con le opzioni di metadati dell'istanza configurate, è possibile avviare un aggiornamento di un'istanza. Per ulteriori informazioni, consulta [Aggiornamento di istanze di dimensionamento automatico](#).

Come limitare l'accesso alle risorse Amazon EC2

L'esempio seguente controlla la configurazione delle istanze che un utente può avviare limitando l'accesso alle risorse Amazon EC2. Per specificare le autorizzazioni a livello di risorsa per le risorse specificate in un modello di avvio, devi includere queste ultime nell'istruzione dell'operazione `RunInstances`.

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "ec2:RunInstances",
      "Resource": [
        "arn:aws:ec2:region:account-id:launch-template/*",
        "arn:aws:ec2:region::image/ami-04d5cc9b88example",
        "arn:aws:ec2:region:account-id:subnet/subnet-1a2b3c4d",
        "arn:aws:ec2:region:account-id:volume/*",
        "arn:aws:ec2:region:account-id:key-pair/*",
        "arn:aws:ec2:region:account-id:network-interface/*",
        "arn:aws:ec2:region:account-id:security-group/sg-903004f88example"
      ]
    },
    {
      "Effect": "Allow",
      "Action": "ec2:RunInstances",

```

```

    "Resource": "arn:aws:ec2:region:account-id:instance/*",
    "Condition": {
      "StringEquals": { "ec2:InstanceType": ["t2.micro", "t2.small"] }
    }
  ]
}

```

In questo esempio, ci sono quattro istruzioni:

- La prima istruzione richiede che gli utenti avviino istanze in una sottorete specifica (**subnet-1a2b3c4d**), utilizzando un gruppo di sicurezza specifico (**sg-903004f88example**) e un AMI specifico (**ami-04d5cc9b88example**). Fornisce agli utenti, inoltre, l'accesso alle risorse specificate in un modello di avvio: interfacce di rete, coppie di chiavi e volumi.
- La seconda istruzione consente agli utenti di avviare le istanze soltanto tramite i tipi di istanza **t2.micro** e **t2.small** per consentire di tenere sotto controllo i costi.

Tuttavia, tieni presente che attualmente non esiste un modo efficace per impedire completamente agli utenti autorizzati ad avviare istanze con un modello di avvio di avviare altri tipi di istanze. Questo perché un tipo di istanza specificato in un modello di avvio può essere sovrascritto per utilizzare tipi di istanza definiti utilizzando la selezione del tipo di istanza basata sugli attributi.

Per un elenco completo delle autorizzazioni a livello di risorse che è possibile utilizzare per controllare la configurazione delle istanze che un utente può avviare, consulta [Operazioni, risorse e chiavi di condizione per Amazon EC2](#) in Informazioni di riferimento sull'autorizzazione del servizio.

Autorizzazioni richieste per assegnare tag a istanze e volumi

L'esempio seguente permette agli utenti di assegnare tag a istanze e volumi alla creazione. Questa policy è necessaria se ci sono tag specificati nel modello di avvio. Per ulteriori informazioni, consulta [Concessione dell'autorizzazione all'assegnazione di tag per le risorse durante la creazione](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

```

{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "ec2:CreateTags",

```

```
    "Resource": "arn:aws:ec2:region:account-id:*/*",
    "Condition": {
      "StringEquals": { "ec2:CreateAction": "RunInstances" }
    }
  ]
}
```

Autorizzazioni aggiuntive per i modelli di avvio

Devi concedere agli utenti della console autorizzazioni per le operazioni `ec2:DescribeLaunchTemplates` e `ec2:DescribeLaunchTemplateVersions`. Senza queste autorizzazioni minime aggiuntive, i dati del modello di avvio e non possono essere caricati nella procedura guidata del gruppo con dimensionamento automatico e gli utenti non possono avviare le istanze utilizzando un modello di avvio tramite la procedura guidata. Puoi specificare le seguenti operazioni aggiuntive nell'elemento `Action` di un'istruzione della policy IAM.

Convalida delle autorizzazioni per `ec2:RunInstances` e `iam:PassRole`

Gli utenti possono specificare quale versione di un modello di avvio utilizza il gruppo con dimensionamento automatico. A seconda delle autorizzazioni, può trattarsi di una versione numerata specifica o versione `$Latest` o `$Default` del modello di avvio. Presta particolare attenzione a quest'ultimo caso. Ciò potrebbe avere la precedenza sulle autorizzazioni per `ec2:RunInstances` e `iam:PassRole` che intendevi limitare.

Questa sezione spiega lo scenario di utilizzo della versione più recente o predefinita del modello di avvio con un gruppo con dimensionamento automatico.

Quando un utente chiama le API `CreateAutoScalingGroup`, `UpdateAutoScalingGroup`, o `StartInstanceRefresh`, Dimensionamento automatico Amazon EC2 verifica le proprie autorizzazioni in base alla versione del modello di avvio che è la versione più recente o predefinita in quel momento prima di procedere con la richiesta. In questo modo vengono convalidate le autorizzazioni per le azioni da completare all'avvio delle istanze, come le azioni `ec2:RunInstances` e `iam:PassRole`. A tale scopo, emettiamo una chiamata Amazon [RunInstancesEC2](#) dry run per verificare se l'utente dispone delle autorizzazioni necessarie per l'azione, senza effettuare effettivamente la richiesta. Quando viene restituita una risposta, essa viene letta da dimensionamento automatico Amazon EC2. Se le autorizzazioni degli utenti non consentono una determinata operazione, il dimensionamento automatico Amazon EC2 invalida la richiesta e restituisce un errore all'utente contenente informazioni sull'autorizzazione mancante.

Una volta completate la verifica iniziale e la richiesta, all'avvio delle istanze, Dimensionamento automatico Amazon EC2 le avvia con la versione più recente o predefinita, anche se è stata modificata, utilizzando le autorizzazioni del suo [ruolo collegato al servizio](#). Ciò significa che un utente che sta utilizzando un modello di avvio può potenzialmente aggiornarlo per passare un ruolo IAM a un'istanza anche se non dispone dell'autorizzazione `iam:PassRole`.

Usa la chiave `autoscaling:LaunchTemplateVersionSpecified` condizionale se desideri limitare chi ha accesso ai gruppi di configurazione per l'utilizzo della versione `$Latest` o `$Default`. Ciò garantisce che il gruppo con dimensionamento automatico accetti solo una versione numerata specifica quando un utente chiama le API `CreateAutoScalingGroup` e `UpdateAutoScalingGroup`. Per un esempio che mostra come aggiungere questa chiave di condizione a una policy IAM, consulta [Richiesta di un modello di avvio e di un numero di versione](#).

Per i gruppi con dimensionamento automatico configurati per utilizzare `$Latest` o `$Default` avviare la versione del modello, è consigliabile limitare gli utenti che possono creare e gestire le versioni del modello di avvio, inclusa l'azione `ec2:ModifyLaunchTemplate` che consente a un utente di specificare la versione predefinita del modello di avvio. Per ulteriori informazioni, consulta [Controllo delle autorizzazioni di controllo delle versioni](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Risorse correlate

Per ulteriori informazioni sulle autorizzazioni per visualizzare, creare ed eliminare i modelli di avvio e le versioni dei modelli di avvio, consulta [Controllare l'accesso ai modelli di avvio con le autorizzazioni IAM](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Per ulteriori informazioni sulle autorizzazioni a livello di risorse che è possibile utilizzare per controllare l'accesso alla chiamata `RunInstances`, consulta [Operazioni, risorse e chiavi di condizione per Amazon EC2](#) nelle Informazioni di riferimento sull'autorizzazione del servizio.

Ruoli IAM per le applicazioni in esecuzione sulle istanze Amazon EC2

Le applicazioni eseguite su istanze Amazon EC2 necessitano di credenziali per accedere ad altri Servizi AWS. Per fornire tali credenziali in modo sicuro, utilizza un ruolo IAM. Il ruolo fornisce le autorizzazioni provvisorie che l'applicazione può utilizzare quando accede ad altre risorse AWS. Le autorizzazioni del ruolo determinano ciò che l'applicazione può fare.

Per le istanze in un gruppo con dimensionamento automatico, è necessario creare un modello o una configurazione di avvio e scegliere un profilo dell'istanza per l'associazione alle istanze. Un

profilo dell'istanza è un container per un ruolo IAM che permette ad Amazon EC2 di trasferire il ruolo IAM a un'istanza quando quest'ultima viene avviata. Innanzitutto, crea un ruolo IAM con tutte le autorizzazioni necessarie per accedere alle risorse. AWS Quindi, è necessario creare il profilo dell'istanza e assegnare il ruolo.

Note

Come best practice, ti consigliamo vivamente di creare il ruolo in modo che disponga delle autorizzazioni minime per gli altri Servizi AWS ruoli richieste dall'applicazione.

Indice

- [Prerequisiti](#)
- [Creazione di un modello di avvio](#)
- [Consulta anche](#)

Prerequisiti

Creazione del ruolo IAM che può essere assunto da un'applicazione in esecuzione su Amazon EC2. Scegli le autorizzazioni appropriate, in modo che l'applicazione a cui successivamente viene assegnato il ruolo possa effettuare le chiamate API necessarie.

Se utilizzi la console IAM anziché uno AWS CLI o uno degli AWS SDK, la console crea automaticamente un profilo di istanza e gli assegna lo stesso nome del ruolo a cui corrisponde.

Creazione di un ruolo IAM (console)

1. Aprire la console IAM all'indirizzo <https://console.aws.amazon.com/iam/>.
2. Nel riquadro di navigazione sulla sinistra, scegliere Roles (Ruoli).
3. Scegli Crea ruolo.
4. In Seleziona tipo di entità attendibile, scegli Servizio AWS .
5. Per il tuo caso d'uso, scegli EC2 quindi scegli Successivo.
6. Se possibile, selezionare la policy delle autorizzazioni da utilizzare o scegliere Crea policy per aprire una nuova scheda del browser e creare una nuova policy da zero. Per ulteriori informazioni, consulta [Creazione di policy IAM](#) nella Guida per l'utente di IAM. Una volta creata

- la policy, chiudere la scheda e tornare alla scheda originale. Selezionare la casella di controllo accanto alle policy di autorizzazione da assegnare al servizio.
- (Facoltativo) Impostare un limite delle autorizzazioni. Questa è una caratteristica avanzata disponibile per i ruoli di servizio. Per ulteriori informazioni, consulta [Limiti delle autorizzazioni per le entità IAM](#) nella Guida per l'utente di IAM.
 - Seleziona Successivo.
 - Nella pagina Name, review and create (Nomina, verifica e crea), in Nome ruolo inserisci un nome per facilitare l'identificazione dello scopo del ruolo. Il nome deve essere univoco all'interno dell'account Account AWS. Poiché altre AWS risorse potrebbero fare riferimento al ruolo, non puoi modificare il nome del ruolo dopo che è stato creato.
 - Verificare il ruolo e quindi scegliere Create role (Crea ruolo).

Autorizzazioni IAM

Utilizza una policy basata sull'identità IAM per controllare l'accesso al tuo nuovo ruolo IAM.

L'autorizzazione `iam:PassRole` è necessaria per l'identità IAM (utente o ruolo) che crea o aggiorna un gruppo con dimensionamento automatico utilizzando un modello di avvio con un profilo dell'istanza specifico.

La seguente policy di esempio concede le autorizzazioni per passare solo i ruoli IAM il cui nome inizia con **qateam-**.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "iam:PassRole",
      "Resource": "arn:aws:iam::account-id:role/qateam-*",
      "Condition": {
        "StringEquals": {
          "iam:PassedToService": [
            "ec2.amazonaws.com",
            "ec2.amazonaws.com.cn"
          ]
        }
      }
    }
  ]
}
```

}

⚠ Important

Per ulteriori informazioni su come Dimensionamento automatico Amazon EC2 convalida le autorizzazioni per l'operazione `iam:PassRole` per un gruppo con dimensionamento automatico che utilizza un modello di avvio, consulta [Convalida delle autorizzazioni per `ec2:RunInstances` e `iam:PassRole`](#).

Creazione di un modello di avvio

Quando crei il modello di lancio utilizzando AWS Management Console, nella sezione Dettagli avanzati, seleziona il ruolo dal profilo dell'istanza IAM. Per ulteriori informazioni, consulta [Crea un modello di avvio utilizzando le impostazioni avanzate](#).

Quando crei il modello di avvio utilizzando il [create-launch-template](#) comando di AWS CLI, specifica il nome del profilo di istanza del tuo ruolo IAM, come mostrato nell'esempio seguente.

```
aws ec2 create-launch-template --launch-template-name my-lt-with-instance-profile --  
version-description version1 \  
--launch-template-data  
'{"ImageId": "ami-04d5cc9b88example", "InstanceType": "t2.micro", "IamInstanceProfile":  
{"Name": "my-instance-profile"}}'
```

Consulta anche

Per ulteriori informazioni su come iniziare a conoscere e a utilizzare i ruoli IAM per Amazon EC2, vedi:

- [Ruoli IAM per Amazon EC2](#) nella Guida per l'utente Amazon EC2 per istanze Linux
- [Utilizzo dei profili delle istanze](#) e [Utilizzo di un ruolo IAM per concedere autorizzazioni ad applicazioni in esecuzione su istanze di Amazon EC2](#) nella Guida per l'utente di IAM

Convalida della conformità per Dimensionamento automatico Amazon EC2

Per sapere se il Servizio AWS è coperto da programmi di conformità specifici, consulta i [Servizi AWS coperti dal programma di conformità](#) e scegli il programma di conformità desiderato. Per informazioni generali, consulta [Programmi per la conformità di AWS](#).

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

La responsabilità di conformità durante l'utilizzo dei Servizi AWS è determinata dalla riservatezza dei dati, dagli obiettivi di conformità dell'azienda e dalle normative vigenti. Per semplificare il rispetto della conformità, AWS mette a disposizione le seguenti risorse:

- [Guide Quick Start per la sicurezza e conformità](#): queste guide all'implementazione illustrano considerazioni relative all'architettura e forniscono la procedura per l'implementazione di ambienti di base su AWS incentrati sulla sicurezza e sulla conformità.
- [Architetture per la sicurezza e la conformità HIPAA su Amazon Web Services](#): questo whitepaper descrive come le aziende possono utilizzare AWS per creare applicazioni conformi alla normativa HIPAA.

Note

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

- [Risorse per la conformità AWS](#): una raccolta di cartelle di lavoro e guide suddivise per settore e area geografica.
- [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 le regole](#) nella Guida per gli sviluppatori di AWS Config: il servizio AWS Config valuta il livello di conformità delle configurazioni delle risorse con pratiche interne, linee guida e regolamenti.

- [AWS Security Hub](#): questo Servizio AWS fornisce una visione completa dello stato di sicurezza all'interno di 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](#): questo Servizio AWS aiuta a verificare continuamente l'utilizzo di AWS per semplificare la gestione dei rischi e della conformità alle normative e agli standard di settore.

Conformità PCI DSS

Dimensionamento automatico Amazon EC2 supporta l'elaborazione, l'archiviazione e la trasmissione di dati di carte di credito da parte di un esercente o di un provider di servizi, oltre a essere conforme allo standard Payment Card Industry (PCI) Data Security Standard (DSS). Per ulteriori informazioni sullo standard PCI DSS, incluse le istruzioni su come richiedere una copia del Pacchetto conformità PCI di AWS, consulta [PCI DSS livello 1](#).

Per informazioni su come ottenere la conformità PCI DSS per i tuoi carichi di lavoro AWS, fai riferimento alla seguente guida alla conformità:

- [Payment Card Industry Data Security Standard \(PCI DSS\) 3.2.1 su AWS](#)

Dimensionamento automatico Amazon EC2 e interfaccia VPC endpoint

Puoi migliorare la posizione di sicurezza del VPC configurando Dimensionamento automatico Amazon EC2 in modo che utilizzi un endpoint VPC di interfaccia. Gli endpoint di interfaccia sono alimentati da AWS PrivateLink, una tecnologia che consente di accedere in modo privato alle API di Amazon EC2 Auto Scaling limitando tutto il traffico di rete tra il tuo VPC e Amazon EC2 Auto Scaling alla rete. 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, ma è consigliata. AWS PrivateLink [Per ulteriori informazioni sugli AWS PrivateLink endpoint VPC, consulta Cos'è? AWS PrivateLink](#) nella Guida AWS PrivateLink

Argomenti

- [Creazione di un endpoint VPC dell'interfaccia](#)

- [Creazione di una policy di endpoint VPC](#)

Creazione di un endpoint VPC dell'interfaccia

Crea un endpoint per Dimensionamento automatico Amazon EC2 utilizzando uno dei seguenti nomi di servizi:

```
com.amazonaws.region.autoscaling
```

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

Non è necessario modificare alcuna impostazione per Dimensionamento automatico Amazon EC2. Amazon EC2 Auto Scaling chiama AWS altri servizi utilizzando endpoint di servizio o endpoint VPC con interfaccia privata, a seconda di quale siano in uso.

Creazione di una policy di endpoint VPC

Puoi allegare una policy all'endpoint VPC per controllare l'accesso alle API Dimensionamento automatico Amazon EC2. 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 una policy 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:DeleteScalingPolicy",
        "Effect": "Deny",
        "Resource": "*",
        "Principal": "*"
    }
]
}
```

Per ulteriori informazioni, consulta [Controllare l'accesso agli endpoint VPC utilizzando le policy degli endpoint](#) nella Guida.AWS PrivateLink

Risoluzione dei problemi di Amazon EC2 Auto Scaling

Amazon EC2 Auto Scaling fornisce messaggi di errore specifici e descrittivi per aiutarti a risolvere i problemi. È possibile trovare i messaggi di errore nella descrizione delle attività di dimensionamento.

Argomenti

- [Recuperare un messaggio di errore dalle attività di dimensionamento](#)
- [Disattiva le attività di scalabilità](#)
- [Altre risorse per la risoluzione dei problemi](#)
- [Risoluzione dei problemi di Amazon EC2 Auto Scaling: errori di avvio dell'istanza EC2](#)
- [Risoluzione dei problemi Amazon EC2 Auto Scaling: problemi AMI](#)
- [Risoluzione dei problemi di Amazon EC2 Auto Scaling: problemi del load balancer](#)
- [Risoluzione dei problemi di Amazon EC2 Auto Scaling: modelli di avvio](#)
- [Risoluzione dei problemi relativi ai controlli dell'integrità in Amazon EC2 Auto Scaling.](#)

Recuperare un messaggio di errore dalle attività di dimensionamento

Per recuperare un messaggio di errore dalla descrizione delle attività di ridimensionamento, utilizzare il [describe-scaling-activities](#) comando. Hai un registro delle attività di dimensionamento che risale a 6 settimane fa. Le attività di dimensionamento vengono ordinate in base all'ora di inizio, con le più recenti elencate per prime.

Note

Le attività di dimensionamento vengono visualizzate anche nella cronologia delle attività nella console Amazon EC2 Auto Scaling sulla scheda Attività del gruppo Auto Scaling.

Per visualizzare le attività di dimensionamento per uno specifico gruppo Auto Scaling, utilizza il comando seguente.

```
aws autoscaling describe-scaling-activities --auto-scaling-group-name my-asg
```

Di seguito è riportato un esempio di risposta, dove `StatusCode` contiene lo stato corrente dell'attività e `StatusMessage` contiene il messaggio di errore.

```
{
  "Activities": [
    {
      "ActivityId": "3b05dbf6-037c-b92f-133f-38275269dc0f",
      "AutoScalingGroupName": "my-asg",
      "Description": "Launching a new EC2 instance: i-003a5b3ffe1e9358e. Status Reason: Instance failed to complete user's Lifecycle Action: Lifecycle Action with token e85eb647-4fe0-4909-b341-a6c42d8aba1f was abandoned: Lifecycle Action Completed with ABANDON Result",
      "Cause": "At 2021-01-11T00:35:52Z a user request created an AutoScalingGroup changing the desired capacity from 0 to 1. At 2021-01-11T00:35:53Z an instance was started in response to a difference between desired and actual capacity, increasing the capacity from 0 to 1.",
      "StartTime": "2021-01-11T00:35:55.542Z",
      "EndTime": "2021-01-11T01:06:31Z",
      "StatusCode": "Cancelled",
      "StatusMessage": "Instance failed to complete user's Lifecycle Action: Lifecycle Action with token e85eb647-4fe0-4909-b341-a6c42d8aba1f was abandoned: Lifecycle Action Completed with ABANDON Result",
      "Progress": 100,
      "Details": "{\"Subnet ID\":\"subnet-5ea0c127\",\"Availability Zone\":\"us-west-2b\"...}",
      "AutoScalingGroupARN": "arn:aws:autoscaling:us-west-2:123456789012:autoScalingGroup:283179a2-f3ce-423d-93f6-66bb518232f7:autoScalingGroupName/my-asg"
    },
    ...
  ]
}
```

Per una descrizione dei campi nell'output, consulta [Attività](#) nella Riferimenti generali alle API Amazon EC2 Auto Scaling.

Visualizzare le attività di dimensionamento per un gruppo eliminato

Per visualizzare le attività di ridimensionamento dopo l'eliminazione del gruppo Auto Scaling, aggiungete `--include-deleted-groups` l'opzione al comando come [describe-scaling-activities](#) segue.

```
aws autoscaling describe-scaling-activities --auto-scaling-group-name my-asg --include-deleted-groups
```

Di seguito è riportata una risposta di esempio, con un'attività di dimensionamento per un gruppo eliminato.

```
{
  "Activities": [
    {
      "ActivityId": "e1f5de0e-f93e-1417-34ac-092a76fba220",
      "AutoScalingGroupName": "my-asg",
      "Description": "Launching a new EC2 instance. Status Reason: Your Spot request price of 0.001 is lower than the minimum required Spot request fulfillment price of 0.0031. Launching EC2 instance failed.",
      "Cause": "At 2021-01-13T20:47:24Z a user request update of AutoScalingGroup constraints to min: 1, max: 5, desired: 3 changing the desired capacity from 0 to 3. At 2021-01-13T20:47:27Z an instance was started in response to a difference between desired and actual capacity, increasing the capacity from 0 to 3.",
      "StartTime": "2021-01-13T20:47:30.094Z",
      "EndTime": "2021-01-13T20:47:30Z",
      "StatusCode": "Failed",
      "StatusMessage": "Your Spot request price of 0.001 is lower than the minimum required Spot request fulfillment price of 0.0031. Launching EC2 instance failed.",
      "Progress": 100,
      "Details": "{\"Subnet ID\":\"subnet-5ea0c127\",\"Availability Zone\":\"us-west-2b\"...}",
      "AutoScalingGroupState": "Deleted",
      "AutoScalingGroupARN": "arn:aws:autoscaling:us-west-2:123456789012:autoScalingGroup:283179a2-f3ce-423d-93f6-66bb518232f7:autoScalingGroupName/my-asg"
    },
    ...
  ]
}
```

Disattiva le attività di scalabilità

Sono disponibili le seguenti opzioni se è necessario esaminare un problema senza interferenze dovute a politiche di ridimensionamento o azioni pianificate:

- Impedisce che tutte le politiche di scalabilità e le azioni pianificate apportino modifiche alla capacità desiderata del gruppo sospendendo i processi `AlarmNotification ScheduledActions`. Per ulteriori informazioni, consulta [Sospendi e riprendi i processi di Amazon EC2 Auto Scaling](#).
- Disabilita le politiche di scalabilità individuali in modo che non modifichino la capacità desiderata del gruppo in risposta alle variazioni di carico. Per ulteriori informazioni, consulta [Disabilitare una policy di dimensionamento per un gruppo con scalabilità automatica](#).
- Aggiorna le politiche di scalabilità di tracciamento dei singoli target in modo che si limitino alla scalabilità orizzontale (aggiunta di capacità) disabilitando la parte di scalabilità della policy. Questo metodo impedisce la riduzione della capacità desiderata del gruppo, ma ne consente l'aumento all'aumentare del carico. Per ulteriori informazioni, consulta [Policy di dimensionamento con monitoraggio degli obiettivi per Dimensionamento automatico Amazon EC2](#).

Altre risorse per la risoluzione dei problemi

Le pagine seguenti forniscono ulteriori informazioni per la risoluzione dei problemi relativi ad Amazon EC2 Auto Scaling.

- [Verificare un'attività di dimensionamento per un gruppo con scalabilità automatica](#)
- [Visualizzare i grafici di monitoraggio nella console Dimensionamento automatico Amazon EC2](#)
- [Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico](#)
- [Considerazioni e limitazioni per hook del ciclo di vita](#)
- [Completare un'operazione del ciclo di vita](#)
- [Fornisci connettività di rete per le istanze Auto Scaling utilizzando Amazon VPC](#)
- [Rimuovere temporaneamente le istanze dal gruppo con scalabilità automatica](#)
- [Disabilitare una policy di dimensionamento per un gruppo con scalabilità automatica](#)
- [Sospendi e riprendi i processi di Amazon EC2 Auto Scaling](#)
- [Verificare quali istanze Auto Scaling vengono terminate durante la riduzione orizzontale](#)
- [Eliminazione dell'infrastruttura Auto Scaling](#)
- [Quote per Amazon EC2 Auto Scaling](#)

Anche le seguenti AWS risorse possono essere utili:

- [Argomenti di Amazon EC2 Auto Scaling nel Knowledge Center AWS](#)

- [Domande su Amazon EC2 Auto Scaling su re:POST AWS](#)
- [Post di Amazon EC2 Auto Scaling nel blog di elaborazione AWS](#)
- [Risoluzione dei problemi CloudFormation nella guida per l'utente AWS CloudFormation](#)

La risoluzione dei problemi richiede spesso query iterative e scoperta da parte di un esperto o di una community di aiutanti. Se continui a riscontrare problemi dopo aver provato i suggerimenti in questa sezione, contatta AWS Support (nella sezione, fai clic su Support, Support Center) o fai una domanda su [AWS re:post](#) utilizzando il tag Amazon EC2 Auto Scaling. AWS Management Console

Risoluzione dei problemi di Amazon EC2 Auto Scaling: errori di avvio dell'istanza EC2

Questa pagina fornisce informazioni sulle istanze EC2 che non si avviano, le possibili cause e le procedure che è possibile eseguire per risolvere i problemi.

Per recuperare un messaggio di errore, consulta [Recuperare un messaggio di errore dalle attività di dimensionamento](#).

Quando le istanze EC2 non si avviano, potresti ricevere uno o più dei seguenti messaggi di errore:

Problemi di avvio

- [La configurazione richiesta non è attualmente supportata.](#)
- [Il gruppo di sicurezza <nome del gruppo di sicurezza> non esiste. Avvio di un'istanza EC2 non riuscito.](#)
- [La coppia di chiavi <coppia di chiavi associata all'istanza EC2> non esiste. Avvio di un'istanza EC2 non riuscito.](#)
- [Il tipo di istanza richiesto \(<tipo di istanza>\) non è supportato nella zona di disponibilità richiesta \(<zona di disponibilità dell'istanza>\)...](#)
- [Il prezzo della richiesta Spot pari a 0,015 è inferiore al prezzo minimo richiesto di gestione di una Richiesta Spot pari a 0,0735...](#)
- [Nome dispositivo <device name> non valido/Caricamento del nome del dispositivo non valido. Avvio di un'istanza EC2 non riuscito.](#)
- [Il valore \(<nome associato al dispositivo di archiviazione dell'istanza>\) per il parametro virtualName non è valido... Avvio di un'istanza EC2 non riuscito.](#)

- [Mappature dei dispositivi a blocchi EBS non supportate per le AMI di instance-store.](#)
- [I gruppi di collocamento non possono essere usati con le istanze di tipo “<instance type>”. Avvio di un'istanza EC2 non riuscito.](#)
- [Cliente. InternalError: errore del client all'avvio.](#)
- [Al momento non disponiamo di una sufficiente capacità di tipo di istanza nella zona di disponibilità richiesta. Avvio di un'istanza EC2 non riuscito.](#)
- [La prenotazione richiesta non ha una capacità compatibile e disponibile sufficiente per questa richiesta. Avvio di un'istanza EC2 non riuscito.](#)
- [La tua prenotazione Blocchi di capacità <reservation id> non è ancora attiva. Avvio di un'istanza EC2 non riuscito.](#)
- [Non è disponibile alcuna capacità Spot che corrisponda alla tua richiesta. Avvio di un'istanza EC2 non riuscito.](#)
- [<numero di istanze> istanza/e già in esecuzione. Avvio di un'istanza EC2 non riuscito.](#)

La configurazione richiesta non è attualmente supportata.

Causa: alcune opzioni nel modello di lancio o nella configurazione di avvio potrebbero non essere compatibili con il tipo di istanza oppure la configurazione dell'istanza potrebbe non essere supportata nella AWS regione o nelle zone di disponibilità richieste.

Soluzione: prova una configurazione di istanza diversa. Per cercare un tipo di istanza che soddisfi i requisiti, consulta [Individuazione di un tipo di istanza Amazon EC2](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Per ulteriori indicazioni su come risolvere il problema, verifica quanto segue:

- Assicurati di aver scelto un AMI supportato dal tipo di istanza. Ad esempio, se il tipo di istanza utilizza un processore AWS Graviton basato su ARM anziché un processore Intel Xeon, è necessaria un'AMI compatibile con ARM. Per ulteriori informazioni sulla scelta di un tipo di istanza compatibile, consulta [Compatibilità per cambiare il tipo di istanza](#) nella Guida per l'utente di Amazon EC2 per istanze Linux.
- Verifica che il tipo di istanza sia disponibile nella Regione e nelle zone di disponibilità richieste. In una determinata Regione o zona di disponibilità, i tipi di istanza di ultima generazione potrebbero non essere ancora disponibili. Nelle Regioni o zone di disponibilità più recenti, i tipi di istanza di ultima generazione potrebbero non essere ancora disponibili. Per cercare i tipi di istanze offerti

per località (regione o zona di disponibilità), usa il comando [describe-instance-type-offerings](#). Per informazioni sulle famiglie e i tipi di istanze Amazon EC2, consulta [Trova tipi di istanza Amazon EC2](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

- Se utilizzi Istanze dedicate o Host dedicati, assicurati di aver scelto un tipo di istanza supportato come Istanza dedicata o Host dedicato.

Il gruppo di sicurezza <nome del gruppo di sicurezza> non esiste. Avvio di un'istanza EC2 non riuscito.

Causa: il gruppo di sicurezza specificato nella configurazione o nel modello di avvio potrebbe essere stato eliminato.

Soluzione::

1. Usa il [describe-security-groups](#) comando per ottenere l'elenco dei gruppi di sicurezza associati al tuo account.
2. Seleziona dall'elenco i gruppi di sicurezza da utilizzare. Per creare invece un gruppo di sicurezza, usa il [create-security-group](#) comando.
3. Crea una nuova configurazione o modello di avvio.
4. Aggiorna il tuo gruppo Auto Scaling con il nuovo modello di avvio o avvia la configurazione utilizzando il [update-auto-scaling-group](#) comando.

La coppia di chiavi <coppia di chiavi associata all'istanza EC2> non esiste. Avvio di un'istanza EC2 non riuscito.

Causa: la coppia di chiavi che è stata usata all'avvio dell'istanza potrebbe essere stata eliminata.

Soluzione::

1. Usa il [describe-key-pairs](#) comando per ottenere l'elenco delle coppie di chiavi disponibili.
2. Dall'elenco, seleziona la coppia di chiavi da utilizzare. Per creare invece una key pair, usa il [create-key-pair](#) comando.
3. Crea una nuova configurazione o modello di avvio.
4. Aggiorna il tuo gruppo Auto Scaling con il nuovo modello di avvio o avvia la configurazione utilizzando il [update-auto-scaling-group](#) comando.

Il tipo di istanza richiesto (<tipo di istanza>) non è supportato nella zona di disponibilità richiesta (<zona di disponibilità dell'istanza>).

Messaggio di errore: il tipo di istanza richiesto (<tipo di istanza>) non è supportato nella zona di disponibilità richiesta (<zona di disponibilità dell'istanza>)...Avvio dell'istanza EC2 non riuscito.

Causa: le zone di disponibilità specificate nel gruppo con dimensionamento automatico non supportano il tipo di istanza scelto.

Soluzione::

1. Verifica quali zone di disponibilità supportano il tipo di istanza scelto utilizzando il [describe-instance-type-offerings](#) comando o dalla console Amazon EC2 controllando il valore delle zone di disponibilità nel riquadro di rete della pagina Tipi di istanza.
2. Aggiorna o rimuovi la sottorete per tutte le zone non supportate nelle impostazioni del gruppo Auto Scaling utilizzando il comando. [update-auto-scaling-group](#) Per ulteriori informazioni, consulta [Aggiungere o rimuovere zone di disponibilità](#).

Il prezzo della richiesta Spot pari a 0,015 è inferiore al prezzo minimo richiesto di gestione di una Richiesta Spot pari a 0,0735...

Causa: il prezzo massimo Spot nella richiesta è inferiore al prezzo istanza Spot per il tipo di istanza selezionato.

Soluzione: invia una nuova richiesta con un prezzo istanza Spot massimo (possibilmente il prezzo on demand). In precedenza, il prezzo istanza Spot da pagare si basava sull'offerta. Oggi, si paga il prezzo istanza Spot corrente. Se imposti il prezzo massimo più alto, questo offre al servizio Amazon EC2 Spot una migliore possibilità di avviare e mantenere la quantità di capacità richiesta.

Nome dispositivo <device name> non valido/Caricamento del nome del dispositivo non valido. Avvio di un'istanza EC2 non riuscito.

Causa 1: le mappature dei dispositivi a blocchi nella configurazione di avvio potrebbero contenere nomi di dispositivi a blocchi non disponibili o attualmente non supportati.

Soluzione::

1. verifica quali nomi di dispositivi sono disponibili per la configurazione specifica dell'istanza. Per ulteriori dettagli sulla denominazione dei dispositivi, consulta [Nomi dei dispositivi su istanze Linux](#) nella Guida per l'utente di Amazon EC2 per istanze Linux.
2. Crea manualmente un'istanza Amazon EC2 che non fa parte del gruppo Auto Scaling e analizza il problema. Se la configurazione dei nomi dei dispositivi a blocchi è in conflitto con i nomi dell'Amazon Machine Image (AMI), l'istanza non andrà a buon fine durante l'avvio. Per ulteriori informazioni, consulta [Mappature dei dispositivi a blocchi](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.
3. Dopo aver confermato che l'istanza sia stata avviata correttamente, utilizzare il comando [describe-volumes](#) per visualizzare il modo in cui i volumi sono esposti all'istanza.
4. Crea una nuova configurazione di avvio o modello di avvio utilizzando il nome del dispositivo elencato nella descrizione del volume.
5. Aggiorna il tuo gruppo Auto Scaling con il nuovo modello di avvio o avvia la configurazione utilizzando il [update-auto-scaling-group](#) comando.

Il valore (<nome associato al dispositivo di archiviazione dell'istanza>) per il parametro `virtualName` non è valido... Avvio di un'istanza EC2 non riuscito.

Causa: il formato specificato per il nome virtuale associato al dispositivo a blocchi non è corretto.

Soluzione::

1. Crea una nuova configurazione di avvio o modello di avvio specificando il nome del dispositivo nel parametro `virtualName`. Per informazioni sui nomi dei dispositivi in istanze Linux, consulta [Denominazione dei dispositivi in istanze Linux](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.
2. Aggiorna il tuo gruppo Auto Scaling con il nuovo modello di avvio o avvia la configurazione utilizzando il [update-auto-scaling-group](#) comando.

Mappature dei dispositivi a blocchi EBS non supportate per le AMI di `instance-store`.

Causa: le mappature dei dispositivi a blocchi specificate nella configurazione o nel modello di avvio non sono supportate sull'istanza.

Soluzione::

1. Crea una nuova configurazione o modello di avvio con mappature dei dispositivi a blocchi supportate dal tipo di istanza. Per ulteriori informazioni, consulta [Mappatura dei dispositivi a blocchi](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.
2. Aggiorna il tuo gruppo Auto Scaling con il nuovo modello di avvio o avvia la configurazione utilizzando il [update-auto-scaling-group](#) comando.

I gruppi di collocamento non possono essere usati con le istanze di tipo “<instance type>”. Avvio di un'istanza EC2 non riuscito.

Causa: il gruppo di collocazione cluster contiene un tipo di istanza non valido.

Soluzione::

1. Per ulteriori informazioni sui tipi di istanza validi supportati dai gruppi di collocazione, consulta [Gruppi di collocazione](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.
2. Per creare un nuovo gruppo di collocazione, Segui le istruzioni dettagliate in [Gruppi di collocazione](#).
3. In alternativa, crea una nuova configurazione o modello di avvio con il tipo di istanza supportato.
4. Aggiorna il tuo gruppo Auto Scaling con un nuovo gruppo di posizionamento, un modello di avvio o una configurazione di avvio utilizzando il [update-auto-scaling-group](#) comando.

Cliente. InternalError: errore del client all'avvio.

Problema: Amazon EC2 Auto Scaling tenta di avviare un'istanza con un volume EBS crittografato, ma il ruolo collegato al servizio non ha accesso alla chiave gestita dal cliente utilizzata per AWS KMS crittografarlo. Per ulteriori informazioni, consulta [Policy AWS KMS chiave richiesta per l'uso con volumi crittografati](#).

Causa 1: è necessario una policy chiave che autorizzi a utilizzare la chiave gestita dal cliente per il ruolo collegato al servizio appropriato.

Soluzione 1: permetti al ruolo collegato ai servizi di utilizzare la chiave gestita dal cliente come segue:

1. Determina quale ruolo collegato ai servizi utilizzare per questo gruppo Auto Scaling.
2. Aggiorna la policy della chiave sulla chiave gestita dal cliente e permetti al ruolo collegato ai servizi di utilizzare la chiave gestita dal cliente.
3. Aggiorna il gruppo Auto Scaling per utilizzare il ruolo collegato ai servizi.

Per un esempio di policy chiave che permetta al ruolo collegato al servizio di utilizzare la chiave gestita dal cliente, vedi [Esempio 1: sezioni delle policy delle chiavi che permettono l'accesso alla chiave gestita dal cliente](#).

Causa 2: se la chiave gestita dal cliente e il gruppo Auto Scaling si trovano in AWS account diversi, è necessario configurare l'accesso tra account alla chiave gestita dal cliente per autorizzare l'uso della chiave gestita dal cliente per il ruolo collegato al servizio appropriato.

Soluzione 2: permetti al ruolo collegato al servizio nell'account esterno di utilizzare la chiave gestita dal cliente nell'account locale come segue:

1. Aggiorna la policy chiave sulla chiave gestita dal cliente per permettere all'account del gruppo Auto Scaling di accedervi.
2. Definisci un utente o ruolo IAM nell'account del gruppo Auto Scaling che possa creare una concessione.
3. Determina quale ruolo collegato ai servizi utilizzare per questo gruppo Auto Scaling.
4. Crea una concessione alla chiave gestita dal cliente con il ruolo collegato ai servizi come assegnatario principale.
5. Aggiorna il gruppo Auto Scaling per utilizzare il ruolo collegato ai servizi.

Per ulteriori informazioni, consulta [Esempio 2: sezioni delle policy delle chiavi che permettono l'accesso multiaccount alla chiave gestita dal cliente](#).

Soluzione 3: utilizza una chiave gestita dal cliente nello stesso account AWS del gruppo Auto Scaling.

1. Copia e crittografa di nuovo lo snapshot con un'altra chiave gestita dal cliente appartenente allo stesso account del gruppo Auto Scaling.
2. Permetti al ruolo collegato ai servizi di utilizzare la chiave gestita dal cliente. Consulta i passaggi alla Soluzione 1.

Al momento non disponiamo di una sufficiente capacità di tipo di istanza nella zona di disponibilità richiesta. Avvio di un'istanza EC2 non riuscito.

Messaggio di errore: attualmente non abbiamo sufficiente capacità <tipo di istanza> nella zona di disponibilità richiesta (<zona di disponibilità richiesta>). Il nostro sistema si attiverà per fornire di capacità di provisioning aggiuntiva. Attualmente è possibile ottenere capacità <tipo di istanza> non

specificando una zona di disponibilità nella richiesta o scegliendo <elenco di zone di disponibilità che attualmente supportano il tipo di istanza>. Avvio di un'istanza EC2 non riuscito.

Causa: al momento, la combinazione di tipo di istanza richiesta e zona di disponibilità non è supportata.

Soluzione: per risolvere il problema, prova quanto segue:

- Attendi qualche minuto affinché Dimensionamento automatico Amazon EC2 trovi la capacità per questo tipo di istanza in altre zone di disponibilità abilitate.
- Distribuisci il gruppo con dimensionamento automatico a zone di disponibilità aggiuntive. Per ulteriori informazioni, consulta [Aggiungere o rimuovere zone di disponibilità](#).
- Segui la best practice per l'utilizzo di un insieme diverso di tipi di istanza in modo da non dipendere da un tipo di istanza specifico. Per ulteriori informazioni, consulta [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#).

La prenotazione richiesta non ha una capacità compatibile e disponibile sufficiente per questa richiesta. Avvio di un'istanza EC2 non riuscito.

Causa 1: hai raggiunto il numero limite di istanze che puoi avviare con una prenotazione della capacità on demand targeted.

Soluzione 1: aumenta il numero di istanze che puoi avviare con la prenotazione della capacità on demand targeted oppure utilizza un gruppo di prenotazioni di capacità in modo che tutto ciò che va oltre la capacità riservata venga avviato come normale capacità on demand. Per ulteriori informazioni, consulta [Utilizza le prenotazioni della capacità on demand per prenotare la capacità in zone di disponibilità specifiche](#).

Causa 2: hai raggiunto il numero limite di istanze che puoi avviare con un blocco di capacità.

Con Blocchi di capacità, sei vincolato dalla quantità di capacità originariamente acquistata. Se avviene un numero di avvii superiore al previsto e utilizzi tutta la capacità disponibile, i avvii non riusciranno. Le istanze terminanti vengono sottoposte a un lungo processo di pulizia prima di essere terminate completamente. Durante questo periodo, non possono essere riutilizzate, poiché, in questo modo, si verifica il fallimento dei lanci. Per ulteriori informazioni, consulta [Usa Capacity Blocks per carichi di lavoro di machine learning](#).

Soluzione 2: per risolvere il problema, provate quanto segue:

- Mantieni la richiesta così com'è. Se un'istanza Capacity Block viene interrotta, è necessario attendere alcuni minuti prima che l'istanza finisca la chiusura e che la capacità torni a essere disponibile. Dimensionamento automatico Amazon EC2 continua a effettuare automaticamente la richiesta di avvio fino a quando la capacità non diventa disponibile.
- Assicurati di acquistare una capacità sufficiente per soddisfare i picchi di carico di lavoro in modo da non riscontrare questo errore frequentemente.

La tua prenotazione Blocchi di capacità <reservation id> non è ancora attiva. Avvio di un'istanza EC2 non riuscito.

Causa: il Blocco di capacità specificato non è ancora attivo.

Soluzione: segui l'approccio consigliato per Blocco di capacità e usa il dimensionamento pianificato. In questo modo puoi assicurarti di aumentare la capacità desiderata del tuo gruppo con dimensionamento automatico solo quando la prenotazione è attiva e di diminuirla prima che la prenotazione sia terminata.

Non è disponibile alcuna capacità Spot che corrisponda alla tua richiesta. Avvio di un'istanza EC2 non riuscito.

Causa: al momento, non c'è abbastanza capacità di riserva per soddisfare la tua richiesta di Istanze Spot.

Soluzione: per risolvere il problema, prova quanto segue:

- Attendi alcuni minuti; la capacità può variare frequentemente. Dimensionamento automatico Amazon EC2 continua a effettuare automaticamente la richiesta di avvio fino a quando la capacità non diventa disponibile.
- Distribuisci il gruppo con dimensionamento automatico a zone di disponibilità aggiuntive. Per ulteriori informazioni, consulta [Aggiungere o rimuovere zone di disponibilità](#).
- Segui la best practice per l'utilizzo di un insieme diverso di tipi di istanza in modo da non dipendere da un tipo di istanza specifico. Per ulteriori informazioni, consulta [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#).

<numero di istanze> istanza/e già in esecuzione. Avvio di un'istanza EC2 non riuscito.

Causa: hai raggiunto il limite del numero di istanze Spot che puoi avviare. Quando crei il tuo AWS account, impostiamo limiti predefiniti al numero di istanze che puoi eseguire in base alla regione.

Soluzione: per risolvere il problema, prova quanto segue:

- Se i tuoi attuali limiti non sono adeguati alle tue esigenze, puoi richiedere un aumento della quota in base alla Regione. Per ulteriori informazioni, consultare [le Service Quotas del servizio Amazon EC2](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.
- Invia una nuova richiesta con un numero ridotto di istanze (che potrai aumentare in un secondo momento).

Risoluzione dei problemi Amazon EC2 Auto Scaling: problemi AMI

Questa pagina fornisce informazioni sui problemi relativi alle AMI, le possibili cause e le procedure che è possibile eseguire per risolvere i problemi.

Per recuperare un messaggio di errore, consulta [Recuperare un messaggio di errore dalle attività di dimensionamento](#).

Quando le istanze EC2 non si avviano a causa di problemi con l'AMI, è possibile ottenere uno o più dei seguenti messaggi di errore.

Problemi relativi ad AMI

- [L'ID AMI <ID dell'AMI> non esiste. Avvio di un'istanza EC2 non riuscito.](#)
- [L'AMI <ID AMI> è in sospeso e non può essere eseguita. Avvio di un'istanza EC2 non riuscito.](#)
- [<Nome dispositivo> Nome dispositivo non valido. Avvio di un'istanza EC2 non riuscito.](#)
- [L'architettura "arm64" del tipo di istanza specificato non corrisponde all'architettura "x86_64" dell'AMI specificata... Avvio dell'istanza EC2 non riuscito.](#)
- [L'AMI "<ID AMI>" è disabilitata e non può essere eseguita. Avvio di un'istanza EC2 non riuscito.](#)

⚠ Important

AWS supporta la condivisione di un'AMI in privato con un altro AWS account modificando le autorizzazioni AMI. Se un'AMI viene resa privata senza essere condivisa, ciò può causare un errore di autorizzazione all'avvio di nuove istanze. Per ulteriori informazioni sulla condivisione di AMI private, consulta [Condividere un'AMI con AWS account specifici](#) nella Amazon EC2 User Guide for Linux Instances.

L'ID AMI <ID dell'AMI> non esiste. Avvio di un'istanza EC2 non riuscito.

- Causa: l'AMI potrebbe essere stata eliminata dopo aver creato la configurazione o il modello di avvio.
- Soluzione::
 1. Crea una nuova configurazione di avvio o un modello di avvio utilizzando un'AMI valida.
 2. Aggiorna il tuo gruppo Auto Scaling con il nuovo modello di avvio o avvia la configurazione utilizzando il [update-auto-scaling-group](#) comando.

L'AMI <ID AMI> è in sospeso e non può essere eseguita. Avvio di un'istanza EC2 non riuscito.

Causa: potresti aver appena creato l'AMI (acquisendo uno snapshot di un'istanza in esecuzione o in qualsiasi altro modo) e potrebbe non essere ancora disponibile.

Soluzione: è necessario attendere che l'AMI diventi disponibile e quindi creare la configurazione o il modello di avvio.

<Nome dispositivo> Nome dispositivo non valido. Avvio di un'istanza EC2 non riuscito.

Causa: quando colleghi un volume EBS a un'istanza EC2, devi fornire un nome di dispositivo valido per il volume. L'AMI selezionata deve supportare questo nome di dispositivo.

Soluzione::

1. Crea una nuova configurazione di avvio o modello di avvio e specifica il nome del dispositivo corretto per l'AMI. La convenzione di denominazione consigliata varia in base al tipo di

virtualizzazione dell'AMI. Per ulteriori informazioni, consulta [Nomi dei dispositivi](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

2. Aggiorna il tuo gruppo Auto Scaling con il nuovo modello di avvio o avvia la configurazione utilizzando il [update-auto-scaling-group](#) comando.

L'architettura "arm64" del tipo di istanza specificato non corrisponde all'architettura "x86_64" dell'AMI specificata... Avvio dell'istanza EC2 non riuscito.

Causa 1: se l'architettura dell'AMI e il tipo di istanza utilizzato nel modello di lancio o nella configurazione di avvio non sono gli stessi, si verifica un errore quando Dimensionamento automatico Amazon EC2 tenta di avviare un'istanza utilizzando una configurazione di istanza incompatibile.

Soluzione 1:

1. Verifica l'architettura della tua AMI utilizzando il comando [describe-images](#) o dalla console Amazon EC2 controllando il valore Architettura nel riquadro dei dettagli della pagina Amazon Machine Images (AMI).
2. Trova un tipo di istanza con la stessa architettura dell'AMI utilizzando il [describe-instance-types](#) comando o dalla console Amazon EC2 controllando la colonna Architettura nella schermata Tipi di istanze. Per ulteriori informazioni sulla scelta di un tipo di istanza compatibile, consulta [Compatibilità per cambiare il tipo di istanza](#) nella Guida per l'utente di Amazon EC2 per istanze Linux.
3. Crea un nuovo modello di avvio o modello o configurazione di avvio utilizzando un tipo di istanza con la stessa architettura dell'AMI.
4. Aggiorna il tuo gruppo Auto Scaling con il nuovo modello di avvio o avvia la configurazione utilizzando il [update-auto-scaling-group](#) comando.

Causa 2: Dimensionamento automatico Amazon EC2 tenta di avviare un tipo di istanza specificato nella policy delle istanze miste per il tuo gruppo con dimensionamento automatico, ma il tipo di istanza non ha la stessa architettura dell'AMI specificata nel modello di lancio.

Soluzione 1: non includere tipi di istanze con architetture diverse nella policy relativa alle istanze miste.

1. Verifica l'architettura della tua AMI utilizzando il comando [describe-images](#) o dalla console Amazon EC2 controllando il valore Architettura nel riquadro dei dettagli della pagina Amazon Machine Images (AMI).
2. Verifica l'architettura di ogni tipo di istanza che intendi includere nella tua policy per le istanze miste utilizzando il [describe-instance-types](#) comando o dalla console Amazon EC2 controllando la colonna Architettura nella schermata Tipi di istanze. Per ulteriori informazioni sulla scelta di un tipo di istanza compatibile, consulta [Compatibilità per cambiare i tipi di istanza](#) nella Guida per l'utente di Amazon EC2 per istanze Linux.
3. Aggiorna o rimuovi i tipi di istanza incompatibili dal gruppo Auto Scaling utilizzando [update-auto-scaling-group](#) il comando.

Soluzione 2: per avviare entrambe le istanze Arm (Graviton2) e x86_64 (Intel) nello stesso gruppo con dimensionamento automatico, è necessario utilizzare i modelli di avvio supportati rispettivamente da un'AMI compatibile con ARM e da un'AMI compatibile con Intel x86, in modo che corrispondano ai tipi di istanze nella policy delle istanze miste.

1. Verifica l'architettura della tua AMI nel tuo modello di lancio esistente utilizzando il comando [describe-images](#) o dalla console Amazon EC2 controllando il valore Architettura nel riquadro dei dettagli della pagina Amazon Machine Images (AMI).
2. Crea un nuovo modello di lancio utilizzando un'AMI che corrisponda all'altra architettura che intendi utilizzare.
3. Aggiorna il tuo gruppo Auto Scaling per sovrascrivere il modello di avvio esistente e specifica il nuovo modello di avvio per ogni tipo di istanza compatibile utilizzando il comando. [update-auto-scaling-group](#) Per ulteriori informazioni, consulta [Utilizza un modello di avvio diverso per un tipo di istanza](#).

L'AMI “<ID AMI>” è disabilitata e non può essere eseguita. Avvio di un'istanza EC2 non riuscito.

Causa: stai tentando di avviare istanze da un'AMI che è stata disattivata. Per ulteriori informazioni, consulta [Disabilita un'AMI](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Soluzione::

1. Crea una nuova configurazione di avvio o un modello di avvio e specifica un'AMI non disattivata.

2. Aggiorna il tuo gruppo Auto Scaling con il nuovo modello di avvio o avvia la configurazione utilizzando il [update-auto-scaling-group](#) comando.

Risoluzione dei problemi di Amazon EC2 Auto Scaling: problemi del load balancer

Questa pagina fornisce informazioni sui problemi causati dal load balancer associato al gruppo Auto Scaling, le possibili cause e le procedure che è possibile eseguire per risolvere i problemi.

Per recuperare un messaggio di errore, consulta [Recuperare un messaggio di errore dalle attività di dimensionamento](#).

Quando le istanze EC2 non si avviano a causa di problemi relativi al load balancer associato al gruppo Auto Scaling, è possibile ottenere uno o più dei seguenti messaggi di errore.

Problemi del load balancer

- [Uno o più gruppi di destinazione non trovati. Convalida della configurazione del load balancer non riuscita.](#)
- [Impossibile trovare Load Balancer <il tuo load balancer>. Convalida della configurazione del load balancer non riuscita.](#)
- [Non vi è alcun load balancer ATTIVO denominato <nome del load balancer>. Aggiornamento della configurazione del sistema di load balancer non riuscito.](#)
- [L'istanza EC2 <ID istanza > non è in VPC. Aggiornamento della configurazione del sistema di load balancer non riuscito.](#)

Note

È possibile utilizzare Reachability Analyzer per risolvere i problemi di connettività verificando se le istanze del gruppo con scalabilità automatica sono raggiungibili tramite il sistema di bilanciamento del carico. Per ulteriori informazioni sui diversi problemi di configurazione errata della rete che vengono rilevati automaticamente da Reachability Analyzer, consulta i [codici di spiegazione di Reachability Analyzer](#) nella Guida per l'utente di Reachability Analyzer.

Uno o più gruppi di destinazione non trovati. Convalida della configurazione del load balancer non riuscita.

Problema: quando il gruppo con scalabilità automatica avvia le istanze, Dimensionamento automatico Amazon EC2 tenta di convalidare l'esistenza delle risorse Elastic Load Balancing associate al gruppo con scalabilità automatica. Quando non è possibile trovare un gruppo di destinazione, l'attività di dimensionamento non va a buon fine e si riceve l'errore `One or more target groups not found. Validating load balancer configuration failed..`

Causa 1: un gruppo di destinazione collegato al gruppo con scalabilità automatica è stato eliminato.

Soluzione 1: [puoi creare un nuovo gruppo Auto Scaling senza il gruppo di destinazione o rimuovere il gruppo target inutilizzato dal gruppo Auto Scaling utilizzando la console Amazon EC2 Auto Scaling o il comando `-groups.detach-load-balancer-target`](#)

Causa 2: il gruppo di destinazione esiste, ma si è verificato un problema durante il tentativo di specificare l'ARN del gruppo di destinazione al momento della creazione del gruppo con scalabilità automatica. Le risorse non vengono create nell'ordine corretto.

Soluzione 2: crea un nuovo gruppo con scalabilità automatica e specifica il nome del gruppo di destinazione alla fine.

Impossibile trovare Load Balancer <il tuo load balancer>. Convalida della configurazione del load balancer non riuscita.

Problema: quando il gruppo con scalabilità automatica avvia le istanze, Dimensionamento automatico Amazon EC2 tenta di convalidare l'esistenza delle risorse Elastic Load Balancing associate al gruppo con scalabilità automatica. Quando non è possibile trovare un Classic Load Balancer, l'attività di dimensionamento non va a buon fine e si riceve l'errore `Cannot find Load Balancer <your load balancer>. Validating load balancer configuration failed..`

Causa 1: il Classic Load Balancer è stato eliminato.

Soluzione 1: puoi creare un nuovo gruppo Auto Scaling senza il bilanciamento del carico o rimuovere il bilanciamento del carico inutilizzato dal gruppo Auto Scaling utilizzando la console Amazon EC2 Auto Scaling o il comando. [detach-load-balancers](#)

Causa 2: il Classic Load Balancer esiste, ma si è verificato un problema durante il tentativo di specificare il nome del load balancer al momento della creazione del gruppo con scalabilità automatica. Le risorse non vengono create nell'ordine corretto.

Soluzione 2: crea un nuovo gruppo Auto Scaling e specifica il nome del load balancer alla fine.

Non vi è alcun load balancer ATTIVO denominato <nome del load balancer>. Aggiornamento della configurazione del sistema di load balancer non riuscito.

Causa: il load balancer specificato potrebbe essere stato eliminato.

Soluzione: è possibile creare un nuovo load balancer e quindi creare un nuovo gruppo Auto Scaling oppure creare un nuovo gruppo Auto Scaling senza load balancer.

L'istanza EC2 <ID istanza > non è in VPC. Aggiornamento della configurazione del sistema di load balancer non riuscito.

Causa: l'istanza specificata non esiste nel VPC.

Solution: puoi eliminare il load balancer associato all'istanza o creare un nuovo gruppo Auto Scaling.

Risoluzione dei problemi di Amazon EC2 Auto Scaling: modelli di avvio

Utilizza le informazioni seguenti per diagnosticare e risolvere i problemi comuni che possono verificarsi durante l'utilizzo di un modello di avvio con un gruppo con scalabilità automatica.

Impossibile avviare le istanze

Se non riesci ad avviare istanze con un modello di avvio già specificato, consulta la sezione seguente per la risoluzione generale dei problemi: [Risoluzione dei problemi di Amazon EC2 Auto Scaling: errori di avvio dell'istanza EC2](#).

È necessario utilizzare un modello di avvio completo valido (valore non valido)

Problema: quando tenti di specificare un modello di avvio per un gruppo con scalabilità automatica, viene visualizzato l'errore You must use a valid fully-formed launch template. È possibile che si verifichi questo errore perché i valori nel modello di avvio vengono convalidati solo quando viene creato o aggiornato un gruppo con scalabilità automatica che utilizza il modello di avvio.

Causa 1: se viene visualizzato l'errore `You must use a valid fully-formed launch template`, sussistono problemi che fanno sì che Dimensionamento automatico Amazon EC2 consideri non valido qualche elemento del modello di avvio. Si tratta di un errore generico che può avere diverse cause.

Soluzione 1: per risolvere i problemi, prova a eseguire i seguenti passaggi:

1. Presta attenzione alla seconda parte del messaggio di errore per reperire ulteriori informazioni. Dopo l'errore `You must use a valid fully-formed launch template`, esamina il messaggio di errore più specifico che identifica il problema che dovrai risolvere.
2. Se non riesci a trovare la causa, testa il modello di avvio con il comando [run-instances](#). Specifica l'opzione `--dry-run` come mostrato nell'esempio seguente. In questo modo puoi riprodurre il problema e fornire informazioni dettagliate sulla sua causa.

```
aws ec2 run-instances --launch-template LaunchTemplateName=my-template,Version='1' --dry-run
```

3. Se un valore non è valido, verifica che la risorsa specificata esista e che sia corretta. Ad esempio, se specifichi una coppia di chiavi Amazon EC2, la risorsa deve esistere nel tuo account e nella regione in cui stai creando o aggiornando il gruppo con scalabilità automatica.
4. Se mancano le informazioni previste, verifica le impostazioni e modifica il modello di avvio secondo necessità.
5. Dopo aver apportato le modifiche, esegui nuovamente il comando [run-instances](#) con l'opzione `--dry-run` per verificare che il modello di avvio utilizzi valori validi.

Per ulteriori informazioni, consulta [Creare un modello di avvio per un gruppo con dimensionamento automatico](#).

Assenza dell'autorizzazione a utilizzare il modello di avvio (autorizzazioni insufficienti)

Problema: quando tenti di specificare un modello di avvio per un gruppo con scalabilità automatica, viene visualizzato l'errore `You are not authorized to use launch template`.

Causa 1: Se stai tentando di utilizzare un modello di avvio e le credenziali di IAM che stai usando non dispongono di autorizzazioni sufficienti, comparirà un errore che indica che non sei autorizzato a utilizzare il modello di avvio.

Soluzione 1: per risolvere il problema, prova quanto segue:

- Verifica che le credenziali IAM che stai utilizzando per effettuare la richiesta dispongano delle autorizzazioni per richiamare le operazioni API EC2 necessarie, inclusa l'operazione `ec2:RunInstances`. Se nel modello di avvio sono stati specificati dei tag, è necessario disporre anche dell'autorizzazione per utilizzare l'operazione `ec2:CreateTags`.
- In alternativa, verifica che le credenziali IAM che stai utilizzando per effettuare la richiesta dispongano della policy `AmazonEC2FullAccess`. Questa policy AWS gestita garantisce l'accesso completo a tutte le risorse di Amazon EC2 e ai servizi correlati, tra cui Amazon EC2 Auto Scaling, CloudWatch ed Elastic Load Balancing.

Per ulteriori informazioni sulle autorizzazioni necessarie per l'utilizzo dei modelli di avvio, comprese le policy IAM di esempio, consulta [Controllare l'accesso ai modelli di avvio con autorizzazioni IAM](#) nella Guida per l'utente Amazon EC2 per istanze Linux. Per altri esempi di policy IAM, consulta [Supporto modello di avvio](#).

Causa 2: se stai tentando di utilizzare un modello di avvio che specifica un profilo di istanza, è necessario che tu disponga dell'autorizzazione IAM per passare il ruolo IAM associato al profilo dell'istanza.

Soluzione 2: verifica che le credenziali IAM che stai utilizzando per effettuare la richiesta dispongano dell'autorizzazione `iam:PassRole` corretta per passare il ruolo specificato al servizio Dimensionamento automatico Amazon EC2. Per ulteriori informazioni e per un esempio di policy IAM, consulta [Ruoli IAM per le applicazioni in esecuzione sulle istanze Amazon EC2](#). Per ulteriori argomenti relativi alla risoluzione dei problemi relativi ai profili di istanza, vedi [Risoluzione dei problemi Amazon EC2 e IAM](#) nella Guida per l'utente IAM.

Causa 3: se stai tentando di utilizzare un modello di avvio che specifica un'AMI in un'altra Account AWS e l'AMI è privato e non condiviso con l'AMI che Account AWS stai utilizzando, ricevi un messaggio di errore che indica che non sei autorizzato a utilizzare il modello di avvio.

Soluzione 3: verifica che le autorizzazioni per l'AMI includano l'account che stai utilizzando. Per ulteriori informazioni, consulta [Condivisione di un'AMI con Account AWS specifici](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Risoluzione dei problemi relativi ai controlli dell'integrità in Amazon EC2 Auto Scaling.

Questa pagina fornisce informazioni sulle istanze EC2 che terminano a causa di un controllo dell'integrità. Descrive le potenziali cause e le operazioni che è possibile eseguire per risolvere i problemi.

Per recuperare un messaggio di errore, consulta [Recuperare un messaggio di errore dalle attività di dimensionamento](#).

Problemi relativi al controllo dell'integrità

- [Un'istanza è stata messa fuori servizio in risposta a un errore di controllo dell'integrità dell'istanza EC2](#)
- [Un'istanza è stata messa fuori servizio in risposta a un riavvio pianificato EC2](#)
- [Un'istanza è stata messa fuori servizio in risposta a un controllo dello stato EC2 che indicava che era stata terminata o interrotta](#)
- [Un'istanza è stata messa fuori servizio in risposta a un errore di controllo dell'integrità del sistema ELB](#)

Note

Puoi ricevere una notifica quando Amazon EC2 Auto Scaling termina le istanze nel tuo gruppo Auto Scaling, anche quando la causa della terminazione dell'istanza non è il risultato di un'attività di ridimensionamento. Per ulteriori informazioni, consulta [Opzioni di notifica Amazon SNS per Amazon EC2 Auto Scaling](#).

Le sezioni che seguono descrivono gli errori di controllo dell'integrità e delle loro cause più comuni. Se si è verificato un problema diverso, consulta quanto segue [Articoli del Portale del sapere AWS](#) per ulteriori informazioni sulla risoluzione dei problemi:

- [Perché Amazon EC2 Auto Scaling ha terminato un'istanza?](#)
- [Perché Amazon EC2 Auto Scaling non ha terminato un'istanza non integra?](#)

Un'istanza è stata messa fuori servizio in risposta a un errore di controllo dell'integrità dell'istanza EC2

Problema: le istanze Auto Scaling non superano i controlli dell'integrità di Amazon EC2.

Causa 1: se ci sono problemi che fanno sì che Amazon EC2 consideri compromesse le istanze del tuo gruppo Auto Scaling, Amazon EC2 Auto Scaling sostituisce automaticamente quelle non funzionanti nell'ambito del controllo dell'integrità. Le verifiche dello stato sono integrate in Amazon EC2 in modo tale da non poter essere disattivate o eliminate. Se l'esito della verifica dello stato di un'istanza è negativo, solitamente devi risolvere direttamente il problema modificandone la configurazione fino a quando l'applicazione non presenta più problemi.

Soluzione 1: per risolvere il problema, esegui questi passaggi:

1. Crea manualmente un'istanza Amazon EC2 che non fa parte del gruppo Auto Scaling e analizza il problema. Per informazioni generali sull'analisi delle istanze compromesse, vedi [Risoluzione dei problemi relativi alle istanze con esito negativo delle verifiche dello stato](#) nella Guida per l'utente di Amazon EC2 per le istanze Linux e [Risoluzione dei problemi delle istanze Windows](#) nella Guida per l'utente di Amazon EC2 per Windows.
2. Dopo aver confermato che l'istanza è stata avviata correttamente ed è corretta, implementa una nuova configurazione di istanza priva di errori nel gruppo Auto Scaling.
3. Elimina l'istanza creata per evitare ulteriori addebiti sull'account AWS .

Causa 2: c'è una mancata corrispondenza tra il periodo di prova del controllo dell'integrità e l'ora di avvio dell'istanza.

Soluzione 2: modifica il periodo di tolleranza del controllo dell'integrità per il gruppo Auto Scaling in un periodo di tempo appropriato per l'applicazione. Le istanze avviate in un gruppo Auto Scaling richiedono un tempo di riscaldamento sufficiente (periodo di grazia) per evitare la chiusura anticipata dovuta alla sostituzione di un test di integrità. Per ulteriori informazioni, consulta [Impostazione del periodo di tolleranza del controllo dell'integrità per un gruppo con scalabilità automatica](#).

Un'istanza è stata messa fuori servizio in risposta a un riavvio pianificato EC2

Problema: le istanze Auto Scaling vengono sostituite immediatamente, quando un evento pianificato indica un problema con l'istanza.

Causa: il Dimensionamento automatico Amazon EC2 sostituisce le istanze con un futuro evento di manutenzione programmata o di ritiro.

Soluzione: questi eventi non si verificano di frequente. In alternativa, se è necessario che qualcosa accada sull'istanza che sta terminando o sull'istanza che si sta avviando, puoi utilizzare gli hook del ciclo di vita. Questi hook ti consentono di eseguire un'operazione personalizzata quando Amazon EC2 Auto Scaling avvia o termina le istanze. Per ulteriori informazioni, consulta [Hook del ciclo di vita di Amazon EC2 Auto Scaling](#).

Se non vuoi sostituire le istanze a causa di un evento programmato, puoi sospendere il processo di controllo dell'integrità per un gruppo Auto Scaling. Per ulteriori informazioni, consulta [Sospendi e riprendi i processi di Amazon EC2 Auto Scaling](#).

Un'istanza è stata messa fuori servizio in risposta a un controllo dello stato EC2 che indicava che era stata terminata o interrotta

Problema: le istanze di Auto Scaling che sono state arrestate, riavviate o terminate vengono sostituite.

Causa 1: un utente ha arrestato, riavviato o terminato manualmente l'istanza.

Soluzione 1: se un controllo dell'integrità non riesce perché un utente ha arrestato, riavviato o terminato manualmente l'istanza, ciò è dovuto al funzionamento dei controlli dell'integrità di Amazon EC2 Auto Scaling. L'istanza deve essere integra e raggiungibile. Se è necessario riavviare le istanze nel gruppo Auto Scaling, suggeriamo di metterle prima in standby. Per ulteriori informazioni, consulta [Rimuovere temporaneamente le istanze dal gruppo con scalabilità automatica](#).

Nota che quando si terminano manualmente le istanze, è necessario completare gli hook del ciclo di vita di terminazione e la deregistrazione dall'Elastic Load Balancing (e lo svuotamento connessione), prima che l'istanza venga effettivamente terminata.

Causa 2: Amazon EC2 Auto Scaling tenta di sostituire le istanze Spot dopo che il servizio Amazon EC2 Spot interrompe le istanze, poiché il prezzo istanza Spot aumenta al di sopra del prezzo massimo o la capacità non è più disponibile.

Soluzione 2: non vi è alcuna garanzia che esista un'Istanza Spot per soddisfare la richiesta in un dato momento. Tuttavia, puoi provare a eseguire le operazioni indicate di seguito:

- Utilizza un prezzo istanza Spot massimo (possibilmente il prezzo on demand). Se imposti il prezzo massimo più alto, questo offre al servizio Amazon EC2 Spot una migliore possibilità di avviare e mantenere la quantità di capacità richiesta.
- Aumenta il numero di pool di capacità diversi da cui è possibile avviare istanze eseguendo più tipi di istanza in più zone di disponibilità. Per ulteriori informazioni, consulta [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#).
- Se utilizzi più tipi di istanza, è consigliabile attivare la caratteristica di ribilanciamento della capacità. Ciò è utile se desideri che il servizio Amazon EC2 Spot tenti di avviare una nuova Istanza Spot prima che venga terminata un'istanza in esecuzione. Per ulteriori informazioni, consulta [Utilizza il ribilanciamento della capacità per gestire le interruzioni spot di Amazon EC2](#).

Causa 3: con Capacity Blocks, Amazon EC2 termina tutte le istanze ancora in esecuzione 30 minuti prima della fine del Capacity Block. Questa interruzione improvvisa fa sì che il gruppo Auto Scaling tenti di avviare nuove istanze per mantenere la capacità desiderata, anche se il Blocco di capacità sta per scadere.

Soluzione 3: per risolvere il problema, provate quanto segue:

- Riduci la capacità desiderata del gruppo Auto Scaling per evitare che tenti di avviare nuove istanze. Per ulteriori informazioni, consulta [Dimensionamento manuale per Dimensionamento automatico Amazon EC2](#).
- Assicurati di eseguire la scalabilità nel gruppo Auto Scaling 30 minuti prima della fine del Blocco di capacità in modo da non riscontrare questo errore frequentemente. Assicurati che tutti gli hook del ciclo di vita siano stati completati 30 minuti prima della fine del Capacity Block. Per ulteriori informazioni, consulta [Usa Capacity Blocks per carichi di lavoro di machine learning](#).

Un'istanza è stata messa fuori servizio in risposta a un errore di controllo dell'integrità del sistema ELB

Problema: le istanze Auto Scaling potrebbero superare i controlli dell'integrità di Amazon EC2. Tuttavia, potrebbero non riuscire i controlli di integrità Elastic Load Balancing per i gruppi di destinazione o i Classic Load Balancers con cui è registrato il gruppo Auto Scaling.

Causa: se il tuo gruppo Auto Scaling si basa sui controlli di stato forniti da Elastic Load Balancing, Amazon EC2 Auto Scaling determina lo stato di integrità delle tue istanze verificando i risultati dei controlli dell'integrità EC2 e dell'Elastic Load Balancing. Il load balancer esegue controlli di

integrità inviando una richiesta a ciascuna istanza e attendendo la risposta corretta o stabilendo una connessione con l'istanza. Un'istanza potrebbe non superare il controllo dell'integrità dell'Elastic Load Balancing perché in un'applicazione in esecuzione sull'istanza si sono verificati problemi per cui il load balancer considera l'istanza fuori servizio. Per ulteriori informazioni, consulta [Controlli dell'integrità per le istanze in un gruppo con dimensionamento automatico](#).

Soluzione 1: come superare i controlli dell'integrità di Elastic Load Balancing:

- Verifica i codici di successo relativi al load balancer e accertati che l'applicazione sia configurata per restituire tali codici in caso di esito positivo.
- Verifica che i gruppi di sicurezza per il load balancer e il gruppo Auto Scaling siano configurati correttamente.
- Verifica che le impostazioni del controllo dell'integrità dei gruppi di destinazione siano configurate correttamente. È possibile definire le impostazioni di controllo dell'integrità per il load balancer per ciascun gruppo di destinazione.
- Prendi in considerazione l'aggiunta di un hook del ciclo di vita di avvio al gruppo Auto Scaling per assicurarti che le applicazioni nelle istanze siano pronte per accettare il traffico, prima che vengano registrate nel load balancer al termine dell'hook del ciclo di vita.
- Imposta il periodo di tolleranza del controllo di integrità per il gruppo Auto Scaling su un periodo di tempo abbastanza lungo per supportare il numero di controlli dell'integrità consecutivi necessari prima che il Elastic Load Balancing consideri integra un'istanza appena avviata.
- Verifica che il load balancer sia configurato nelle stesse zone di disponibilità del gruppo Auto Scaling.

Per ulteriori informazioni, consulta i seguenti argomenti:

- [Controlli dell'integrità per i gruppi di destinazione](#) nella Guida per l'utente di Application Load Balancer
- [Controlli dell'integrità per i gruppi di destinazione](#) nella Guida per l'utente di Network Load Balancer
- [Controlli dell'integrità per i gruppi di destinazione](#) nella Guida per l'utente di Gateway Load Balancer
- [Configura i controlli dell'integrità per il Classic Load Balancer](#) nella Guida per l'utente per Classic Load Balancer

Soluzione 2: aggiorna il gruppo Auto Scaling per disattivare i controlli di integrità Elastic Load Balancing.

Informazioni correlate

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

Risorsa	Descrizione
Documentazione di riferimento dell'API Dimensionamento automatico Amazon EC2	La documentazione per ogni operazione API mostra i parametri richiesti e la risposta XML e fornisce collegamenti ad argomenti di riferimento SDK specifici per lingua.
autoscaling in Riferimento ai comandi della AWS CLI	Descrizioni dei comandi della AWS CLI che possono essere utilizzati con i gruppi con dimensionamento automatico.
Riferimenti dei cmdlet di AWS Tools for PowerShell	Gli AWS strumenti PowerShell consentono di eseguire operazioni di script sulle AWS risorse dalla PowerShell riga di comando.
Creazione di gruppi Auto Scaling con AWS CloudFormation	La risorsa AWS::AutoScaling::AutoScalingGroup consente di creare, modellare e gestire i gruppi di Auto Scaling senza azioni manuali.
Endpoint e quote di Dimensionamento automatico Amazon EC2 in Riferimenti generali di AWS	Informazioni su regioni ed endpoint di Dimensionamento automatico Amazon EC2.
Pagina del prodotto	La pagina Web principale che include informazioni su Dimensionamento automatico Amazon EC2.
AWS re:Post	Servizio di domande e risposte gestito da AWS che offre risposte collettive e recensite da esperti alle tue domande tecniche.
Creare un'AMI nella Guida per l'utente di Amazon EC2 per le istanze Linux	Scopri come creare un'Amazon Machine Image (AMI) da un'istanza personalizzata.

Risorsa	Descrizione
Come connetterti all'istanza Linux nella Guida per l'utente di Amazon EC2 per le istanze Linux	Informazioni su come connettersi a un'istanza Linux che hai avviato.
Come connettersi alla tua istanza Windows nella Guida per l'utente di Amazon EC2 per le istanze Windows	Informazioni su come connettersi a un'istanza Windows che hai avviato.
Creazione di un allarme di fatturazione per monitorare gli AWS addebiti stimati nella Amazon CloudWatch User Guide	Scopri come monitorare i costi stimati utilizzando CloudWatch.
Application Auto Scaling User Guide	Scopri come configurare la scalabilità automatica per risorse scalabili per Amazon Web Services oltre Amazon EC2.

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 step-by-step i tutorial](#) 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 per informazioni su AWS Support one-on-one, un canale di supporto a risposta rapida per aiutarti 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

Nella seguente tabella sono descritte le aggiunte significative apportate alla documentazione di Dimensionamento automatico Amazon EC2 a partire da luglio 2018. Per ricevere notifiche sugli aggiornamenti della documentazione, puoi sottoscrivere il feed RSS.

Modifica	Descrizione	Data
Aggiornamento IAM di sicurezza	La policy AutoScalingServiceRolePolicy gestita ora concede autorizzazioni aggiuntive ad Amazon EC2 <code>ec2:GetSecurityGroupsForVpc</code> (e) <code>ec2:GetInstanceTypesFromInstanceRequirements</code>	29 febbraio 2024
L'ibernazione della piscina calda è supportata in aggiunta Regioni AWS	Ora puoi ibernare le istanze in un pool caldo in due regioni aggiuntive: AWS GovCloud (Stati Uniti orientali) e (Stati Uniti occidentali). AWS GovCloud Per ulteriori informazioni sulle warm pool, consulta Warm pools per il Dimensionamento automatico Amazon EC2 nella Guida per l'utente di Dimensionamento automatico Amazon EC2.	26 febbraio 2024
L'ibernazione in piscina calda è supportata in aggiunta Regioni AWS	Ora puoi ibernare le istanze in un pool caldo in due regioni aggiuntive: Europa (Zurigo) e Medio Oriente (Emirati Arabi Uniti). Per ulteriori informazioni sulle warm pool, consulta Warm pools per il Dimensionamento automatico Amazon EC2	21 febbraio 2024

[amento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

[Support per l'utilizzo di parametri tra account diversi](#)

Ora puoi usare un AWS Systems Manager parametro condiviso da un altro Account AWS con Amazon EC2 Auto Scaling. Per ulteriori informazioni, consulta [Utilizzare AWS Systems Manager i parametri anziché gli ID AMI nei modelli di avvio nella Guida per l'utente di Amazon EC2 Auto Scaling](#).

21 febbraio 2024

[Nuova opzione di protezione del prezzo Spot](#)

Ora puoi definire la soglia di protezione del prezzo per le istanze Spot come percentuale del prezzo on demand quando utilizzi la selezione del tipo di istanza basata sugli attributi . Per ulteriori informazioni, consulta la sezione [Protezione e dei prezzi](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

29 gennaio 2024

[Policy di manutenzione delle istanze](#)

Ora puoi utilizzare una policy di manutenzione delle istanze per definire se le istanze vengono avviate prima o dopo la chiusura delle istanze esistenti durante eventi che causano la sostituzione delle istanze, incluso un aggiornamento dell'istanza. Per ulteriori informazioni, consulta [policy di manutenzione delle istanze](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

15 novembre 2023

[Blocchi di capacità per ML](#)

Ora puoi avviare le istanze in un blocco di capacità specificando l'ID di prenotazione del blocco di capacità quando crei un modello di avvio. Con Blocchi di capacità puoi prenotare istanze GPU in date future per supportare i tuoi carichi di lavoro di machine learning (ML) di breve durata. Per ulteriori informazioni, consulta [Use Capacity Blocks per carichi di lavoro di machine learning](#) nella Guida per l'utente di Amazon EC2 Auto Scaling.

31 ottobre 2023

[Nuove funzionalità di aggiornamento dell'istanza](#)

Ora puoi configurare un aggiornamento dell'istanza per impostarne lo stato su fallito e, facoltativamente, ripristinarlo quando rileva che uno specifico CloudWatch allarme è entrato nello stato. ALARM Per ulteriori informazioni, consulta [Annullamento delle modifiche tramite un rollback](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

31 luglio 2023

[Modifiche alla Guida](#)

Alla guida è stato aggiunto un nuovo argomento sulle Istanze on demand nelle prenotazioni della capacità. Per ulteriori informazioni sulle prenotazioni di capacità, consulta [Utilizzo delle prenotazioni di capacità on demand per la prenotazioni della capacità in zone di disponibilità specifiche](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

28 luglio 2023

Modifiche alla Guida

Alla guida è stato aggiunto un nuovo argomento sulla migrazione AWS CloudFormation degli stack dalle configurazioni di avvio ai modelli di avvio. Per ulteriori informazioni, consulta [Esecuzione della migrazione degli stack AWS CloudFormation dalle configurazioni di avvio ai modelli di avvio](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

18 aprile 2023

[Supporto per nuove operazioni API](#)

Questa versione aggiunge tre nuove operazioni API `AttachTrafficSources` , `DetachTrafficSources` e `DescribeTrafficSources` . Inoltre, un nuovo campo, `TrafficSources` , è stato aggiunto ai risultati delle operazioni `DescribeAutoScalingGroups` . Inoltre, un nuovo stato di attività, `WaitingForConnectionDraining` , è stato aggiunto ai risultati delle operazioni `DescribeScalingActivities` . Il Dimensionamento automatico Amazon EC2 supporta anche un nuovo valore `VPC_LATTICE` per il campo `HealthCheckType` nelle operazioni `CreateAutoScalingGroup` , `UpdateAutoScalingGroup` e `DescribeAutoScalingGroups` . Per ulteriori informazioni, consulta [Riferimento delle API di Dimensionamento automatico Amazon EC2](#).

31 marzo 2023

[Supporto per Amazon VPC Lattice](#)

Questa è la versione di disponibilità generale di VPC Lattice per il Dimensionamento automatico Amazon EC2. Per maggiori informazioni, consulta [Indirizza il traffico verso il tuo gruppo con dimensionamento automatico o con un gruppo target VPC Lattice](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

31 marzo 2023

[Modifiche alla Guida](#)

La sezione con AWS CLI esempi per lavorare con Elastic Load Balancing ora include esempi nuovi e aggiornati. Per ulteriori informazioni, consulta [Esempi di utilizzo di Elastic Load Balancing with the AWS Command Line Interface \(AWS CLI\)](#) nella Amazon EC2 Auto Scaling User Guide.

31 marzo 2023

[Support per la scalabilità
predittiva in aggiunta Regioni
AWS](#)

Ora puoi creare politiche di scalabilità predittiva nelle regioni del Medio Oriente (Emirati Arabi Uniti) e AWS GovCloud (Stati Uniti orientali). Per ulteriori informazioni, consulta [Dimensionamento predittivo per Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

16 marzo 2023

[Nuove funzionalità di aggiornamento dell'istanza](#)

Ora puoi scegliere di terminare o ignorare le istanze in standby e sostituire o ignorare le istanze protette dalla riduzione orizzontale, invece di aspettare che diventino sostituibili. Puoi inoltre eseguire il rollback delle modifiche di un aggiornamento dell'istanza non riuscito. Nell'ambito di questo aggiornamento, la documentazione è stata ampliata per includere argomenti relativi al rollback dell'aggiornamento di un'istanza, all'annullamento di tale aggiornamento e alla comprensione dei valori predefiniti per i parametri configurabili di un aggiornamento dell'istanza. Per ulteriori informazioni, consulta [Sostituzione delle istanze di dimensionamento automatico in base a un aggiornamento di istanza](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

10 febbraio 2023

[Support per l'utilizzo di un AWS Systems Manager parametro per un ID AMI](#)

Ora puoi utilizzare un parametro di Systems Manager anziché un ID AMI nel modello di avvio. Per ulteriori informazioni, consulta [Utilizzo di parametri AWS Systems Manager al posto di ID AMI nei modelli di avvio](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

19 gennaio 2023

[Suggerimenti per il dimensionamento predittivo](#)

Ora puoi ottenere suggerimenti al fine di valutare e scegliere la policy di dimensionamento predittivo o adatta dalla console Dimensionamento automatico Amazon EC2. Per ulteriori informazioni, consulta [Valutazione delle policy di dimensionamento predittivo](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

18 gennaio 2023

[Previsioni di dimensionamento predittivo](#)

Le previsioni generate dal dimensionamento predittivo ora vengono aggiornate ogni sei ore anziché una volta al giorno. Per ulteriori informazioni, consulta [Dimensionamento predittivo per Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

6 gennaio 2023

[Support per la CloudWatch matematica metrica](#)

Ora puoi utilizzare la matematica dei parametri durante la creazione di policy di dimensionamento con monitoraggio degli obiettivi. Con la matematica a metrica, puoi interrogare più CloudWatch metriche e utilizzare espressioni matematiche per creare nuove serie temporali basate su queste metriche. Per ulteriori informazioni, consulta [Creazione di una policy di dimensionamento con monitoraggio degli obiettivi per Dimensionamento automatico Amazon EC2 tramite la matematica dei parametri](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

8 dicembre 2022

[Aggiornamento alle autorizzazioni del ruolo collegato ai servizi di IAM](#)

La policy `AutoScalingServiceRolePolicy` ora concede autorizzazioni aggiuntive a Dimensionamento automatico Amazon EC2. Per ulteriori informazioni, consulta [policy gestite da AWS per Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

6 dicembre 2022

[Nuova strategia di allocazione spot](#)

Ora è possibile utilizzare la strategia di allocazione ottimizzata per prezzo e capacità per richiedere istanze spot dai pool Spot con meno probabilità di interruzioni e il prezzo più basso possibile. Per ulteriori informazioni, consulta [Strategie di allocazione](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

10 novembre 2022

[Supporto per il dimensionamento predittivo nella regione Asia Pacifico \(Giacarta\)](#)

Ora puoi creare policy di dimensionamento predittivo nella regione Asia Pacifico (Giacarta). Per ulteriori informazioni, consulta [Dimensionamento predittivo per Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

13 ottobre 2022

[Supporto per i parametri personalizzati per il dimensionamento predittivo nella console](#)

Ora puoi utilizzare i parametri personalizzati per la creazione di policy di dimensionamento predittivo dalla console Dimensionamento automatico Amazon EC2. Per ulteriori informazioni, consulta [Dimensionamento predittivo per Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

13 ottobre 2022

[CloudWatch monitoraggio delle metriche di scalabilità predittiva](#)

Ora puoi accedere ai dati di monitoraggio per la scalabilità predittiva utilizzando CloudWatch. Ciò consente di utilizzare la matematica dei parametri per creare nuove serie temporali che visualizzano l'accuratezza dei dati di previsione. Per ulteriori informazioni, consulta [Monitora i parametri di scalabilità predittiva a CloudWatch nella Guida per l'utente di Amazon EC2 Auto Scaling](#).

7 luglio 2022

[Supporto per il dimensionamento predittivo nella regione Asia Pacifico \(Osaka-Locale\)](#)

Ora puoi creare policy di dimensionamento predittivo nella regione Asia Pacifico (Osaka-Locale). Per ulteriori informazioni, consulta [Dimensionamento predittivo per Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

6 luglio 2022

[Ibernazione warm pool supportata in altre regioni](#)

Ora è ora possibile ibernare le istanze in un pool di istanze in quattro regioni aggiuntive: Africa (Città del Capo), Asia Pacific (Giacarta), Asia Pacifico (Osaka-Locale) ed Europa (Milano). Per ulteriori informazioni sulle warm pool, consulta [Warm pools per il Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

5 luglio 2022

[Aggiornamento dei controlli dell'integrità](#)

Quando si eseguono controlli dell'integrità, il Dimensionamento automatico Amazon EC2 ora ti aiuta a ridurre al minimo i tempi di inattività che possono verificarsi a causa di problemi temporanei o di controlli dell'integrità configurati in modo errato. Per ulteriori informazioni, consulta [Come il Dimensionamento automatico Amazon EC2 riduce al minimo i tempi di inattività](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

21 maggio 2022

[Riscaldamento predefinito dell'istanza](#)

Ora puoi unificare tutte le impostazioni di riscaldamento e raffreddamento per un gruppo Auto Scaling e ottimizzare le prestazioni delle politiche di scalabilità che scalano continuamente abilitando il riscaldamento predefinito dell'istanza. Per ulteriori informazioni, consulta [Impostazione del riscaldamento predefinito dell'istanza per un gruppo con scalabilità automatica](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

19 aprile 2022

[Modifiche alla Guida](#)

Alla guida è stato aggiunto un nuovo capitolo sull'integrazione con altri AWS servizi. Per ulteriori informazioni, consulta [servizi AWS integrati con Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

29 marzo 2022

[Aggiornamento alle autorizzazioni del ruolo collegato ai servizi di IAM](#)

La policy `AutoScalingServiceRolePolicy` ora concede autorizzazioni di lettura aggiuntive a Dimensionamento automatico Amazon EC2. Per ulteriori informazioni, consulta [policy gestite da AWS per Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

28 marzo 2022

[I metadati dell'istanza forniscono lo stato del ciclo di vita di destinazione](#)

È possibile recuperare lo stato del ciclo di vita di un'istanza Auto Scaling dai metadati dell'istanza. Per maggiori informazioni, consulta la sezione [Recupero dello stato del ciclo di vita tramite i metadati dell'istanza](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

24 marzo 2022

[Supporto per la nuova funzionalità dei warm pool](#)

Ora è possibile ibernare le istanze in un warm pool per arrestare le istanze senza eliminare i contenuti della memoria (RAM). Inoltre è possibile restituire le istanze al warm pool durante la riduzione orizzontale, invece di terminare sempre la capacità dell'istanza di cui si ha bisogno in un secondo momento. Per ulteriori informazioni, consulta [Warm pools per Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

24 febbraio 2022

[Modifiche alla Guida](#)

La console Dimensionamento automatico Amazon EC2 è stata aggiornata con opzioni aggiuntive per avviare un aggiornamento delle istanze con l'opzione Salta corrispondenza abilitata e una configurazione desiderata specificata. Per ulteriori informazioni, consulta [Start or cancel an instance refresh \(console\)](#) (Avvio o annullamento di un aggiornamento delle istanze [console]) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

3 febbraio 2022

[Parametri personalizzati per policy di dimensionamento predittivo](#)

Ora puoi scegliere se utilizzare parametri personalizzati quando crei policy di dimensionamento predittivo. Puoi anche utilizzare la matematica dei parametri per personalizzare ulteriormente i parametri inclusi nella policy. Per ulteriori informazioni, consulta [Advanced predictive scaling policy configurations using custom metrics](#) (Configurazioni avanzate delle policy di dimensionamento predittivo utilizzando parametri personalizzati).

24 novembre 2021

[Nuova strategia di allocazione on demand](#)

Ora puoi scegliere se avviare istanze on demand in base al prezzo (prima i tipi di istanza con prezzi più bassi) quando si crea un gruppo con dimensionamento automatico che utilizza una policy di istanze miste. Per ulteriori informazioni, consulta [Strategie di allocazione](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

27 ottobre 2021

[Selezione del tipo di istanza basata su attributi](#)

Dimensionamento automatico Amazon EC2 aggiunge il supporto per la selezione del tipo di istanza basata su attributi. Invece di scegliere manualmente i tipi di istanza, è possibile esprimere i requisiti di istanza come un insieme di attributi, come vCPU, memoria e archiviazione. Per ulteriori informazioni, consulta [Creazione di un gruppo con dimensionamento automatico o utilizzando la selezione del tipo di istanza basata sugli attributi](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

27 ottobre 2021

[Supporto per il filtraggio dei gruppi per tag](#)

Ora è possibile filtrare i gruppi Auto Scaling utilizzando i filtri dei tag quando si recuperano informazioni sui gruppi Auto Scaling utilizzando il comando `describe-auto-scaling-groups`. Per ulteriori informazioni, consulta [Utilizzo dei tag per filtrare i gruppi Auto Scaling](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

14 ottobre 2021

Modifiche alla Guida

La console Amazon EC2 Auto Scaling è stata aggiornata per aiutarti a creare politiche di terminazione personalizzate con AWS Lambda. La documentazione della console è stata rivista di conseguenza. Per ulteriori informazioni, consulta [Utilizzo di diversi criteri di terminazione \(console\)](#). 14 ottobre 2021

Supporto per la copia delle configurazioni di avvio nei modelli di avvio

Ora puoi copiare tutte le configurazioni di avvio in una AWS regione in nuovi modelli di lancio dalla console Amazon EC2 Auto Scaling. Per ulteriori informazioni, consulta [Copia di una configurazione di avvio in un modello di avvio](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2. 9 agosto 2021

[Espande la funzionalità di aggiornamento dell'istanza](#)

Ora è possibile includere aggiornamenti, ad esempio una nuova versione di un modello di avvio, quando si sostituiscono le istanze aggiungendo la configurazione desiderata al comando `start-instance-refresh`. È inoltre possibile ignorare la sostituzione delle istanze che hanno già la configurazione desiderata abilitando l'opzione `skip-matching` (salta corrispondenza). Per ulteriori informazioni, consulta [Sostituzione delle istanze Auto Scaling in base a un aggiornamento di istanza](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

5 agosto 2021

[Supporto per policy di terminazione personalizzate](#)

Ora puoi creare politiche di terminazione personalizzate con AWS Lambda. Per ulteriori informazioni, consulta [Creazione di una policy di terminazione personalizzata con Lambda](#). La documentazione per specificare le policy di terminazione è stata aggiornata di conseguenza.

29 luglio 2021

[Modifiche alla Guida](#)

La console Dimensionamento automatico Amazon EC2 è stata aggiornata e migliorata con caratteristiche aggiuntive per semplificare le operazioni pianificate con un fuso orario specificato. La documentazione del [Dimensionamento programmato](#) è stata aggiornata di conseguenza.

3 giugno 2021

[Volumi gp3 nelle configurazioni di avvio](#)

È ora possibile specificare i volumi gp3 nelle mappature dei dispositivi a blocchi per le configurazioni di avvio.

2 giugno 2021

[Supporto per il dimensionamento predittivo](#)

Ora puoi utilizzare il dimensionamento predittivo per dimensionare in modo proattivo i gruppi Dimensionamento automatico Amazon EC2 con una policy di dimensionamento. Per ulteriori informazioni, consulta [Dimensionamento predittivo per Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2. Con questo aggiornamento, la policy [AutoScalingServiceRolePolicy](#) gestita ora include l'autorizzazione a richiamare l'azione dell'`cloudwatch:GetMetricData` API.

19 maggio 2021

[Modifiche alla Guida](#)

Ora puoi accedere a modelli di esempio per i lifecycle hook da GitHub. Per ulteriori informazioni, consulta [Hook del ciclo di vita di Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

9 aprile 2021

[Supporto per warm pool](#)

È ora possibile bilanciare le prestazioni (ridurre al minimo gli avvii a freddo) e i costi (interrompere l'over-provisioning della capacità di istanza) per le applicazioni con tempi di avvio lunghi aggiungendo warm pool ai gruppi Auto Scaling. Per ulteriori informazioni, consulta [Warm pools per Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

8 aprile 2021

[Supporto per i punti di controllo](#)

È ora possibile aggiungere e punti di controllo a un aggiornamento delle istanze per sostituirle per fasi ed eseguire verifiche sulle istanze in punti specifici. Per ulteriori informazioni, consulta [Aggiunta di punto di controllo a un aggiornamento di istanza](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

18 marzo 2021

[Modifiche alla Guida](#)

Documentazione migliorata per l'utilizzo EventBridge con gli eventi e gli hook del ciclo di vita di Amazon EC2 Auto Scaling. Per ulteriori informazioni, consulta [Using Amazon EC2 Auto Scaling EventBridge with and Tutorial: Configure a lifecycle hook che richiama una funzione Lambda nella Amazon EC2 Auto Scaling User Guide.](#)

18 marzo 2021

[Supporto per i fusi orari locali](#)

È ora possibile creare operazioni pianificate ricorrenti nel fuso orario locale aggiungendo l'opzione `--time-zone` al comando `put-scheduled-update-group-action`. Se il fuso orario osserva l'ora legale, si adegua automaticamente in funzione dell'ora legale. Per ulteriori informazioni, consulta [Dimensionamento programma](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

9 marzo 2021

[Espansione della funzionalità per le policy di istanze miste](#)

È ora possibile dare la priorità ai tipi di istanza per la capacità Spot quando si utilizza una policy di istanze miste. Dimensionamento automatico Amazon EC2 tenta di rispettare e le priorità del tipo di istanza sulla base del best effort, ma prima ottimizza la capacità. Per ulteriori informazioni, consultare la sezione relativa ai [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

8 marzo 2021

[Dimensionamento delle attività per i gruppi eliminati](#)

È ora possibile visualizzare le attività di dimensionamento per i gruppi Auto Scaling eliminati aggiungendo l'opzione `--include-deleted-groups` al comando `describe-scaling-activities`. Per ulteriori informazioni, consulta [Risoluzione dei problemi di Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

23 febbraio 2021

[Miglioramenti della console](#)

È ora possibile creare e allegare un Application Load Balancer o un Network Load Balancer dalla console Dimensionamento automatico Amazon EC2. Per ulteriori informazioni, consulta [Come creare e allegare un nuovo Application Load Balancer o un Network Load Balancer \(console\)](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

24 novembre 2020

[Interfacce di rete multiple](#)

È ora possibile configurare un modello di avvio per un gruppo con scalabilità automatica che specifica più interfacce di rete. Per ulteriori informazioni, consulta [Interfacce di rete in un VPC](#).

23 novembre 2020

[Modelli di avvio multipli](#)

È ora possibile utilizzare più modelli di avvio con i gruppi Auto Scaling. Per maggiori informazioni, consulta [Specifica di un modello di lancio diverso per un tipo di istanza](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

19 novembre 2020

[Bilanciatori del carico Gateway](#)

Guida aggiornata su come allegare un load balancer Gateway a un gruppo con scalabilità automatica per garantire che le istanze di appliance avviate dal Dimensionamento automatico Amazon EC2 vengano automaticamente registrate e deregistrate dal load balancer. Per ulteriori informazioni, consulta [Tipi di Elastic Load Balancing](#) e [Come allegare un load balancer al gruppo con scalabilità automatica](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

10 novembre 2020

[Durata massima dell'istanza](#)

È ora possibile ridurre la durata massima dell'istanza a un giorno (86.400 secondi). Per ulteriori informazioni, consulta la sezione [Sostituzione delle istanze di Auto Scaling in base alla durata massima dell'istanza](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

9 novembre 2020

[Ribilanciamento della capacità](#)

Puoi configurare il tuo gruppo con dimensionamento automatico affinché avvii un'istanza sostitutiva Spot quando Amazon EC2 emette un suggerimento di ribilanciamento. Per maggiori informazioni, consulta [Ribilanciamento della capacità di Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

4 novembre 2020

[Servizio di metadati dell'istanza versione 2](#)

Quando usi le configurazioni di avvio, puoi utilizzare Instance Metadata Service Version 2, che è un metodo orientato alla sessione per richiedere metadati dell'istanza. Per maggiori informazioni, consulta [Configurazione delle opzioni dei metadati dell'istanza](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

28 luglio 2020

[Modifiche alla Guida](#)

Vari miglioramenti e nuove procedure da console in [Verifica di quali istanze di Auto Scaling terminano durante la riduzione orizzontale](#), [Monitoraggio delle istanze e dei gruppi Auto Scaling](#) e le sezioni [Modelli di avvio](#) e [Configurazioni di avvio](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

28 luglio 2020

[Aggiornamento istanza](#)

Quando apporti una modifica alla configurazione, avvia l'aggiornamento di un'istanza per aggiornare tutte le istanze del gruppo con dimensionamento automatico. Per ulteriori informazioni, consulta [Sostituzione delle istanze Auto Scaling in base a un aggiornamento di istanza](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

16 giugno 2020

[Modifiche alla Guida](#)

Vari miglioramenti nelle sezioni relative alle sezioni [Sostituzione delle istanze in base alla durata massima dell'istanza](#), [Gruppi Auto Scaling con più tipi di istanza e opzioni di acquisto](#), [Dimensionamento basato su Amazon SQS](#) e [Assegnazione di tag a gruppi e istanze Auto Scaling](#) della Guida per l'utente di Dimensionamento automatico Amazon EC2.

6 maggio 2020

[Modifiche alla Guida](#)

Vari miglioramenti alla documentazione IAM. Per ulteriori informazioni, consulta [Supporto per il modello di avvio](#) ed [Esempi di policy basate su identità di Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

4 marzo 2020

[Disabilitazione delle policy di dimensionamento](#)

Ora puoi disabilitare e riabilitare le policy di dimensionamento. Questa caratteristica consente di disabilitare temporaneamente una policy di dimensionamento mantenendo i dettagli della configurazione in modo che sia possibile abilitare nuovamente la policy in un secondo momento. Per ulteriori informazioni, consulta [Disabilitazione di una policy per un gruppo con dimensionamento automatico](#) nella Guida dell'utente di Dimensionamento automatico Amazon EC2.

18 febbraio 2020

[Aggiunta della funzionalità di notifica](#)

Amazon EC2 Auto Scaling ora invia eventi quando i gruppi di Auto Scaling non riescono a AWS Health Dashboard scalare orizzontalmente a causa della mancanza di un gruppo di sicurezza o di un modello di lancio. Per ulteriori informazioni, consulta [Notifiche AWS Health Dashboard per Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

12 febbraio 2020

[Modifiche alla Guida](#)

Vari miglioramenti e correzioni nelle sezioni [Funzionamento di Dimensionamento automatico Amazon EC2 con IAM](#), [Esempi di policy basate su identità di Dimensionamento automatico Amazon EC2](#), [Policy chiave CMK necessarie per l'utilizzo con volumi crittografati e Monitoraggio delle istanze e dei gruppi Auto Scaling](#)) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

10 febbraio 2020

Modifiche alla Guida

Documentazione migliorata per i gruppi Auto Scaling che utilizzano la ponderazione delle istanze. Informazioni su come utilizzare le policy di dimensionamento quando calcoli la capacità desiderata con le "unità di capacità". Per ulteriori informazioni, consulta [Funzionamento delle policy di dimensionamento](#) e [Tipi di adeguamenti di dimensionamento](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

6 febbraio 2020

Nuovo capitolo sulla sicurezza

Un nuovo capitolo sulla [Sicurezza](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2 consente di capire come applicare il [modello di responsabilità condivisa](#) quando si utilizza Dimensionamento automatico Amazon EC2. Nell'ambito di questo aggiornamento, il capitolo della guida per l'utente "Controllo dell'accesso alle risorse Dimensionamento automatico Amazon EC2" è stato sostituito da una nuova sezione più utile, [Identity and Access Management per Dimensionamento automatico Amazon EC2](#).

4 febbraio 2020

[Suggerimenti per i tipi di istanza](#)

AWS Compute Optimizer fornisce consigli sulle istanze Amazon EC2 per aiutarti a migliorare le prestazioni, risparmiare denaro o entrambi. Per ulteriori informazioni, consulta [Suggerimenti sui tipi di istanza](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

3 dicembre 2019

[Host dedicati e gruppi di risorse host](#)

Guida aggiornata per mostrare come creare un modello di avvio che specifica un gruppo di risorse host. Ti permette di creare un gruppo con dimensionamento automatico o con un modello di avvio che specifica un'AMI BYOL da utilizzare su host dedicati. Per ulteriori informazioni, consulta [Creazione di un modello di avvio per un gruppo con dimensionamento automatico](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

3 dicembre 2019

[Supporto per gli endpoint Amazon VPC](#)

Ora puoi stabilire una connessione privata tra il VPC e Dimensionamento automatico Amazon EC2. Per ulteriori informazioni, consulta [Dimensionamento automatico Amazon EC2 ed endpoint VPC di interfaccia](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

22 novembre 2019

[Durata massima dell'istanza](#)

È ora possibile sostituire automaticamente le istanze specificando il periodo massimo di tempo in cui un'istanza può essere in servizio. Se qualche istanza si sta avvicinando a questo limite, Dimensionamento automatico Amazon EC2 gradualmente la sostituisce. Per ulteriori informazioni, consulta la sezione [Sostituzione delle istanze di Auto Scaling in base alla durata massima dell'istanza](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

19 novembre 2019

Ponderazione delle istanze

Per i gruppi con dimensionamento automatico con più tipi di istanza, è ora possibile specificare, facoltativamente, il numero di unità con cui ogni tipo di istanza contribuisce alla capacità del gruppo. Per ulteriori informazioni, consulta [Ponderazione istanza per Dimensionamento automatico Amazon EC2](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

19 novembre 2019

Numero minimo di tipi di istanza

Non è più necessario specificare altri tipi di istanza per gruppi di istanze Spot, on demand e riservate. Per tutti i gruppi con dimensionamento automatico, il minimo è ora un tipo di istanza. Per ulteriori informazioni, consultare la sezione relativa ai [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

16 settembre 2019

[Supporto per una nuova strategia di allocazione Spot](#)

Dimensionamento automatico Amazon EC2 ora supporta una nuova strategia di allocazione Spot "ottimizzata per la capacità" che soddisfa la richiesta utilizzando pool di istanze Spot scelti in modo ottimale in base alla capacità Spot disponibile. Per ulteriori informazioni, consultare la sezione relativa ai [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

12 agosto 2019

[Modifiche alla Guida](#)

Documentazione Dimensionamento automatico Amazon EC2 migliorata negli argomenti [Ruoli legati al servizio](#) e [Policy necessaria per le chiavi CMK per l'utilizzo con volumi crittografati](#).

1 agosto 2019

[Supporto per il miglioramento dell'assegnazione dei tag](#)

Dimensionamento automatico Amazon EC2 ora aggiunge i tag alle istanze Amazon EC2 nell'ambito della stessa chiamata API che avvia le istanze. Per ulteriori informazioni, consulta [Assegnazione di tag ai gruppi e alle istanze Auto Scaling](#).

26 luglio 2019

Modifiche alla Guida

Documentazione Dimensionamento automatico Amazon EC2 più esaustiva sull'argomento [Sospensione e ripresa dei processi di dimensionamento](#). Aggiornata la sezione [Esempi di policy gestite dal cliente](#) con l'inclusione di una policy di esempio che permette agli utenti di trasferire a Dimensionamento automatico Amazon EC2 solo ruoli personalizzati specifici collegati ai servizi.

13 giugno 2019

Supporto per le nuove funzionalità di Amazon EBS

Aggiunto il supporto per le nuove funzionalità di Amazon EBS nell'argomento sul modello di avvio. Cambia lo stato di crittografia di un volume EBS durante il ripristino da parte di uno snapshot. Per ulteriori informazioni, consulta [Creazione di un modello di avvio per un gruppo con dimensionamento automatico](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

13 maggio 2019

Modifiche alla Guida

Documentazione di Dimensionamento automatico Amazon EC2 migliorata nelle seguenti sezioni: [Verifica delle istanze Auto Scaling che vengono terminate durante la riduzione orizzontale](#) , [Gruppi Auto Scaling](#), [Gruppi Auto Scaling con più tipi di istanze e opzioni di acquisto](#), e [Dimensionamento dinamico per Dimensionamento automatico Amazon EC2](#).

12 marzo 2019

Supporto per combinare tipi di istanza e opzioni di acquisto

Come eseguire il provisioning e scalare automaticamente le istanze tra le opzioni d'acquisto (Spot, on demand e riservate) e i tipi di istanze all'interno di un singolo gruppo con dimensionamento automatico. Per ulteriori informazioni, consultare la sezione relativa ai [Gruppi con dimensionamento automatico con più tipi di istanze e opzioni di acquisto](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

13 novembre 2018

[Argomento aggiornato per il dimensionamento basato su Amazon SQS](#)

Guida aggiornata a illustrazione dell'utilizzo dei parametri personalizzati per dimensionare un gruppo con dimensionamento automatico in risposta al variare delle esigenze da una coda Amazon SQS. Per ulteriori informazioni, consulta [Dimensionamento basato su AMAZON SQS](#) nella Guida per l'utente di Dimensionamento automatico Amazon EC2.

26 luglio 2018

Nella tabella seguente sono descritte importanti modifiche apportate alla documentazione di Dimensionamento automatico Amazon EC2 prima di luglio 2018.

Funzionalità	Descrizione	Data di rilascio
Supporto per le policy di dimensionamento con monitoraggio degli obiettivi	Configurazione del dimensionamento dinamico per un'applicazione in poche fasi. Per ulteriori informazioni, consulta Policy di dimensionamento con monitoraggio degli obiettivi per Dimensionamento automatico Amazon EC2 .	12 luglio 2017
Supporto per le autorizzazioni a livello di risorsa	Creazione di policy IAM per controllare l'accesso a livello di risorsa. Per ulteriori informazioni, consulta Controllo dell'accesso alle risorse Dimensionamento automatico Amazon EC2 .	15 maggio 2017
Monitoraggio dei miglioramenti	I parametri del gruppo con dimensionamento automatico non richiedono più l'attivazione del monitoraggio dettagliato. È ora possibile abilitare la raccolta dei parametri di gruppo e visualizzare grafici dei parametri dalla scheda Monitoraggio nella console. Per ulteriori informazioni,	18 agosto 2016

Funzionalità	Descrizione	Data di rilascio
	<p>consulta Monitoraggio dei gruppi e delle istanze di Auto Scaling con Amazon. CloudWatch</p>	
<p>Supporto per Application Load Balancer</p>	<p>Come allegare uno o più gruppi di destinazione a un gruppo con dimensionamento automatico nuovo o esistente. Per ulteriori informazioni, consulta Come allegare un load balancer a un gruppo con dimensionamento automatico.</p>	<p>11 agosto 2016</p>
<p>Eventi per hook del ciclo di vita</p>	<p>Amazon EC2 Auto Scaling invia eventi quando chiama i EventBridge lifecycle hook. Per ulteriori informazioni, consulta Ottenere EventBridge la scalabilità del gruppo Auto Scaling.</p>	<p>24 febbraio 2016</p>
<p>Protezione delle istanze</p>	<p>Come impedire a Dimensionamento automatico Amazon EC2 di selezionare istanze specifiche per la terminazione durante il dimensionamento. Per ulteriori informazioni, consulta Protezione delle istanze.</p>	<p>07 dicembre 2015</p>
<p>Policy di dimensionamento per fasi</p>	<p>Crea una policy di dimensionamento che ti consente di dimensionare in base alle dimensioni dell'utilizzo fuori limite dell'allarme. Per ulteriori informazioni, consulta Tipi di policy di dimensionamento.</p>	<p>06 luglio 2015</p>
<p>Aggiornamento del load balancer</p>	<p>Come allegare o distaccare un load balancer da un gruppo con scalabilità automatica esistente. Per ulteriori informazioni, consulta Come allegare un load balancer a un gruppo con scalabilità automatica.</p>	<p>11 giugno 2015</p>
<p>Support per ClassicLink</p>	<p>Collega istanze EC2-Classical nel gruppo con dimensionamento automatico a un VPC, consentendo la comunicazione tra queste istanze collegate EC2-Classical e le istanze nel VPC utilizzando indirizzi IP privati. Per ulteriori informazioni, consulta Collegamento delle istanze EC2-Classical a un VPC.</p>	<p>19 gennaio 2015</p>

Funzionalità	Descrizione	Data di rilascio
Hook del ciclo di vita	Mantieni le istanze appena avviate o in fase di terminazione in stato di attesa mentre esegui operazioni su di esse. Per ulteriori informazioni, consulta Hook del ciclo di vita di Dimensionamento automatico Amazon EC2 .	30 luglio 2014
Distacco delle istanze	Distacco delle istanze da un gruppo con dimensionamento automatico. Per ulteriori informazioni, consulta Distacco delle istanze EC2 dal gruppo con dimensionamento automatico .	30 luglio 2014
Come porre le istanze in stato di standby	Come porre le istanze che sono in stato InService in stato Standby. Per ulteriori informazioni, consulta Rimozione temporanea delle istanze dal gruppo con dimensionamento automatico .	30 luglio 2014
Gestione dei tag	Gestisci i gruppi Auto Scaling utilizzando la AWS Management Console. Per ulteriori informazioni, consulta Assegnazione di tag ai gruppi e alle istanze Auto Scaling .	01 maggio 2014
Supporto per istanze dedicate	Avvia istanze dedicate specificando un attributo di tenancy di posizionamento al momento della creazione di una configurazione di avvio. Per ulteriori informazioni, consulta Tenancy di posizionamento delle istanze .	23 aprile 2014
Crea un gruppo o una configurazione di avvio da un'istanza EC2	Crea un gruppo con dimensionamento automatico o una configurazione di avvio utilizzando un'istanza EC2. Per ulteriori informazioni sulla creazione di una configurazione di avvio utilizzando un'istanza EC2, consulta Creazione di una configurazione di avvio utilizzando un'istanza EC2 . Per ulteriori informazioni sulla creazione di un gruppo con dimensionamento automatico utilizzando un'istanza EC2, consulta Creazione di un gruppo con dimensionamento automatico utilizzando un'istanza EC2 .	02 gennaio 2014

Funzionalità	Descrizione	Data di rilascio
Come allegare le istanze	Abilita la scalabilità automatica per un'istanza EC2 collegando l'istanza a un gruppo Auto Scaling esistente . Per ulteriori informazioni, consulta Come allegare le istanze EC2 al gruppo con dimensionamento automatico.	02 gennaio 2014
Visualizzazione delle limitazioni dell'account	Visualizza le limitazioni sulle risorse di Auto Scaling per il tuo account. Per ulteriori informazioni, consulta Limitazioni di Auto Scaling.	02 gennaio 2014
Supporto per la console per Dimensionamento automatico Amazon EC2	Accedi ad Amazon EC2 Auto Scaling utilizzando. AWS Management Console Per ulteriori informazioni, consulta Nozioni di base su Dimensionamento automatico Amazon EC2.	10 dicembre 2013
Assegnazione di un indirizzo IP pubblico	Assegna un indirizzo IP pubblico a un'istanza avviata in un VPC. Per ulteriori informazioni, consulta Avvio di istanze di Auto Scaling in un VPC.	19 settembre 2013
Policy di terminazione istanza	Specifica una policy di terminazione dell'istanza per Dimensionamento automatico Amazon EC2 da utilizzare e per terminare le istanze EC2. Per ulteriori informazioni, consulta Controllo di quali istanze Auto Scaling terminare durante la riduzione orizzontale.	17 settembre 2012
Supporto per ruoli IAM;	Avvia istanze EC2 senza un profilo dell'istanza IAM. Puoi utilizzare questa funzione per assegnare ruoli IAM alle istanze, permettendo alle tue applicazioni di accedere ad altri servizi Amazon Web Services in modo sicuro. Per ulteriori informazioni, consulta Avvio di istanze di Auto Scaling con un ruolo IAM.	11 giugno 2012
Supporto per istanze Spot	Avvio di istanze Spot con una configurazione di avvio. Per maggiori informazioni, consulta la sezione Richiedere istanze spot per applicazioni flessibili e tolleranti ai guasti.	7 giugno 2012

Funzionalità	Descrizione	Data di rilascio
Assegnazione di tag a gruppi e istanze	<p>Assegna tag a gruppi Auto Scaling e specifica che il tag si applica anche a istanze EC2 avviate dopo la sua creazione. Per ulteriori informazioni, consulta Assegnazione di tag ai gruppi e alle istanze Auto Scaling.</p>	26 gennaio 2012
Supporto per Amazon SNS	<p>Utilizza Amazon SNS; per ricevere notifiche ogni volta che Dimensionamento automatico Amazon EC2 avvia o termina istanze EC2. Per ulteriori informazioni, consulta Ricezione di notifiche SNS quando il gruppo con dimensionamento automatico viene dimensionato.</p> <p>Inoltre, Dimensionamento automatico Amazon EC2 ha aggiunto le seguenti nuove caratteristiche:</p> <ul style="list-style-type: none"> • La possibilità di configurare attività di dimensionamento ricorrenti utilizzando una sintassi cron. Per ulteriori informazioni, consulta l'operazione API PutScheduledUpdateGroupAction. • Una nuova impostazione di configurazione che consente la scalabilità orizzontale senza aggiungere e l'istanza lanciata al load balancer (). LoadBalancer Per ulteriori informazioni, consulta il tipo di dati API ProcessType. • Il flag ForceDelete nell'operazione DeleteAutoScalingGroup che indica a Dimensionamento automatico Amazon EC2 di eliminare il gruppo con dimensionamento automatico con le istanze associate senza attendere queste siano terminate. Per ulteriori informazioni, consulta l'operazione API DeleteAutoScalingGroup. 	20 luglio 2011
Operazioni di dimensionamento pianificate	<p>Aggiunto il supporto per operazioni di dimensionamento pianificate. Per ulteriori informazioni, consulta Dimensionamento programmato per Dimensionamento automatico Amazon EC2.</p>	2 dicembre 2010

Funzionalità	Descrizione	Data di rilascio
Supporto per Amazon VPC	Aggiunto supporto per Amazon VPC. Per ulteriori informazioni, consulta Avvio di istanze di dimensionamento automatico in un VPC .	2 dicembre 2010
Supporto per cluster HPC	Aggiunto il supporto per cluster high performance computing (HPC).	2 dicembre 2010
Supporto per controlli dell'integrità	Aggiunto il supporto per utilizzare i controlli dell'integrità dell'Elastic Load Balancing con istanze EC2 Auto Scaling gestite da Dimensionamento automatico Amazon EC2. Per ulteriori informazioni, consulta Health checks for instances in an Auto Scaling group .	2 dicembre 2010
Supporto per CloudWatch allarmi	È stato rimosso il vecchio meccanismo di attivazione e riprogettato Amazon EC2 Auto Scaling per utilizzare la funzionalità di allarme. CloudWatch Per ulteriori informazioni, consulta Dynamic scaling for Dimensionamento automatico Amazon EC2 (Dimensionamento dinamico per Dimensionamento automatico Amazon EC2) .	2 dicembre 2010
Sospensione e ripresa del dimensionamento	Aggiunto il supporto per sospendere e riprendere i processi di dimensionamento.	2 dicembre 2010
Supporto per IAM;	Aggiunto il supporto per IAM. Per ulteriori informazioni, consulta Controllo dell'accesso alle risorse Dimensionamento automatico Amazon EC2 .	2 dicembre 2010

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.